Environmental Influences on Patterns of Atmospheric Particulate Matter: a Quantitative Study Using Ground- and Satellite-Based Observations

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der KIT-Fakultät für

Bauingenieur-, Geo- und Umweltwissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Roland Stirnberg

aus Traunstein

Tag der mündlichen
Prüfung: 18.06.2020

Referent: Professor Dr. Jan Cermak

Korreferent: PD Dr. Andreas Philipp

Karlsruhe 2020

Danksagung

Kurzfassung

Luftverschmutzung, insbesondere hohe Konzentrationen von mikroskopischen Partikeln in der Atmosphäre, sogenannte Feinstaubpartikel (PM), haben schwerwiegende Auswirkungen auf die menschliche Gesundheit. Partikel mit einem aerodynamischen Durchmesser von weniger als $10\,\mu\mathrm{m}$ (PM10) können in die Atemwege gelangen und bereits eine kurzzeitige Exposition gegenüber hohen PM-Konzentrationen kann zu unmittelbaren negativen Auswirkungen wie Asthmaanfällen führen. Sind Menschen über einen längeren Zeitraum erhöhten PM-Konzentrationen ausgesetzt, kann die Lunge geschädigt werden und das Risiko von Herz-Kreislauf-Erkrankungen und Diabetes steigt. Diese gesundheitlichen Auswirkungen können die Lebenserwartung senken. Obwohl in den letzten Jahrzehnten ein rückläufiger Trend der PM-Konzentrationen in Europa zu verzeichnen ist, liegen die aktuellen PM-Konzentrationen in vielen Mitgliedsstaaten immer noch über den WHO-Empfehlungen, was zur Folge hat, dass die derzeitigen PM-Konzentrationen in vielen Gebieten Europas mit hoher Wahrscheinlichkeit für Menschen schädlich sind. Infolgedessen wurden bereits einige Maßnahmen gegen die Luftverschmutzung umgesetzt, darunter städtische Umweltzonen und andere Einschränkungen für den privaten Autoverkehr. Es sind jedoch weitere Anstrengungen erforderlich, um gesundheitlich unbedenkliche PM-Konzentrationen zu ermöglichen. Um effizientere Strategien für eine bessere Luftqualität zu entwickeln, müssen den Entscheidungsträgern zusammenhängende Informationen über räumlich-zeitliche Muster der PM-Konzentrationen zur Verfügung stehen. Die sogenannte Aerosol Optische Dicke (AOD), die aus satellitengestützten Messungen gewonnen wird, hat das Potenzial, diese Informationen zu liefern. Die AOD stellt das Integral der Partikelbelastung in einer Atmosphärensäule dar, die mit den bodennahen PM-Konzentrationen in Beziehung gesetzt werden kann. Dies ist notwendig, da bodennahe PM-Konzentrationen von besonderer Relevanz sind für die Bestimmung schädlicher Auswirkungen auf den Menschen. Die Verwendung der AOD zur Annäherung der PM-Konzentrationen in Bodennähe bringt jedoch einige Herausforderungen mit sich, da die Beziehung zwischen AOD und PM durch eine Reihe von meteorologischen Parametern beeinflusst wird. Daher wird in dieser Arbeit das Potenzial satellitengestützter AOD zur Bestimmung bodennaher PM-Konzentrationen analysiert und eine Grundlage für die genaue Ableitung zusammenhängender Informationen zur bodennahen Luftverschmutzung durch satellitengestützte AOD geschaffen. Darüber hinaus ist bekannt, dass verschiedene Umweltfaktoren PM-Konzentrationen beeinflussen und die Luftverschmutzung verstärken können.

Um die Wirksamkeit von Strategien zur Verbesserung der Luftqualität wissenschaftlich beurteilen zu können, müssen die Auswirkungen von Umwelteinflüssen auf die PM-Konzentrationen von anthropogenen Emissionen getrennt werden. In dieser Arbeit wird das wissenschaftliche Verständnis der Umwelteinflüsse auf die PM-Konzentrationen und die Entwicklung von Phasen starker Verschmutzung in Bezug auf die atmosphärischen Umgebungsbedingungen erweitert.

In dieser Arbeit werden drei zusammenhängende Studien vorgestellt, die sich jeweils mit einer Hauptforschungsfrage befassen. Diese Hauptforschungsfragen zusammen mit den Hauptergebnissen sind:

Wie beeinflussen die meteorologischen Bedingungen die statistische Beziehung zwischen AOD und PM?

Eine für den Nordosten Deutschlands durchgeführte Studie zeigt einen nichtlinearen Zusammenhang zwischen AOD und PM10 in Bodennähe auf, was auf den Einfluss der meteorologischen Parameter relative Luftfeuchtigkeit (RH), Grenzschichthöhe (BLH), Windrichtung und Windgeschwindigkeit zurückzuführen ist. Wenn eine relativ trockene Atmosphäre ($30\,\% < \mathrm{RH} \leq 50\,\%$) mit einer mittelhohen BLH (600–1,200 m) zusammenfällt, sind die normalisierten Werte von AOD und PM10 bei Anwendung einer semiquantitativen Skala ähnlich. Dies deutet auf eine Übereinstimmung von AOD und PM10 hin. Mit zunehmender Umgebungsfeuchte (>80 %) erhöht sich die AOD durch die Feuchtigkeitsaufnahme der Partikel und dem dadurch verursachten hygroskopischen Partikelwachstum. Dies führt zu einer relativen Überschätzung der trockenen Partikelkonzentration in Bodennähe, wenn diese auf Basis der AD approximiert wird. Dieser Effekt kann jedoch durch höhere PM10-Messwerte bei niedrigen Grenzschichten (<600 m) kompensiert werden, was schließlich zu AOD- und PM10-Satellitenmessungen in ähnlicher Größenordnung führt. Die Windrichtung beeinflusst die Beziehung zwischen AOD und PM10 durch den Transport von Luftmassen mit unterschiedlichen Eigenschaften in das Untersuchungsgebiet. Unter Bedingungen, die von westlichen Luftmassen dominiert werden, ist die Wahrscheinlichkeit vergleichsweise hoch, dass die AOD bei Anwendung einer semiquantitativen Skala relativ höher ist als die PM10-Beobachtung. Dies deutet auf eine Überschätzung der PM10-Konzentrationen auf Basis der AOD hin. Westliche Luftmassen sind häufig marinen Ursprungs und haben damit tendenziell eine höhere RH und enthalten einen hohen Gehalt an Meersalzen. Meersalze sind hydrophil und fördern das hygroskopische Wachstum von Partikeln, wodurch wiederum die AOD erhöht wird.

Die Analyse des Zusammenhangs zwischen AOD und PM10 zeigt, dass die Berücksichtigung der Parameter RH, BLH und Wind notwendig ist, wenn Schätzungen von PM10 auf Basis von satellitengestützter AOD angestrebt werden. Das in dieser Studie vorgestellte Konzept der Normalisierung der AOD / PM10-Datenpaare eignet sich für die Anwendung in anderen Untersuchungsgebieten. Die Erkenntnisse dieser Studie haben das Potenzial, zukünftige

Abschätzungen bodennaher PM-Konzentrationen auf Basis von Satelliten-AOD zu verbessern.

Was sind die bestimmenden Einflussfaktoren auf PM10-Konzentrationen, wenn diese auf Basis der vorherrschenden Umweltbedingungen und AOD abgeschätzt werden?

Ein statistisches Modell wird zur Vorhersage in Deutschland gemessener PM10-Konzentrationen auf Basis satellitengestützter AOD und unter Berücksichtigung der Umweltbedingungen aufgesetzt. Sensitivitätsanalysen an diesem Modell zeigen, dass die wichtigsten Einflussfaktoren auf die modellierten PM10-Konzentrationen die Ost-West-Windströmung, die BLH und die Temperatur sind. Einströmung von Luftmassen aus östlichen Richtungen über mehrere Tage hinweg erhöht die modellierten PM10-Konzentrationen im Durchschnitt um $\sim$10 μg/m^3 im Vergleich zu Situationen, die von westlichem Einstrom dominiert werden. Dies ist auf grenzüberschreitenden Partikeltransport aus Ländern östlich von Deutschland zurückzuführen. Modellierte PM10-Konzentrationen für niedrige BLHs (<800 m) erhöhen sich um durchschnittlich $\sim$10 μg/m^3 aufgrund der Akkumulation von Partikeln in Bodennähe. Dieser Mechanismus erweist sich im Winter und Herbst als besonders wichtig. Im Sommer zeigen die Modellergebnisse eine deutliche Erhöhung der PM10-Vorhersagen (bis zu $\sim$12 μg/m^3 bei um 15 K erhöhten Temperaturen). Dies ist auf eine verstärkte biogenene Aktivität und erhöhte Staubaufwirbelung aufgrund ausgetrockneter Böden zurückzuführen. Im gleichen Modell-Setup zeigen Sensitivitätsanalysen, dass die AOD positiv mit PM10 korreliert, aber BLH und die Ost-West-Windkomponente die Beziehung zwischen AOD und PM10 wesentlich beeinflussen. AOD und PM10 korrelieren im Sommer schwächer, da dann die Partikel innerhalb einer höheren Grenzschicht stärker verteilt sind und die AOD weitgehend von den höher in der Atmosphäre befindlichen Partikeln bestimmt wird. Die Ergebnisse deuten darauf hin, dass die AOD zur Beurteilung der Luftqualität am Boden verwendet werden kann, wenn sie durch eine statistische Modellierung mit meteorologischen Umgebungsbedingungen verknüpft wird. Darüber hinaus wird der starke Einfluss der meteorologischen Bedingungen auf die PM10-Konzentrationen aufgezeigt.

Wie bestimmen atmosphärische Prozesse die Konzentration verschiedener chemischer Bestandteile von PM1 auf lokaler Ebene?

Atmosphärische Einflussfaktoren auf Konzentrationen von Feinstaubpartikeln, die kleiner als 1 μm (PM1) sind, und deren chemischen Hauptbestandteile, werden analysiert. Der Fokus liegt dabei auf den Prozessen, die zu Phasen mit hoher Schadstoffbelastung führen. Ein statistisches Modell wird aufgestellt, um die täglichen Schwankungen der PM1-Konzentrationen unter Berücksichtigung der meteorologischen Bedingungen zu reproduzieren. Dafür werden Daten eines suburban-geprägten Standorts südwestlich von Paris, Frankreich, verwendet. Sensitivitätsanalysen des Modells deuten darauf hin, dass Spitzenkonzentrationen von PM1 im Winter durch die bodennahe Partikelakkumulation bei niedrigen BLHs in Kombination mit der Bildung neuer Partikel und

erhöhten Heizungsemissionen bei niedrigen Temperaturen ($<\sim 5\,°C$) verursacht werden. Im Sommer sind die Spitzenkonzentrationen von PM1 im Allgemeinen niedriger als im Winter. Erhöhte PM1-Konzentrationen treten jedoch auf, wenn windstille Bedingungen mit hohen Temperaturen zusammentreffen, die zu photochemisch induzierten Partikelbildungsprozessen führen. Der Transport verschmutzter Luft aus der Pariser Region führt in beiden Jahreszeiten zu einem deutlichen Anstieg der PM1-Konzentrationen.

Hochaufgelöste Fallstudien zeigen eine große Variabilität der Prozesse, die zu Phasen starker Verschmutzung führen. Die Prozesse variieren nicht nur zwischen, sondern auch innerhalb der Jahreszeiten. Die Modellergebnisse zeigen beispielsweise für eine Phase starker Luftverschmutzung im Januar 2016, dass diese durch einen Temperaturabfall verursacht wurde, was die modellierten PM1 Konzentrationen um bis zu $11\,\mu g/m^3$ erhöht. Dies wird auf eine verstärkte Bildung von sekundären anorganischen Aerosolen (SIA) und einen Anstieg der lokalen Heizungsemissionen zurückgeführt. Im Gegensatz dazu werden im Dezember 2016 hohe PM1-Konzentrationen hauptsächlich durch eine niedrige BLH und Partikeladvektion aus dem Raum Paris verursacht. Ein beobachteter Rückgang der Schadstoffkonzentrationen während dieser Phase hängt mit einer Änderung der Windrichtung zusammen, die weniger belastete, maritime Luftmassen herantransportiert, was zu einem Rückgang der PM1-Konzentrationen von $\sim$-$4\,\mu g/m^3$ führt. Obwohl sich diese Ergebnisse auf lokale Ereignisse beziehen, sind die Ergebnisse verallgemeinerbar und auch für andere Regionen relevant. Dies betrifft beispielsweise die Relevanz der Bildung neuer Partikel während kalter Temperaturen oder die Akkumulation von Partikeln in bodennahen Atmosphärenschichten durch eine niedrige BLH. Der Einfluss transportierter Partikel unterstreicht die Notwendigkeit großräumiger Maßnahmen zur Senkung des atmosphärischen Partikelgehalts.

Diese Arbeit liefert ein quantitatives Verständnis der Beziehung zwischen AOD und PM10 und schafft damit eine Grundlage für Abschätzungen von PM auf Basis von AOD. Diese PM-Abschätzungen sind von großem Nutzen für die Identifizierung von stark verschmutzten Gebieten und zur langfristigen Beobachtung der Luftqualität auf großen räumlichen Skalen. Darüber hinaus ist das wissenschaftliche Verständnis der Umweltprozesse, die PM-Konzentrationen beeinflussen, wichtig, um atmosphärische Prozesse bei der Entwicklung von Strategien zur Schadstoffminderung berücksichtigen zu können.

Abstract

Air pollution, in particular high concentrations of suspended particles in the atmosphere, so-called particulate matter (PM), has serious implications on human well-being. Particles with an aerodynamic diameter below $10\,\mu$m (PM10) can enter the human respiratory system, and even short-term exposures to high PM concentrations can lead to immediate negative effects such as an increased numbers of asthma seizures. If humans are exposed to increased PM concentrations for a longer period, the lungs can be damaged and the risk of cardiovascular diseases and diabetes increases. These adverse health effects can lead to a loss in life expectancy. Although the last decades showed a declining trend of PM concentrations in Europe, PM levels still remain above WHO recommendations, meaning that there is a high probability of current PM concentrations being harmful to humans in many areas of Europe. As a consequence, several measures against air pollution have been proposed, including urban low emission zones or other partial driving bans. Still, more efforts are needed to ensure PM levels are below harmful levels. To develop more efficient plans towards better air quality, policymakers need to be provided with coherent information on spatiotemporal patterns of PM concentrations. Aerosol optical depth (AOD) retrieved from satellite-based measurements has the potential to provide this information. AOD constitutes a columnar-integrated approximation of the atmospheric particle burden and can be related to near-ground concentrations of PM, which is the more relevant quantity concerning harmful effects on humans. However, using AOD to approximate PM concentrations at ground level presents several challenges, as the relationship between AOD and PM is influenced by a number of meteorological parameters. Hence, this thesis analyses the applicability of columnar, satellite-based AOD to derive ground-level PM. Furthermore, it is known that environmental conditions are important drivers of PM concentrations and can amplify high pollution situations. For a scientifically sound assessment of the efficiency of strategies towards better air quality, the effect of environmental influences on PM concentrations needs to be separated from anthropogenic emissions. It is, therefore, a further goal of this thesis to advance the scientific understanding of environmental drivers on PM concentrations and of how high pollution episodes evolve with respect to ambient atmospheric conditions.

Based on these motivations, three studies are presented in this thesis, each addressing a main research question. These main research question along with the main findings are:

How do meteorological conditions influence the statistical relationship between AOD and PM?

For a study area in northeastern Germany, the relationship between AOD and PM10 is shown to be highly nonlinear, which is attributed to the influence of the meteorological parameters relative humidity (RH), boundary layer height (BLH), wind direction, and wind speed. Generally, when a relatively dry atmosphere (30 % <RH≤ 50 %) coincides with a medium BLH (600–1200 m), AOD and PM10 are in the same range on a semi-quantitative scale, indicating agreement of AOD and PM10. AOD increases with increasing ambient RH (>80 %) due to hygroscopic particle growth, leading to a relative overestimation of the dry particle concentration near the ground. However, this effect can be compensated if a low boundary layer (<600 m) is present, which in turn significantly increases PM10, eventually leading to satellite AOD and PM10 measurements of similar magnitude. Wind direction influences the relationship between AOD and PM10 by transporting air masses with varying characteristics to the study area. During conditions dominated by western air masses, there is a relatively high probability that the satellite AOD is higher than the PM10 observation on a semi-quantitative scale, indicating an overestimation of PM10 concentrations by AOD. Western air masses are more likely to be of marine origin, meaning that they are wetter, i.e., tend to have a higher RH, and carry a high content of sea salts. Sea salts are hydrophilic and promote the hygroscopic growth of particles, thus increasing the AOD.

The analysis of the relationship between AOD and PM10 shows that if estimates of PM10 from AOD are targeted, the consideration of the parameters RH, BLH, and wind is necessary. The concept presented in this research chapter can be generalized for use in other study areas. Insights of this study have the potential to improve future estimations of near-ground PM concentrations based on satellite AOD.

What are the main drivers of PM10 when environmental conditions and AOD are used to estimate PM10 concentrations?

A statistical model is set up to predict hourly mean PM10 concentrations for measurement sites in Germany, taking into account satellite AOD and ambient environmental conditions pertaining to meteorology and the land surface. Sensitivity analyses on this model show that important meteorological drivers of modeled PM10 concentrations include multi-day east-west wind flow, BLH, and temperature. Multiday inflow from eastern directions increases PM10 concentrations on average by $\sim$10 μg/m^3 when compared to situations dominated by western inflow, indicating trans-boundary particle transport from countries east of Germany. An increase of modeled PM10 concentrations for low BLHs (<800 m) is estimated to be on average $\sim$10μg/m^3 due to confinement of particles close to the ground. This mechanism is shown to be particularly important in winter and autumn. An increase in de-seasonalised temperatures in summer causes a marked increase in PM10 predictions (up to $\sim$12μg/m^3 for a temperature increase of 15 K) due to enhanced biogenic activity and increased dust resuspension due to dried-up soils. In the same modelling framework,

sensitivity analyses show that AOD is positively related to PM10, but BLH and the east-west wind component substantially influence the relationship between AOD and PM10. AOD and PM10 are weaker correlated in summer when particles are more dispersed within a well-mixed boundary layer, and the AOD is largely determined by particles higher up in the atmosphere. These results suggest that AOD can be used to assess air quality at ground level in a statistical modelling framework that links it with meteorological conditions. Furthermore, the strong influence of meteorological conditions on variations of PM10 concentrations is shown.

How do atmospheric processes drive concentrations of different PM1 species at a local scale?
Atmospheric drivers of concentrations of speciated fine-mode particles, i.e., particles smaller than $1\,\mu$m (PM1), and processes leading to high-pollution episodes are analysed. A statistical model is set up to predict day-to-day variations of PM1 concentrations under consideration of meteorological conditions using data from a suburban site southwest of Paris, France. Sensitivity analyses on the model suggest that the particle accumulation effect of low BLHs in combination with the formation of new particles and increased heating emissions during low-temperatures ($<\sim5\,°$C) drive peak PM1 concentrations in winter. Summer peak PM1 concentrations are generally lower than in winter. Still, elevated PM1 concentrations in summer occur when calm conditions coincide with high temperatures that lead to photochemically-induced particle formation processes. Advection of polluted air from the Paris region cause substantially increased PM1 concentrations in both seasons. High-resolution case studies show a large variability of processes that can lead to the development of high pollution events. Processes vary between and even within seasons. A high pollution episode in January 2016 is shown to be driven by a drop in temperature, increasing PM1 predictions by up to $11\,\mu$g/m^3. This is attributed to enhanced formation of secondary inorganic aerosols (SIA) and an increase in local wood-burning emissions. In contrast, during December 2016, high PM1 concentrations are caused mainly by a shallow BLH and particle advection from the Paris area. It is shown that an observed decrease in pollution levels is linked to a change in wind direction, advecting cleaner, maritime air to the PM measurement site, leading to a decrease of $\sim$-4$\,\mu$g/m^3. Although these results refer to local events, the results are generalizable and relevant for other regions as well. This concerns for example the relevance of new particles formation processes during cold conditions, or the accumulation of particles in near-ground layers of the atmosphere during shallow BLHs. The influence of transported particles underlines the need for large-scale measures to reduce atmospheric PM concentrations.

This thesis provides a quantitative understanding of the relationship between AOD and PM10, which constitutes a valuable basis for future efforts to accurately estimate PM from AOD. These PM estimates are highly beneficial for the identification of pollution hotspots and the long-term monitoring of air quality over a large spatial scale. Furthermore, the advancement of the

scientific understanding of environmental processes driving PM concentrations is crucial for the development of pollution mitigation strategies and to scientifically test the efficiency of these measures.

Contents

List of Figures

List of Tables

List of Acronyms

ACSM   Aerosol chemical speciation monitor

ACTRIS   Aerosols, Clouds and Trace gases Research Infrastructure

AERONET   Aerosol Robotic Network

AiRDI   Air quality rank difference index

AirPres   Air pressure

AOD   Aerosol optical depth

BC   Black carbon

BCff   Black carbon from fossil-fuel burning

BCwd   Black carbon from wood burning

BLH   Boundary layer height

BRDF   Surface bidirectional reflectance distribution function

BVOCs   Biogenic volatile organic compounds

CABAM   Characterising the atmospheric boundary layer based on automatic lidar and ceilometer measurements

CALIOP   Cloud-Aerosol Lidar with Orthogonal Polarization

CALIPSO   Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations

CAPE   Convective available potential energy

CLC   Coordination of information on the environment land cover

Conti   Continentality

CPA   Change point analysis

CTM   Chemical transport model

DB   Deep blue

DEM   Digital elevation model

DOY	Day of the year
DT	Dark target
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecasts
EEA	European Environment Agency
Elev	Elevation
ERA	ECMWF Re-Analysis
GBRT	Gradient boosted regression trees
GOES	Geostationary Operational Environmental Satellite
ICE	Individual conditional expectation
LUTs	Look-up tables
MAIAC	Multi-angle implementation of atmospheric correction
MISR	Multiangle Imaging Spectroradiometer
ML	Machine learning
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
NDVI	Normalized difference vegetation index
NIR	Near infrared
NO3_frac	NO_3 fraction of the previous day
NRMSE	Normalized RMSE
Org	Organic matter
PD	Partial dependence
PDF	Probability density function
PM	Particulate matter
PM1	Particulate matter below $1\,\mu$m
PM10	Particulate matter below $10\,\mu$m
PM10em	Annual emission of PM10
PM2.5	Particulate matter below $2.5\,\mu$m

POLDER Polarization and Directionality of the Earth's Reflectance

Precip_tsince Time since last precipitation

Precip_magn Magnitude of last precipitation

Precip_acc Cumulative precipitation last 24h

RADOLAN Radar Online Adjustment Project

RF Random forests

RH Relative humidity

RMSE Root mean squared error

SEVIRI Spinning Enhanced Visible and Infrared Imager

SIA Secondary inorganic aerosols

SIRTA Site Instrumental de Recherche par Télédétection Atmosphérique

SOA Secondary organic aerosols

SRCs Spectral regression coefficients

SSRD Surface solar radiation downward

SWIR Shortwave infrared

T Air temperature

Tan Air temperature anomaly

TISR Total incoming solar radiation

TOA Top of the atmosphere

TPI Topographic position index

u u wind vector

UBA Umweltbundesamt

umean Mean u wind vector

v v wind vector

VIIRS Visible Infrared Imaging Radiometer Suite

vmean Mean v wind vector

VOCs Volatile organic compounds

WHO World Health Organisation

ws Wind speed

1 Introduction and general motivation

1.1 Small particles, large impact: atmospheric PM

Air pollution has serious implications on human well-being. In particular, high concentrations of microscopic particles in the atmosphere, so-called particulate matter (PM[1]), are known to have adverse effects on human health. Especially particles with an aerodynamic diameter below $10\,\mu$m (PM10) can enter the human respiratory system, causing short-term effects such as increased numbers of asthma seizures (Hughes et al., 2018) and an exacerbation of chronic obstructive pulmonary diseases (Sinharay et al., 2018). Furthermore, the lungs can be damaged due to inflammation, cytotoxicity, and emphysema (Atkinson et al., 2001; Pope et al., 2002; Huang et al., 2012; Badyda et al., 2016). Chronic exposure to increased levels of PM poses an important risk factor by increasing the chance of cardiovascular diseases (Hennig et al., 2018; Lelieveld et al., 2019) and possibly diabetes (Weaver et al., 2019).

Adverse effects on human health can be quantified by coupling risk estimations for the population from large-scale and long-term cohort studies with PM concentration data from in situ measurements (Pope et al., 2002) or atmospheric chemistry models (Lelieveld et al., 2015, 2019). Based on the number of civilians that are exposed to air pollution, an estimation of premature deaths and loss in life expectancy due to ambient PM concentrations can be provided. Lelieveld et al. (2015), estimated ambient air pollution to cause 3.3 million premature deaths worldwide per year. Estimates for Europe amount to 790,000 premature deaths per year (Lelieveld et al., 2019). For Germany, the German Federal Environment Agency (Umweltbundesamt, UBA) has quantified 47,000 premature deaths per year (Kallweit and Wintermeyer, 2013). The use of premature deaths as an index for adverse health effects of PM has recently attracted methodological criticism in the scientific community, with Morfeld and Erren (2019) now proposing to quantify the disease burden in terms of loss of life expectancy. In Europe, the estimated number of premature deaths would be equivalent to a reduction of the average life expectancy of 2.2 years (Lelieveld et al., 2019), while for Germany, the estimation is given as loss of longevity of 10 years per 1,000 inhabitants per year in Kallweit and Wintermeyer (2013). Assuming a life expectancy of 80 years, this would correspond to an overall loss in life expectancy of 0.8 years per capita. People living in urban areas are particularly affected by poor air quality (see Izquierdo et al., 2020) and with ongoing

---

[1] Note that the terms "PM" and "aerosol" are treated as synonyms in this thesis and both refer to particles suspended in the atmosphere.

urbanization, the number of urban dwellers is projected to grow (Baklanov et al., 2016; Li et al., 2019b). In addition to the negative effects on human well-being, it is known that the increase in morbidity rates has negative effects on national economies by reducing the number of the available working population and placing a burden on the health system (Turnock et al., 2016; Bayat et al., 2019; Jenniches and Worrell, 2019).

In order to protect European citizens against the harmful effects of air pollution, the EU has introduced a legal framework requiring its member states to comply with certain thresholds of atmospheric particle concentrations. The thresholds are regulated in the directive 2008/50/EC of the European Parliament and of the Council (see Tab. 1.1) for particles smaller than $10\,\mu$m and particles smaller than $2.5\,\mu$m (PM2.5). Regarding PM10, the limit of $50\,\mu$g/m^3 may not be exceeded more than 35 times a year and the annual average may not exceed $40\,\mu$g/m^3. For PM2.5, the annual threshold was set to $25\,\mu$g/m^3 as of 2015 and was decreased to $20\,\mu$g/m^3 as of January 2020 (EU, 2008). Presumably, this framework has motivated European states to intensify their efforts towards cleaner air, as PM concentrations have been decreasing in the last decades (Turnock et al., 2016; Stafoggia et al., 2017; Grange et al., 2018). Political changes have led to stricter standards for road transport, energy, and industrial sectors, in particular in eastern European countries, and, together with technological advances, have caused a decrease of annual mean PM2.5 concentrations in Europe of $\sim$35 % between 1970 and 2010 (Turnock et al., 2016). Annual mean PM10 and PM2.5 concentrations have also been declining in Germany and have largely complied with EU regulations in recent years (Umweltbundesamt, 2019). While this is undoubtedly an enormous success, it is important to note that the thresholds imposed by the EU are the result of political decision making processes and hence represent the lowest threshold concentrations possible in the context of capabilities, public health priorities, and political willingness (see World Health Organization, 2006). There is no body of evidence suggesting that below these thresholds no adverse health effects are to be expected. The World Health Organisation (WHO) argues that the concentration level at which adverse health effects are to be expected is very close to the natural background concentration (World Health Organization, 2006). Following this argumentation, the WHO has set much stricter thresholds as a recommendation for policymakers (see Tab. 1.1). The WHO recommendation for the PM2.5 annual average concentration is set at a maximum of $10\,\mu$/m^3. This threshold was exceeded at 93 % of all measurement sites in Germany in 2018 (Umweltbundesamt, 2019). Furthermore, the WHO recommends that daily average PM2.5 concentrations do not exceed $25\,\mu$/m^3 more than 3 times a year. This recommendation was not met at any measurement site in 2018. Concerning PM10, the WHO recommendation for the annual average concentration ($10\,\mu$/m^3) was missed at 23 % of all sites (Umweltbundesamt, 2019). No legal threshold or WHO recommendation exists for PM1 particles, although studies show that particles with a diameter smaller than $1\,\mu$m (PM1) are particularly dangerous to human health (Chen et al., 2017; Yang et al., 2018).

Table 1.1: Overview of EU legal thresholds and WHO recommendations for maximum PM10, PM2.5, and PM1 concentrations. Numbers taken from World Health Organization (2006), EU (2008), and Umweltbundesamt (2019),

	PM10	**PM2.5**	**PM1**
EU threshold	Annual average: $40\,\mu g/m^3$ Daily average: $50\,\mu g/m^3$, 35 times a year	Annual average: $20\,\mu g/m^3$ Daily average: no threshold	Annual average: no threshold Daily average: no threshold
WHO recommen- dation	Annual average: $20\,\mu g/m^3$ Daily average: $50\,\mu g/m^3$, 3 times a year	Annual average: $10\,\mu g/m^3$ Daily average: $25\,\mu g/m^3$, 3 times a year	Annual average: no threshold Daily average: no threshold

This thesis acknowledges the adverse effects of increased PM concentrations to human well-being and the challenges that accompany efforts towards cleaner air. To develop efficient plans towards cleaner air, policymakers need to be provided with coherent information on spatiotemporal patterns of PM concentrations. Aerosol optical depth (AOD) retrieved from satellite-based measurements has the potential to provide this kind of information. Providing columnar-integrated approximations of atmospheric particle burdens, AOD measurements pose a key parameter for aerosol studies (Sogacheva et al., 2020). Satellite-based AOD has the potential to provide spatially-coherent information on PM concentrations, which is necessary because of the limited number of PM measurement sites that have been put up in Germany cover only a small fraction of the national territory (see Fig. 1.1). The importance of spatially coherent information on PM concentrations has been furthermore emphasized by the European Court of Justice, which ruled in 2019 that thresholds must be met at any site and that averaging over several measurement sites is an invalid procedure (Euopean Court of Justice, 2019). This means that the assessment of air quality currently relies heavily on the choice of the measurement sites, as it is referring to point measurements and not to agglomerated areas (e.g., cities). To overcome this problem, the European Directive 2008/50 on ambient air quality and cleaner air for Europe specifically mentions that fixed measurements should be supplemented by further techniques indicative for the geographical distribution of concentrations (EU, 2008). Satellite measurements have the potential to fulfill this task.

Furthermore, the effect of environmental influences on particle transport and on the emission, formation, and depletion of particles need to be separated from anthropogenic emissions to scientifically assess the efficiency of political measures. For example, it is known that, although anthropogenic emissions play an important role for PM concentrations in the atmosphere,
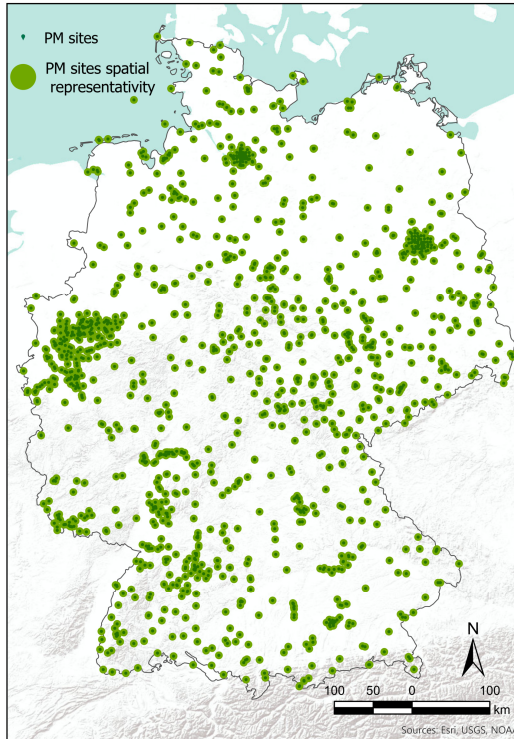
Figure 1.1: Location of PM measurement sites of the UBA and area covered assuming a spatial representativity of 5 km. In total, only ∼18 % of the national territory are covered by PM measurements. Note that the availability of measurements varies as some sites were abandoned, meaning that in reality, the percentage covered might be substantially smaller.

meteorological conditions related to large-scale circulation patterns as well as local-scale boundary layer processes are important drivers of PM concentrations. Therefore, it is a further goal of this thesis to advance the scientific understanding of how high pollution episodes evolve with respect to ambient atmospheric conditions. Given that there is a number of proposed efforts to reduce anthropogenic particle emissions, such as the setting up of low emission zones or other driving restrictions in Berlin (Berlin Senate, 2013; Bonn et al., 2016), Munich (Qadir et al., 2013; Fensterer et al., 2014), or London (Ellison et al., 2013), it is mandatory to develop a more quantitative understanding of drivers of air pollution. Eventually, this would enable a scientific assessment of the efficiency and the usefulness of plans to ensure improved air quality.

This thesis is structured in five main parts. An introductory chapter describes the current state of knowledge of processes related to the formation and depletion of PM, with regard to meteorology and land surface parameters (section 1.2). A technical description of in situ PM measurements and of AOD and the associated challenges of the application of AOD within the context of on-site air pollution are given in section 1.3. In the first research chapter, different atmospheric conditions that influence the suitability of columnar, satellite-based AOD to derive ground-level PM are analysed and a basis to accurately derive coherent information on near-ground air pollution from satellite-based AOD is set (chapter 2). The subsequent research chapter builds on these insights to quantitatively link satellite AOD and PM measurements under consideration of environmental conditions (chapter 3). In the final research chapter, site-specific meteorological drivers of PM are quantified, allowing a detailed analysis of conditions that lead to high-pollution events (chapter 4). The findings are concluded and contrasted with the current state of knowledge in a final discussion (chapter 5). The study areas are northeastern Germany (chapter 2), the whole German territory (chapter 3), and suburban Paris (chapter 4).

1.2   PM in the atmosphere – sources, composition and environmental controls

Atmospheric concentrations of PM are often the result of complex interactions between natural and anthropogenic particle sources and sinks. The atmospheric boundary layer usually contains a mixture of particles that differ in terms of characteristics such as chemical constituents, size, lifetime, and hygroscopicity, to name a few. The major chemical constituents of PM include mineral dust, sea salt, organic matter (Org), black carbon (BC), and inorganic ions such as nitrate (NO_3^-), sulfate (SO_4^{2-}), and ammonium (NH_4^+) (Putaud et al., 2004, 2010; Fuzzi et al., 2015). The typical size of atmospheric particles ranges from some nanometers to some micrometers (see Fig. 1.2, Fuzzi et al., 2015). In the following subsections, PM sources and transformation processes, typical size distributions, and important environmental controls of PM concentrations are highlighted.

Primary sources and transformation processes of PM

Suspended atmospheric particles are either directly emitted to the atmosphere as primary particles or formed as secondary particles by gas-particle conversion processes such as nucleation and subsequent condensation (Kanakidou et al., 2005; Hallquist et al., 2009; Fuzzi et al., 2015). Human activity influences the atmospheric concentration of particles by providing a net particle source (Klimont et al., 2017). Still, the bulk amount of PM is emitted by natural sources.
Large natural emitters of primary particles are the oceans, deserts, wildfires, volcanoes, and the

biosphere (Dentener et al., 2006; de Leeuw et al., 2011; Fuzzi et al., 2015). Primary marine aerosols emitted from the ocean surfaces constitute one of the largest particle source to the atmospheric particle budget. These particles reach the atmosphere when water bubbles burst due to breaking waves on the water surface. These particles either consist purely of sea salt or but can also contain substantial amounts of organic matter (de Leeuw et al., 2011; Fuzzi et al., 2015). Deserts are strong particle emitters due to the vast areas of bare soil, which allow an efficient, wind-driven transport of mineral dust (Fuzzi et al., 2015). Saharan dust events, for example, can have substantial impact on PM concentrations in Europe (Knippertz and Todd, 2012; Hande et al., 2015). Wildfires frequently occur during the dry seasons and emit large quantities of BC and Org (Dentener et al., 2006). Primary particles emitted by the biosphere can be comprised of very heterogeneous constituents including microorganisms, pollen, spores, and other biological fragments (Möhler et al., 2007; Fuzzi et al., 2015).

Anthropogenic emissions are related to transport and agriculture as well as residential heating. Transport-related particle emissions can be due to exhaust or non-exhaust sources. Non-exhaust emissions due to the abrasion of tyres, breaks, and roads or due to dust resuspension are of growing importance as transport activities generally increase in Europe. Conversely, exhaust emissions have been decreasing during the last two decades due to cleaner technologies (Fuzzi et al., 2015). Agricultural activities such as tillage operations (Funk et al., 2008) or grazing (Hoffmann et al., 2008) are known to deteriorate air quality, in particular during dry summer episodes (Hoffmann and Funk, 2015). Overall, primary particles emitted by agricultural activities account for around 25 % of global dust emissions (Ginoux et al., 2012). Residential heating based on wood or coal combustion is still a widespread source of BC not only in developing countries, but also in Europe. Due to ineffective combustion, residential heating based on solid fuels can be also an important source of organic compounds (Chafe et al., 2015).

Emissions of precursor gases from natural and anthropogenic sources indirectly contribute to the atmospheric particle budget after being transformed to secondary aerosols in the atmosphere. Important precursor gases are sulfur dioxide (SO_2), nitrogen oxides (NO_x), ammonia (NH_3), and volatile organic compounds (VOCs). Natural emitters of SO_2 are volcanoes and wildfires (Dentener et al., 2006). NO_x is emitted during microbial decomposition processes in soils (Hall et al., 1996) and, in conjunction with NH_3, by agricultural activities due to the deployment of fertilizer (Bertram et al., 2005; Petetin et al., 2014). VOCs emitted from biogenic sources (Biogenic Volatile Organic Compounds, BVOCs) have been identified as a major precursor for secondary organic aerosols (SOA), which in turn contribute substantially to the organic particle fraction (Atkinson, 2000; Kanakidou et al., 2005; Li et al., 2020). Anthropogenic sources of the precursors gases SO_2 and NO_x are mainly the combustion of fuels and fossil fuels (Chafe et al., 2015; Fuzzi et al., 2015; Li et al., 2017). Anthropogenic sources of VOCs include diesel engines or gasoline vehicles (Gentner et al., 2012).

Transformation processes that form secondary inorganic aerosols (SIA) based on NH_3, NO_x and SO_2 are relatively well understood (see Pay et al., 2012; Fuzzi et al., 2015; Alanen et al., 2017) and the availability of NH_3, NO_x, and SO_2 in the atmosphere is known to substantially impact PM2.5 concentrations (Megaritis et al., 2013). The formation of SIA is a two-step process, starting with the oxidation of gaseous NO_x or SO_2 to form nitric acid (HNO_3) or sulfuric acid (H_2SO_4), respectively. Subsequently, HNO_3, H_2SO_4, and NH_3 partition between the gas and particle phase in a thermodynamic equilibrium, which is determined by ambient temperature and humidity. Based on the availability of SO_4^{2-} or NO_3^-, either $(NH_4)_2SO_4$ or NH_4NO_3 is then formed (Pay et al., 2012). Secondary formation pathways through heterogeneous aqueous reactions have been observed during haze conditions (Wang et al., 2012). Given an abundance of SO_2, NO_x, and NH_3, an uptake of these precursor gases is facilitated, allowing the formation of droplet-mode NO_3^- and SO_4^{2-}. The importance of this formation pathway for the total SIA formation rates increases with increasing relative humidity (RH). Given a combination of low temperatures and high RH, the droplet-mode formation of SIA can be the dominating formation pathway (Wang et al., 2012).

SOA are formed when VOCs such as isoprene, terpene, and aromatic species, are oxidized to form semi-volatile or low volatility organic species (Hoffmann et al., 1997; Carlton et al., 2007). The oxidized VOCs are transformed to particulate SOA either by condensation on existing (organic) particles or formation of new particles via nucleation (Hallquist et al., 2009; Perraud et al., 2012; Fuzzi et al., 2015). Gaseous VOCs are degraded via oxidation reactions with hydroxyl (OH) radicals, ozone (O_3), NO_3^- radicals or via photolysis. With progressing oxidation states, the carbon chains are fragmented. The formation of SOA is also known to be effective in aqueous environments (clouds, fog, or wet aerosols). Water-soluble organics are dissolved and undergo oxidation processes in the aqueous environment. After the subsequent evaporation of water, aqueous SOA particles are added to the atmospheric concentration (Blando and Turpin, 2000; Ervens et al., 2011). In contrast to SIA formation processes, SOA gas-particle transformation processes are still not well understood, owing to an estimated number of 10,000–100,000 different VOCs occurring in the atmosphere (Goldstein and Galbally, 2007; Hallquist et al., 2009).

Size distribution and particle characteristics on different scales

Typical size distributions of atmospheric particle numbers and particle mass of aerosol populations are shown in Fig. 1.2. Several modes of high particle numbers or particle concentrations exist, which are largely the consequence of the particle sources and secondary formation pathways presented in the previous section. Particles are shifted along the size distribution towards larger particle modes by coagulation of and condensation on existing particles or towards smaller particle modes by abrasion or evaporation. Although smaller particles occur in higher numbers, they make up only a small part of the particle mass (Fuzzi et al., 2015).
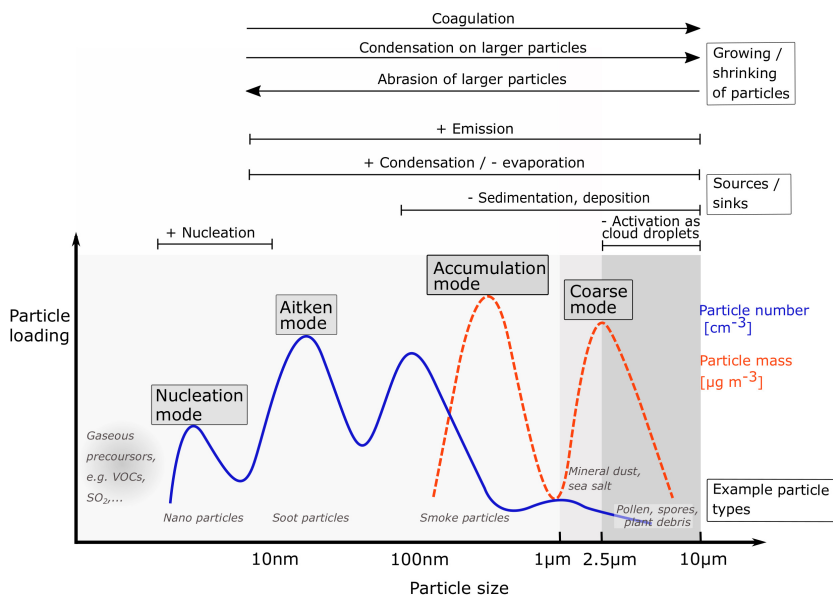
Figure 1.2: Schematic example of particle number and particle mass size distribution for a characteristic aerosol population. Size-dependent processes governing particle number and mass are outlined in the top part of the figure. Processes that cause particles to grow or shrink, thus shifting them along the size distribution axis are marked with an arrow. In the bottom part, example particle types are given. Adapted from Fuzzi et al. (2015) and Deutscher Wetterdienst (2020).

Ultrafine particles, i.e., particles with a size of a few nanometers, are formed in large numbers from condensable vapors such as H_2SO_4 or HNO_3, which start to form critical clusters and condense (Curtius, 2006; Hallquist et al., 2009; Fuzzi et al., 2015). These particles make up the nucleation mode (see Fig. 1.2). Newly-formed particles quickly grow by coagulation or condensation, hence the shift of the size distribution towards the Aitken mode (Curtius, 2006; Fuzzi et al., 2015). Further condensation of vapors (e.g., H_2SO_4, HNO_3, NH_3, VOCs) on existing particles effectively increases the particle mass (Hallquist et al., 2009). This way, particles from the smaller-sized modes eventually contribute to the accumulation mode. Particles from the larger-sized modes shrink by abrasion processes and can thus also contribute to the accumulation mode. In the coarse mode, the emission of primary particles dominates. For particles with a size of hundreds of nanometers or some micrometers, mass becomes increasingly important and deposition becomes the most important removal mechanism. Particles can be removed from the atmosphere by dry deposition, which requires particles to be transported towards the surface by atmospheric turbulence and gravitation. For particles to interact with the surface, they have to overcome a

quasi-laminar sub-layer that forms over all surface elements. For small particles (<100 nm), the only possible process to achieve this is by random motion, i.e., Brownian diffusion, while coarser particles overcome this layer by gravitation. The interception of particles with the surface mainly takes place on vegetation (Fuzzi et al., 2015). A further, more effective removal mechanism is the wet deposition, which includes rainout, washout, and nucleation scavenging. Rainout refers to in-cloud processes in which PM is incorporated. Washout describes the interception of particles by falling precipitation. Nucleation scavenging refers to the activation of particles as cloud condensation nuclei (Garland, 1978; Fowler et al., 2009; Fuzzi et al., 2015). Larger particles act as cloud droplet activators more effectively as they reduce the magnitude of supersaturation needed to form droplets (Andreae and Rosenfeld, 2008).

Following the particle mass distribution shown in Fig. 1.2, the majority of studies in this field use the indicators PM10, PM2.5, and PM1. PM10 represents the quasi-total particle mass and includes coarse (particle size between 2.5 μm and 10 μm) and fine particles (size <2.5 μm). On the contrary, the size fraction PM2.5 does not include the fraction of coarse particles. The size fraction PM1 includes only the finest particles.

Not only the size distribution, but also the distribution of PM concentrations in space is influenced by the occurrence of particle sources and formation processes. Lenschow et al. (2001) introduced a conceptual differentiation of regional background, urban increment and kerbside increment particles. The concept is based on the observation that secondary aerosols, which constitute the major fraction to the regional background, are relatively homogeneously distributed as their concentration is less source-dependant (see Fig. 1.3, Lenschow et al., 2001; Fuzzi et al., 2015). The urban increment (or urban background) is defined as the increase in concentrations in urban areas relative to its rural surroundings. At the kerbside, an additional increase due to high traffic emissions is to be expected (Lenschow et al., 2001). This concept has been widely accepted and also confirmed by measurements (Putaud et al., 2010). However, there is some evidence that the effect of the urban increment is rather low and that urban background concentrations often do not differ substantially from the regional background at least for European cities (see Crippa et al., 2013; Keuken et al., 2013; Petetin et al., 2014; Fuzzi et al., 2015). Still, an increase at the kerbside due to traffic emissions and dust resuspension results in impaired air quality (Fuzzi et al., 2015).

Expanding the concept of Lenschow et al. (2001) by the factor time allows to conceptually describe the processes altering the particle composition and their size distribution (see Fig. 1.3). Close to the source, particles maintain the corresponding source characteristics. With progressive residence time in the atmosphere, the transformation of particles tends to make them more homogeneous. This is because new particles that enter the atmosphere are coated by sulfates, secondary organics, nitrates, and therefore become increasingly similar, at least on the outside (Fuzzi et al., 2015). At local scales, both primary and secondary particles are important while at large spatiotemporal

scales, the fraction of secondary aerosols dominates as the importance of primary sources is decreased.
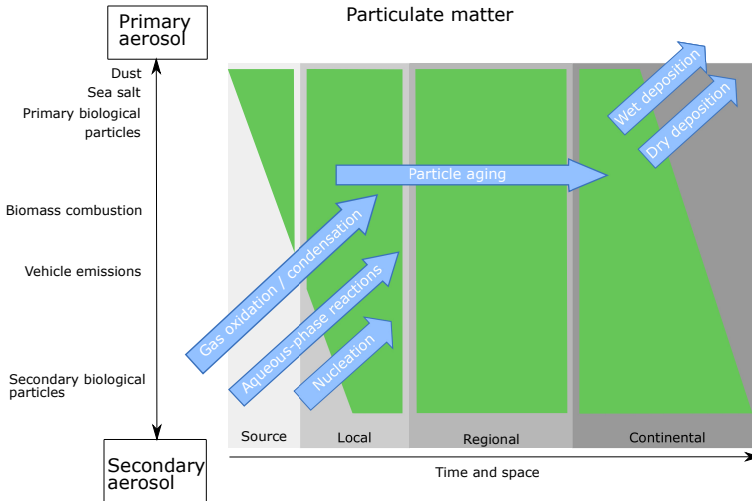
Figure 1.3: Processes affecting atmospheric PM composition on different scales. On the y-axis, the relative importance of primary (high importance on the top) and secondary particles (high importance at the bottom) is indicated and typical sources are given. Green reflects the importance of PM sources for the corresponding spatiotemporal scale (x-axis), while grey colors indicate the absence of PM sources. The grey shading reflects the increasing spatiotemporal scale. Relevant processes altering the particle composition are shown as blue arrows. Adapted from Fuzzi et al. (2015).

PM data gathered at more than 60 rural, urban, and kerbside sites over the European continent indicate that in general, the main constituents of both PM10 and PM2.5 are Org, SO_4^{2-}, and NO_3^-. Sea salt can also be a major constituent for the coarse PM fraction for stations closer to the oceans, while mineral dust can dominate at kerbside sites due to dust resuspension (Putaud et al., 2004, 2010). Moving from rural to urban to kerbside stations, the relative contributions of SO_4^{2-} and NO_3^- are decreasing, while the fraction of carbonaceous particles increases (Putaud et al., 2010). This observation follows the scale-dependant PM concentration patterns that were described previously, since carbonaceous particles are mainly associated with primary emissions close to the source (traffic and other combustion processes), while secondary particles dominate background concentrations due to the transformation processes in the atmosphere (Putaud et al., 2004). Typical PM annual average concentrations in central Europe for PM10 (PM2.5) at kerbside sites are $\sim$40–55 μg/m^3 ($\sim$22–39 μg/m^3). Concentrations of $\sim$25–45 μg/m^3 ($\sim$18–40 μg/m^3) are reported for urban stations, concentrations of $\sim$11–23 μg/m^3 ($\sim$8–20 μg/m^3) for rural stations.

Natural continental background concentrations are in the range of $\sim$5–8 μg/m^3 ($\sim$3–5 μg/m^3) (Putaud et al., 2004).

Environmental controls of PM concentrations

While atmospheric PM concentrations are largely governed by the interaction of the outlined sources, sinks, and atmospheric transformation processes, PM concentrations are also substantially influenced by meteorological conditions. These conditions impact the transport of particles and local-scale processes in the boundary layer, e.g., by changing the vertical distribution of particles in the atmosphere, the effectiveness of deposition processes, or the effectiveness of transformation processes (Cermak and Knutti, 2009; Rost et al., 2009; Pearce et al., 2011; Bressi et al., 2013; Megaritis et al., 2014; Dupont et al., 2016; Petäjä et al., 2016; Yang et al., 2016; Li et al., 2017). In addition, land surface characteristics can influence emissions of natural primary particles and precursor gases, as well as dry deposition rates (De Meij et al., 2015; Fuzzi et al., 2015; Bonn et al., 2016; Zhao et al., 2016). An overview of important environmental variables and how they impact PM concentrations is given in the following.

The boundary layer height (BLH) determines the height up to which particles are distributed in the atmosphere (Gupta and Christopher, 2009b; Schäfer et al., 2012). Limiting the vertical exchange of air masses, it regulates the vertical component of the volume of air in which particles typically disperse. A lower BLH typically increases near-ground PM concentrations as particles are not mixed into higher atmospheric levels and accumulate near the ground (Gupta and Christopher, 2009b; Schäfer et al., 2012; Liu et al., 2018), although some argue that BLH cannot be related directly to concentrations of pollutants and that other meteorological parameters and local sources need to be considered (Geiß et al., 2017). Higher BLHs in combination with high wind speeds increase atmospheric ventilation processes, thus decreasing near-surface particle concentrations (Sujatha et al., 2016; Wang et al., 2018). Wind generally regulates particle transport and turbulent mixing away from the surface (Li et al., 2017).

Depending on the location, air masses from certain wind directions can be affected by major emission sources and carry an increased amount of particles (Lenschow et al., 2001; Petetin et al., 2014; Li et al., 2015; Petit et al., 2015; Srivastava et al., 2018). Wind speed and direction also govern the type of particles in the atmosphere (Adamopoulos et al., 2007; Levy and Hanna, 2011; Chudnovsky et al., 2013a; Zheng et al., 2017). Air temperature can influence PM concentrations in multiple ways. While some authors suggest that PM concentrations generally decrease with increasing temperatures (Li et al., 2015; Wang and Ogawa, 2015), the influence of temperature on PM concentrations depends on the chemical composition of a PM population, as PM species react differently to changes in temperature (Megaritis et al., 2013, 2014). Air temperature modifies the emission of biological secondary PM precursors such as VOCs (Fowler et al., 2009; Megaritis

et al., 2013). Furthermore, increased oxidation rates of SO_2, which were found for rising air temperatures, again increase the SO_4^{2-} formation rates (Dawson et al., 2007). Decreasing temperatures, on the other hand, enhance the condensation of high saturation vapour pressure compounds, such as nitric acid and sulfuric acid, also leading to increasing formation of new particles (Hueglin et al., 2005; Pay et al., 2012; Bressi et al., 2013; Megaritis et al., 2014). Solar irradiation is a prerequisite for many transformation processes of SOA (Guenther et al., 1993; Petetin et al., 2014) and SIA particles (e.g., for the oxidation of SO_2 to form SO_4^{2-}, Dawson et al., 2007; Pay et al., 2012; Gen et al., 2019). Precipitation is know to remove particles (see section 1.2, Radke et al., 1980; Rost et al., 2009; Bressi et al., 2013; Li et al., 2015) and their gaseous precursors from the atmosphere (Megaritis et al., 2014). However, as mentioned previously, under increasingly wet atmospheric conditions, the importance of secondary particle formation through aqueous reactions increases (Ervens et al., 2011; Wang et al., 2012).

Furthermore, land-surface characteristics are an important control of PM concentrations. Various land cover types can act as sinks or sources for particles (Bonn et al., 2016; Churkina et al., 2017; Beloconi et al., 2018). For example, vegetation can both enhance (Churkina et al., 2017) and decrease PM concentrations (see Bonn et al., 2016), depending on the type of vegetation and other meteorological conditions. In addition, topography can be of importance for PM concentrations as particles tend to accumulate in valleys (Emili et al., 2011b). This is because stagnant conditions, which lead to the accumulation of particles in the atmosphere, are more likely to occur in basins (Wang et al., 2018).

These environmental conditions and processes have been identified as drivers of local PM concentrations, and it is the complex interaction of the physical and chemical processes related to the environmental conditions that govern atmospheric PM concentrations (e.g., Petit et al., 2015; Dupont et al., 2016).

1.3 In situ measurements and satellite-based approximation of PM concentrations

1.3.1 In situ measurements of PM

The analyses conducted in the scope of this thesis rely on in situ measurements of hourly mean concentrations of PM10 (chapters 2 and 3) and PM1 (chapter 4). In the following, a description of the principles behind these measurements is provided.

PM10 concentrations can be determined either by measuring the attenuation of β-radiation by a dust-coated filter (Umweltbundesamt, 2004) or by the principle of oscillating micro weighing (TÜV Rheinland, 2012). For the former measurement principle, a constant airflow is channeled through a strip of filter paper, on which particles accumulate. The particles are irradiated by β-

radiation from a radioactive source. The PM concentration is then determined by applying a linear regression on the change of the measured radiation signal divided by the time steps. A moving average is applied for 60 individual measurements to avoid noisy measurements. Filter strips are automatically changed when the maximum particle accumulation is reached (Umweltbundesamt, 2004). The principle of oscillating micro weighing is a continuous gravimetric procedure, i.e., it determines particle mass deposited on a filter. The weighing of the particle mass works by measuring the change in frequency of oscillation of a cantilever element within the detection system, which holds the filter. The particle mass is related to a predefined and continuous volume of sampled air, thereby the particle concentration is computed (TÜV Rheinland, 2012). To avoid condensation, particle inlets of the measurement devices are heated permanently. Thus, measurements are largely uninfluenced by temperature and humidity (Umweltbundesamt, 2004). The uncertainty of the measurements varies between measurement devices, but should not be larger than 25 % of the measured particle concentration (VDI, 2002; Bundesministerium der Justiz und für Verbraucherschutz, 2010). The locations of the measurement sites are chosen following EU prescriptions. Site locations are classified as background, industrial, or traffic and can be representative for urban, rural, or suburban areas. Traffic sites are generally situated relatively close to main roads or intersections, while industrial stations are close to strong emitters. Urban background sites are assumed to capture the contribution of all sources near the site without one particular source dominating. Suburban background stations need to be placed downwind (referring to the main wind direction) of emission sources. Rural background stations should not be influenced by agglomerations or industrial sites closer than five kilometres. All site locations exclude micro-scale effects (EU, 2008). The German PM monitoring network is maintained and quality-checked by the German federal environmental agencies and collected by the UBA.

Speciated PM1 concentrations can be determined with an aerosol chemical speciation monitor (ACSM). The ACSM provides continuous, near real-time and artifact-free measurements of the major chemical composition of non-refractory submicron PM1 (Ng et al., 2011; Petit et al., 2014, 2015). It allows differentiating PM in its chemical constituents, including organic compounds, NH_4^+, SO_4^{2-}, NO_3^-, and chloride (Cl^-). The ACSM sucks in air at a constant rate and larger particles are cut off. Non-refractory particles are then flash-vaporized and ionized. Ions are then detected by a mass spectrometer (Ng et al., 2011; Petit et al., 2014). Complementary to the ACSM, BC can be monitored with a seven-wavelength aethalometer (Petit et al., 2015), which measures light extinction on a collecting filter. A differentiation into black carbon from fossil fuels (BCff) and black carbon from wood burning (BCwd) is achieved using differences in the absorption spectrum (Favez et al., 2009; Sciare et al., 2010; Healy et al., 2012; Petit et al., 2014). For the data set used here, the consistency of ACSM and aethalometer measurements is cross-checked by comparing the sum of all monitored species with a nearby tapered element oscillating microbalance with a filter dynamic measurement system. While there is some inconsistency,

in particular at low PM1 concentrations, the measurements are considered to be robust (Petit et al., 2014). PM1 data used in this thesis is measured at the Site Instrumental de Recherche par Télédétection Atmosphérique (SIRTA) supersite in the framework of the Aerosols, Clouds and Trace gases Research Infrastructure (ACTRIS) project.

1.3.2 Remote sensing measurements of atmospheric aerosols

While in situ measurements provide data with high accuracy and high temporal resolution, they suffer from limited horizontal and vertical coverage and are time-, cost-, and labour-intensive in maintenance. By contrast, ground-based remote sensing techniques have the potential to provide information on the vertical distribution of particles within the atmosphere and allow continuous measurements with relatively little effort. Available instruments use both active and passive remote sensing techniques to approximate atmospheric PM concentrations. Active sensors provide their own source of energy, e.g., a lidar beam, while passive remote sensing sensors rely on the sun as an energy source. For example, a single-wavelength lidar signal is used by a ceilometer to provide information on the aerosol backscatter coefficient and to detect aerosol layers below $\sim$5 km height (Wiegner et al., 2014). Lidar pulses are emitted upwards in the near infrared (NIR) between $0.9\,\mu$m and $1.1\,\mu$m and the reflected light is measured by the instrument at a temporal resolution of up to $1\,\text{min}^{-1}$ (Wiegner et al., 2014). More advanced lidar-systems such as Raman lidars can derive information on aerosol properties by measuring the frequency shift of backscattered light after being affected by the lidar signal (Ansmann et al., 1990, 1992; Wiegner et al., 2014). Sun photometers, on the other hand, measure the atmospheric transmittance of solar radiation (Hoff et al., 2009). The instrument orientates itself towards the sun and measures the radiation at a certain wavelength (between $0.34\,\mu$m–$1.02\,\mu$m). The measured solar radiation is modified by the distance between the Earth and the sun, the optical air mass, the AOD, and gaseous absorption, which needs to be corrected for to derive an aerosol signal. In addition to direct sun measurements, measurements of the clear sky allow retrievals of the aerosol optical properties or size distributions. This measurement principle is used within the Aerosol Robotic Network (AERONET), which provides AOD measurements from automatically-operating sunphotometers on a global basis (Holben et al., 1998).

Active and passive satellite-based remote sensing instruments provide large-scale and spatially coherent estimates of the atmospheric particle characteristics, particle concentrations, and also on the vertical distribution of particles within the atmosphere. Since the launch of the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite in 2006, the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) instrument provides lidar measurements from space. CALIOP is a nadir-looking, two-wavelength lidar, which is sensitive to different polarization states. The instrument sends out light pulses at $0.532\,\mu$m and $1.064\,\mu$m wavelengths

simultaneously. The lidar backscatter is captured by a receiver telescope. Retrieved aerosol properties include aerosol layer height (base and top), aerosol extinction profile, AOD, and aerosol type. As the satellite is constantly moving forward, the retrieved signals provide a curtain-like snapshot of the troposphere (Winker et al., 2009). Aside from the active CALIOP sensor, a wide range of passive, satellite-based instruments conduct measurements mostly in the visible to NIR spectrum and potentially allow for an AOD retrieval. Some of these instruments are briefly presented here. The Moderate Resolution Imaging Spectroradiometer (MODIS) instruments onboard the NASA polar-orbiting twin-satellites Terra and Aqua make two measurements per day - one during the morning hours (Terra, local solar Equatorial crossing time at 10:30), and one during the afternoon (Aqua, local solar Equatorial crossing time at 1:30) (Sayer et al., 2014). The MODIS instrument provides observations with a broad swath width of 2330 km, enabling each instrument to provide near-global coverage of radiance measurements on a daily basis (Levy et al., 2007, 2013; van Donkelaar et al., 2016). The Multiangle Imaging Spectroradiometer (MISR) instrument onboard the Terra satellite has nine different viewing angles, which allows for the retrieval of aerosol properties. However, MISR has a smaller swath of ∼380 km and therefore takes about a week to achieve global coverage (Bruegge et al., 2007; Zhang and Reid, 2010; van Donkelaar et al., 2016). Complementary to MODIS and MISR retrievals, AOD retrievals are available based on measurements from the Visible Infrared Imaging Radiometer Suite (VIIRS, Hsu et al., 2019) and the Polarization and Directionality of the Earth's Reflectance radiometer (POLDER, Dubovik et al., 2019). Furthermore, AOD can be retrieved from sensors onboard geostationary platforms, e.g., from the Geostationary Operational Environmental Satellite (GOES)-West and GOES-East imagers (Zhang et al., 2013) or from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard the Meteosat Second Generation (MSG) satellites (Clerbaux et al., 2017). Recently, a merged AOD product based on multiple satellite-based AOD retrievals has been introduced by Sogacheva et al. (2020), with the goal to provide a long-term climatology of mean monthly AOD. The research conducted in this thesis relies on AOD retrieved from MODIS measurements. MODIS AOD retrievals have been shown to be of high quality when validated against AERONET (Lyapustin et al., 2018; Jethva et al., 2019) and offer a large spatial coverage with a continuous, almost 20 year long time series. Because both Aqua and Terra carry the MODIS instrument, two retrievals per day are possible over the same region resulting in a large and valid data basis. The concept of the AOD, the basics behind the MODIS AOD retrieval algorithms, and challenges when approximating PM based on satellite-based AOD will be discussed in the following sections.

1.3.3   AOD – concept, application and challenges for PM approximations

The concept of AOD

As solar radiation passes through the atmosphere, parts of its spectrum are attenuated along the way towards the satellite sensor by interaction with particles and gases. The simplest case of the radiative transfer in the visible spectrum is described by Beer's Law:

$$I(\lambda) = I_0(\lambda)e^{-\alpha(\lambda)z} \tag{1.1}$$

where I is the measured radiance at a specific wavelength λ after the passing of the source radiance I_0 (here, it refers to the radiation emitted by the sun) through the atmospheric path z. α is the extinction coefficient of the atmosphere $[m^{-1}]$. The optical depth τ of the atmosphere at a certain wavelength is the integral of the atmospheric extinction coefficient down to height z:

$$\tau_\lambda = \int_z^\infty \alpha(\lambda, z)dz \tag{1.2}$$

where z = 0 at the surface and τ = 0 at the top of the atmosphere (TOA). The AOD is the total columnar optical depth of the atmosphere corrected for absorption of gases in the atmosphere (e.g., water vapour or ozone) and hence accounts only for scattering and absorption by atmospheric particles (Hoff et al., 2009; Sogacheva et al., 2020). The transmittance of the atmosphere that can be derived from the difference between I and I_0 in Equ. 1.1, is directly related to the one-way optical depth (I_3 in Fig. 1.4) as measured by sunphotometers from the ground (Holben et al., 1998; Hoff et al., 2009). For a satellite to retrieve AOD, solar radiation scattered in the field of view of the satellite by aerosols (term I_2 as shown in Fig. 1.4) needs to be related to the reflected radiation from the Earth's surface (I_4), taking into account the attenuation of the radiance reflected from the surface after scattering and absorption in the atmosphere (I_5) (Hoff et al., 2009).
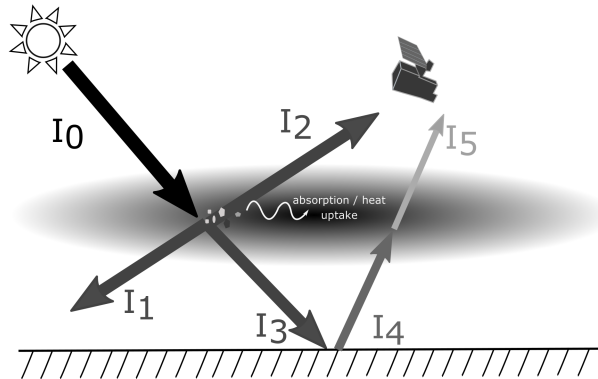
Figure 1.4: Important radiative processes and sources of radiation seen by a satellite. I_0 denotes the incoming solar radiation at the TOA, I_1 is the scattered radiation in the atmosphere away from the satellite, I_2 is the scattered radiation into the field of view of the satellite. I_3 is the radiation that is transmitted through the atmosphere towards the Earth's surface. I_4 denotes the reflected radiation from the Earth's surface, which is further attenuated by scattering and absorption in the atmosphere and then reaches the satellite (I_5). Adapted from Hoff et al. (2009).

Satellite-based AOD retrievals from MODIS measurements

Deducing AOD based on multispectral radiance signals of the satellite poses an inverse problem. It requires the radiance signal to be separated into atmospheric and surface contributions. The basic assumption is that in a clear, i.e., non-cloudy situation, the atmospherically backscattered solar radiation scattered by aerosols can be differentiated from the Earth's surface backscatter utilizing information on surface albedo in the retrieval algorithm (see Fig. 1.5, Levy et al., 2013). Therefore, assumptions need to be made about surface characteristics and the size- and shape-distribution of particles (Holzer-Popp et al., 2002; Hoff et al., 2009). In short, aerosol retrievals have to address the main tasks of surface reflectance determination, the modelling of aerosol microphysical and optical properties, atmospheric correction of gas absorption, cloud screening, and identification of snow- or ice-covered surfaces.

AOD retrieval algorithms usually focus on the measured reflectance in the visible range spectrum (0.4-0.7 μm) where the radiative transfer is dominated by aerosol extinction with only a few gases having absorption features. The wavelength-dependent extinction efficiency is highest for particles of size just below $\sim$1 μm, which coincides with the particle accumulation mode. Hence, a large mass of particles of this size is present and because Mie-scattering is strongest where particle size

and wavelength are equal (Hoff et al., 2009), the extinction is enhanced and the largest AOD signal can be expected (Hsu et al., 2013).
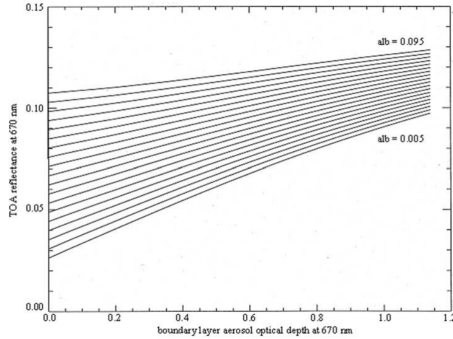
Figure 1.5: Influence of the surface albedo on TOA reflectances at 670 nm for different AODs. TOA reflectances at 670 nm are simulated for an aerosol population as a function of AOT at 670 nm for different surface albedo values ranging from 0.005 (lowest curve) to 0.095 (uppermost curve) with albedo steps of 0.005. Connecting values of TOA reflectance with boundary layer AOD strongly depends on the surface albedo (Holzer-Popp et al., 2002).

The inversion process from the multispectral signal measured by the satellite is approached based on precomputed, so-called look-up tables (LUTs). LUTs summarize simulated reflectances for a range of expected gases as well as aerosol sizes, forms, and optical properties and allow a quick approximation of the solution that fits best to the observed radiance. LUTs are calculated using radiative transfer algorithms with specified surface, aerosol, and atmospheric properties. Once a pixel is identified as cloud-free and passes initial quality checks, AOD is retrieved from the the multispectral radiances obtained by the satellite based on the LUTs, taking into the account solar zenith angle, viewing angle, and single scattering albedo (Levy et al., 2007; Hoff et al., 2009; Hsu et al., 2013; Levy et al., 2013; Sayer et al., 2014). Since different retrieval schemes are necessary depending on the Earth's surface parameter, various algorithms have been developed to retrieve AOD based on calibrated and geolocated MODIS radiance measurements (Levy et al., 2013).

The MODIS dark target (DT) algorithm, for example, retrieves AOD over dark land surfaces (e.g., vegetation, Kaufman et al., 1997; Levy et al., 2007, 2013) and over the oceans (Tanré et al., 1997; Levy et al., 2013). The DT algorithm over land builds upon spectral relationships between measured radiances in the shortwave infrared (SWIR) at $2.1\,\mu$m in the visible wavelengths at $0.47\,\mu$m and $0.66\,\mu$m, respectively (Levy et al., 2007; Sayer et al., 2014; Gupta et al., 2016). The surface reflectance can be derived from the SWIR channel, which is then correlated to radiances measured in the $0.47\,\mu$m and $0.66\,\mu$m channels. Applying a variable, angle- and surface-type dependent correlation factor, the surface reflectance in the visible can be derived

and separated from the atmospheric signal (Levy et al., 2007). After separating the surface from the atmosphere, TOA spectral reflectances are matched to location- and season-dependant aerosol models, representing atmospheric optical properties for a set of likely aerosol conditions as prescribed in the LUTs (Levy et al., 2013; Sayer et al., 2014). Various aerosol types are characterized by their size distribution, which in turn influences the spectral response. Aerosol types are categorized into several aerosol-type models, which may be fine-dominated or coarse-dominated (Levy et al., 2013). The DT algorithm only works over dark surfaces (i.e., low reflectance in the visible wavelengths), where a stark contrast to the aerosol signal is given (Levy et al., 2013) and the spectral relationship between the SWIR and visible channels is valid (Levy et al., 2007).

The deep blue (DB) algorithm is used to retrieve AOD over bright land surfaces, i.e., surfaces with high reflectances in the visible wavelengths (Sayer et al., 2014). The DB algorithm uses, amongst others, the near-ultraviolet band at $0.41\,\mu$m wavelength, where the reflectance of the bright surfaces (relating to wavelengths in the visible range) is actually relatively low, enabling the detection of an aerosol retrieval signal (Hsu et al., 2013; Levy et al., 2013). For the latest overhaul of the DB algorithm, a hybrid approach was invented to determine the surface reflectance. It is either determined by applying a geolocation-based static surface reflectance database (e.g., over bright desert areas), by a dynamic surface reflectance method including information on vegetation (also over areas with changing vegetation cover), or by a combination of the former two methods (e.g., over urban and other built-up areas, Hsu et al., 2013).

In this thesis, AOD data retrieved by the multi-angle implementation of atmospheric correction (MAIAC) algorithm based on MODIS measurements are used (Lyapustin et al., 2011a,b, 2012, 2018). The MAIAC algorithm builds upon concepts used in the DT and DB algorithms, but has a different workflow. First, MODIS level 1 data (i.e., calibrated and geolocated reflectances, Levy et al., 2007) are regridded to a fixed, pseudo-geostationary 1 km grid. Before the actual AOD retrieval, the columnar atmospheric water vapour content is determined and corrected for using the MODIS NIR ($0.94\,\mu$m), and a cloud-screening process is executed (Lyapustin et al., 2018). Based on the fixed grid, a time series of each pixel allows to observe the same grid cell over time, using multi-angle observations from different orbits. This procedure improves the discrimination of atmospheric and surface contributions and enables an accurate computation of the surface bidirectional reflectance distribution function (BRDF), which describes the anisotropic spectral surface reflectance (Shell, 2004; Lyapustin et al., 2018). The calculation of the AOD relies on the assumption that the surface BRDF remains largely constant over time (considering a time series of 16 consecutive days), allowing the MAIAC algorithm to more accurately approximate the radiatve transfer through the atmosphere (Lyapustin et al., 2011b). MAIAC determines the BRDF in the MODIS SWIR channel ($2.1\,\mu$m), which is normally relatively transparent in non-cloudy atmospheric situations. Assuming a similar-shaped BRDF between the MODIS SWIR and

blue bands (0.47 μm), the BRDF for the blue band is determined based on the SWIR band using so-called spectral regression coefficients (SRCs), which dynamically describe the spectral ratios between the MODIS SWIR and blue bands (Lyapustin et al., 2011b). The most recent version of MAIAC determines the SRCs using the minimum ratio of surface reflectance for the blue and SWIR bands for a 2-month period. The idea behind this approach is that unaccounted aerosols in the blue band would increase the reflectance, whereas most surfaces are dark at this wavelength. Thus, the minimum ratio is assumed to represent the most reliable SRC estimate (Lyapustin et al., 2018). AOD at 0.47 μm and 0.55 μm wavelength is then determined by matching measured reflectance with precomputed LUT-values. MAIAC LUTs are based on regional AERONET AOD climatologies (Holben et al., 1998) and a seperate dust model (Lyapustin et al., 2018). AOD retrievals based on MAIAC provide a high-resolution (1x1 km^2 pixel size) and robust estimate of the atmospheric columnar aerosol content. The high resolution of AOD retrievals is valuable as the high-resolution AOD is representative for a smaller area and therefore more strongly correlated to in situ PM measurements, which represent local conditions. Validation studies also have shown that MAIAC can provide increased coverage (see Fig. 1.6, Superczynski et al., 2017) and has comparable or better accuracy when compared to the DT and DB algorithms (Lyapustin et al., 2018; Jethva et al., 2019). For these reasons, MAIAC AOD retrievals are used for the analyses of this thesis.
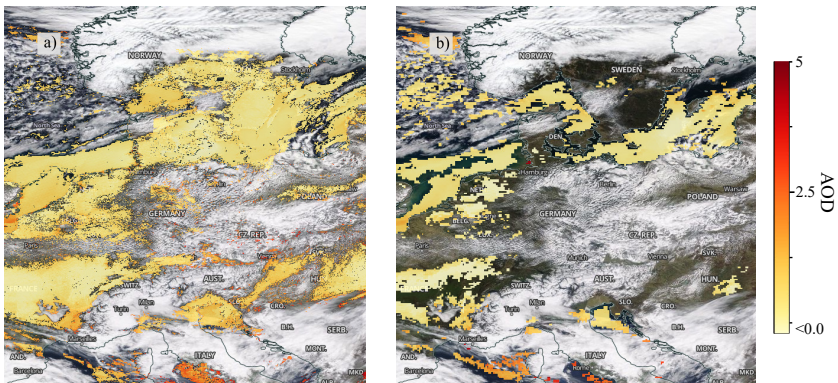
Figure 1.6: *Comparison of AOD retrieved by MAIAC (left panel) versus AOD retrieved by both DB and DT (right panel) at 0.47 μm on March 31st, 2020 (NASA, 2020). The MAIAC retrieval has a higher resolution of 1 km compared to the AOD retrieved by DB and DT, which has a resolution of 10 km. MAIAC AOD covers most of the cloud-free areas over Switzerland, the Benelux and northeast France, where the DB and DT algorithms do not retrieve AOD.*

Approximating near-ground PM concentrations based on satellite AOD

The AOD provides a measure of the atmospheric aerosol loading and can be used to estimate PM concentrations near the ground, which is the most relevant quantity concerning harmful effects on humans. However, the use of AOD to approximate PM concentrations at ground level presents several challenges. As outlined in the previous sections, AOD retrievals from satellite-based measurements and in situ measurements of PM concentrations represent different scales and different types of measurements. AOD reflects the extinction of radiation by aerosols in a vertical column influenced by ambient atmospheric conditions, while PM measurements reflect a localized dry mass concentration of particles of a certain size distribution measured near ground under relatively stable conditions (see Wang, 2003). Generally, it is known that ambient weather conditions can exert a strong influence on the relationship between AOD and PM, superimposing small-scale effects that occur at PM measurement sites (Gupta and Christopher, 2009b). The relationship between these two entities is therefore highly variable and often non-linear (Wang, 2003; Barnaba et al., 2010; Kloog et al., 2011; Chudnovsky et al., 2013b). As a consequence, relying on satellite-based AOD observations as proxy for near-ground air pollution can be misleading in some situations. On a seasonal scale, the relationship can even be inverse (Koelemeijer et al., 2006; Arvani et al., 2016). This is illustrated in Fig. 1.7, which shows monthly distributions of measured PM10 concentrations near ground and MAIAC AOD for the Po Valley, Italy. The yearly courses of PM10 and AOD differ considerably: PM10 concentrations peak in winter, while retrieved AOD is highest during summer months. The main reason for these contrasting seasonal cycles is the varying vertical distribution of aerosols within the atmosphere: in situations with a high boundary layer, particles dissipate in the atmosphere, decreasing PM concentrations near ground. During low-BLH-situations, vertical exchange is limited and particles are confined in the lower atmosphere, which leads to an increase in near-ground PM concentrations. BLH tends to be higher in summer due to pronounced turbulence (Koelemeijer et al., 2006; Gupta and Christopher, 2009b; Geiß et al., 2017).

21
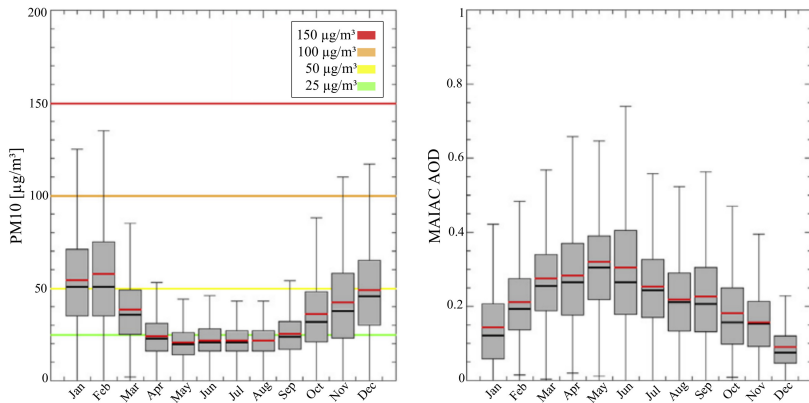
Figure 1.7: PM10 (left panel) and MAIAC AOD (right panel) monthly trends for the PO valley, Italy. PM10 concentrations are based on data from the Italian Regional Agency for Environmental Protection (ARPA) monitoring stations (Arvani et al., 2016). Boxplot whiskers show the 10th and 90th percentile, respectively. Outliers out of this range are not shown for visual clarity. The light grey box shows the 25th to 50th percentile. The median (mean) is shown as black (red) horizontal lines. The yearly course of PM10 and AOD is inverse: while PM10 concentrations peak in winter, retrieved AOD is highest during summer months. Adapted from Arvani et al. (2016).

Therefore, information about the vertical distribution of aerosols, approximated for example by the BLH, is necessary for understanding variations in the relationship between AOD and PM (van Donkelaar et al., 2006; Gupta et al., 2006; Boyouk et al., 2010). Furthermore, moisture in the atmosphere, e.g., approximated by ambient RH measurements, exerts influence on the AOD. With increasing RH levels, hygroscopic growth of particles (aerosol swelling) is triggered, which enhances the scattering coefficient and consequently increases the AOD (Wang, 2003; Gupta et al., 2006; Gupta and Christopher, 2009b). The magnitude of hygroscopic growth depends on the size and chemical species of particles (Wang and Martin, 2007; Zieger et al., 2013; Titos et al., 2014). Because in situ PM measurements usually refer to dry conditions, they are not directly affected by ambient relative humidity (see section 1.3.1). Furthermore, the source of emissions changes with wind direction, influencing the relationship between AOD and PM by determining important particle characteristics such as the size distribution or the hygroscopy (Zieger et al., 2013).

Since the publication of MODIS AOD products, widespread efforts have been made to analyse and apply systematic links between satellite AOD and PM concentrations, contributing to a growing body of literature (see Wang, 2003; Engel-Cox et al., 2004; Wang and Martin, 2007; Gupta and Christopher, 2009a,b; van Donkelaar et al., 2010; Chudnovsky et al., 2013b; Stafoggia et al., 2020). A large variety of methods has been applied for this task. In Koelemeijer et al. (2006), for example, a physical description of AOD as a function of RH, idealized aerosol extinction coefficients, and

BLH is proposed, assuming the latter is well-mixed, aerosols have similar optical properties and the sky is cloud-free:

$$AOD = PM_{2.5} \, BLH \, f(RH) \frac{3Q_{ext,dry}}{4\rho r_{eff}} \tag{1.3}$$

where f(RH) is the ratio of ambient and dry extinction coefficients, ρ is the aerosol mass density, $Q_{ext,dry}$ is the dry Mie extinction efficiency, and r_{eff} is the particle effective radius. The results by Koelemeijer et al. (2006) are encouraging with a spatial correlation (R) between yearly average PM10 and AOT of 0.6 for European rural background stations. The approach has been further advanced by Emili et al. (2010). Still, it requires many assumptions and produces reliable PM estimates only if a longer period is analysed, e.g., monthly mean PM concentrations. While Koelemeijer et al. (2006) based their study on previously conducted measurements of aerosol characteristics, a study by van Donkelaar et al. (2010) used a chemical transport model (CTM) to retrieve information on aerosol size, aerosol type, RH, and the vertical distribution of the aerosol extinction. The information was then used to calculate a local conversion factor that relates AOD to PM2.5 on a global basis. This approach has been further extended including additional satellite retrievals in van Donkelaar et al. (2016). A common approach is to first establish statistically-derived, empirical relationships between AOD and PM and then to extend these relationships to areas not covered by PM measurements. This way, estimates of PM based on AOD can be produced. Several studies have trained statistical models on the relationship between AOD and PM, accounting for a range of additional parameters (see reviews by Rybarczyk and Zalakeviciute, 2018; Shin et al., 2020). Methods include linear regression models (Arvani et al., 2016), multiple-additive regression models (Gupta and Christopher, 2009b; Zhang et al., 2018), land-use regression models (Kloog et al., 2011; Nordio et al., 2013) or a combination of the latter two (Kloog et al., 2012; Chudnovsky et al., 2014; Hu et al., 2014; Lee, 2019). With increasing data availability and computational power, machine learning (ML) methods, e.g., artificial neural networks (Gupta and Christopher, 2009a; Di et al., 2016), random forest (RF, Brokamp et al., 2018; Chen et al., 2018a; Grange et al., 2018; Stafoggia et al., 2020), or a combination of multiple ML methods (Yazdi et al., 2020) have been successfully produced PM estimates in recent years. ML models proved to be beneficial as they efficiently reproduce the variable relationship between AOD and PM and interactions of input features (Elith et al., 2008; Brokamp et al., 2017) and can thus produce high-accuracy estimates of PM concentrations, given the input data are of adequate quality.

1.4 Overview of applied data sets

The analyses conducted in the context of this thesis span different scales, ranging from regional (Germany, chapter 3), sub-regional (parts of Germany, chapter 2), to local scales (suburban

Paris, chapter 4). Therefore, different requirements are placed on the applied data sets. For the sub-regional and regional scale analyses, data from satellite sensors (MODIS, VIIRS), in situ measurements from the German Meteorological service (Deutscher Wetterdienst, DWD) and the UBA, and ECMWF Re-Analysis (ERA) data are assimilated and collocated in time and space. The local scale analysis focuses on in situ measurements from the SIRTA supersite. In general, the choice of the data sets is motivated by the data accuracy, its spatiotemporal coverage, and by the addressed research questions. An overview of the data sets used is given in Tab. 1.2 while further information on the meteorological and land surface data are provided in the respective chapters.

Table 1.2: Overview of data sets used in this thesis. See also Tab. 3.1 for more information on the data basis of chapter 3. NDVI stands for normalized difference vegetation index, CLC stands for coordination of information on the environment land cover, EEA stands for European Environment Agency, DEM stands for digital elevation model.

Data set	Parameter	Chapter
MODIS MAIAC	AOD	2, 3
MODIS	NDVI	3
DWD meteorological measurements	Air pressure, RH Continentality factor, precipitation	2, 3
Copernicus CLC	Land cover types	3
EEA EU-DEM, v1.1	Elevation	3
ERA-Interim	Wind vectors u & v, BLH, temperature, convective available potential energy Surface solar radiation downwards	2, 3
EEA emission data base	Annual mean emission of NH_3, PM10, SO_2, NO_x	3
VIIRS earth at night	Lights at night	3
UBA air quality measurements	PM10	2, 3
SIRTA air quality measurements	Speciated PM1	4
SIRTA meteorological measurements	Temperature, RH, BLH air pressure, wind vectors u & v, precipitation, total incoming solar radiation	4

1.5 Scientific aims and outline

As outlined in section 1.3, the application of AOD for air quality studies is promising and provides information on the atmospheric particle loading with large spatial coverage and long, continuous time series (Gupta et al., 2016; de Hoogh et al., 2018; Lyapustin et al., 2018), also in remote areas with no ground measurements available (Edwin and Mölders, 2020). However, there are still challenges in current techniques to estimate PM from AOD. Therefore, the quantification

of the influence of meteorological conditions on the relationship between AOD and PM10 is targeted in this thesis. Furthermore, it is known that environmental factors are major drivers of PM concentrations and need to be considered for reliable PM estimates (see e.g., Pearce et al., 2011; Megaritis et al., 2014; De Meij et al., 2015; Li et al., 2017). The determination of the type and magnitude of the influence of important environmental factors on atmospheric PM concentrations is a further objective of this thesis. Finally, atmospheric processes governing PM concentrations will be analysed with a particular focus on the development of high-pollution episodes. These are of particular scientific and public interest, as even short phases of exposure to elevated particle concentrations can have damaging effects on the human health (Stieb et al., 2017; Hughes et al., 2018). The outlined objectives are covered by three research chapters:

(1) Meteorological influences on the relationship between PM10 and AOD

(2) Mapping and understanding PM10 patterns using AOD and environmental conditions

(3) Meteorology-driven, local-scale variability of PM1

The research questions and the underlying hypotheses for each research chapter are presented in the following and also shown in Fig. 1.8.

(1) Meteorological influences on the relationship between PM10 and AOD

Possible influences on the relationship between AOD and PM are described in chapter 1.3.3, along with physical and empirical approaches to derive PM concentrations based on AOD. Although the influence of BLH (Boyouk et al., 2010; Emili et al., 2010; Arvani et al., 2016) and RH (Arvani et al., 2016; Dupont et al., 2016; Zheng et al., 2017) is largely established, current literature lacks a quantitative understanding of how meteorological conditions influences the relationship and in which situations AOD is a reliable proxy for PM. Gupta et al. (2006) and Zheng et al. (2017), for example, reported a strong linear correlation between AOD and PM2.5 for low-BLH-situations, while Chudnovsky et al. (2013a) found low correlations between AOD and PM2.5 for similar conditions. In Arvani et al. (2016), the authors reported that correcting for different BLHs substantially improves the correlation between AOD and PM10. Concerning the influence of RH, Emili et al. (2010) and Chudnovsky et al. (2013a) did not find a significant influence on the relationship between AOD and PM. Others reported that water uptake by particles can lead to hygroscopic growth, thus increasing AOD (see Barnaba et al., 2010; Crumeyrolle et al., 2014; Dupont et al., 2016). Further studies indicated an influence of wind speed (Zheng et al., 2017) and wind direction (Zieger et al., 2013) on the relationship between AOD and PM. An advancement of the scientific understanding of influences on the relationship between instantaneous AOD and

hourly mean PM10 concentrations is targeted and motivated by the main research question:

How do meteorological conditions influence the statistical relationship between AOD and PM?
The guiding hypothesis is that the relationship between AOD and PM varies with respect to hygroscopic particle growth and the vertical distribution of particles in the atmosphere. Both of these aspects are largely governed by ambient meteorological conditions, such as BLH, RH, and wind.

The main research question gives rise to the following, more specific research questions:

- *What are typical numerical characteristics of the relationship between AOD and PM?*

- *What meteorological conditions have an increased probability to cause an overestimation or underestimation of PM by AOD?*

Agreement or divergence of PM10 and AOD measurements cannot be determined directly, as the variables represent different quantities. Therefore, a semi-quantitative, comparative index is introduced, which normalizes AOD and PM10 values before comparison and enables an assessment of whether AOD and PM10 agree. Situations in which these parameters are of similar magnitude can be identified under consideration of ambient meteorological conditions. This way, the capability of satellite AOD observations to map and monitor surface PM concentrations is assessed and a basis is set to increase the accuracy of PM estimations from AOD. The study area covers a sub-regional scale ($\sim 12{,}000\,\mathrm{km}^2$) around the cities of Berlin and Potsdam in northeastern Germany. Choosing a relatively small study area helps to put the focus on meteorological drivers, as the area is expected to be relatively homogeneous in terms of terrain and climatology. Satellite-based AOD retrievals, PM observations, and meteorological data from reanalysis and ground measurements are used for the analysis.

(2) Mapping and understanding PM10 patterns using AOD and environmental conditions

Building upon the scientific understanding gained within the first research chapter, empirical links between AOD and environmental conditions are established to map hourly PM10 concentrations using a multivariate statistical modelling framework. Besides meteorological parameters, the considered parameters used to predict PM10 concentrations cover characteristics of the land surface, which have been found to influence PM concentrations in previous studies (Bonn et al., 2016; Beloconi et al., 2018). Sensitivity analysis conducted within the statistical modelling framework allow identifying and quantifying important drivers of PM10 concentrations. In the same modelling setup, the capability of satellite AOD observations for air quality studies is

assessed. The research is conducted following the main research question:

What are the main drivers of PM10 when environmental conditions and AOD are used to estimate PM10 concentrations?

The guiding hypothesis is that using a statistical modelling framework, hourly PM10 concentrations can be estimated from satellite AOD retrievals with high accuracy if land surface characteristics in the vicinity of the PM measurement sites and meteorological conditions at the time of the AOD acquisition are taken into account. It is hypothesized that changes in these environmental conditions drive variations of PM10 concentrations.

On that basis, the following specific questions are addressed:

- *In such a multivariate statistical prediction framework, are meteorological or land surface parameters of higher importance to determine PM10 concentrations?*

- *How does satellite-retrieved AOD interact with other parameters in the multivariate statistical prediction framework?*

While previous studies showed that ML models are powerful tools for the prediction of PM concentrations from AOD under consideration of environmental conditions, (van Donkelaar et al., 2010; Kloog et al., 2015; Hu et al., 2017; Stafoggia et al., 2017; Zhang et al., 2018; Stafoggia et al., 2020), the potential of these models to provide insights into physical processes involved has been largely neglected. Only recently, studies have approached this idea for air quality surveys (Hu et al., 2017; Grange et al., 2018; Park et al., 2019), without fully exploiting its potential. Here, the multivariate influences on the statistical model are quantitatively analysed by isolating the influence of individual parameters (see Cermak and Knutti, 2009; Andersen et al., 2017; Fuchs et al., 2018). The study merges satellite-based AOD retrievals, measurements of the German PM-network, meteorological measurements, reanalysis data, and other satellite measurements of land-surface characteristics.

(3) Meteorology-driven, local-scale variability of PM1

As pointed out in previous studies, not only anthropogenic emissions but also meteorological conditions exert a strong influence on PM concentrations (see also section 1.2 Rost et al., 2009; Pearce et al., 2011; Bressi et al., 2013; Megaritis et al., 2014). Still, much remains unknown concerning the meteorological processes driving PM concentrations because an analysis of influences of individual meteorological parameters is complicated due to their highly interconnected nature and also their synchronous effect on various, partly contradicting processes (e.g., temperature might influence formation as well as destruction of particles). This is particularly

true for high pollution situations, which can vary substantially in terms of timing, processes, and sources (Petit et al., 2017). It is the objective of this research chapter to quantify the contribution of critical atmospheric variables to day-to-day variations of PM concentrations at a local scale, i.e., without the use of AOD. The focus is on particles in the fine mode fraction (PM1) with individual analyses for different chemical constituents. By excluding the coarse particle fraction, transformation processes and new particle formation processes can be analysed with higher accuracy. The following main research question is targeted:

How do atmospheric processes drive concentrations of different PM1 species at a local scale?
The guiding hypothesis is that particle transport from other regions and local-scale processes in the boundary layer govern the main fraction of variations of local PM1 concentrations. The amount of advected particles is determined by the particle loading of air masses, which can be approximated by the direction of inflow. Changes of local meteorological conditions, e.g., temperature, BLH, and humidity initiate processes pertaining to particle formation, particle accumulation, or changes in natural emissions. It is further hypothesized that different PM1 species show varying responses to changes in meteorological conditions.

Specific questions addressed in this study are:

- *What are season-specific characteristics of high-pollution situations?*

- *How do different meteorological parameters contribute to the development of specific high-pollution situations?*

Mean afternoon PM1 concentrations from the SIRTA supersite southwest of Paris (Haeffelin et al., 2005) are analysed. Data from an ACSM and an aethalometer are combined with meteorological measurements from the same site, allowing processes to be resolved at a high degree of detail. The study area has been chosen as the metropolitan area of Paris is one of the most densely populated and industrialised areas in Europe and air quality is a recurring issue. Several studies focused on this area in the past years (Bressi et al., 2014; Petetin et al., 2014; Petit et al., 2015; Dupont et al., 2016; Petit et al., 2017; Srivastava et al., 2018). Based on the results of these studies, a better scientific understanding of the complex and varying mechanisms leading to high pollution events is approached. Valuable insights are generated that help to tackle the problem of harmful levels of PM for the Paris region and other areas suffering from severe air pollution. Building upon the ideas presented in the previous research chapter, the research questions are approached using explainable ML techniques. A further statistical model is set up, linking meteorological conditions with hourly mean PM1 species concentrations. Given that the statistical model explains a satisfying amount of occurring variability of PM1 concentrations (see Grange et al., 2018), empirical patterns extracted

from the statistical model can be related to physical processes. High-resolution case studies are conducted by quantifying the contribution of the meteorological variables to high pollution situations using SHapley Additive exPlanation (SHAP) regression values (Lundberg and Lee, 2017; Lundberg et al., 2020).
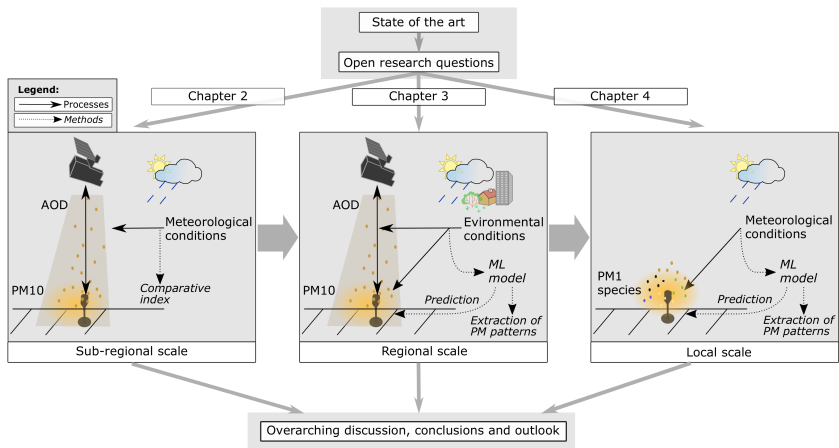
Figure 1.8: Schematic illustration of the three research chapters of this thesis. The first research chapter analyses meteorological influences the relationship between satellite-based AOD (grey box in the background) and near-ground PM concentrations (brown region). Building upon the insights gained in the first research chapter, the second research chapter establishes links between AOD and PM taking into account meteorological and land surface parameters on a regional scale. The third research chapter provides insights on atmospheric controls of PM at a local level without including AOD observations.

The research questions and underlying hypotheses outlined here constitute the frame of this thesis. The main structure of this thesis, the hypotheses, and the methodology are schematically illustrated in Fig. 1.8. The studies presented in chapters 2–4 are set up to test the hypotheses and answer the open research questions. An overarching discussion in chapter 5 merges the results of the individual research chapters. The research presented in this thesis has already been published or is ready for peer-reviewing in scientific journals. The original publications are attached at the end of this thesis.

2 Meteorological influences on the relationship between PM10 and AOD

2.1 On the need for a quantitative analysis of conditions for deriving PM from AOD

The availability of accurate, extensive, and spatially continuous information on PM concentrations is important for the development of measures against air pollution, epidemiological exposure studies, and the information of the public. However, this information is not readily available as networks of PM monitoring stations are not dense enough to, e.g., reflect interactions between urban areas and their surroundings. In addition, measuring stations are much more frequent in urban areas than in rural areas (see Fig. 1.1), so that there is often no estimation of PM exposure for rural areas. In this regard, satellite-based AOD has the potential to fill in the blanks.

The challenges of deriving ground-level PM concentrations based on AOD retrieved from satellite measurements have been highlighted in chapter 1.3.3, and include, amongst others, the unknown vertical distribution of particles in the atmosphere (van Donkelaar et al., 2006; Gupta et al., 2006; Boyouk et al., 2010) and the hygroscopic growth of particles (Crumeyrolle et al., 2014). AOD has frequently been used as a proxy for PM with varying success (Gupta et al., 2006; Kloog et al., 2011; Nordio et al., 2013; Arvani et al., 2016; Stafoggia et al., 2017). Studies that analysed the relationship between AOD and PM pointed out the importance of including RH (Wang, 2003; Gupta et al., 2006; Wang and Martin, 2007) and BLH (Gupta et al., 2006; Zheng et al., 2017) when approximating PM based on AOD. Still, a thorough and detailed treatment of the specific roles of meteorological conditions in determining the relationship between AOD and PM is missing. Accordingly, the aim of this study is to analyse meteorological influences on the relationship between AOD and PM, with the ultimate goal of constraining conditions under which satellite AOD observations can be used to reliably approximate PM. These conditions are contrasted with conditions that have a higher probability to cause an underestimation or overestimation of ground PM concentrations. Knowledge of these conditions could potentially improve the accuracy of air quality estimates based on AOD, in particular over remote areas that lack a dense network of in situ PM measurement stations. Using MAIAC AOD and PM10 measurements in combination with ERA-Interim reanalysis and station data from the DWD, numerous collocated data pairs are analysed in an area around Berlin, Germany. A novel approach is introduced, which includes a

monthly ranking of AOD and PM10 observations for the period 2001–2015. Ranked AOD/PM10 data pairs are linked to surrounding meteorological parameters. This allows to characterise the relationship between AOD and PM10 for each data pair individually and to perform in-depth analyses.

2.2   Study domain, data and methods

Study domain

The study focuses on a rectangular area ($\sim$12,000 km^2) around the cities of Berlin and Potsdam in northeastern Germany. The area Berlin is chosen as a study site because of its continental location in flat terrain and excellent data availability. It is dominated by the cities in its centre and is surrounded by agricultural land and large vegetated areas to the north and south (see Fig. 2.1). A great number of PM10 measurements in the city area and its surroundings are available. The area is mainly characterized by continental climate with only sparse, episodic maritime influence when advection of air masses from northern and northwestern directions occurs. Thus, the area is suitable for studying the influence of wind direction to discriminate between these regimes.

Figure 2.1: Geographical extent of the study area. Different land cover classes are colored, positions of meteorological stations and PM10-measurement sites are indicated.
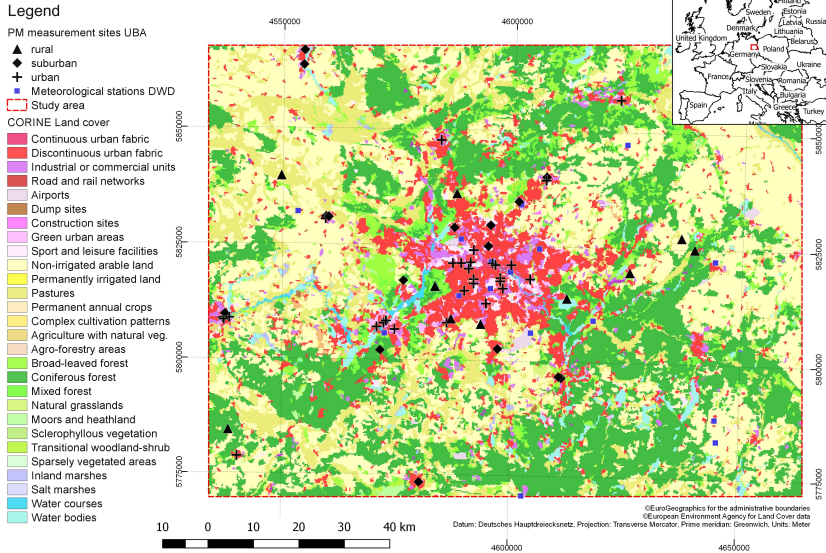
Local PM10 sources dominate in the region, with traffic the major contributor. Vegetated areas surrounding Berlin influence the mixture of particles in and around the urban area. Agricultural areas can have pollution levels similar to the urban background (Bonn et al., 2016). It has been argued that particle transport from the east can markedly contribute to overall pollution levels (Kerschbaumer, 2007; Bonn et al., 2016). In general, two different air mass regimes can be discriminated. Western air masses tend to transport lower particle concentrations, whereas continental, eastern air masses accumulate particles and thus carry elevated particle concentrations (Zieger et al., 2014).

Materials

PM-measurements

PM10 is analysed in this study instead of PM2.5 due to a greater availability of data for the study region. PM10 measurements might be more adequate for the analysis, as larger particles are less easily transported and more representative of local conditions. Additionally, PM2.5 excludes the

fraction of larger particles, which are nevertheless accountable for light extinction and are thus measured by the satellite. Hourly mean concentrations [$\mu g/m^3$] of PM10 measured by the UBA are used.

The locations of PM measurement sites and station types are shown in Fig. 2.1. Traffic sites, which are generally close to main roads or busy intersections (EU, 2008), are excluded from the analysis as they are by design not representative for conditions beyond the specific traffic situation and thus the area covered by a satellite pixel.

Satellite AOD data

High-resolution AOD computed by application of the MAIAC technique is used (Lyapustin et al., 2011a,b, 2012, 2018). Uncertainty estimates as included in the product are applied for filtering here. Pixels with an uncertainty larger 0.4 were discarded. For valid daytime AOD data used in this study, Aqua and Terra satellite overpass times range from ~ 9 am UTC to ~ 1 pm UTC.

Meteorological data

To represent regional-scale meteorological conditions, ERA-Interim reanalysis data generated by the European Centre for Medium-Range Weather Forecasts (ECMWF) are used with a resolution of 0.125° (Dee et al., 2011). These include BLH [m] and wind speed [m/s] in u and v direction (10 m height). Wind speed and wind direction [°] are calculated using u and v wind components. BLH data are employed as a proxy for the vertical concentration of aerosols in the lower troposphere, assuming particles are well-mixed within the boundary layer (Ansmann et al., 2000; Gupta and Christopher, 2009b) and AOD mostly represents attenuation in the boundary layer (Schäfer et al., 2008). Depending on the closest satellite overpass time, ERA-Interim data at 9 am UTC, 12 noon UTC or 3 pm UTC are used.

Ground measurements were obtained from the DWD. One hour mean values of RH [%] measured in 2 m height are used (DWD Climate Data Center (CDC), 2017). The locations of DWD measurement sites are given in Fig. 2.1 as blue dots. This study uses near-ground RH measurements, as these are more widely available than measurements of the vertical distribution of water vapour. Under dry conditions (RH < 55 %), near-ground RH measurements show reasonable agreement with vertically resolved measurements. This might not be the case at higher RH (Crumeyrolle et al., 2014).

Methods

UBA PM station coordinates are used as spatial reference, i.e., AOD pixels, ERA-Interim pixels and meteorological station data coordinates are collocated with the position of these stations. Only

the nearest pixels below a threshold of 0.015° ($\sim$ 1 km) distance from the PM measurement were collocated with the corresponding closest UBA station. The relatively small distance threshold was chosen to ensure that the MAIAC AOD reflects the situation in close vicinity of the UBA station. Terra/Aqua satellite overpass time is used as temporal reference for the PM10 and meteorological data, i.e., values closest to the overpass time are used. The work flow of this study is summarized in Fig. 2.2.

Figure 2.2: Overview of the work flow of this study. Methods are shown in curved boxes, inputs and outputs in rectangular boxes.

For the analysis of factors influencing the relationship between AOD and PM10, this study introduces the air quality rank difference index (AiRDI). The AiRDI allows for a direct, semi-quantitative assessment of whether AOD and PM10 agree, or whether the satellite qualitatively over- or underestimates PM10. Accordingly, situations in which these parameters are of similar magnitude can be identified systematically. The AiRDI is dimensionless and is computed for each spatially and temporally collocated AOD-PM10 data pair. First, AOD and PM10 values are sorted from lowest to highest on a monthly basis. The sorting is performed for each month individually to eliminate seasonal effects as outlined in the introduction. The sorted AOD (PM10) values are then assigned ranks corresponding to their positions on a scale from 0 to the number of AOD (PM10) observations per month (n), yielding AOD_{ranked} and $PM10_{ranked}$. The rank of AOD (PM10) within

each month is then divided by n and multiplied by 100, yielding normalized ranks greater than 0 and up to 100 (see equ. 2.1 and equ. 2.2). Finally, the AiRDI is computed by subtracting $PM10_{rank}$ from AOD_{rank} (see equ. 2.3).

$$AOD_{rank}(month) = \frac{AOD_{ranked}}{n(month)} * 100 \qquad (2.1)$$

$$PM10_{rank}(month) = \frac{PM10_{ranked}}{n(month)} * 100 \qquad (2.2)$$

$$AiRDI(month) = AOD_{rank}(month) - PM10_{rank}(month) \qquad (2.3)$$

Positive AiRDI means that AOD_{rank} is higher than $PM10_{rank}$. In other words, the satellite observation is relatively higher than the PM10 concentration, indicating a relative overestimation of PM10 by the MAIAC AOD. The opposite is true for a negative AiRDI; in this case, lower AOD_{rank} coincide with higher $PM10_{rank}$, indicating an underestimation of PM10 by the MAIAC AOD. In cases where the AiRDI is close to zero, AOD and PM10 are in the same range.

2.3 Results and discussion

2.3.1 Nonlinearity of the relationship between AOD and PM10

Several studies report a robust linear correlation between daily means of AOD and PM, e.g., for the southern U.S. (Wang, 2003), the Po Valley in Italy (Arvani et al., 2016) or selected cities worldwide (Gupta et al., 2006). In the Berlin study area, the relationship between AOD and PM10 is highly nonlinear. This is suggested by patterns shown in Figs 2.3a and 2.3b, which depict AiRDI in relation to AOD and 10. A strong positive linearity would result in a slope of zero, as increasing AOD would coincide with increasing PM10, resulting in a constant AiRDI. In fact, the Figs show an almost linear increase (decrease) of AiRDI for increasing AOD_{rank} ($PM10_{rank}$). Higher (lower) AOD is likely to overestimate (underestimate) PM10 concentrations; AiRDI is more likely to be positive for high AOD, but more likely to be negative for low AOD. The pattern is inverse for the comparison of PM10 and AiRDI.
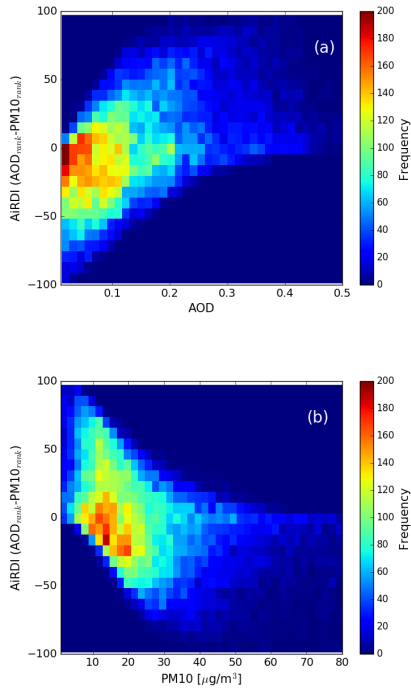
Figure 2.3: 2D-histograms of AiRDI, AOD_{rank} (a) and AiRDI, $PM10_{rank}$ (b). The frequency of occurrence of associated data pairs is color-coded.

2.3.2 Meteorological conditions of AOD-PM10 agreement and divergence

In this section, meteorological factors governing the relationship between AOD and PM10 are identified and connected to the AiRDI. An in-depth analysis concerning the role of RH and BLH is conducted. Wind speed and direction is investigated to determine the role of weather conditions and the prevailing inflow direction. AiRDI patterns related to wind are partly driven by the redistribution of RH and BLH.

To identify possible factors influencing AiRDI, data was split in positive (AiRDI >25), negative values (AiRDI < -25) and AiRDI values close to zero ($25 \geq$ AiRDI $\geq$ -25). Subsequently, the median of each parameter for each season was calculated. Then, the relative deviation of each data

point from this median for each AiRDI-class was determined, resulting in the distribution shown in Fig. 2.4.

Figure 2.4: *Determination of factors influencing AiRDI including RH, BLH, wind velocity in u and v direction (u/2,v/2), AOD and PM10 (shown here as PM). Data are split on a seasonal basis and divided in positive (a–d), negative values (e–h) and AiRDI values close to zero (i–l). Winter months are denoted as DJF, spring months as MAM, summer months as JJA, fall months as SON. Box plots show the relative deviation of each parameter observation to the seasonal mean. Deviations for u and v were multiplied by a factor of 0.5 to prevent a distortion of the y-axes. The box plots extend from the lower to the upper quartile value, with a red line at the median. Whiskers show the range with outliers marked as short, horizontal black lines. The number of observations for each class is denoted as n.*

The class AiRDI >25 is characterized by strong positive deviations of BLH and slightly positive RH deviations. This is consistent with what was introduced above: a higher BLH tends to decrease PM10 concentrations near ground because particles disperse within the atmospheric column, while elevated ambient RH tends to initiate hygroscopic growth, increasing AOD (Wang, 2003; Gupta and Christopher, 2009b). Both effects combined result in a positive AiRDI. Additionally, a high BLH might increment the number of particles distributed in the lower troposphere due to enhanced turbulence. Subsequently, AOD would also increase (Barnaba et al., 2010). BLH does not show a marked deviation in summer, pointing to a decreased importance of BLH during the summer months.

Patterns of the class AiRDI<-25 display almost inverse characteristics. Here, BLH and RH

37

show negative deviations. Negative AiRDI values result from low BLHs, which increase PM10 concentrations near ground, whereas low ambient RHs decrease the influence of hygroscopic growth. PM10 concentrations appear to have very high positive deviations in winter months, when particles are often confined to layers close to the ground (Barnaba et al., 2010). The frequent occurrence of stable stratification situations along with low turbulent atmospheric exchange during stagnant synoptic situations could lead to very high particle concentrations (Rost et al., 2009).

Parameters in the class $25 \geq \text{AiRDI} \geq -25$ do not show strong deviations of any parameter. The best agreement between AOD and PM10 is achieved when RH and BLH are near their seasonal median values.

In summary, AOD observations coinciding with high (low) ambient RH and high (low) BLH have a higher probability to cause an overestimation (underestimation) of PM10. Patterns are largely constant throughout the year. Wind speed and direction also play an important role. But like RH and BLH, the relationship is complex and is treated in more detail in the following sections.

2.3.3 The role of RH and BLH

This section quantifies the interconnected effects of RH and BLH on the relationship between AOD and PM10. To this end, a multivariate analysis was conducted using ranges of RH and BLH. These were related to AOD_{rank}, PM10_{rank} and AiRDI.

Three AiRDI-regimes can be discriminated in Fig. 2.5a. Low RH coinciding with low BLH leads to negative AiRDI values, relatively high RH coinciding with high BLH leads to positive AiRDI values. A curved area of near-zero AiRDI values separates negative and positive AiRDI patterns, indicating meteorological conditions where the magnitudes of AOD and PM10 match closely. The shape of this patch emphasizes the dependency of the AiRDI on both RH and BLH. With decreasing BLH, the near-zero values are shifted towards higher RH. Obviously, the effects on AiRDI of both parameters can compensate each other, leading to a AiRDI close to zero.
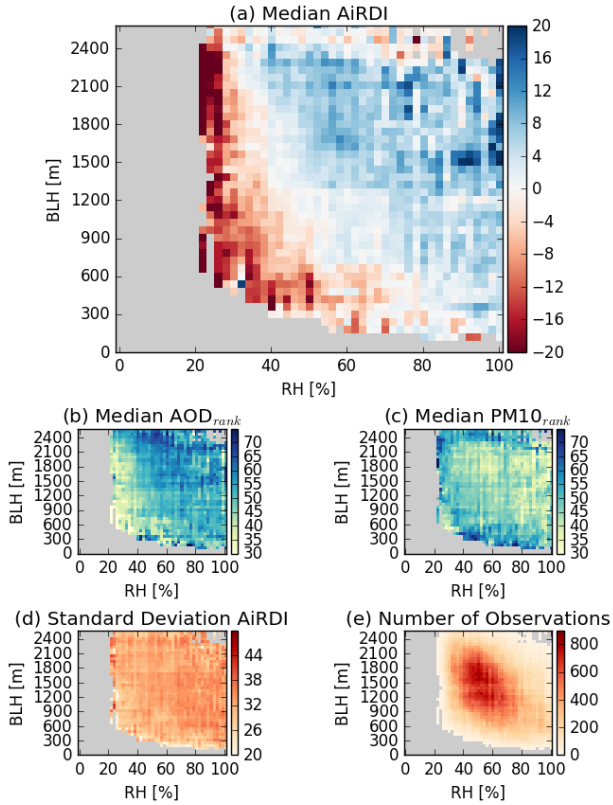
Figure 2.5: a: Relationship between AiRDI (AOD_{rank} - $PM10_{rank}$), RH and BLH. BLH is gridded in steps of 60 m, RH in steps of 2 %, each grid cell represents the median AiRDI.
b: Relationship between AOD_{rank}, RH and BLH, same grouping of RH and BLH as (a).
c: Relationship between $PM10_{rank}$, RH and BLH, same grouping of RH and BLH as (a).
d: Standard deviation of AiRDI for each class, same grouping of RH and BLH as (a).
e: Number of observations for each grid cell, same grouping of RH and BLH as (a).

Fig. 2.5b highlights the effect of hygroscopic growth on AOD, with increased RH leading to higher AOD_{rank}. Furthermore, the data suggest that AOD_{rank} is positively correlated not only to RH, but also to BLH (Fig. 2.5b). Relatively high AOD_{rank} coincides with relatively high BLH. This effect might be due to convection processes, which increase the BLH and cause particles to be lifted up into the atmosphere more easily. In these situations, the deposition of particles would be prevented and particles would remain in the atmosphere, causing the AOD to increase. Low BLH, as an

indicator of low turbulence, would lead to reduced particle release and thus relatively low AOD. Additionally, when particles are confined in a shallow boundary layer, a reduced path length for satellite measurements might reduce the sensitivity to AOD (Gupta and Christopher, 2009a).

PM10$_{rank}$ is elevated for BLH <600 m (Fig. 2.5c) owing to limited dissipation of particles into higher atmospheric layers and subsequent accumulation of particles near the ground. At very high BLH (>2,400 m), PM10$_{rank}$ is also slightly elevated. Around 3 % of all observations are related to this pattern. PM10$_{rank}$ at these high BLH values should be interpreted with care: For mid-latitudes, BLH >2,400 m appears to be very high (Von Engeln and Teixeira, 2013) and might not properly reflect real conditions. A weak dependence of PM10$_{rank}$ on RH exists: PM10$_{rank}$ decreases with increasing RH. This might be due to the fact that RH is elevated when precipitation influenced the atmosphere prior to the measurement, leading to increased RH and decreased particle concentrations due to wet deposition of particles (Rost et al., 2009). Another possible reason might be that the influence of maritime air is overrepresented in situations with high RH. These air masses are humid and clean, i.e., carry fewer particles. Finally, this pattern could be due to measurement errors of the UBA PM sites caused by condensation on the particle inlet of the PM sensors, despite the heating of the sensor inlet. This way, fewer particles would reach the sensor, decreasing reported particle concentrations. When BLH is greater than 600 m, the AiRDI is mostly driven by RH. In a very moist atmosphere (RH >90 %), the median AiRDI is near zero only at very low BLH. The influence of hygroscopic growth on AOD is apparent when AiRDI approaches positive values. In these situations, AOD is increased while PM10 remains unaffected. This effect can be seen from about RH >40 %, depending on BLH. Previous studies report that hygroscopic growth starts at RH $\sim$ 50 % (Lee and Kim, 2007).

Results presented in this section partly contradict earlier studies, which identify situations with low BLH and low RH as favorable to derive PM estimates based on AOD (Gupta et al., 2006). Fig. 2.5a suggests that a very low BLH tends to cause an underestimation of PM10 concentrations by the satellite (negative AiRDI). RH and BLH clearly have an interconnected influence on the relationship between AOD and PM10, which emphasizes the need to examine both parameters simultaneously. Generally, AOD and PM10 match best when the atmosphere is relatively dry (40 % < RH $\leq$ 70 %) and BLH is at medium levels (600–1,200 m). High ambient RH, on the other hand, can be compensated by a low BLH, leading to a good agreement of AOD and PM10. RH measurements used in this study only capture near-ground ambient RH. Nevertheless, Fig. 2.5a suggests that near-ground RH has a large influence on the relationship between AOD and PM10.

2.3.4 The role of wind direction and wind speed

In this section, patterns of AiRDI, AOD$_{rank}$ and PM10$_{rank}$ in relation to large scale patterns of wind speed and wind direction are analysed and connected to results of previous chapters. Patterns

of wind speed and direction give hints on particle origins. Knowledge on the direction of origin of air masses, for example, helps characterize particle species and concentrations (Zieger et al., 2013; Titos et al., 2014). Wind speeds help to distinguish between the dominance of local sources or particles originating further away. To distinguish between the effects of wind speed and direction, wind speed data are split in four classes 0–3 m/s, >3–5 m/s, >5–7 m/s and >7 m/s. The resulting polar plots are comprised of 16 wind direction classes with a range of 22.5°, starting from 0°. Each class is colored with the corresponding median AiRDI value, computed of all associated AOD_{rank} and $PM10_{rank}$ that appear in the corresponding class. Radii of the polar bars indicate the frequency of occurrence.
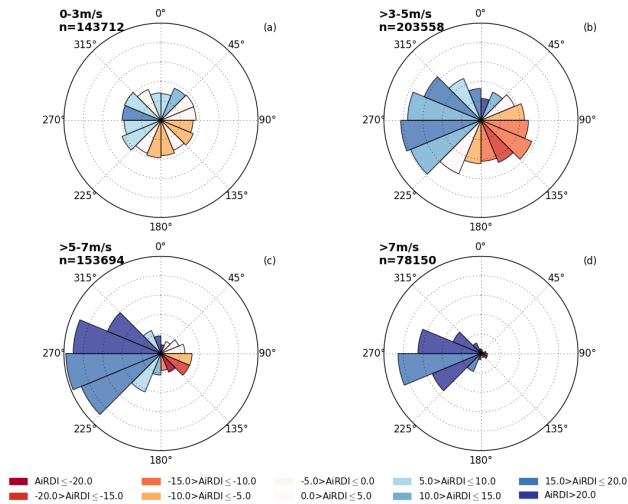
Figure 2.6: Relationship between the AiRDI (AOD_{rank} - $PM10_{rank}$), wind direction and wind speed for the study area. Panels a–d represent a wind speed class as shown on the upper left of each panel. Polar bars enclose a wind direction class, each 22.5° in range. Colors indicate the mean AiRDI for each polar bar, the radius of each bar indicates the frequency of occurrence.

A clear AiRDI pattern can be identified in Fig. 2.6. Air masses from northwestern or western direction tend to cause a positive AiRDI, while air masses from the east and south-east tend to coincide with negative AiRDI values. This pattern is most distinctive for wind speed classes >3–5 m/s and >5–7 m/s. For lower wind speeds (0–3 m/s), the pattern is less clearly defined. This is likely due to the dominance of local particle sources, superimposing the effects of large-scale air mass transports. Results suggest that satellite AOD tends to overestimate particle concentrations

near the ground when strong northwestern or western winds prevail, whereas an underestimation occurs when eastern and southeastern winds dominate. There is a higher chance of AiRDI close to zero at lower wind speeds, as local emissions can be captured more easily by both ground stations and the satellite.

Figure 2.7: Relationship between AOD_{rank}, wind direction and wind speed. Setup of panels a–d is identical to previous plot.

Figure 2.8: Relationship between PM10$_{rank}$, wind direction and wind speed. Setup of panels a–d is identical to previous plot. PM10$_{rank}$ is generally higher when continental, i.e., easterly air masses dominate.

PM10$_{rank}$ (Fig. 2.8a–d) is generally highest during conditions with low wind speeds (0–3 m/s). This is not surprising, as transport of pollutants takes place less effectively and particles from local sources inside the urban area of Berlin accumulate in the atmosphere. Additionally, dry deposition of particles might take place more effectively, as particles reach the surface more easily due to decreased vertical winds. This might lead to increased PM10 concentrations at the measurement sites. Winds from eastern directions show elevated PM10$_{rank}$, indicating advection of continental air masses with high particle numbers near the ground. Air masses from the west, possibly of maritime origin, tend to be cleaner than continental air, because particles are scavenged by wet deposition (Rost et al., 2009; Zieger et al., 2013). At higher wind speeds, this pattern is most distinct.

Generally, patterns for AOD$_{rank}$ are less pronounced (Figs 2.7a–d). In contrast to PM10$_{rank}$, AOD$_{rank}$ is slightly elevated when high wind speeds occur from the west. The inverse pattern of AOD$_{rank}$ versus PM10$_{rank}$ in terms of wind speed and wind direction demonstrates that satellite derived AOD is not suited for directly deriving ground-based PM10 in all situations.

Several explanations for these observations are possible. Advection of western air masses transports higher amounts of moisture to the Berlin study area, indicated by higher values of RH in Fig. 2.9. Increased availability of moisture in the atmosphere could promote particle water uptake

and cause hygroscopic swelling of particles. Increased particle sizes would lead to rising AOD_{rank} but not affect $PM10_{rank}$, which ultimately leads to positive AiRDI. Eastern winds, on the other hand, transport continental air masses with a potentially larger particle loading (Kerschbaumer, 2007; Bonn et al., 2016). This might increase PM10 concentrations.

A further possible explanation might be the varying hygroscopicity of particles carried by different air masses. Western air masses are thought to carry a higher amount of hydrophilic particles (e.g., sea salts) which tend to be more prone to hygroscopic growth due to more effective water uptake (Zieger et al., 2013; Titos et al., 2014; Granados-Muñoz et al., 2015). The opposite might be true for air masses from eastern and northeastern origins carrying anthropogenic and industrial aerosols. For the measurement station Melpitz, located around 130 km south of Berlin, it was found that the scattering enhancement of particles was lowest when air masses were advected from the east and northeast (Zieger et al., 2013, 2014). Thus, these particles are less likely to increase the AOD by hygroscopic growth (Zieger et al., 2013; Titos et al., 2014). Finally, eastern wind directions might indicate a stable atmospheric situation with lower BLH (see Fig. 2.10), causing pollution levels to increase near the ground (Gupta and Christopher, 2009b; Geiß et al., 2017). As can be seen in Fig. 2.10, advection of air masses from the west at higher wind speeds tends to increase BLH.

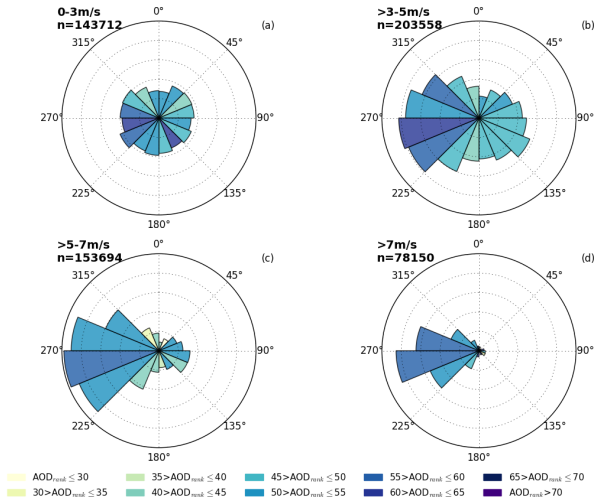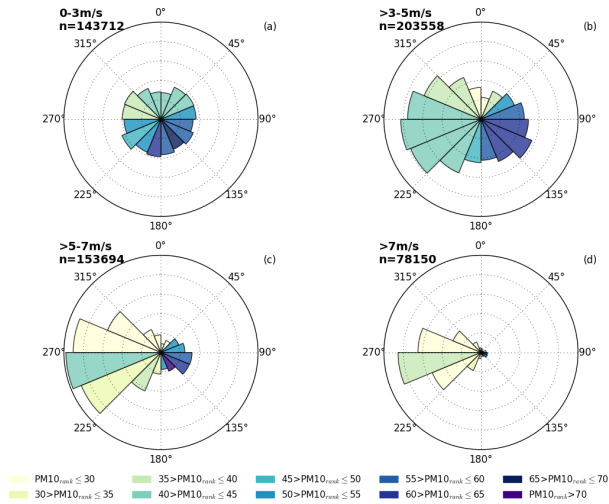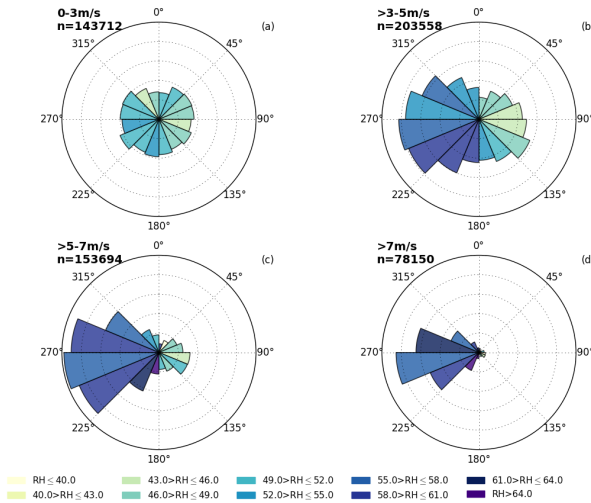Figure 2.9: Relationship between RH, wind direction and speed. Setup of panels a–d is identical to previous plot.
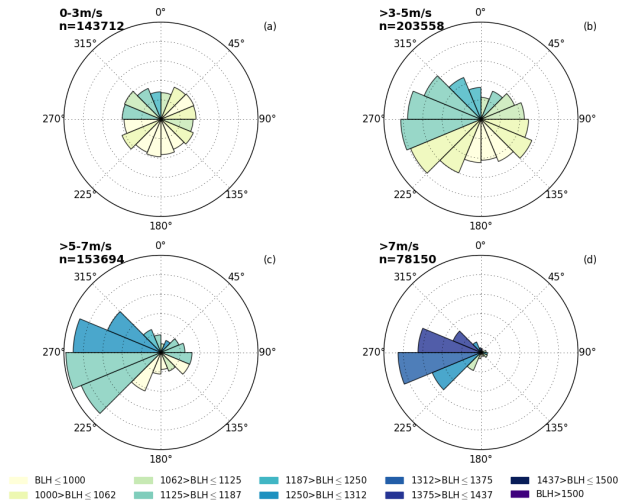
Figure 2.10: Relationship between BLH, wind direction and speed. Setup of panels a–d is identical to previous plot.

In summary, advection of western air masses tends to increase RH and BLH, coinciding with slightly elevated AOD_{rank} and low $PM10_{rank}$. Eastern and southeastern air masses tend to show lower RH and BLH, coinciding with slightly decreased AOD_{rank} and elevated $PM10_{rank}$. These patterns are generally more pronounced for higher wind speeds, whereas dependence on wind direction decreases with lower wind speeds. To conclude, the following statements summarize the findings of this chapter:

- Western air masses with higher wind speeds establish a relatively high probability that the satellite AOD overestimates PM10.

- The opposite is true for eastern direction of origin. In these cases, a higher probability of the satellite AOD underestimating PM10 levels exists.

- Differences in RH and BLH patterns seem to be the main drivers behind the dependence of AiRDI on wind direction besides the transport of pollutants and varying hygroscopicity of particles.

- Agreement between AOD and PM10 is more likely to occur when wind speeds are low.

45

2.4 Specific conclusions

Information on RH, BLH, wind direction and speed proved to be of importance, driving the relationship between AOD and PM10. Generally, AOD serves as a good proxy for PM10 concentrations near ground when a very dry atmosphere ($30\% < RH \leq 50\%$) coincides with a high BLH (>1,200 m), when moderate ambient RH (50–80 %) coincides with moderate BLH (600–1,200 m) or when humid conditions (RH >80 %) coincide with a shallow BLH (<600 m). In case of a dry atmosphere, AOD is not influenced by hygroscopic growth and thus relatively low. This can be compensated by low PM10 levels, which appear more frequently in situations with a high BLH. When ambient RH is high, satellite-derived AOD tends to overestimate street-level PM10 concentrations. This effect can be compensated by a low BLH, which tends to increase PM10 levels near the ground. In combination, these effects can lead to an AiRDI close to zero, indicating a good match of satellite-retrieved AOD and ground-based PM10 measurements. When moderate RH and BLH prevail, AOD is only marginally affected by hygroscopic growth due to limited availability of moisture, and particle concentrations in the boundary layer seem to agree well with particle concentrations near the ground. These results suggest that it might be misleading to examine RH and BLH separately.

In terms of wind speed and wind direction, it was found that situations with prevailing western direction of origin have a higher probability to cause a positive AiRDI, whereas winds from the east tend to cause a negative AiRDI. The AiRDI is more likely to be near-zero in situations with low wind speeds

The AiRDI proved to be a useful approach to a perform semi-quantitative comparison of the parameters AOD and PM10 and to better understand their interactions with surrounding meteorological conditions. Knowledge of the effect of these conditions can help future approaches to derive high-quality PM estimates based on AOD. The AiRDI approach can be generalized for application in other study areas. Still, this study did not account for some sources of uncertainties, for example effects of cloud contamination or aerosol layers above the BLH (Crumeyrolle et al., 2014; Bourgeois et al., 2018). Future analyses could further focus on the role of different aerosol species or the role of precipitation events on the relationship between AOD and PM.

3 Mapping and understanding PM10 patterns using AOD and environmental conditions

3.1 Benefits of using machine learning models to derive PM from AOD

As shown in chapter 2, PM can be estimated with higher accuracy from satellite-based AOD, if environmental conditions are explicitly considered. Environmental conditions are not only expected to influence the relationship between AOD and PM10, but also to also to directly influence PM concentrations near ground (see section 1.2, conceptual study design in Fig. 3.1). In this chapter, PM10 concentrations are estimated using a ML model, which is trained to predict PM10 under consideration of AOD and environmental conditions. In previous studies, ML models have been successfully leveraged to predict PM from AOD, accounting for a range of additional parameters and mostly with a focus on accurately predicting PM concentrations (van Donkelaar et al., 2010; Kloog et al., 2015; Hu et al., 2017; Stafoggia et al., 2017; Rybarczyk and Zalakeviciute, 2018; Zhang et al., 2018). ML models are powerful statistical models, which efficiently reproduce non-linear relationships and interactions of input features (Elith et al., 2008; Brokamp et al., 2017). Furthermore, in contrast to physical models, machine learning approaches do not require extensive prior process knowledge and rely on empirical relationships (Knüsel et al., 2019). This means that ML models have the potential to reveal or quantify processes that are as yet undetermined. A ML model extracts empirical patterns between input features and the target variable during the training process. Depending on how well this training process is conducted, the model can transfer the extracted patterns to unseen data and predict an unknown target variable. The model's ability to transfer its learnings on unseen data determines the generalizability of the extracted patterns.

Only recently, air quality studies approached the idea of analysing the information content of explanatory variables in statistical models to infer processes (e.g., Hu et al., 2017; Grange et al., 2018; Park et al., 2019). Here, these approaches are expanded and an in-depth analysis of model-inherent relationships is provided and related to physical processes. Multivariate processes are investigated by isolating certain variables and studying inputs and responses for dominant patterns (see Cermak and Knutti, 2009; Andersen et al., 2017; Fuchs et al., 2018). Based on the predictions of hourly PM10 concentrations by a gradient boosted regression trees (GBRT) model, important drivers of PM10 concentrations are identified. Furthermore, the capability of satellite observations
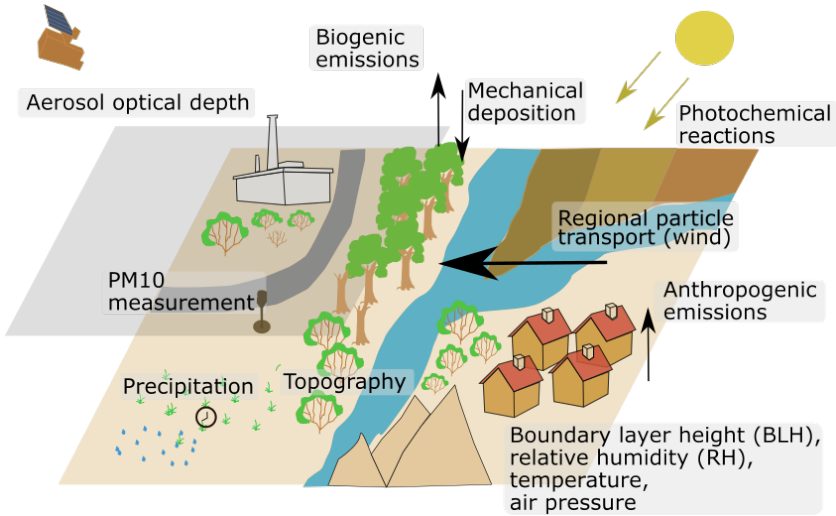
Figure 3.1: Conceptual design of potential influences on variations in hourly PM10 concentrations. The Machine Learning model is set up to reproduce these variations.

to map and monitor surface air pollution is assessed by investigating the relationship between AOD and PM10 in the same modelling setup.

3.2 Data and methods

The data basis of this study is comprised of 8 years (2007–2015) including satellite observations from MODIS and others, model output from the ECMWF and station data from the German Meteorological Service DWD and the UBA. AOD observations are based on the MAIAC algorithm. A statistical model is set up to predict hourly PM10 concentrations based on a variety of input features, which are summarized in Tab. 3.1. Data with high temporal resolution are selected to reflect the atmospheric situation close to the satellite overpass times. Generally, station measurements were preferred for parameters, which have adequate coverage and which are not expected to vary on small scales (e.g., air pressure).

Table 3.1: List of input features stating data origin, data representation and abbreviation used in chapter 3.

Data set	Variable	Abbreviation
Input Features		
MODIS MAIAC	Aerosol optical depth	AOD
MODIS	Normalized difference vegetation index	NDVI
NASA Earth at night	lights at night	Nightlights
EU-DEM, v1.1	Elevation [m],	Elev,
	Topographic Position Index [m]	TPI
CORINE Land Cover	Land cover types	1CLC, 2CLC, 3CLC, 4CLC 5CLC, 6CLC
DWD meteorological measurements	Air pressure (sea level) [hPa]	AirPres
	Relative humidity [%]	RH
	Continentality	Conti
RADOLAN	Time since last precip. [h],	Precip_tsince,
	Magnitude of last precip. [mm/h],	Precip_magn,
	Cumulative precip. last 24 h [mm],	Precip_acc
ERA-Interim	wind vectors u,v [m/s],	u,v
	mean wind vectors of previous hours [m/s]	umean, vmean
	Boundary layer height [m]	BLH
	Temperature [°],	T
	temperature anomalies [K],	Tan
	Convective available potential energy [J/kg]	CAPE
	Surface solar radiation downwards [J/m^2]	SSRD
EEA emission data base	Annual emission of NH_3, PM10, SO_2, NO_x [t/year]	NH_3, PM10em, SO_2, NO_x
other	Day of year	DOY
	Day of week	Weekday
Model Outcome		
UBA air quality measurements	PM10 concentrations [μg/m^3]	PM10

Satellite data

MAIAC AOD

MAIAC Collection 6 AOD is used (Lyapustin et al., 2011b,a,b, 2018). Quality flags are applied for filtering. Pixels were only incorporated when clear conditions are reported, i.e., no contamination of the data due to clouds is to be expected (Lyapustin et al., 2018). An additional filter was set up to exclude AOD close to clouds to avoid increased AOD near cloud fringes due to aerosol swelling effects (Várnai et al., 2013; Schwarz et al., 2017). Therefore, the distance to the nearest cloud as classified by the MAIAC algorithm (Lyapustin et al., 2018) was calculated and a threshold of $0.1°$ was set, which corresponds to $\sim$7 km. This is in the range of what previous studies used as threshold (Koren et al., 2007; Emili et al., 2011a). Terra or Aqua satellite overpass times were used as temporal reference for data collocation, i.e., other input feature values closest to the overpass times were used. Valid daytime AOD acquisition times as used in this study range from $\sim$9 UTC to $\sim$1 UTC. AOD is an important input to the statistical model as it provides implicit information on the total aerosol loading in the atmosphere, reflecting natural as well as anthropogenic sources.

MODIS NDVI

16-day NDVI averages obtained from the MODIS MOD13Q1 Version 6 product are used (DOI: 10.5067/MODIS/MOD13Q1.006) to approximate photosynthetically active vegetation (Tucker, 1979). The NDVI was found to influence PM concentrations in several previous works (Chudnovsky et al., 2014; Stafoggia et al., 2017; Beloconi et al., 2018). Vegetation acts as a sink by increasing the aerodynamic roughness and available surface for mechanical deposition (Fuzzi et al., 2015). On the other hand, vegetation can increase background particle concentrations by emitting pollen in spring (Fuzzi et al., 2015) and by enhancing the formation of SOA Churkina et al. (2017). Around each PM station coordinate, a window with an edge size of 20 km is established to reflect the local contribution of vegetation to the total particle concentration. An edge size of 20 km is comparable to the mean representativeness of the PM stations (EU, 2008). The mean NDVI of each window was used as a predictor.

NASA night lights

The NASA night lights data set is included as a surrogate for population density. Based on data from the VIIRS, the night-time lights product is available at 500 m resolution (Román et al., 2018). Population density is an important factor contributing to PM10 concentrations as it reflects human activity (Beloconi et al., 2018; Park et al., 2019). The mean night light intensity of a 20 km-window around each PM station was used as a predictor.

EEA DEM

Topography can have a marked influence on the accumulation of particles (Emili et al., 2011a). Here, the v1.1 EU-DEM is used, which is a hybrid product produced by the EEA based on SRTM and ASTER GDEM fused by a weighted averaging approach. It has a spatial resolution of 25*25 m (more information and download at https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata). Station altitude and the dominant topography around each PM station are incorporated as predictors. A Topographic Position Index (TPI) is computed to provide information on the topography of a pixel relative to its surrounding pixels. It employs the station altitude and subtracts the mean altitude of surrounding pixels in its vicinity. Positive values indicate a summit position, whereas negative values indicate a valley position (Egli et al., 2018). Again, a window size of 20 km was chosen.

EEA Corine Land Cover (CLC)

Data from the CLC inventory for the years 2006 and 2012 are used (Bossard et al., 2000), accessed via https://land.copernicus.eu/pan-european/corine-land-cover/view. In its finest thematic accuracy, the CLC data set consists of 44 classes. For this study, a more simplistic classification in 6 land cover types was chosen, which represent the most relevant land cover types influencing air quality (see Tab. 3.2). The data has a spatial resolution of 250 m. Because the CLC data are categorical, a window of 20 km edge size was set up and the number of occurrences of each type in that window was calculated. The number of occurrences of each feature is used as predictor for the model.

Reanalysis data

The Era-Interim reanalysis data by the ECMWF provides full spatial coverage for the whole study period with an interpolated spatial resolution of 0.125 ° (Dee et al., 2011). The data set has been successfully used in previous air quality studies (Stafoggia et al., 2017; Zheng et al., 2017; Chen et al., 2018b). To capture regional transport of particles, ERA-Interim reanalysis wind components [m/s] in east-west and north-south direction (10 m height) are used. Wind direction and speed can influence both particle concentrations (Chudnovsky et al., 2013b; Li et al., 2015; Beloconi et al., 2018) and the relationship between AOD and PM10 (Zheng et al., 2017; Stirnberg et al., 2018). Wind direction and speed are included as instantaneous values and as the mean of precedent days (72 h). BLH data are employed to approximate the vertical distribution of aerosols in the lower troposphere, implying that particles are well-mixed within the boundary layer (Ansmann et al., 2000; Gupta and Christopher, 2009b). Dispersion of particles within a high and well-mixed boundary layer leads to reduced particle numbers near ground. If the BLH is low, particles

Table 3.2: Original labels and aggregation of corine land cover classes used in this study.

Code	Label	Original Labels Included
1	Artificial surfaces	continuous urban fabric, discontinuous urban fabric, industrial or commercial units, road and rail networks and associated land, port areas, airports, mineral extraction sites, dump sites, construction sites, green urban areas
2	Agricultural areas	non-irrigated arable land, permanently irrigated land, rice fields, vineyards, fruit trees and berry plantations, olive groves, pastures, annual crops associated with permanent crops, complex cultivation patterns, land principally occupied by agriculture with significant areas of natural vegetation, agro-forestry areas
3	Forest and semi natural areas	broad-leaved forest, coniferous forest, mixed forest, natural grasslands, moors and heathland, sclerophyllous vegetation, transitional woodland-shrub, beaches & dunes & sands, bare rocks, sparsely vegetated areas, burnt areas, glaciers and perpetual snow
4	Wetlands	inland marshes, peat bogs, salt marshes, salines, intertidal flats
5	water bodies	water courses, water bodies, coastal lagoons, estuaries, sea and ocean
6	Nodata, unclassified	nodata, unclassified land surface, unclassified water bodies, unclassified

accumulate and increase PM concentrations near ground as they are constrained to a smaller volume (Gupta and Christopher, 2009b; Wagner and Schäfer, 2017). Era-Interim temperature in 2 height is included as instantaneous values and as temperature anomaly. Temperature anomalies are determined as the deviation of the expected value for each day of the year. Expected values are calculated by averaging daily values over 30 years, then calculating a running mean over 30 days to achieve a smooth sequence of the expected temperature. In addition to temperature, surface solar radiation downward (SSRD) and convective available potential energy (CAPE) are included to capture potential secondary aerosol formation based on photochemical transformation processes (Wang and Martin, 2007) and to capture convective mixing in the atmosphere (Chudnovsky et al., 2013b).

Ground-based data

UBA PM10

The focus here is on PM10 rather than PM2.5 because the latter excludes the fraction of larger particles, which are nevertheless accountable for light extinction and thus contribute to AOD measured by the satellite (Emili et al., 2011b). In addition, PM10 measurements are more widely available. The PM stations are maintained by the UBA and measure the hourly mean PM10 concentration. The spatial distribution of station types and their altitudes are shown in the scientific appendix, Fig. A.1 and Fig. A.2. UBA PM station coordinates are used as spatial reference for data collocation, i.e., pixels from continuous data grids are collocated with the position of these stations if below a distance threshold of 0.01 ° (∼0.7 km). Urban industrial stations are not considered in this study, as they are primarily influenced by local emissions from point sources that cannot be adequately represented with the available data. PM10 concentrations were checked for sudden peaks, which could be due to localized events. These are filtered out by eliminating situations, where the PM10 concentration was more than double the mean of the previous and following hours.

DWD meteorological data

Air pressure and RH data were obtained from the DWD (DWD Climate Data Center (CDC), 2017). Both parameters potentially influence PM10 concentrations (see section 1.2). and the relationship bewteen AOD and PM10 (see section 1.3.3).

Continentality factor

The relationship between AOD and PM10 as well as driving factors of PM10 concentrations may vary in different climatic regions (Di et al., 2016). This is accounted for by the inclusion of a dimensionless continentality factor k, which is calculated following the formula by Conrad (1946):

$$k = (1.7 * A / sin(\phi + 10°)) - 14, \tag{3.1}$$

where A corresponds to the difference between the hottest and coldest mean monthly temperature and ϕ to the latitude in decimal degree. Here, k was calculated using the mean temperature of July (1980–2010) and the mean temperature of January (1980–2010). Temperature data are provided by the DWD (DWD Climate Data Center (CDC), 2018).

DWD Radolan precipitation

The influence of precipitation on PM10 concentrations includes wash-out effects and the limitation of movement of particles after precipitation events (Rost et al., 2009; Fuzzi et al., 2015; Li et al., 2015). In this study, data from the Radar Online Adjustment Project (RADOLAN) are used. The dataset is produced by the DWD and merges radar measurements with rain gauge data, also including orographic correction (Bartels et al., 2004; Weigl, 2017). The data are available with high spatial coverage and are expected to be better suited for the present analysis than the Era-Interim precipitation product, which has some known biases (de Leeuw et al., 2015). In the statistical model, the time since the last precipitation event [h] and its magnitude [mm/h] as well as the accumulated precipitation of the last 24 h [mm] are included. Hourly means of precipitation around each PM-Station are averaged within a window of edge size 5 km. The chosen window size is smaller than the previous ones, as precipitation effects typically vary on small scales.

Other input data

EEA emission database

Annual emissions of NO_x, PM10, SO_2, and NH_3 (in tonnes, based on the year 2008) are included to approximate background pollution levels. The data are gathered by the EEA and described in Theloke et al. (2009). Diffuse air releases from traffic, agricultural, industrial, and residential sources are covered. Strong emitters that fall under the European Pollutant Release and Transfer Register are not included in this data set.

Spatiotemporal information

Seasonality was shown to be an important factor in previous studies (Grange et al., 2018). Here, DOY is used as seasonal proxy. To mirror the seasonal cycle, DOY was converted to a sine curve with +1 representing summer solstice and -1 representing winter solstice (Stolwijk et al., 1999; Park et al., 2019). To further approximate variability in emission strengths based on human activity, day of the week is included.

Gradient Boosted Regression Trees

Model specifications

GBRT as implemented in python's scikit-learn module are used (Friedman, 2002; Pedregosa et al., 2012). GBRT merge several statistical approaches found in machine learning applications: decision trees (1.) and boosting (2.) with gradient descent (3.).

1. Decision trees use decision nodes to split the predictor space in subsets, which provide the most homogeneous distribution, i.e., the subsets, variance is minimized. For each subset, regression trees fit the mean response of the observations that go into the model (Elith et al., 2008).

2. Similar to the RF method, GBRT consist of an ensemble of decision trees. In GBRT models, however, the construction of the ensemble is different, as decision tree regressors are sequentially added to the ensemble (Friedman, 2001; Elith et al., 2008; Rybarczyk and Zalakeviciute, 2018). Each new tree that is added to the ensemble boosts its predecessor with the goal to minimize a loss function and existing trees are not changed when new trees are added. The trees are fitted on a subset of the complete data set, which induces a random component to the model to reduce overfitting (Friedman, 2001; Elith et al., 2008).

3. Each new predictor is fitted to the predecessor's previous residual error using gradient descent (Friedman, 2001; Elith et al., 2008).

GBRT capture complex interactions and interactive effects between individual predictors (Elith et al., 2008; Brokamp et al., 2017), which nevertheless need to be considered when interpreting the model outcome. An advantage of tree-based methods such as RF or GBRT is that, compared to deep learning methods, model decisions can be retraced and dependencies of the model outcome to input features can be quantified, allowing for conclusions regarding physical processes. This makes GBRT an interpretable machine learning method. GBRT theoretically produce results more effectively than the RF method, as trees are built systematically and less iterations are required (Elith et al., 2008). GBRT have shown to have good predictive power in previous studies (Elith et al., 2008; Fuchs et al., 2018; Just et al., 2018). The general framework of setting up the model is shown in Fig. 3.2 and includes input feature selection, hyperparameter tuning, model training, and model validation.

Feature selection: recursive feature elimination

Redundant input features lack information and potentially degrade model performance by inducing misleading information, thus weakening the target orientation of the model (Meyer et al., 2018). Therefore, a feature selection is conducted to eliminate redundant predictors. A base model is initialized with a fixed number of trees (500), a fixed learning rate (0.1) and all other hyperparameters on default settings. One feature is then excluded and the data set is randomly split into a training data set (2/3) and a test data set (1/3) fifty-fold. After fifty repetitions, the decrease in model performance on the test data set due to the exclusion of the feature was determined using R^2 as indicator. In the final model, the magnitude of the last precipitation event, the accumulated

Table 3.3: List of hyperparameters and parameter grid that is applied during the grid search. Hyperparameters determine the architecture of the GBRT model.

Hyperparameter	Description	Parameter grid
loss	loss function to be optimized	fixed: least squares regression
learning_rate	contribution of trees to ensemble	[0.08,0.05,0.01]
n_estimators	number of boosting iterations	[800,1800,2500,4000]
subsample	fraction of samples used for individual tree	[0.6,0.8]
min_samples_split	minimum number of samples used to split a decision node	[6,10,14]
min_samples_leaf	minimum number of samples required for a leaf node	[14,18,20]
max_depth	maximum depth for individual tree, i.e., maximum number of node layers	[6,10,14]
max_features	fraction of features to be considered when searching for best split	[0.5,0.8]

rainfall of the last 24 hours and the annual mean emission of SO_2 are excluded, as their exclusion did not lead to a decrease in model performance.

Hyperparameter tuning, model training, and model evaluation

Hyperparameters refer to the model architecture, e.g., the number of trees or the number of decision nodes. The determination of adequate model hyperparameters is essential to avoid overfitting of the model, but at the same time ensure that the model is able to generalize. A grid search is executed, where several parameter combinations are tested. A list of tested parameters is provided in Tab. 3.3.

There are trade-offs between *max_depth*, *n_estimators*, and *learning_rate*. A lower *learning_rate* requires a higher number of trees, as the contribution of each tree is decreased. The contribution of each tree is effectively the step size of the gradient descent. Thus, if *learning_rate* is too large, the model cannot adapt to the training data. The number of decision nodes in a tree (max_depth) also affects *n_estimators*: increasing *max_depth* reduces the number of necessary trees but increases the risk of overfitting.

The model penalizes erroneous predictions to improve accuracy. The calculation of the penalty value is determined by the loss function. In this study, a least squares loss function is chosen. The least squares loss function is sensitive to very high (low) values as it strongly penalizes large
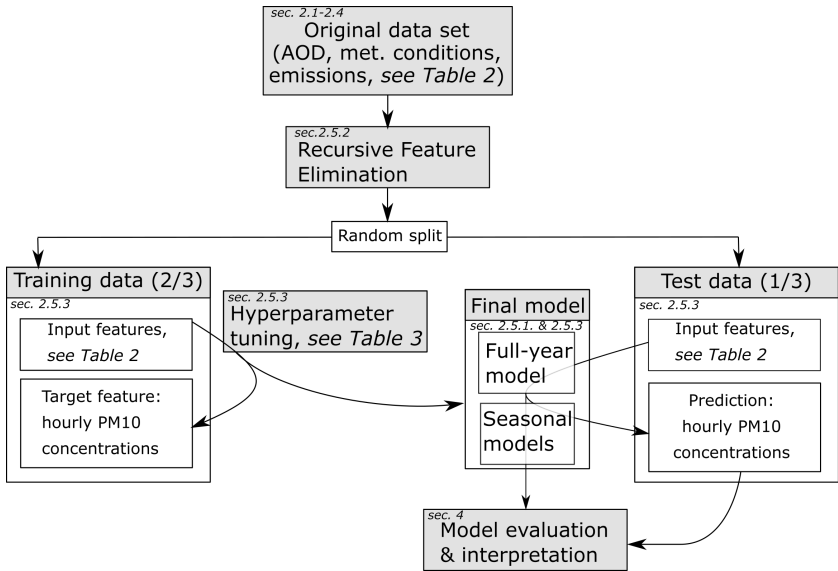
Figure 3.2: Framework of chapter 3: from input feature selection (i.e., separating important features from noise), hyperparameter tuning, model training, and model evaluation to the interpretation of the results.

deviations between predictions and observations and will adjust the model accordingly. This is desirable, as the model should be able to reproduce high concentrations of PM10. Model performance is validated using two kinds of validation strategies. The integrated scikit-learn split function creates a random training (2/3) and test data set (1/3), ensuring a comparable distribution of both data sets. To test the spatial generalizability of the model, a leave-location-out approach is conducted. Therefore, one third of randomly chosen stations are restrained for validation. If the model performance is lower compared to the random approach, the spatial generalizability of the model is limited (Meyer et al., 2018).

For further analysis, four seasonal models are trained on seasonal subsets of the data (see also Fig. 3.2) using the hyperparameters determined in the full-year grid search. One input feature set representative for all seasonal models and the full-year model is used to ensure their comparability. The only exception is the EEA emission data set, which is only included in the yearly model, since it represents yearly emissions.

Model interpretation: isolation of feature contributions

Feature importance

The relative feature importance reflects the explanatory power a feature provides to the model. Assuming that the model is able to capture physical processes well, the feature importance represents a valuable qualitative measure to determine the relative magnitude of the influence of input features to predicted PM10 concentrations. The feature importance is calculated by repeated permutation of one feature (Strobl et al., 2007).

Partial dependence

To quantify the influence of input features on the model, the partial dependence (PD) of modelled PM10 concentrations on input features is calculated. PDs express the average effect of one input feature on the modelled PM10 outcome while accounting for average effects of complement input features (Friedman, 2001; Elith et al., 2008; Goldstein et al., 2015). The investigated input feature is gridded and the corresponding average PM10 prediction is calculated with respect to complement features, which are varied over their marginalized distributions (e.g., 1st–99th percentile). Thus, PD plots reflect the mean change of average predicted PM10 concentrations based on one input feature. The isolated effects of input features on the model response can be evaluated and put into context regarding their significance to physical and chemical processes determining PM10 concentrations (Grange et al., 2018).

Individual conditional expectation

PD plots reflect the mean model response and neglect model heterogeneity. Model responses to single data instances (i.e., one set of input features related to one PM10 observation) can be unwrapped and bundled in one plot using the individual conditional expectation (ICE) approach. ICE plots reflect individual predicted responses as a function of one data instance depending on correspondent feature observations. Model responses are computed by keeping the complement features constant while the investigated feature varies, thus creating new data instances and predictions from the model. The average over all ICE lines yields the PD plot, allowing model heterogeneity and mean model response to be depicted simultaneously (Goldstein et al., 2015).

3.3 Results and discussion

3.3.1 Model performance

The overall model performance is shown in Fig. 3.3, depicting observed PM10 vs. predicted PM10 for the validation data. The full-year model explains 77 % of the variance. The slope (0.76)

shows a slight overestimation of low PM10 concentrations as well as an underestimation of high PM10 concentrations. Presumably, the underestimation is due to processes not captured by the input features, i.e., street-scale processes not covered by AOD observations but still influencing PM10 observations, such as increased PM10 emissions due to traffic jams or localized dust resuspension. In addition, the model possibly tends to underestimate higher PM10 observations, because most valid data points are available for medium to low PM10 concentrations. Thus, the model is optimized to best reproduce these observations. This tendency was reduced by the choice of the least squares loss function as described in section 3.2, but likely still continues to affect the model accuracy. The model performance is comparable to similar studies, also in its underestimation of PM (Hu et al., 2017; Stafoggia et al., 2017; Grange et al., 2018; Zhang et al., 2018). Tenfold random train/test splits were conducted, resulting in ten models. Validation of these models revealed very similar performances, which is why only one model is shown and used for subsequent model analysis. Applying the leave-location-out split approach slightly deteriorated model performance. Depending on the restrained stations, R^2 ranged from 0.5-0.7. The model is considered as adequate to be used for further investigations. Nevertheless, note that high PM10 values tend to be underestimated by the model (see Fig. 3.3). The spatial distribution of the model skill is shown in the scientific appendix B (Fig. B.1 and Fig. B.2).

Model performances vary seasonally (see Fig. 3.4). The model performs best in winter and spring with high R^2 values (0.77) and low NRMSEs (0.32 and 0.3), while R^2 is lowest in summer with an R^2 of 0.63 and a slightly increased NRMSE of 0.34. PM10 concentrations generally show less variance in summer, which reduces the RMSE, but possibly provides the model less variance to learn from and deteriorates its skill (i.e., R^2). Obviously, processes governing PM10 concentrations in summer are not as well captured by the model. This will be further addressed in the following chapters.

3.3.2 Information content of input features

Temperature (anomaly and absolute), AOD, 3-day mean east-west wind component, DOY and BLH are of high importance to the model (see Fig. 3.5). The importance of the DOY suggests that the model captures the seasonality of PM10 concentrations, which are higher in winter and lower in summer. The relatively high importance of AOD and the good model performance emphasize the suitability of AOD to infer on PM10 concentrations when additional parameters are taken into account. A comparison to similar studies, e.g., by Grange et al. (2018); Park et al. (2019), reveals comparable relative feature importances of AOD, solar radiation (Park et al., 2019), DOY and BLH (Grange et al., 2018; Park et al., 2019). The high importance of the 3-day mean of the east-west wind component found in this study aligns with the high importance of the back trajectory clusters in the study by Grange et al. (2018). Both parameters reflect regional particle transport. However,
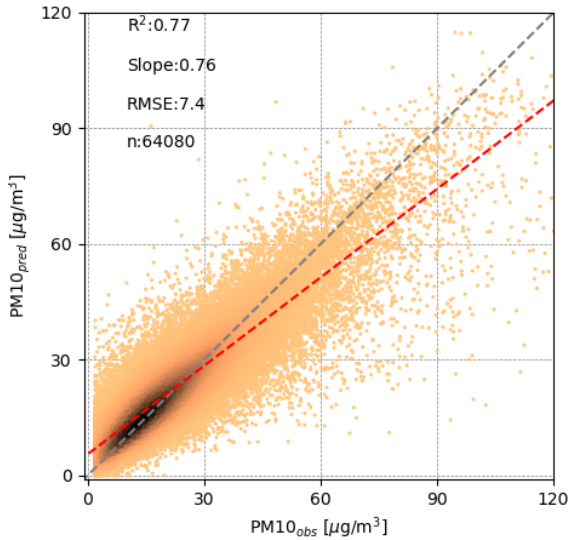
Figure 3.3: *Scatter plot showing the full-year model predictions for hourly PM10 concentrations. Also shown are coefficient of determination (R^2), slope (red dotted line) and root mean squared error (RMSE). The color range from black (high) to orange (low) indicates the frequency of occurrence. The relatively high R^2 shows that the model covers the majority of occurring variance. However, an underestimation of higher PM10 concentrations leads to a lower slope.*

there is a discrepancy regarding the importance of wind speed and temperature. Wind speed is considered as instantaneous wind components in this study, which have limited importance. While the importance of absolute air temperature is comparable, the high importance of temperature anomalies found here shows that the approach of splitting temperature information into absolute values and anomalies as pursued in this study provides additional information to the model. Since comparing feature importance values can provide only limited insights into processes behind air pollution patterns, further quantitative analyses are presented in the following chapters using the ICE approach.
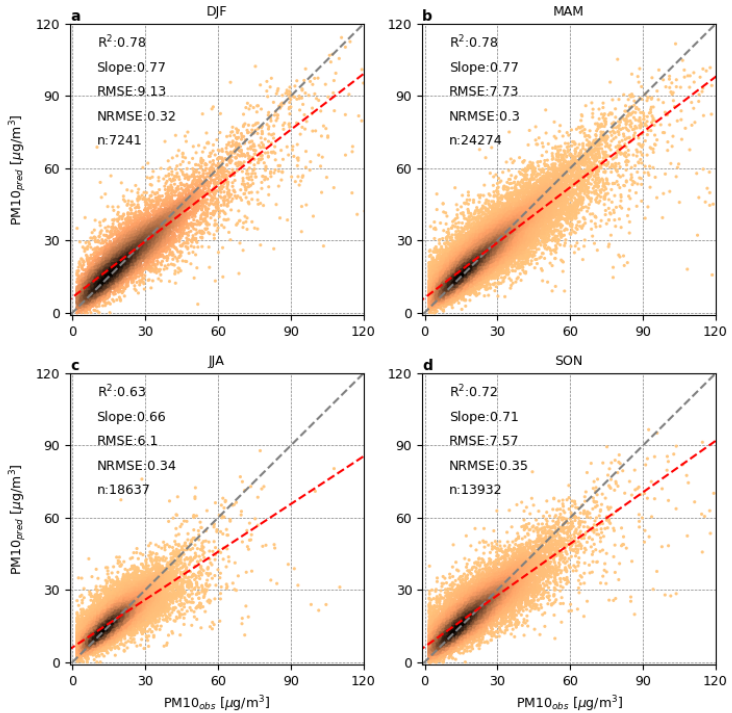
Figure 3.4: Scatter plot showing seasonal model predictions for hourly PM10 concentrations. Also shown are seasonal R^2, slope (red dotted line), normalised RMSE (NRMSE) and RMSE. The NRMSE was calculated by dividing the RMSE by the mean of the corresponding seasonal subgroup of PM10 observations. Colors as in Fig. 3.3.
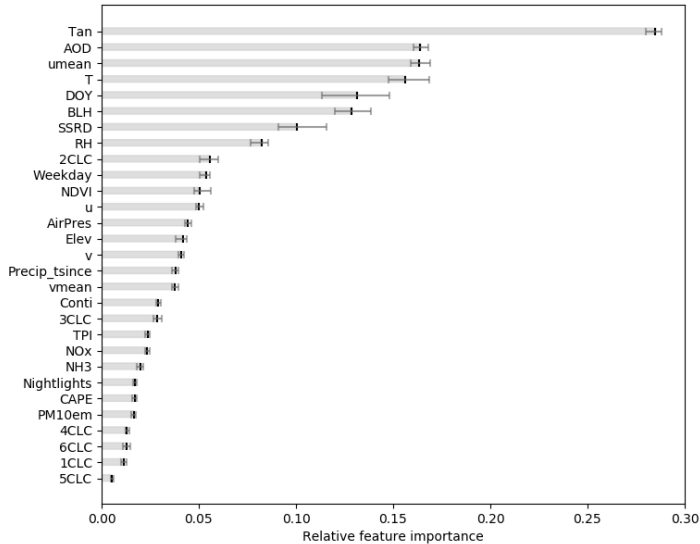
Figure 3.5: Relative importance of input features based on repeated permutation (see section 3.2). The range represents the standard deviation. Abbreviations correspond to those shown in Tab. 3.1.

3.3.3 Model sensitivity

Mesoscale wind information

The PD of the 3-day mean east-west wind component shows a consistent pattern throughout all seasons. Positive values (i.e., prevailing western direction of inflow) are associated with reduced concentrations of PM10, whereas a negative east-west wind component (i.e., winds from the east) is associated with increased model PM10 concentrations (Fig. 3.6). For the full-year model the maximum difference in mean PM10 predictions is $\sim$10 μg/m^3, while in winter, the maximum difference is $\sim$20 μg/m^3. These numbers agree well with results from van Pinxteren et al. (2019), who quantified the influence of eastern air-masses on eastern Germany. They found the contribution of trans-boundary transport from eastern European countries to be 13 μg/m^3 on average, depending on meteorological conditions. Air masses from continental eastern Europe tend to transport higher amounts of particles, whereas western, more maritime air tends to be cleaner

due to precipitation along the trajectories of air masses. Source regions of PM10 include industrial and residential areas in Poland and the Czech Republic with heavy industries or extensive usage of solid fuels for residential heating (Kiesewetter et al., 2015; Reizer and Juda-Rezler, 2016; Beloconi et al., 2018; van Pinxteren et al., 2019). Results by Grange et al. (2018) also show increased values of PM10 for northern and northeastern wind directions, although to a lesser extent. In winter, the effect of particle transportation is strongest, presumably due to increased emissions from domestic heating in eastern Europe (Reizer and Juda-Rezler, 2016; van Pinxteren et al., 2019). A slight increase in modelled PM10 at low values of instantaneous east-west wind components is visible (see Fig. 3.7). This could indicate insufficient mixing of the atmosphere, which would lead to an accumulation of particles near ground (Chudnovsky et al., 2013b). Overall, instantaneous wind information has little influence on the model, causing the PD to remain relatively constant. The relatively constant PD of instantaneous wind information implies little influence on PM10 predictions. This suggests that wind information needs to be extended to a longer time scale to influence PM predictions. As shown in Grange et al. (2018), wind speed aggregated for a daily period can substantially influence PM10 concentrations, with lower speeds causing higher concentrations. Park et al. (2019) also found the maximum wind speed of previous three hours to be of importance for their statistical predictions of PM10. The north-south wind component PDs (instantaneous and 3-day) do not provide clear trends.

BLH and CAPE

The PD of BLH shows that the model is able to reproduce the pattern of decreasing particle concentrations with increasing BLH (Gupta and Christopher, 2009b; Wagner and Schäfer, 2017). The shape of the full-year PD of BLH shown in Fig. 3.8d is similar to that provided by Grange et al. (2018) for observations in Switzerland. They found a reduction of $\sim 8\,\mu\text{g/m}^3$ for daily PM10 predictions. A reduction in mean PM10 predictions of $\sim 10\,\mu\text{g/m}^3$ for situations with higher BLH can be seen in Fig. 3.8a. This similarity is encouraging and proves the robustness of the modelling approach, since both studies use ERA-Interim BLH in a comparable geographic setting.
For BLH values above $\sim 800\,\text{m}$, the PD remains constant, i.e., the influence of the boundary layer on PM concentrations stagnates (Liu et al., 2018). Mean modelled PM10 concentrations increase slightly in conditions with very high BLH ($>2,000\,\text{m}$). This pattern could be related to the formation of a deep, convective boundary layer coinciding with high temperatures, enhancing the formation of secondary aerosols (Grange et al., 2018). The abundance of radiation, high temperatures, and precursor gases at excess concentrations would be needed therefore (Fuzzi et al., 2015). Indeed, the pattern is most prominent in summer, when these prerequisites are most likely to be met. The PD of BLH in summer is almost constant, i.e., little information is provided to the model. Fig. 3.8d reflects a shift of the frequency of occurrence of data points towards higher
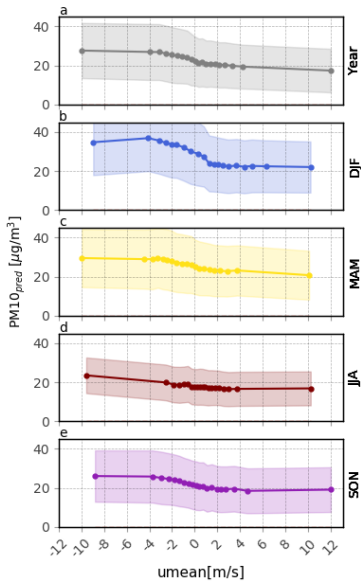
Figure 3.6: Partial Dependence plot showing the mean model response to changes in 3-day mean east-west wind component [m/s] as bold lines for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, SON). Shaded areas show the range of the Individual Conditional Expectation (ICE) lines (10th to 90th percentiles). The horizontal distance of dots on the bold line indicates the distribution of valid data points. Negative (positive) values represent dominant inflow of eastern (western) air masses.

BLH: during summer months, medium to high BLHs are more likely to occur due to enhanced convection. As mentioned before, a BLH above 800 m provides only little information to the model, thus reducing its predictive ability. The accumulation effect of a lower BLH is most pronounced in winter with increased mean PM10 predictions of almost $20\,\mu\mathrm{g/m^3}$. In addition, PM10 emissions are expected to be higher in wintertime due to combustion of solid fuels for domestic heating. This has been shown for Eastern European countries (Reizer and Juda-Rezler, 2016; van Pinxteren et al., 2019), and likely influences PM10 concentrations in Germany. Thus, with higher locally produced or advected PM10 emissions, the accumulation effect of a low BLH is more distinct in winter than in summer, where reduced emissions are expected and an accumulation
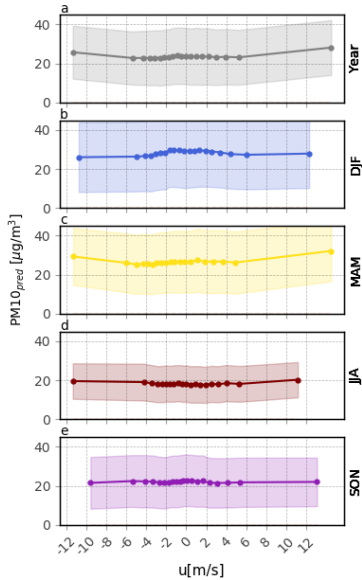
Figure 3.7: PD plot showing the mean model response to changes in east-west wind component [m/s] as bold lines for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, SON). Negative (positive) values represent dominant inflow of eastern (western) air masses. Description as in Fig. 3.6.

of particles is not expected (Wagner and Schäfer, 2017). A dependence of model PM10 on CAPE could not be identified using the PD approach.

Two-way PD: umean and BLH

Depicting the PD of BLH and the 3-day mean east-west wind component simultaneously allows for the quantification of particularly high-polluted situations, considering combined effects of both features. A probability density function (PDF) using Gaussian kernel density estimates is added to provide a qualitative estimate of the frequency of occurrence of PD values. Values outside the PDF-estimation as shown in Fig. 3.9 are extrapolated based on the trained model and do not represent observed data.

The two-way PD suggests mean modelled PM10 concentrations double due to changes in BLH and
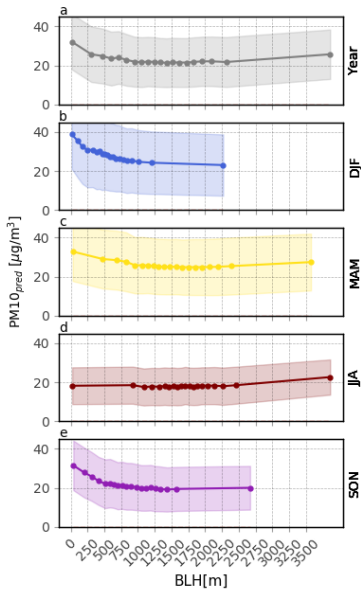
Figure 3.8: PD plot showing the mean model response to changes in instantaneous BLH [m] for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, SON). Description as in Fig. 3.6.

wind flow, referring to the full-year model. Highest mean predictions ($\sim$35 μg/m^3) are modelled when eastern winds coincide with shallow boundary layers, whereas lowest mean predictions occur during medium BLH and western winds ($\sim$17 μg/m^3). Note that due to the tendency of the model to underestimate high PM10 levels, concentrations could be higher in reality. Patterns differ seasonally. In winter, highest mean PM10 predictions surpass 45 μg/m^3 during shallow BLH-conditions and wind flow from the east. In summer, this is not the case. As mentioned in chapter 3.3.3, there is indication of elevated PM10 concentrations during very high BLH ($\sim$>2,000 m) conditions, coinciding with eastern wind flow. Highest mean predictions in summer do not surpass $\sim$22 μg/m^3 (within the limits of the PDF estimation).

Figure 3.9: Two-way PD of umean and BLH, full-year model (a) and seasonal models (b–e, DJF, MAM, JJA, SON). Similar to the previously shown PD, the two-way PD shows the mean modelled PM10 (color-coded from light orange to black) to the isolated effects of umean and BLH. On top of the PD-isolines, a probability density function (PDF) using Gaussian kernel density estimates is added to indicate the qualitative frequency of occurrence of values from high (yellow) to low (dark blue).

67

Thermal influence on PM10

In spring, summer, and autumn, positive temperature anomalies cause a marked increase of mean model PM10 predictions (Fig. 3.10). Presumably, higher temperatures in spring and summer (Figs 3.10c, 3.10d) reflect enhanced biogenic activity, as vegetation is generally more active at higher temperatures. Consequently, emissions of primary particles such as debris or pollen and the emission of BVOCs are stimulated (Laothawornkitkul et al., 2009). With higher BVOC emission, an enhancement of secondary SOA formations is leveraged (Megaritis et al., 2013; Churkina et al., 2017). In addition to increased biogenic activities, higher temperatures cause the soil to dry up more quickly, thus increasing dust emissions (Hoffmann and Funk, 2015). In winter, positive anomalies have very little effect on predicted PM10 concentrations (Fig. 3.10b). This supports these hypotheses, since neither increased biogenic activity nor dried-up soils are to be expected in winter. Higher temperatures could reflect increased photochemical oxidation processes, which trigger photochemical reactions leading to new particle formation processes (Birmili and Wiedensohler, 2000; Größ et al., 2018). However, there was no trend in the partial dependence of SSRD, which shows a weak influence of SSRD on PM10 predictions (see scientific appendix C, Fig. C.1). Note however that the lacking influence of SSRD could also be due to the fact that this study is confined to cloud-free situations due to the availability of AOD. The variation of SSRD would be higher when including cloudy days, which would likely improve the information content provided to the model by including SSRD. Another possible explanation for increased PM10 concentrations at higher temperatures could be that these situations are associated with stable synoptic conditions (at least in spring, summer, and autumn), causing particles to accumulate in the atmosphere. PD trends for instantaneous temperature (Fig. 3.11) are inverse to those described for temperature anomalies. The full-year model PD shows a decrease in predicted PM10 concentrations for higher temperatures. Presumably, instantaneous temperature reflects the annual cycle of PM10 (similar to DOY), which is why the seasonal PDs show no trends. PD of air temperature as presented in Grange et al. (2018) reveal a different pattern. In their study, temperatures below freezing are associated with high PM10 concentrations, medium temperature in the range of 0–15 °C with low PM10 concentrations and temperature above 15 °C with high PM10 concentrations. However, when combining the effects of temperature anomalies and temperature presented in Figs 3.10 and 3.11, the emerging pattern would be similar. This suggests that the model presented in this study is able to discriminate between the seasonal component of temperature and the immediate effect of temperature on particle emissions, e.g., due to new particle formation (Birmili and Wiedensohler, 2000; Bressi et al., 2013; Petetin et al., 2014; Größ et al., 2018).

Figure 3.10: *PD plot showing the mean model response to changes in temperature anomalies [K] for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, SON). Description as in Fig. 3.6.*

RH, precipitation

Increased RH is associated with higher PM10 predictions (see Fig. 3.12). This is likely related to an increase in AOD due to aerosol swelling in humid conditions (Wang, 2003; Crumeyrolle et al., 2014)

Other than increasing AOD, the importance of RH could also be related to a correlation between BLH and RH (see PDF-estimate in Fig. 3.13). Higher RH reduces the magnitude of turbulent vertical flux and subsequently reduces the BLH (Adamopoulos et al., 2007; Petäjä et al., 2016), which in turn could increase PM10 predictions. On the other hand, vertical transport of water vapor is impeded by a low BLH, which increases RH and stimulates formation of aqueous secondary aerosols (Liu et al., 2018). To analyse the influences of BLH and RH on predicted PM10, the two-way PD of RH and BLH is calculated (see Fig. 3.13). It shows a changing pattern with increasing RH, visible in a change of line structure orientation from vertical to more horizontal in the two-
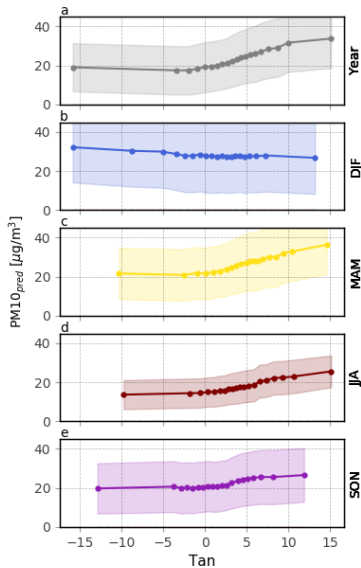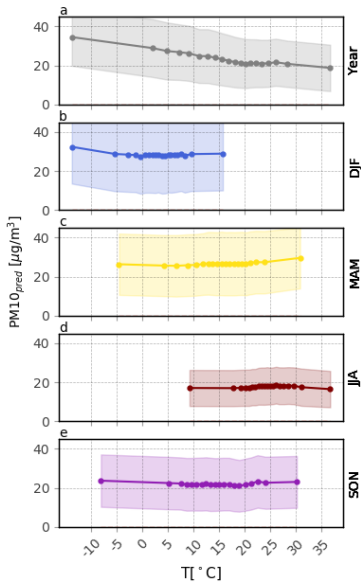
Figure 3.11: PD plot showing the mean model response to changes in instantaneous temperature [°C] for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, SON). Description as in Fig. 3.6.

way PD plot. While BLH dominates during shallow boundary layer conditions (vertical lines), the influence of RH is more prominent at higher BLH values (horizontal lines). This pattern points to an influence of RH on PM10 predictions, which is decoupled from the BLH, as the influence of BLH above 800 m is marginal. A similar pattern is found for all seasons except for summer. In summer, the model outcome does not show any response to changes in BLH, as there are rarely any BLH values below 800 m (see scientific appendix D, Fig. D.1). A study by Belle et al. (2017) conducted in the United States found RH to have positive impact on PM2.5 concentrations during cloud-free conditions due to an increase in sulfate and nitrate masses. However, they found PM2.5 to decrease with increasing RH during cloudy conditions.

The more time passed since the last precipitation event, the higher the PM10 prediction tends to be, reflecting the accumulation of local emissions in the atmosphere. The influence of this effect on the model is not pronounced and stagnates from about 100 hours (see scientific appendix C,

Figure 3.12: PD plot showing the mean model response to changes in RH [%] for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, SON). Description as in Fig. 3.6.

Fig. C.2). The magnitude of the last precipitation event and accumulated precipitation of last 24 h were not included in the model due to lacking importance as determined in the feature selection (see chapter 3.2). Note however that the low importance of precipitation could be related to the consideration of only cloud-free days in this study. Thus, the immediate effect of rainfall on particles in the atmosphere cannot be investigated. In addition, possible effects of precipitation along the trajectories of advected air masses are not covered.

NDVI, Corine Land Cover and spatiotemporal factors

The NDVI was of minor importance for the prediction of PM10 concentrations. No trends or seasonal differences were found by application of the PD approach. However, with increasing number of pixels in the vicinity of a PM station classified as agricultural areas (2CLC), PM10 concentrations tend to be higher. Likely, this is related to primary emission of dust from arable
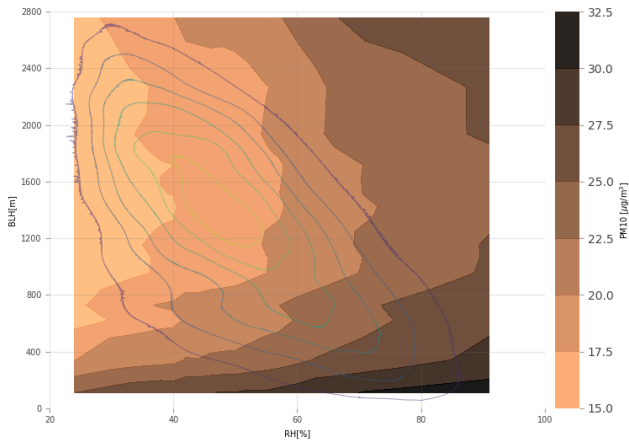
Figure 3.13: Two-way PD of RH and BLH, full-year model. Description as in Fig. 3.9.

lands and the application of fertilizers (NH_3), which constitute important precursors for secondary particle formation (Hoffmann and Funk, 2015; Wagner et al., 2015). For the other land cover classes, no trends were found using the PD approach. Lower PM10 concentrations are predicted on Saturdays and Sundays, indicating that reduced anthropogenic activity (less traffic, reduced industrial production) has an immediate effect on PM10 concentrations. Increasing altitude slightly reduces the mean PM10 prediction, possibly due to lower population density and more effective pollution dispersion processes (Hu et al., 2014; Beloconi et al., 2018). The PDs of NH_3, continentality, PM10 mean annual emissions and of the TPI do not show a distinct trend. SO_2 was excluded from the model during the feature selection process.

3.3.4 Determinants of the relationship between AOD and PM10

The full-year model and the seasonal models associate increasing AOD with increasing PM10 (see Fig. 3.14). This pattern is less distinct in summer (except for very high AOD) when particles are generally more dispersed within a well-mixed boundary layer and the AOD is largely determined by particles higher up in the atmosphere, thus weakening the relation between AOD and PM10. The relationship between AOD and PM10 is not bivariate and can be modified by ambient meteorology (Guo et al., 2009; Gupta and Christopher, 2009b; Sorek-Hamer et al., 2017; Stirnberg et al., 2018). A quantification of this effect is approached here by using the two-way PD method.
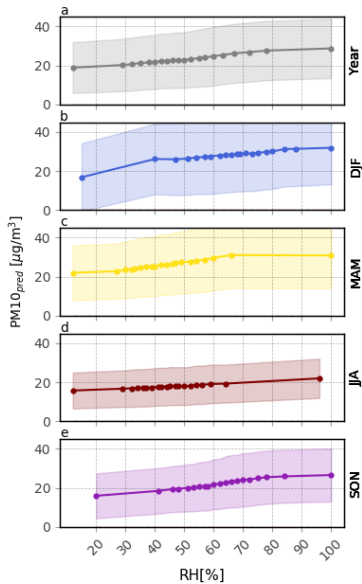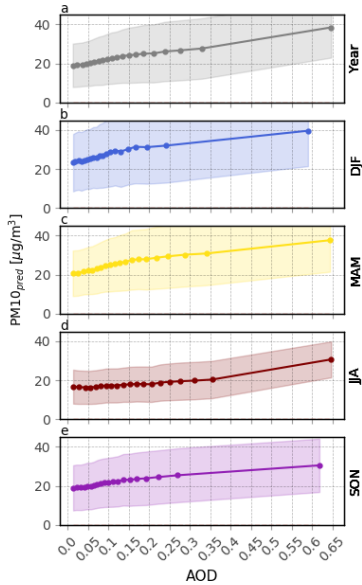
Figure 3.14: PD plot showing the mean model response to changes in AOD for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, SON). Description as in Fig. 3.6.

The two-way PD of AOD and BLH reveals a dependence of the model on both AOD and BLH (see Fig. 3.15). The importance of interactive effects of these features can be illustrated by the following example: assume an AOD of 0.2 and BLH of 2,000 m versus an AOD of 0.2 and BLH of 200 m. In the latter case, the mean predicted PM10 concentration is $\sim$10 µg/m^3 higher as the aerosol content determining AOD is closer to the ground and thus more relevant for the PM10 prediction. In other words, a prediction based on AOD (assuming that AOD is largely determined by attenuation in the boundary layer (Schäfer et al., 2008)) alone would lead to erroneous PM10 predictions, as AOD does not fully capture the particle accumulation effect of a shallow boundary layer (Stirnberg et al., 2018). Similar effects can be observed for the two-way PD of AOD and the 3-day mean east-west wind component (see Fig. 3.16) and for the two-way PD of AOD and temperature anomalies (plot not shown). The two-way PD of AOD and BLH shows only a minor seasonal pattern, which is mostly driven by BLH (see scientific appendix D, Fig. D.2).

Figure 3.15: Two-way PD of AOD and BLH, full-year model. Description as in Fig. 3.9.

Figure 3.16: Two-way PD of AOD and umean, full-year model. Description as in Fig. 3.9.

Westerly wind flow (positive umean) leads to substantially lower PM10 predictions when compared to similar AOD values in situations dominated by easterly wind flow (negative umean). The two-way PD suggests this effect to be as large as $\sim 8\,\mu g/m^3$ (see Fig. 3.16). Air masses from the east possibly carry a higher amount of near-ground particles (Bonn et al., 201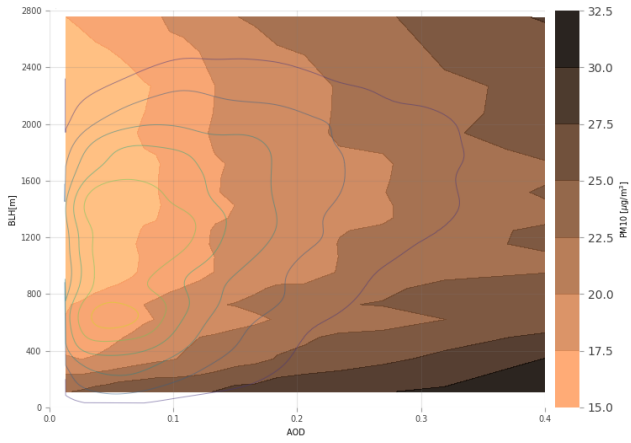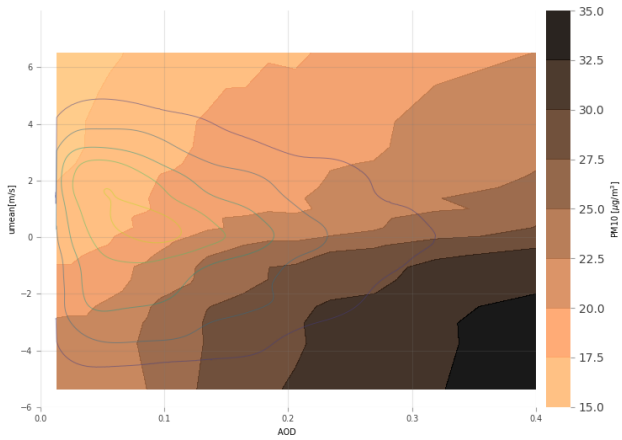6; Reizer and Juda-Rezler, 2016; Beloconi et al., 2018), affecting PM10 observations more strongly than AOD. Another reason for the observed effect could be that western air masses carry a relatively large amount of sea salts with a high hygroscopic growth factor. By effectively taking up water, these constituents enhance light scattering, thus increasing AOD without increasing PM10 measurements (Zieger et al., 2013, 2014; Stirnberg et al., 2018). The seasonality for the two-way PD of AOD and umean is weak (see scientific appendix D, Fig. D.3)

3.4 Specific conclusions

A machine learning model is used to advance the understanding of drivers of near-ground PM10 and the capability to use satellite AOD to infer on PM10. Parameters pertaining to meteorology, land cover, and satellite-based AOD are considered and related to hourly PM10 concentrations. The model performs well (overall R^2 of 0.77, RMSE $= 7.44\,\mu g/m^3$) and provides a basis to assess sensitivities. These allow for the isolation and quantification of effects of ambient conditions on PM10. Overall, the model is more sensitive to meteorological conditions than to land cover parameters. BLH, east-west winds, DOY, temperature, and RH are identified as the important driving factors of PM10 variations. Representing regional particle transport, the 3-day mean of the east-west wind component substantially modifies PM10 concentrations, depending on the direction of inflow. Eastern inflow generally increases PM10 concentrations. Modelled PM10 concentrations were also increased during higher than average temperatures. Possibly, this is due to stimulated vegetation activity, increasing primary particle and precursor gas emissions. The influence of BLH is most prominent at very low ($\sim<500\,m$) values. However, there is indication that very high BLH values ($\sim<2,500\,m$) influence PM10 concentrations as well. While the former threshold marks the effects of particle accumulation within a shallow boundary layer, the latter threshold could indicate the formation of a deep boundary layer with stimulated formation of secondary aerosols as suggested by Grange et al. (2018). If BLH is between these thresholds, its explanatory power is limited. In these situations, other processes determine PM10 concentrations. Overall, the model outcome suggests that there are different meteorological boundary conditions that potentially cause elevated PM10 concentrations in winter and summer (Fig. 3.17).

In winter, very shallow boundary layers, coinciding with multi-day easterly wind flow cause highest mean PM10 predictions ($>30\,\mu g/m^3$). Mean PM10 predictions for these conditions are as high as $40\,\mu g/m^3$ (see Fig. 3.9). This is probably related to higher anthropogenic emissions in winter and frequently low BLH.
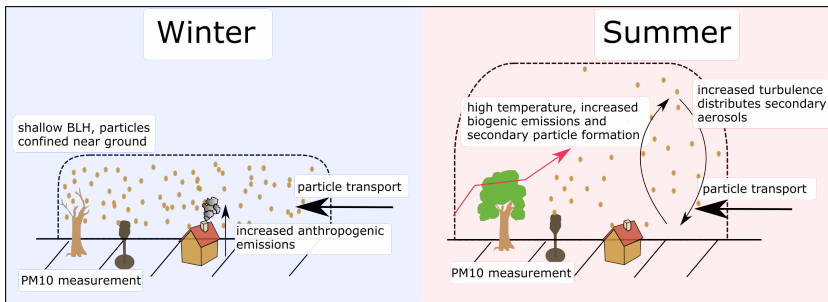
Figure 3.17: Schematic representation of different processes driving high pollution situations in winter (left) and summer (right).

In summer, higher temperatures associated with increased formation of secondary aerosols and coinciding with multi-day easterly wind flow lead to mean predicted PM10 concentrations $>27\,\mu g/m^3$ (plot not shown). High PM10 concentrations in summer appear to be largely uncoupled from changes in BLH (see Fig. 3.8d). The model R^2 decreases in summer, suggesting that the statistical model does not as well resolve these processes. In addition, the relationship between AOD and PM10 is weaker in summer.

Results presented in this study suggest that meteorology plays a substantial role in the development of high pollution situations. This has potential implications for plans towards better air quality in high-polluted areas, as meteorological conditions need to be taken into account, e.g., for temporary traffic bans. In addition, there is a need to introduce measures to reduce air pollution on a regional scale. Measures limited to city scales can only decrease pollution levels associated with local emission sources, which can be superimposed by transported particles.

The importance of AOD for the statistical model highlights the suitability of AOD for air quality studies. However, potential implications and limitations for the use of satellite AOD for air quality studies are described. This study has shown that satellite-derived AOD can be used to infer street-level PM10 concentrations, if ambient meteorological conditions are taken into account explicitly. In particular, temperature anomalies, the east-west regional wind component and BLH modify the relationship between PM10 and AOD. A drawback of including AOD is the restriction to cloud-free situations, which potentially introduces a bias due to non-random data gaps (Belle et al., 2017). Depending on the situation and location, both an over- or underestimation of PM10 could be the consequence (Belle et al., 2017). In addition, the influence of certain meteorological variables could be underestimated due to important processes under cloudy conditions, which are not covered (Belle et al., 2017; Brokamp et al., 2018).

The use of GBRT proved fruitful to understand interconnected processes and the approach presented here can be potentially expanded to other research questions focusing on the understanding multivariate processes. Future efforts will further address the determination of mechanisms leading to high pollution events using machine learning not only for total PM10 concentrations, but for individual aerosol species.

4 Meteorology-driven, local-scale variability of PM1

4.1 A case study on drivers of air pollution in suburban Paris

In chapter 3 it is shown that meteorology is a major driver of PM10 concentrations on a regional scale. In this chapter, an advancement of the scientific understanding of processes leading to high pollution events is targeted. The scale is narrowed to a case-study setup, allowing to resolve atmospheric processes at a high level of detail. Fine-mode particle concentrations (PM1) are related to local meteorological conditions, and differences between the main chemical constituents are shown. Excluding the coarse particle fraction, the importance of transformation processes and new particle formation processes can be analysed with high accuracy under various atmospheric conditions, as these are more relevant for the fine-mode fraction (see section 1.2).

Measurements at the SIRTA supersite southwest of Paris are used for the analysis. The site provides both meteorological and PM-data and is representative for suburban conditions (Haeffelin et al., 2005). The metropolitan area of Paris is one of the most densely populated and industrialised areas in Europe. Thus, air quality is a recurring issue and has been at the focus of studies in the past years (Bressi et al., 2014; Petetin et al., 2014; Petit et al., 2015; Dupont et al., 2016; Petit et al., 2017; Srivastava et al., 2018). Results indicate that the Paris metropolitan region is often affected by mid- to long-range transport of pollutants, as due to the city's flat terrain, an efficient horizontal exchange of air masses is frequent (Bressi et al., 2013; Petit et al., 2015). High-pollution events often occur in late autumn, winter, and early spring. Often, these episodes are characterised by stagnant atmospheric conditions and a combination of local contributions, e.g., traffic emissions, residential emissions, or regionally transported particles, e.g., nitrates from manure spreading, or sulfates from point sources (Petetin et al., 2014; Petit et al., 2014, 2015; Srivastava et al., 2018). Air masses originating from continental Europe (Belgium, Netherlands, western Germany) generally enhance particle concentrations, especially of SIA (Bressi et al., 2013; Srivastava et al., 2018). Hence the variability between high-pollution episodes in terms of timing, sources and meteorological boundary conditions is considerable (Petit et al., 2017).

A GBRT model is set up to predict PM1 and localised and individualised attributions of feature contributions are performed using SHapley Additive exPlanation regression (SHAP) values (Lundberg and Lee, 2017; Lundberg et al., 2018a, 2020). With the use of SHAP values, a

detailed insight to the decisions of the statistical model can be provided (Lundberg et al., 2018a). The attribution of situation-based statistical feature contributions enables quantitative estimates of drivers of PM concentrations and allows to infer on processes driving PM concentrations at high temporal resolution. Building on existing literature, the contribution of meteorological parameters to high PM1 concentrations are quantified and different processes driving peak particle concentrations are highlighted. Although the analysis is conducted on a local scale, the results are also relevant for other regions and have implications for large-scale studies, e.g., by improving the description of meteorology-dependent variations of PM concentrations in physical models.

4.2   Data and methods

Data sets

Seven years (2012–2018) of meteorological and air quality data from the Site Instrumental de Recherche par Télédétection Atmosphérique (SIRTA, Haeffelin et al., 2005) supersite are the basis of this study. The SIRTA Atmospheric Observatory is located about $25\,km$ southwest of Paris ($48.713\,°N$ and $2.208\,°E$, Fig. 4.1). This study focuses on day-to-day variations of total and speciated PM1. To separate diurnal effects e.g., the development of the boundary layer during morning hours (Petit et al., 2014; Dupont et al., 2016; Kotthaus and Grimmond, 2018a) from day-to-day variations of PM1, mean concentrations of total and speciated PM1 for the afternoon period 12–15 UTC are considered. The diurnal cycle of PM concentrations is thus not included in the data. In section 4.2, the PM1 and meteorological data and preprocessing steps before setting up the machine learning model are described. A description of the applied machine learning model and data analysis techniques follows.

Submicron particle measurements

Aerosol chemical speciation monitor (ACSM, Ng et al., 2011) measurements are conducted at SIRTA. Species measured include Org, NH_4^+, SO_4^{2-}, NO_3^-, and Cl^-. In addition, BC is monitored by a seven-wavelength Magee Scientific aethalometer (model AE31, Petit et al., 2015), which provides measurements of BC, divided into BCff and BCwd. As an additional input to the machine learning model, the fraction of NO_3^- of the previous day is added (NO3_frac). Pollution events dominated by NO_3^- are often linked to regional-scale events, which depend on anthropogenically-influenced processes in the source regions of NO_3^- precursors (Petit et al., 2017). This is approximated by the inclusion of the fraction of NO_3^- of the previous day.

79
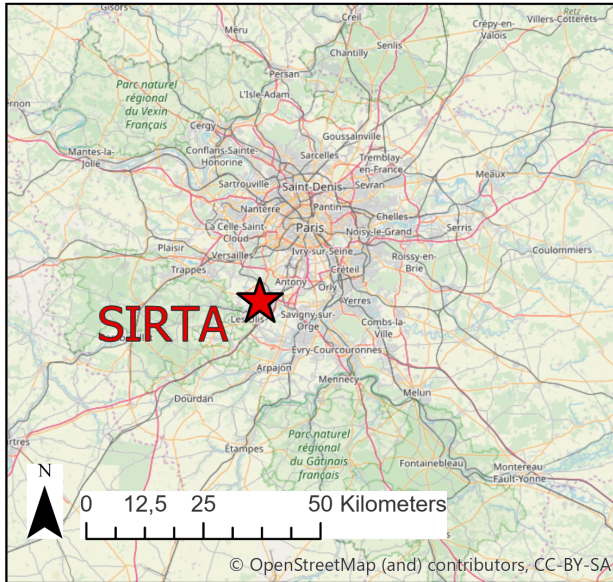
Figure 4.1: Location of the SIRTA supersite southwest of Paris. © OpenStreetMap contributors 2020. Distributed under a Creative Commons BY-SA License.

Meteorological data

Meteorological variables included in this study are air temperature (in 2 m height, T), relative humidity (RH), air pressure (AirPres), precipitation, wind speed (ws), wind direction (u, v) and total incoming solar radiation (TISR). Data are taken from a quality-controlled and 1 h averaged dataset. Further information on the instrumentation used for the acquisition of these variables is provided in Chiriaco et al. (2018). BLH is derived from automatic lidar and ceilometer measurements of a Vaisala CL31 ceilometer using the CABAM algorithm (characterising the atmospheric boundary layer based on automatic lidar and ceilometer measurements, Kotthaus and Grimmond, 2018a,b). Due to an instrument failure, during the period July to mid-November 2016, SIRTA ALC measurements had to be replaced with measurements conducted at the Paris Charle de Gaulle Airport, which is located northeast of Paris.

Meteorological factors are chosen as input features for the statistical model based on findings of previous studies (see section 1.2). Meteorological observations are converted to suitable input information for the statistical model (see section 4.2). Wind speed (ws) is calculated from the u and

v components [m/s] and the maximum wind speed of the afternoon period is included in the model. U and v wind components are then normalised to values between 0 and 1, thus only depicting the direction information. To reduce the impact of short-term fluctuation in wind direction, the 3-day mean wind direction is calculated based on the normalised u and v wind components (umean and vmean). Hours since the last precipitation event (Precip_tsince) are counted and used as input to capture the particle accumulation effect between precipitation events (Rost et al., 2009; Petit et al., 2017).

Methods

Machine learning model: technique and application

GBRT (used in a python 3.6.4 environment with the scikit-learn module, Friedman, 2002; Pedregosa et al., 2012) are applied to predict daily total and speciated PM1 concentrations. A detailed description of the functionally of GBRT models is provided in section 3.2.

Characteristics of the meteorological training data set with respect to observed total and speciated PM1 concentrations are conveyed to the statistical model. The learned relationships are then used for model interpretation and to produce estimates of PM1 based on unseen meteorological data to test the model. Model hyperparameters are tuned by executing a grid search, systematically validating testing previously defined hyperparameter combinations and determining the best combination via a three-fold cross validation. Note that PM1 data are not normally distributed, i.e., there is more data available for mid-range concentrations. To avoid that the model primarily optimizes its predictions on these values, a least-squares loss function was chosen, which is more sensitive to higher PM1 values (i.e., outliers of the PM1 data distribution).

For each PM species, a specific GBRT model is set up and used for the analysis of meteorological influences on individual PM1 species (see section 4.3.2). Additionally, a quasi-total PM1 model is used to predict the sum of all species at once, which is used for an analysis of meteorological drivers of high-pollution events (see sections 4.3.3 and 4.3.4). Train and test data sets to evaluate each model are created by randomly splitting the full data set. These splits, however, are the same for the species models and the full PM1 model to ensure comparability between the models. Three quarters of the data are used for training and hyperparameter tuning with cross-validation (n=1086), and one quarter for testing (n=363). In addition, the robustness of the model results is tested by repeating this process ten times, resulting in ten models with different train- / test-splits and different hyperparameters.

Explaining model decisions: SHapley Additive exPlanation (SHAP) values

While being powerful predictive models, tree-based machine learning methods also have a high interpretability (Lundberg et al., 2020). In order to understand physical mechanisms on the basis of model decisions, the contributions of the meteorological input features to the model outcome are analysed. Feature contributions are attributed using SHAP values, which allow for an individualised, unique feature attribution for every prediction (Shapley, 1953; Lundberg and Lee, 2017; Lundberg et al., 2018a, 2020). SHAP values provide a deeper understanding of model decisions than the relatively widely used partial dependence plots (Friedman, 2001; Goldstein et al., 2015; Fuchs et al., 2018; Lundberg et al., 2018a; McGovern et al., 2019; Stirnberg et al., 2020). Partial dependence plots show the global mean effect of an input feature to the model outcome, while SHAP values quantify the feature contribution to each single model prediction, accounting for multicollinearity. Feature contributions are calculated from the difference in model predictions with that feature present, versus predictions for a retrained model, without the feature. Since the effect of withholding a feature depends on other features in the model due to interactive effects between the features, differences are computed for all possible feature subset combinations of each data instance (Lundberg and Lee, 2017).

Summing up SHAP values for each input feature at a single time step yields the final model prediction. SHAP values can be negative since SHAP values are added to the base value, which is the mean prediction when taking into account all possible input feature combinations. Negative (positive) SHAP values reduce (raise) the prediction below (above) the base value. The higher the absolute SHAP value of a feature, the more distinct is the influence of that feature on the model predictions. The sum of all SHAP values at one time step yields the final prediction of PM1 concentrations. An example of breaking down a model prediction into feature contributions using SHAP values is shown schematically in Fig. 4.2. The computation of traditional Shapley Regression values is time consuming, since a large number of all possible feature combinations have to be included. The SHAP framework for tree-based models allows a faster computation compared to full shapley regression values while maintaining a high accuracy (Lundberg and Lee, 2017; Lundberg et al., 2018a) and is therefore used here. The shap python implementation is used for the computation of SHAP values (https://github.com/slundberg/shap).

Pairwise interactive effects between input features can be estimated using the SHAP approach. Interactive effects are defined as the difference between the SHAP values for one feature when a second feature is present and the SHAP values for the one feature when the other feature is absent (Lundberg et al., 2018a). SHAP values are a novel tool to better understand multivariate natural systems, in particular when applied in state-of-the-art machine learning models as GBRT. So far, SHAP values have been used in the field of medical science (Lundberg et al., 2018b; Li et al., 2019a; Lundberg et al., 2020), but have yet to be applied to study environmental systems.
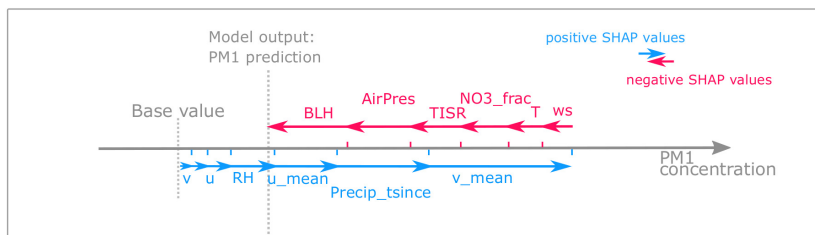
Figure 4.2: Conceptual figure illustrating the interaction of SHAP values and model output. Starting with a base value, which is the mean prediction if all data points are considered, positive SHAP values (blue) increase the final prediction of total and speciated PM1 concentrations, while negative SHAP values (red) decrease the prediction. The sum of all SHAP values for each input feature yields the final prediction. Depending on whether positive or negative SHAP values dominate, the prediction is higher or lower than the base value (Lundberg et al., 2018b). Adapted from https://github.com/slundberg/shap.

4.3 Results and discussion

4.3.1 Model performance

The performance of the ten model iterations is assessed by comparing the coefficient of determination (R^2) and normalised root mean square error (NRSME) for independent test data (Fig. 4.3). While the models for BCwd, BCff and total PM1 show small spread, Cl^- and NO_3^- exhibit larger variations between model runs (indicated by horizontal and vertical lines in Fig. 4.3). This suggests that while drivers of BCff concentration variations are well covered by the model, this is less so in the case of Cl^- and NO_3^-. Possible reasons for this are that no explicit information on anthropogenic emissions or chemical formation pathways are included in the models. Still, the model performance indicators highlight that a large fraction of the variations in particle concentrations are explained by the meteorological variables used as model inputs. Performances of model iterations of the species-specific and total PM1 are generally similar, suggesting a robust model outcome.
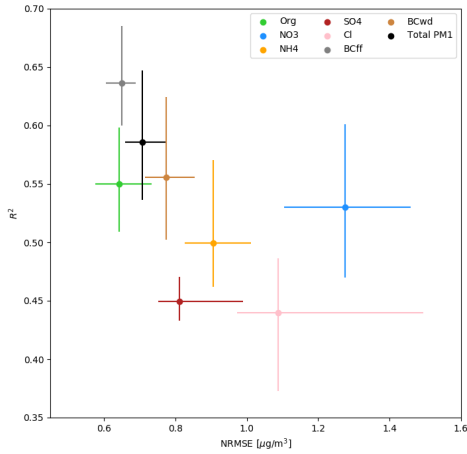
Figure 4.3: Performance indicators for ten model iterations: coefficient of determination R^2 against normalised Root Mean Squared Error (NRMSE) for the separate species models and the total PM1 model. Vertical and horizontal lines indicate the maximum spread in R^2 and NRMSE, respectively.

The mean input feature importance, ordered from high to low, of the total PM1 model run by means of the SHAP feature attribution values is shown in Fig. 4.4, The NO_3^- fraction of the previous day has the highest impact on the model, followed by temperature, wind direction information, and BLH. Hence, the impact of these meteorological variables on model decisions is analysed in more detail in the following.
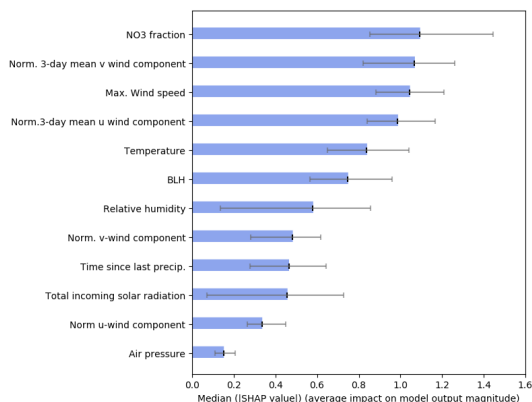
Figure 4.4: Ranked median SHAP values of the model input features, i.e., the average absolute value that a feature adds to the final model outcome, referring to the total PM1 model [µg/m³] (Lundberg et al., 2018b). Horizontal lines indicate the variability between model runs.

4.3.2   Sensitivity of PM1 species concentrations to changes in meteorology

To gain insights on relevant processes governing particle concentrations in the Paris region, the contribution of input features on species and total PM1 concentration predictions from the statistical model, i.e., the SHAP values, are plotted as a function of absolute feature values (Figs 4.5–4.7). The contribution of an input feature to each prediction of the species or total PM1 concentrations is shown while taking into account intra-model variability. Intra-model variability of SHAP values, i.e., different SHAP value attributions for the same feature value within one model, is shown by the vertical distribution of dots for absolute input feature values. Intra-model variability is caused by interactions of the different model input features.

Contribution of temperature

The impact of ambient air temperature on modelled species concentrations is highly non-linear (Fig. 4.5). All species show increased contributions to predictions at temperatures below ~4 °C while the contribution of high temperatures on model predictions differs substantially between species. The statistical model is able to reproduce well-known characteristics of species concentration variations related to temperature. For example, sulfate formation is enhanced with

increasing temperatures (Fig. 4.5d) due to an increased oxidation rate of SO_2 (Dawson et al., 2007; Li et al., 2017) and strong solar irradiation due to photochemical oxidation (Gen et al., 2019). Dawson et al. (2007) reported an increase of 34 ng/m^{3}K for PM2.5 concentrations using a CTM. The increase in sulfate at low ambient temperatures is not reported in this study. It is likely linked to increased aqueous phase particle formation in cold and foggy situations (Rengarajan et al., 2011; Petetin et al., 2014; Cheng et al., 2016). Considerable local formation of nitrate at low temperatures (Fig. 4.5b) is consistent with results from a previous study in the area by Petetin et al. (2014). Enhanced formation of ammonium nitrate at lower temperatures (Fig. 4.5c) by the shifting gas-particle equilibrium has been previously reported by Clegg et al. (1998); Pay et al. (2012); Bressi et al. (2013); Petit et al. (2015). The increase in organic matter and BCwd concentrations at low temperatures (Fig. 4.5g) is likely related to human behaviour, as biomass burning is often used for domestic heating in the study area (Favez et al., 2009; Sciare et al., 2010; Healy et al., 2012; Jiang et al., 2019). In addition, organic matter concentrations are linked to the condensation of semi-volatile organic species at low temperatures (Putaud et al., 2004; Bressi et al., 2013). The sharp increase in modelled concentrations of organics above 25 °C (Fig. 4.5a) could be due to enhanced biogenic activity leading to a rise in biogenic emissions and secondary aerosol formation (Guenther et al., 1993; Churkina et al., 2017; Jiang et al., 2019).

The contribution of temperature on modelled total PM1 concentrations (Fig. 4.5h) is consistent with the response patterns to changes in temperatures described for the individual species in panels 4.5a–g, with positive contributions at both low (<4 °C) and high air temperatures (>25 °C). For temperatures below freezing, the model allocates maximum contributions to modelled total PM1 concentrations of up to 12 μg/m^3. The spread of SHAP values between model iterations is generally higher for low temperatures (vertical grey bars in Fig. 4.5), where absolute values are of greater magnitude, but in all cases the signal contained in the feature contributions far exceeds the spread between model runs. The inter-model variability of the feature contribution depends to some extent on the distribution of the different training data sets. This effect is particularly pronounced for BCff, which shows very high variability at low temperatures. As BCff concentrations depend on anthropogenic emissions, BCff predictions highly depend on the diversity of emissions included in the training data.
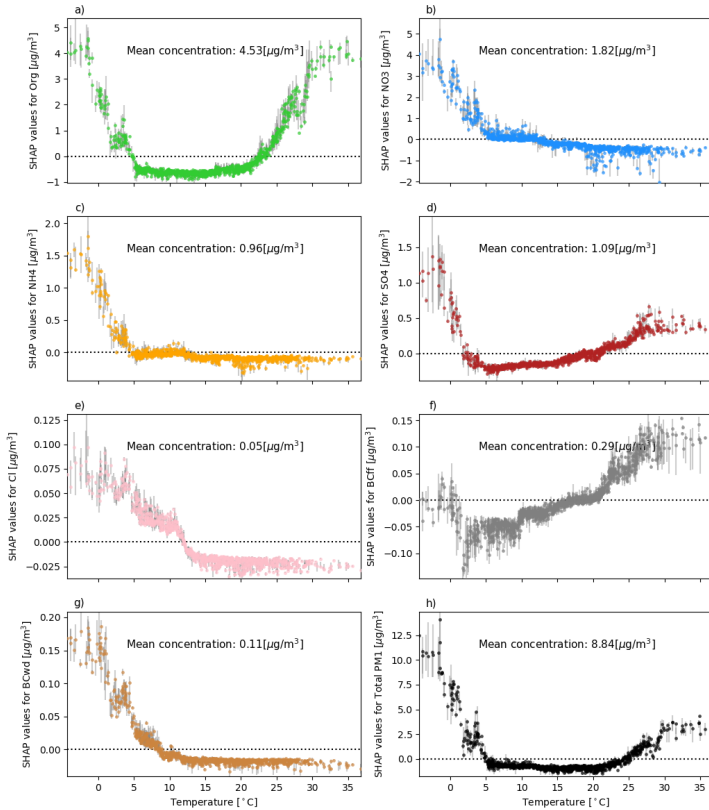
Figure 4.5: Air temperature SHAP values (contribution of temperature to the prediction of species and total PM1 concentrations [µg/m³] for each data instance) vs. absolute air temperature [°C]. Inter-model variability of allocated SHAP values is shown as the variance of predicted values between the ten model iterations and plotted as vertical grey bars. The dotted horizontal line indicates the transition from positive to negative SHAP values.

Contribution of BLH

Variations in BLH can have a substantial impact on near-surface particle concentrations, as the mixed layer is the atmospheric volume in which the particles are dispersed (Klingner and Sähn, 2008; Dupont et al., 2016; Wagner and Schäfer, 2017). The effect of BLH variations on modelled particle concentrations is highly nonlinear for all species (Fig. 4.6), with the magnitude of the contribution varying by species. Possible reasons for this will be discussed in the

following. Similar to the pattern observed for temperature SHAP values, the inter-model variation of predictions is highest for low BLHs since predicted particle concentrations then have the highest variation. Contributions of BLH to predicted particle concentrations are highest for very shallow mixed layers due to the accumulation of particles close to the ground under shallow BLH conditions (Dupont et al., 2016; Wagner and Schäfer, 2017). A relatively distinct transition from positive contributions during shallow boundary layer conditions ($\sim$0–800 m) towards negative contributions at high BLHs is evident for all species except SO_4^{2-}. Modelled SO_4^{2-} concentrations (Fig. 4.6d) show a less distinct response to low BLHs as they are largely driven by point sources and thus susceptible to particle advection, independent of BLH (Pay et al., 2012; Petit et al., 2014, 2015). The accumulation effect under low BLH conditions is hence less important. Furthermore, an increase of SO_4^{2-} concentrations with higher BLHs ($>\sim 1,500$ m) could be due to a more effective transport of SO_4^{2-} and its precursor SO_2 under high BLH conditions (Pay et al., 2012).

In agreement with results from previous studies focusing on PM10 (Grange et al., 2018; Stirnberg et al., 2020) or PM2.5 (Liu et al., 2018), SHAP values do not change much for BLH above $\sim$800–900 m, i.e., boundary layer height variations above this level do not strongly influence submicron particle concentrations. Positive contributions for BLHs above $\sim$800–900 m have been previously reported by Grange et al. (2018), who relate this pattern to enhanced secondary aerosol formation in a very deep and dry boundary layer. The positive contributions of high BLHs on SO_4^{2-} and organics could be explained following this argumentation. Processes driving the positive contribution of high BLHs on BCff, however, remain inconclusive. For predicted total PM1 concentrations, the maximum positive contribution of the BLH is as high as 5.5 $\mu g/m^3$ while negative contributions can amount to -2 $\mu g/m^3$. While the maximum contribution by BLH is lower than the maximum contribution determined for air temperature, the frequency of shallow BLH is far greater than that of the minimum temperatures that have the largest effect (Figs 4.5d and 4.6d). In addition to causing particles to accumulate near the surface, low BLH can also provide effective pathways for local new particle formation. Secondary pollutants, such as nitrate and ammonium, are increased at low BLHs as conditions favorable to the formation of nitrate usually coincide with shallow BLH conditions (i.e., low temperatures, often in combination with high RH, Pay et al., 2012; Petetin et al., 2014; Dupont et al., 2016; Wang et al., 2016). BC concentrations, however, are mainly associated with primary emissions, as is a substantial fraction of organic matter (Petit et al., 2015). Hence, the accumulation of these particles during low turbulence conditions can explain the strong influence of BLH on BCwd and BCff.
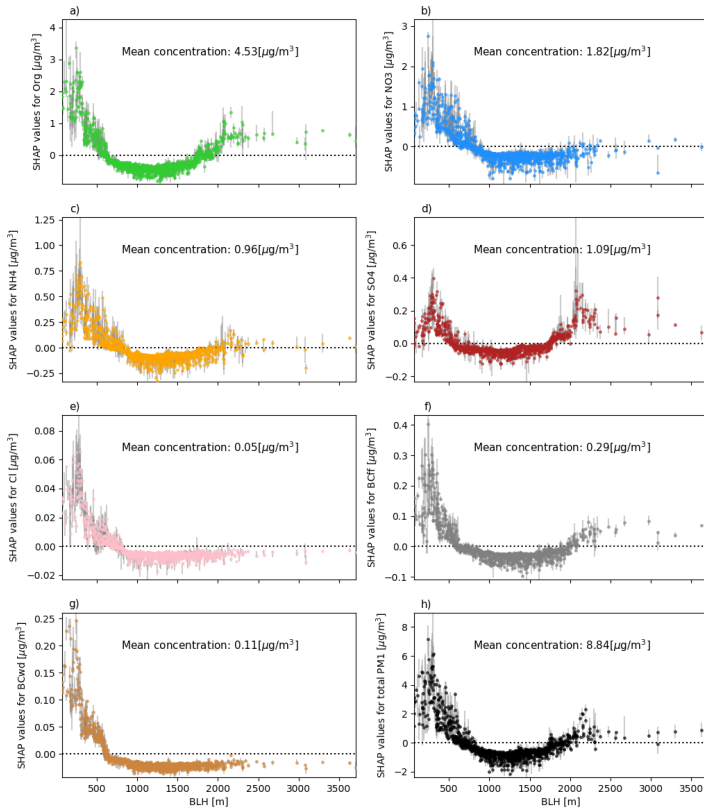
Figure 4.6: As Fig. 4.5 for BLH SHAP values (contribution of BLH to the prediction of species and total PM1 for each data instance) vs. absolute BLH values [m].

Contribution of wind direction

To analyse the contribution of wind direction to predicted particle concentrations, SHAP values of normalised 3-day mean u and v wind components were added up and transformed to degrees (Fig. 4.7). Generally, contributions of wind are highest when winds from the northern to northeastern sectors prevail, while negative values are evident for southwesterly directions. Given the location of the measurement site, this pattern undoubtedly reflects the advection of particles emitted in the urban area of Paris and from continental Europe versus cleaner marine air masses

during southwestern inflow (Bressi et al., 2013; Petetin et al., 2014; Petit et al., 2015; Srivastava et al., 2018). Increased concentrations of organic matter are predicted for northerly, northeasterly and easterly winds (Fig. 4.7a). These patterns suggest a significant contribution of advected organic particles from a specific wind sector. This is in agreement with the findings of Petetin et al. (2014) who estimated that 69 % of the PM2.5 organic matter fraction is advected by northeasterly winds, which is related to advected particles from wood burning sources in the Paris area and SOA formation along the transport trajectories. While Petit et al. (2015) did not find a wind direction dependence of organic matter measured at SIRTA using wind regression, they reported the regional background of organic matter to be of importance. Comparing upwind rural stations to urban sites, Bressi et al. (2013) concluded organic matter is largely driven by mid- to long-range transport. Contributions to the SO_4^{2-}-model (Fig. 4.7d) are highest for northeastern and eastern wind direction, which aligns with previous findings by Pay et al. (2012); Bressi et al. (2014); Petit et al. (2017), who identified the Benelux region and western Germany as strong emitters of sulfur dioxide (SO_2). SO_2 can be transformed to particulate SO_4^{2-} (Pay et al., 2012) while being transported towards the measurement site. A further possible source is the Orly Airport to the east of SIRTA as jet exhausts also contain SO_2 (Kapadia et al., 2016). Nitrate concentrations are affected by long-range transport from continental Europe (Benelux, western Germany), which are imported towards SIRTA from northeastern directions (Petetin et al., 2014; Petit et al., 2014). It is to be expected that the influence of mid- to long-range transport on the particle observations at SIRTA is rather substantial, with most high pollution days affected by particle advection from continental Europe (Bressi et al., 2013). As shown by Petetin et al. (2014), SIA species like NO_3^-, SO_4^{2-} or NH_4^+ are rarely formed within the Paris city area, but advected from further away, at least during the spring and summer months. Concerning BCff and BCwd, model results suggest a dependence on wind direction during northwestern to northeastern inflow (Figs 4.7f and g). Although BC concentrations are expected to be largely determined by local emissions (Bressi et al., 2013), e.g., from local residential areas, a substantial contribution of imported particles from wood burning and traffic emissions from the Paris region (Laborde et al., 2013; Petetin et al., 2014) and continental sources is likely (Petetin et al., 2014).
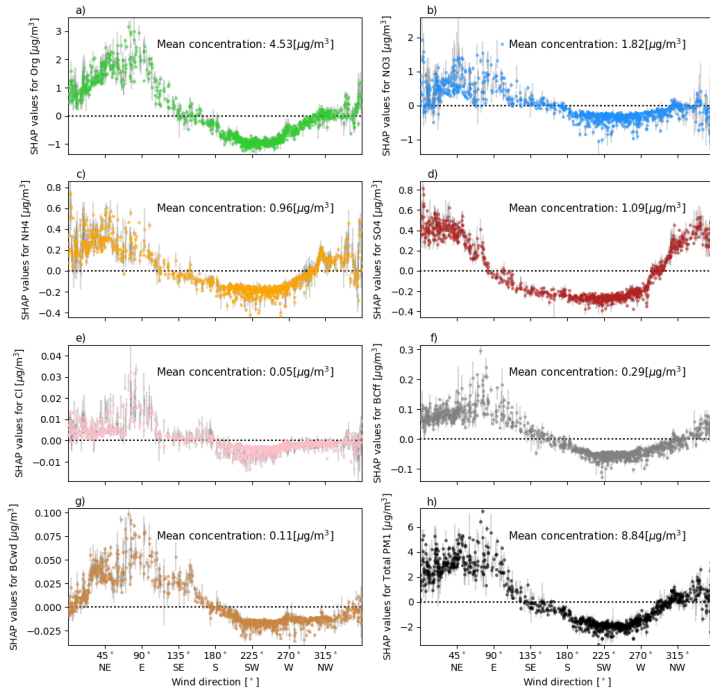
Figure 4.7: As Fig. 4.5 for wind direction SHAP values (contribution of 3-day mean wind direction to the prediction of species and total PM1 for each data instance) vs. absolute wind direction [°].

Contribution of feature interactions

Pairwise interactive effects between BLH vs. time since last precipitation and BLH vs. maximum wind speed are investigated in Figs 4.8a and 4.8b. SHAP interaction effects between BLH and time since last precipitation are most pronounced for BLHs below $\sim 500\,\mathrm{m}$ (Fig. 4.8a). Interaction values are negative for low BLHs paired with time since last precipitation close to zero hours. With increasing time since last precipitation, interaction effects become positive, thus increasing the contribution of Precip_tsince and BLH to the model outcome. An explanation of this pattern concerning underlying processes could be that due to the lack of precipitation, a higher number of particles is available in the atmosphere for accumulation, hence increasing the accumulation effect of a shallow BLH. In case of recent precipitation, the accumulation effect of a shallow BLH

91

is weakened. For higher BLHs, interactive effects with time since the last precipitation event are marginal. Interactive effects between BLH and wind speed are shown in Fig. 4.8b. Positive SHAP values for maximum wind speeds below ~2 m/s reflect stable situations, favoring the accumulation of particles, whereas high wind speeds enhance the ventilation of particles (Sujatha et al., 2016). This can also be deduced from Fig. 4.8b, which shows increased SHAP values for low wind speeds in combination with a low BLH. Low wind speeds combined with a high BLH (>~1,000 m), on the other hand, result in decreased SHAP values. Similarly, low BLHs combined with higher wind speeds (>~2 m/s) also decrease predictions of total PM1 concentrations.
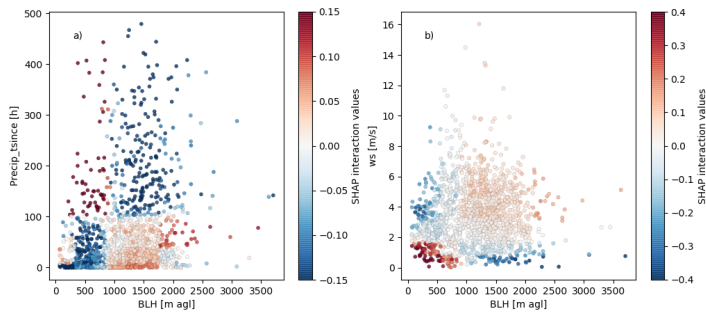
Figure 4.8: BLH vs. a) time since last precipitation and BLH vs. b) maximum wind speed, respectively, colored by the SHAP interaction values for the respective features.

4.3.3 Meteorological conditions of high-pollution events

To further identify conditions that favor high pollution episodes, the data set is split into situations with exceptionally high total PM1 concentrations (>95th percentile) and situations with typical concentrations of total PM1 (interquartile range, IQR) for the meteorological summer and winter seasons. Mean SHAP values refer to the total PM1 model, corresponding input feature distributions and species fractions for the two subgroups are aggregated seasonally. This allows for a quantification of seasonal feature contributions to average or polluted situations. Figs 4.9 and 4.10 show mean SHAP values for typical (left) and high-pollution (right) situations in the upper panel. The distribution of SHAP values are shown as box plots for each feature. Absolute feature value distributions are given in the bottom of the figure. In the lowest subpanel, the chemical composition of the total PM1 concentration for each subgroup is shown. The largest contributor to high pollution situations in winter is air temperature (Fig. 4.9). SHAP values for temperature are substantially increased during high pollution situations, when temperatures are systematically

lower. Further contributing factors to high pollution situations are the BLH, wind speed, NO_3^- fraction of the previous day, and the wind direction components. In winter, the PM1 composition shows a relatively large fraction of nitrates, which is increased during high pollution situations (Fig. 4.9, lower panel). High concentrations of nitrate in winter can be linked to advection or to enhanced formation due to the temperature-dependent low volatility of ammonium nitrate (Petetin et al., 2014). The organic matter fraction is slightly decreased during high pollution situations. BLH and maximum wind speed contributions to high pollution situations are linked to low ventilation conditions which are very frequent in winter (Dupont et al., 2016). Positive contributions of wind direction for inflow from the northern and eastern sectors are dominant during high pollution situations while inflow from the southern and western sectors prevails during average pollution situations (see Fig. 4.7, Bressi et al., 2013; Petetin et al., 2014; Srivastava et al., 2018). Note that the time since the last precipitation is increased during high pollution situations, but the effects on the model outcome is weak. This suggests that lacking precipitation is not a direct driver of modelled total PM1 concentrations, but increases the contribution of other input features (see Fig. 4.8a) or is a meaningful factor in only some situations.

Figure 4.9: *Mean feature contributions (i.e., SHAP values) for situations with low total PM1 concentrations (left) and situations with high pollution (right), respectively, during winter (December, January, February). Respective range of SHAP values by species are shown as box plots, with median (bold line), 25–75th percentile range (boxes), and 10–90th percentile range (whiskers). Both training and test data are included. Absolute feature value distributions (given as normalised frequencies) as well as the chemical composition of the total PM1 concentration are shown in the subpanels. Colors of the box plots correspond to colors in the feature distribution subpanels. SHAP values of the input features u_norm_3d and u_norm as well as v_norm_3d and v_norm were merged to u_norm, merged and v_norm, merged to achieve better transparency.*

Summer total PM1 composition (Fig. 4.10) is characterised by a larger fraction of organics compared to the winter season (Fig. 4.9). As a considerable fraction of organic matter is formed locally (Petetin et al., 2014), the increased proportion of organics could be due to more frequent stagnant synoptic situations that may limit the advection of transported SIA. In addition, the positive SHAP values of solar irradiation and temperature highlight that the solar irradiation stimulates transformation processes and increases the number of biogenic SOA particles (Guenther et al., 1993; Petetin et al., 2014). As mean temperatures are highest in summer, positive temperature SHAP values are associated with increased organic matter concentrations (Fig. 4.5). The higher importance (i.e., higher SHAP values) of time since the last precipitation event during high pollution situations points to an accumulation of particles in the atmosphere. Dry situations can also enhance the emission of dust over dry soils (Hoffmann and Funk, 2015). The negative contribution of BLH during both typical and high pollution situations reflects seasonality, as afternoon BLHs in summer are usually too high to have a substantial positive impact on total PM1 concentrations (see Fig. 4.6). BLH is thus not expected to be a driver of day-to-day variations of summer total PM1 concentrations. Note that the average BLH is higher during high pollution situations, which likely points to increased formation of SO_4^{2-} (see Fig. 4.6).
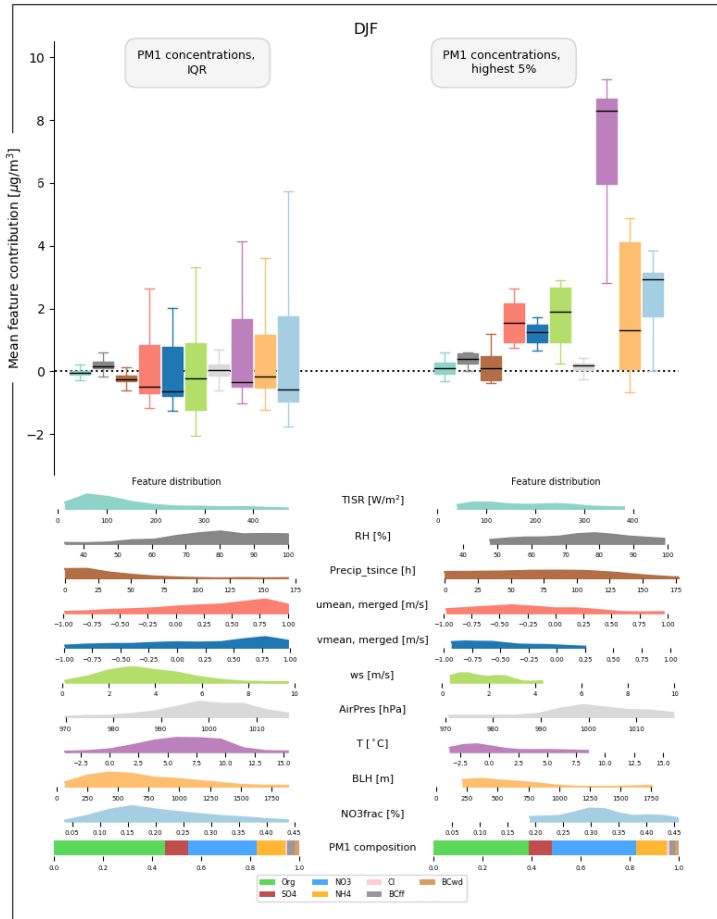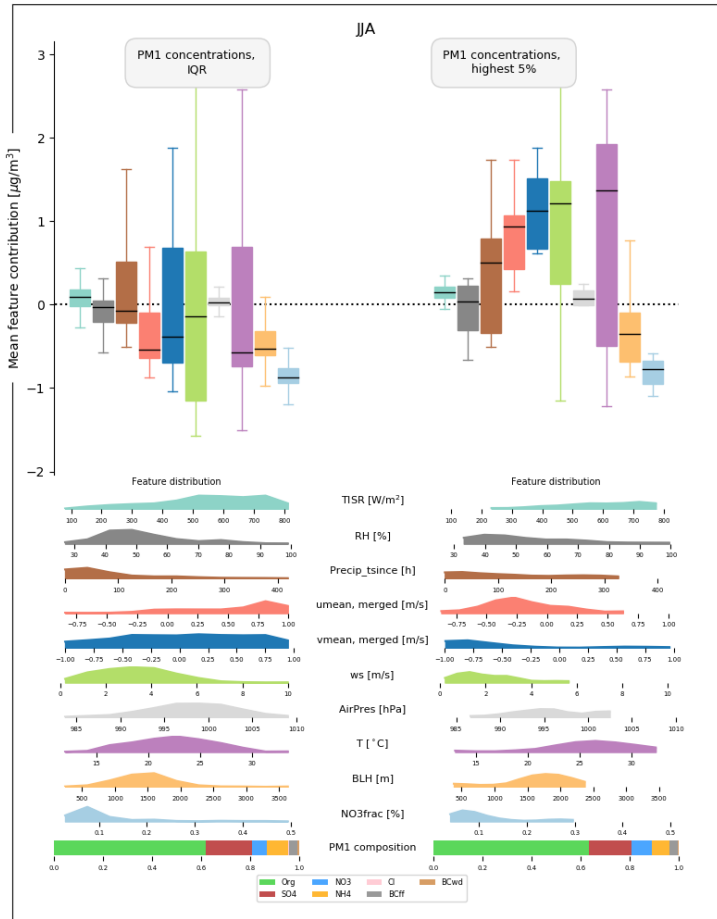
Figure 4.10: As Fig. 4.9 for mean feature contributions (i.e., SHAP values) for situations with low total PM1 concentrations (left) and situations with high pollution (right), respectively, during summer (July, June, August).

4.3.4 Day-to-day variability of selected high pollution events

Analysing the combination of SHAP values of the various input features on a daily basis allows for direct attribution of the respective implications for modelled total PM1 concentrations (Lundberg et al., 2020). Here, four particular pollution episodes are selected to analyse the model outcome

with respect to physical processes (Figs 4.11–4.14). The examples highlight the advantages but also the limitations of the interpretation of the statistical model results. The high pollution episodes took place in winter 2016 (10th–30th January and 25th November–25th December), spring 2015 (11th–31st March) and summer 2017 (8th–28th June). The upper panels in Figs 4.11–4.14 indicate the total PM1 prediction as horizontal black line with vertical black lines denoting the range of predictions of all 10 models. The observed species concentrations are shown in the corresponding colors. The subsequent panels show absolute values and SHAP values for the most relevant meteorological input features.

January 2016

Prior to the onset of the high-pollution episode in January 2016 (Fig. 4.11), the situation is characterised by BLHs in the range of 1,000 m, temperatures above freezing ($\sim$5–10 °C), frequent precipitation and winds from the southwest. The organic matter fraction dominates the particle speciation. The episode itself is reproduced well by the model. According to the model results, the event is largely temperature-driven, i.e., SHAP values of temperature explain a large fraction of the total PM1 concentration variation (note the adjusted y-axis of the temperature SHAP values). On 18th January, temperatures drop below freezing, coupled with a decrease in BLH. As a consequence, both modelled and observed PM1 concentrations start to rise. A further increase in total PM1 concentrations is driven by a sharp transition from stronger southwestern to weaker northeastern winds (strong negative u component, weak negative v component) on January 19th. The combined effects of these changes lead to a marked increase in total PM1 concentrations, peaking at $\sim$37 μg/m^3 on 20th January. On the following days, temperatures increase steadily, thus the contribution of temperature decreases. At the same time, although values of BLH remain almost constant, the contribution of BLH drops substantially from $\sim$5 μg/m^3 to $\sim$2 μg/m^3. This is due to interactive effects between BLH and the features wind speed, time since last precipitation and normalised v-wind-component. All of these features increase the contribution of BLH on 20th January, but decrease its contribution on 21st–23rd January. The high pollution episode ceases after a shift to southeastern winds and the increasing temperatures. The pollution episode is characterised by a relatively large fraction of NO$_3^-$ and NH$_4^+$, which explains the strong feature contribution of temperature to the modeled total PM1 concentration, as the abundance of these species is temperature dependent (see Fig. 4.5) and points to a large contribution of locally formed inorganic particles. Still, the contribution of wind direction and speed also suggests that advected secondary particles and their build-up in the boundary layer are relevant factors during the development of the high pollution episode (Petetin et al., 2014; Petit et al., 2014; Srivastava et al., 2018).
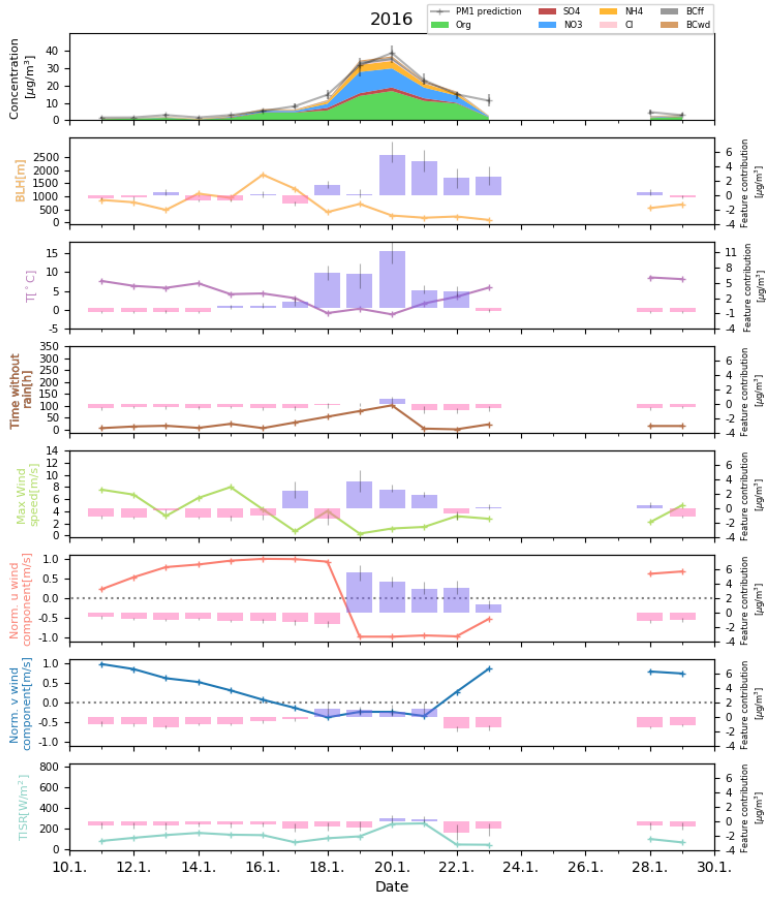
Figure 4.11: Winter pollution episode in January 2016. (a): predicted total PM1 and observed PM1 species concentrations, with absolute input feature values and corresponding SHAP values of (b) BLH, (c) temperature, (d) hours after rain, (e) maximum wind speed (f) normalised u wind, and (g) normalised v wind component.

December 2016

A high-pollution episode with several peaks of total PM1 is observed in November and December 2016. The first peak on 26th December is followed by an abrupt minimum in total PM1 concentrations on 28th November, and a build-up of pollution in a shallow boundary layer towards the second peak on 2nd December with total PM1 concentrations exceeding $40\,\mu g/m^3$. In the following days, total PM1 concentrations continuously decrease, eventually reaching a second minimum on 11th December. A gradual increase in total PM1 concentrations follows, resulting in a third (double-)peak total PM1 concentration on 17th December. Total PM1 concentrations drop to lower levels afterwards. Throughout the 3.5 week-long episode, high pollution is largely driven by shallow BLH ($<\sim$500 m), and weak north-northeasterly winds, i.e., a regime of low ventilation and advection of continental air. During the brief periods with lower total PM1 concentrations, these conditions are disrupted by a higher BLH ($\sim$28th November), or a change in prevailing winds ($\sim$11th December). In contrast to the pollution episode in January 2016, this December 2016 episode is not driven by temperature changes. Temperatures range between $\sim$5–12 °C and have a minor contribution to predicted total PM1 concentrations (see also Fig. 4.5), emphasizing the different processes causing air pollution in the Paris region. The large contributions of wind direction and speed to the modeled total PM1 concentration highlight the importance of advected particles during this period that continuously build-up in the shallow boundary layer to produce these highly polluted conditions. Note that the model is not able to fully reproduce the pollution peak on 2nd December, which may be indicative of missing input features in the model. Judging from the PM1 species composition during this time (relatively high fraction of NO_3^- and BC), it seems likely that missing information on particle emissions may be the reason for the difference between modeled and observed total PM1 concentration.
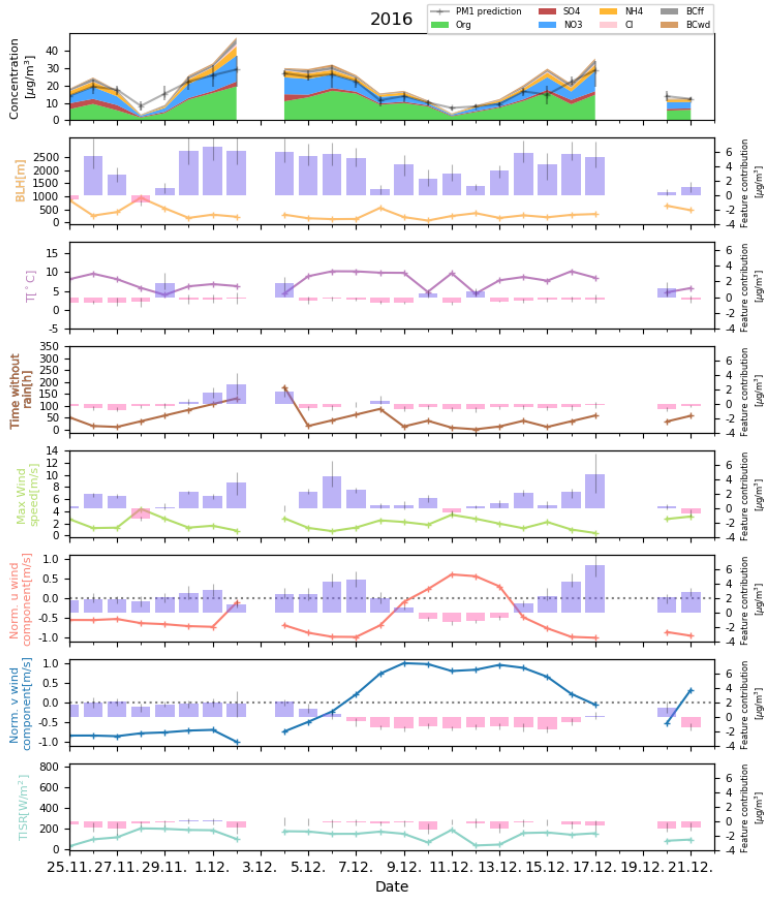
Figure 4.12: As Fig. 4.11 for a further winter pollution episode in December 2016.

June 2017

A period of above average total PM1 concentrations occurred in June 2017. The episode is very well reproduced by the model, suggesting a strong dependence of the observed total PM1 concentration to meteorological drivers. Although absolute total PM1 concentrations are substantially lower than during the previously described winter pollution episodes, the event is still above average for summer pollution levels. Organic matter particles dominate the PM1 fraction throughout the episode, with a relatively high SO_4^{2-} fraction. Conditions during this episode are characterised by strong solar irradiation (positive SHAP values) and high BLHs (mostly negative SHAP values), which show low day-to-day variability and reflect characteristic summer conditions. A lack of precipitation (no rain for a period of more than 2 weeks) and high temperatures also contribute to the total PM1 concentrations during this episode. While solar irradiation and time since last precipitation are associated with positive SHAP values throughout this period, air temperature only has a positive contribution when exceeding $\sim$25 °C. This aligns with patterns shown in Fig. 4.5, where increased concentrations of organic matter and SO_4^{2-} are identified for high temperatures. Peak total PM1 concentrations of $\sim$17 µg/m^3 are observed on June 20th and 21st. A change in the east-west wind component from western to eastern inflow directions in conjunction with an increase in temperatures to above 30 °C are the drivers of the modeled peak in total PM1 concentrations. BLH is also increased with values $\sim$2,000 m, which are associated with slightly positive SHAP values. This observation fits with findings described previously and is likely linked to enhanced secondary particle formation (Megaritis et al., 2014; Jiang et al., 2019). As suggested by response patterns of species to changes in BLH shown in Fig. 4.6, this effect is linked to an increase in SO_4^{2-} concentrations. The main fraction of the peak total PM1 values, however, is linked to an increase in organic matter concentrations due to the warm temperatures (see Fig. 4.5).
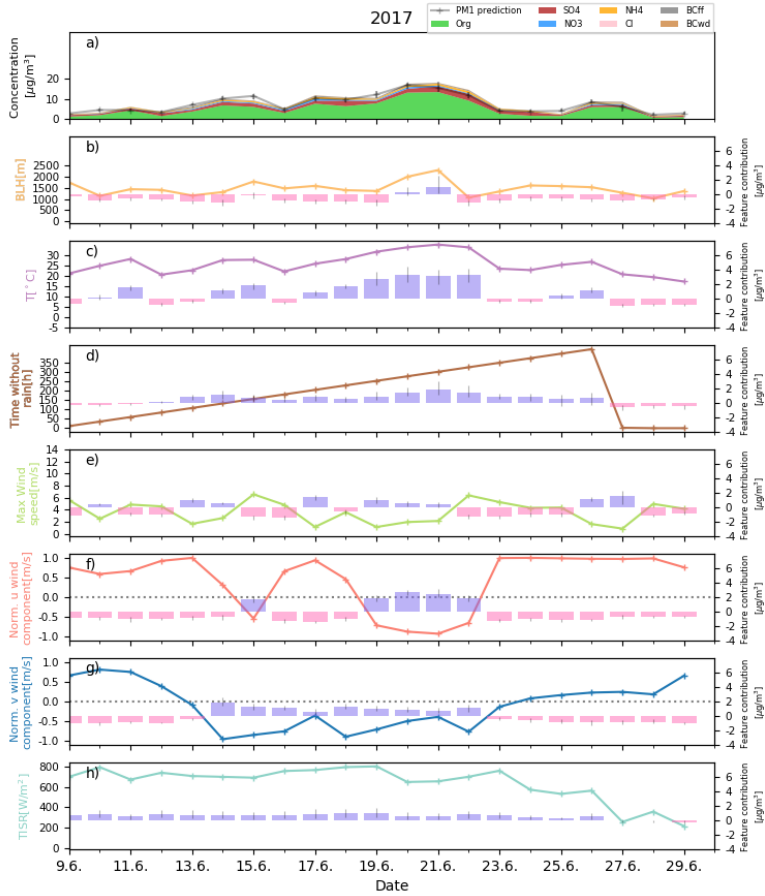
Figure 4.13: As Fig. 4.11 for an exemplary summer pollution episode in June 2017.

March 2015

High particle concentrations are measured in early March 2015 with high day-to-day variability. This modelled course of the pollution episode is chosen to compare results to previous studies focusing on the evolution of this episode (Petit et al., 2017; Srivastava et al., 2018). The episode is of particular interest, since it also shows some limitations of the statistical model setup. The episode is characterised by high fractions of SIA, in particular SO_4^{2-}, NH_4^+ and NO_3^- (Fig. 4.14, upper panel) and similar concentrations observed at multiple measurement sites in France (Petit et al., 2017). Contributions of local sources are low and much of the episode is characterised by winds blowing in from the northwest, advecting aged SIA towards SIRTA (Petit et al., 2017; Srivastava et al., 2018). A widespread scarcity of rain probably enhanced the large-scale formation of secondary pollution across western Europe (in particular western Germany, the Netherlands, Luxemburg, Petit et al., 2017), which were then transported towards SIRTA. This is reflected by the SHAP values of the u and v wind components, which are positive throughout the episode (see Figs 4.14g and 4.14h). Concentration peaks of total PM1 are measured on 18th and 20th March. Both peaks are characterised by a rapid development of total PM1 concentrations. As described in Petit et al. (2017), these strong daily variations of total PM1, which are mainly driven by the SIA fraction, could be due to varying synoptic cycles, especially the passage of cold fronts. The contribution of BLH and temperature is relatively small, which is consistent with the high influence of advection on total PM1 concentrations during the episode. The exceptional character of the episode (see Petit et al., 2017) partly explains the bad performance of the model in capturing total PM1 variability during the event. Unusual rain shortage is observed in large areas of western Europe prior to the episode (Petit et al., 2017). While time since precipitation at the SIRTA-site is a large positive contributor to the model predictions (see Fig. 4.14d), it is not driving the day-to-day variations. The unusual nature of this event and lacking information on emission in the source regions and formation processes along air mass trajectories in the model likely explain why the model has difficulties in reproducing this pollution episode. While this has implications for the application of explainable machine learning models for rare events, this is not expected to be an issue for the other cases and seasonal results presented here.
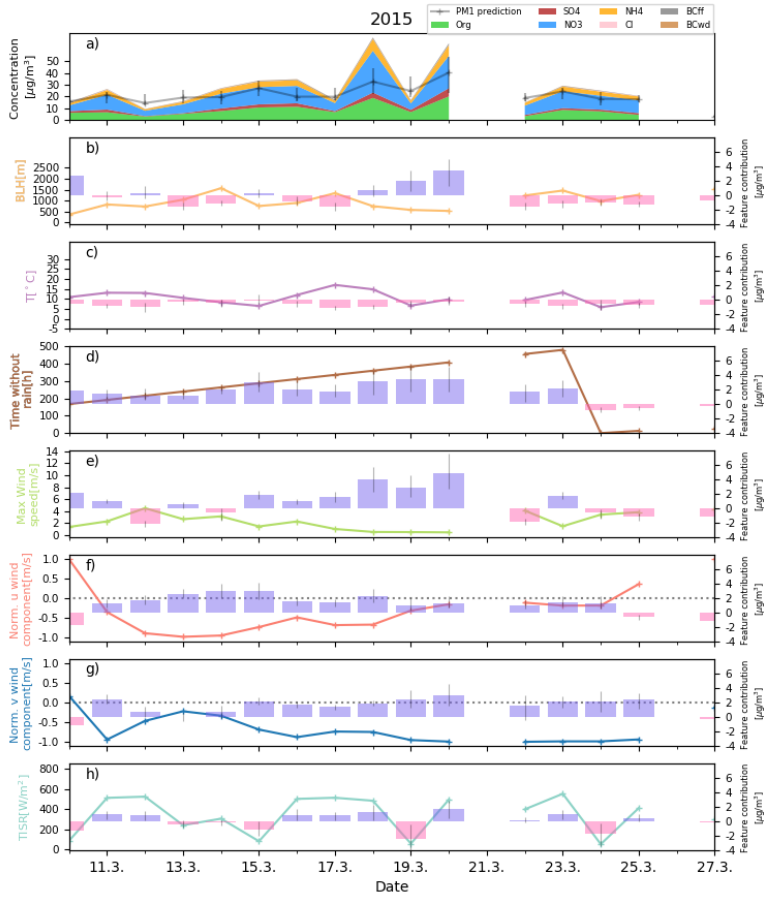
Figure 4.14: As Fig. 4.11 for an exemplary spring pollution episode in March 2015.

4.4   Specific conclusions

In this study, dominant patterns of drivers of PM1 species and total PM1 concentrations are identified and analysed using a novel, data-driven statistical approach. A machine learning model is set up to predict measured speciated and total PM1 concentrations based on meteorological measurements from the SIRTA supersite, southwest of Paris. The statistical model is able to reproduce daily variability of particle concentrations well, and is used to analyse and quantify the atmospheric processes causing high-pollution episodes during different seasons using a SHAP-value framework. Comparison of the results based on the statistical model with findings of previous studies on air quality patterns in the Paris region shows good agreement. Peak concentrations of total PM1 in winter are mainly driven by shallow BLHs, low temperatures and low wind speeds. These conditions are often amplified by northeastern wind inflow. A detailed analysis revealed different drivers of high-pollution episodes in winter. For an episode in January 2016, model results show a strong contribution of temperature to the elevated PM1 concentrations during this episode (up to $11\,\mu\mathrm{g/m^3}$ are attributed to temperature), suggesting enhanced local, temperature-dependant particle formation. During a different, prolonged pollution episode in December 2016, temperature levels were relatively stable and had no influence. Here, contributions of BLHs ($<500\,\mathrm{m}$) were quantified to be the main driver of PM1 peaks with contributions up to $\sim 6\,\mu\mathrm{g/m^3}$, along with wind direction contributions of also $\sim 6\,\mu\mathrm{g/m^3}$. Total PM1 concentrations in spring can be as high as $50\,\mu\mathrm{g/m^3}$. These peaks in spring are not as well reproduced by the model as winter episodes and are likely due to new particle formation processes along the air mass trajectories, in particular of nitrate. Summer PM1 concentrations are lower than in other seasons. Model results suggest that summer peak concentrations are largely driven by high temperatures, particle advection from Paris and continental Europe with low wind speeds and prolonged periods without precipitation. For an example episode in June 2017, temperatures above $30\,^\circ\mathrm{C}$ contribute $\sim 3\,\mu\mathrm{g/m^3}$ to the total PM1 concentration. On site scarcity of rain increases air pollution, but does not appear to be a main driver of strong day-to-day variations in particle concentrations. Main drivers of day-to-day variability of predicted PM1 concentrations are changes in wind direction, air temperature, and BLH. These changes often superimpose the influence of time without precipitation. Individual PM1 species are shown to respond differently to changes in temperature. While SO_4^{2-} and organic matter concentrations are increased during both high and low temperature situations, NH_4^+ and NO_3^- are substantially increased only at low temperatures. Model results indicate that SIA formation is enhanced during shallow BLH conditions.

Many of the results presented here hold true for regions other than suburban Paris and are thus beneficial for the general understanding of drivers of air pollution. This includes the importance of formation processes of secondary pollutants as well as the dominant role of the BLH for PM1 concentrations. The contribution of wind direction furthermore highlights the role of advected

pollution and emphasizes the need for large-scale measures against air pollution. The GBRT approach in combination with the SHAP regression values presented here provides an intuitive tool to assess meteorological drivers of air pollution and to advance the understanding of high pollution events by uncovering different physical mechanisms leading to high-pollution episodes. To our knowledge, this is the first time that the SHAP-framework for explainable machine learning is extensively applied in atmospheric sciences. The results of this study are highly relevant for policy makers, e.g., by providing a basis for future clean air programs or by providing the potential of a statistically-based early warning system for high pollution episodes. Statistical models such as the one developed in the present study can be used to attribute changes in air quality to policy measures (Cermak and Knutti, 2009), and may help political decision makers develop and implement effective clean-air policies. This study could be extended in the future, e.g., by including information on anthropogenic emissions or further stations down- and upwind of SIRTA, which would allow further analysis of dominant advection patterns.

5 Concluding discussion and outlook

5.1 Concluding discussion

Air pollution continues to impact the lives of numerous people across Europe and poses a major health issue. Therefore, it is essential to improve the scientific understanding of patterns of air quality to develop efficient mitigation strategies. The results of this thesis contribute to this goal by improving the applicability of satellite-based AOD for estimating regional-scale patterns of PM concentrations and by providing new insights on environmental processes that amplify air pollution. In this context, three main scientific topics have been explored in this work:

- **(1) Meteorological influences on the relationship between PM10 and AOD (chapter 2)**
  Deriving accurate estimations of near-ground PM concentrations from satellite-based AOD is challenging, as influences on the relationship between AOD and PM, in particular pertaining to meteorology, are not yet fully understood. In this study, a quantitative analysis of these meteorological influences was provided, setting a basis to accurately estimate near-ground PM from satellite-based AOD.

- **(2) Mapping and understanding PM10 patterns using AOD and environmental conditions (chapter 3)**
  Based on the results of chapter 2, links were established between PM10 concentrations, AOD, and ambient environmental conditions using a statistical modelling framework. Analyses of PM10 patterns were provided in this study by identifying and quantifying its major drivers.

- **(3) Meteorology-driven, local-scale variability of PM1 (chapter 4)**
  Approaching the complex mechanisms leading to high pollution episodes is challenging, as the effects of interacting meteorological variables are not easy to separate and quantify. In-depth analyses of local-scale atmospheric processes that cause high concentrations of fine-mode PM were performed in chapter 4 based on a statistical modelling framework. Differences between the main chemical PM1 constituents are highlighted.

Based on the existing literature, three main open research questions were identified and addressed in chapters 2–4:

How do meteorological conditions influence the statistical relationship between AOD and PM?

What are the main drivers of PM10 when environmental conditions and AOD are used to estimate PM10 concentrations?

How do atmospheric processes drive concentrations of different PM1 species at a local scale?

The results of this thesis under the guidance of these questions are discussed in the following.

(1) Meteorological influences on the relationship between PM10 and AOD

In a sub-regional study in northeastern Germany, satellite-based AOD retrievals were collocated with hourly, near-ground PM10 measurements. The dimensionless AiRDI is introduced, which allows for a pairwise, in-depth characterisation of the relationship between AOD and PM10. The AiRDI provides a semi-quantitative estimation of the agreement of each AOD-PM10 data pair under consideration of ambient meteorological conditions. A positive (negative) AiRDI indicates a relative overestimation (underestimation) of PM10 by the AOD.

AiRDI-patterns, as shown in Fig. 2.3 on page 36 suggest that the relationship between AOD and PM10 is highly nonlinear. Low AODs ($\sim$<0.15) have a relatively high chance to underestimate PM10 concentrations, while there is a higher chance of PM10 concentrations to be overestimated from high AODs ($\sim$>0.2). It follows that a linear association between these parameters cannot be the best statistical representation to accurately estimate PM from AOD as suggested by previous studies (cf. Wang, 2003; Gupta et al., 2006; Dinoi et al., 2010; Harbula and Kopacková, 2011; Arvani et al., 2016). The findings of this thesis suggest that the nonlinearity of the relationship is caused by the influences of meteorological parameters, which are discussed in the following.

Positive AiRDI-values are more likely to occur when above average RH coincides with above average BLHs (see Fig. 2.4a–d on page 37). This pattern agrees with previous observations that under humid atmospheric conditions, i.e., a high RH, hygroscopic particle growth increases satellite-derived AOD without influencing PM10 measurements (see Fig. 2.5b, Barnaba et al., 2010; Crumeyrolle et al., 2014). Conversely, AiRDI patterns suggest an underestimation of PM10 by AOD when below average RH coincides with below average BLH (see Fig. 2.4i–l). Less hygroscopic growth is to be expected during these conditions. In addition, near-ground PM10 concentrations tend to be increased under low-BLH-conditions, when particles are confined to lower atmospheric layers and accumulate (see Fig. 2.5c, Gupta and Christopher, 2009b). During average atmospheric conditions, i.e., when ambient RH (50–80 %) coincides with a moderate BLH (600–1,200 m, see Fig. 2.5a), the AiRDI has an increased probability of being close to zero, indicating good agreement of ranked AOD and PM10 data pairs. Furthermore, it was found that the influences of RH and BLH on the relationship between AOD and PM10 can cancel each other out

(see Fig. 2.5), which is indicated by an AiRDI close to zero. This is, for example, the case when a dry atmosphere (RH<50 %) coincides with a high BLH (>1,200 m) or when humid atmospheric conditions (RH>80 %) coincide with a shallow BLH (<600 m). A relatively high number of the AOD-PM10-data pairs that were analysed in this study fall in the AiRDI-range close to zero (good agreement of ranked AOD and PM10), but there are nonetheless many observations in the positive AiRDI-range, indicating a relative overestimation of PM10 from AOD (see Fig. 2.5a and 2.5e on page 39).

Although BLH and RH appear to be the main governing factors of the relationship between AOD and PM10, wind direction also influences the relationship between AOD and PM10 (Fig. 2.6 on page 41) by transporting air masses with varying characteristics to the study area (Figs 2.9 and 2.10). It was shown that western air masses establish a higher probability of an overestimation of PM10 concentrations by AOD. This fits well with previous observations, which showed that western, marine air masses are generally cleaner (i.e., carry a smaller particle burden) and wetter (i.e., have a higher RH) (Zieger et al., 2013). These air masses are thus more likely to increase the AOD by hygroscopic growth, while at the same time carrying less particles (see Zieger et al., 2013; Titos et al., 2014). In addition, particles within maritime air masses have a higher content of sea salts, which are hydrophilic and further promote hygroscopic growth of particles (Titos et al., 2014; Granados-Muñoz et al., 2015). On the contrary, air masses from eastern and southeastern wind sectors have a higher probability of the satellite AOD underestimating PM10 concentrations. These air masses generally carry more anthropogenic and industrial aerosols, which are less hydrophilic (Zieger et al., 2014) and thus less likely to induce hygrosopic particle growth. Patterns shown in Fig. 2.6 indicate that there is a higher chance of an AiRDI close to zero at lower wind speeds, when local sources dominate both AOD and PM10.

Previous studies have indicated the nonlinearity of the relationship between AOD and PM10 and emphasized the day-to-day variability of this relationship (Gupta et al., 2006; Nordio et al., 2013; Chudnovsky et al., 2014) due to changes in atmospheric RH and BLH (Gupta et al., 2006; Gupta and Christopher, 2009b). Concerning the influence of RH, Barnaba et al. (2010) and Crumeyrolle et al. (2014) demonstrated the positive influence of hygroscopic growth on AOD using in situ measurements in the Po valley, Italy, and Baltimore, U.S.A., respectively. Gupta et al. (2006) approached the relationship between AOD and PM2.5 in a case study in Texas, collocating AOD and PM2.5 data pairs and binning the data in different ranges of meteorological parameters. Computing linear correlation coefficients for the subgroups, they found a higher correlation coefficient between PM2.5 and MODIS AOD for BLHs below 200 m. They proposed that during low BLH-conditions, the amount of particles observed by the in situ instrument will be similar as observed by the satellite, i.e., AOD better represents near-ground PM concentrations. On the contrary, Chudnovsky et al. (2013a) found only minor influence of the BLH on the relationship between AOD and PM2.5 over the Boston area, U.S.A., when values were between 250–500 m.

Furthermore, for a study conducted in the southeast U.S.A., Gupta and Christopher (2009b) set up a multiple regression approach taking into account temperature, RH, BLH, wind speed, and cloud fraction. They found the model to perform best when a well-mixed boundary layer ($\sim$2,000 m) coincides with high temperatures and a dry RH (40–50 %). Zheng et al. (2017) emphasize the influence of both RH and BLH on the relationship between PM2.5 concentrations and AOD for the Bejing area, China. A conversion factor η is introduced, which they defined as the quotient of PM2.5 divided by AOD. They showed that the quotient η decreases with increasing RH, which indicates an increase in AOD. η also decreases with increasing BLH, which suggests increased PM2.5 concentrations.

Overall, the previous literature agrees that RH and BLH influence the relationship between AOD and PM, the former by increasing the scattering coefficient via hygroscopic growth, the latter by changing the vertical distribution of particles in the atmosphere,. This appears to be largely independent of the geographic location and was observed in studies covering Europe (Barnaba et al., 2010), U.S.A. (Gupta et al., 2006; Gupta and Christopher, 2009b; Crumeyrolle et al., 2014), and China (Zheng et al., 2017). However, conclusions on the best conditions to derive near-ground PM concentrations from satellite-based AOD differ. One possible reason therefore could be differences in methodology. Approaching the relationship between AOD and PM with linear correlations on binned parameter ranges (see Gupta et al., 2006; Chudnovsky et al., 2014) relies strongly on the bin sizes and limits. Furthermore, as indicated by Zheng et al. (2017), the influences of BLH and RH need to be considered simultaneously to determine agreement or divergence of PM and AOD. In this regard, the AiRDI provides new insights as it offers a coherent view on meteorological influences on the relationship between AOD and PM10, and allows to pinpoint conditions of relative agreement between AOD and PM. The application of such an approach to remote areas that lack a dense network of in situ PM measurement stations would be highly beneficial. For example, combining satellite-based AOD with globally available meteorological reanalysis data would possibly provide more accurate PM estimates than relying solely on AOD for estimating PM concentration exposure. Still, depending on the area, other factors would have to be taken into account, e.g., desert dust transport (Yahi et al., 2013). In areas with good availability of PM measurement sites, the use of statistical methods to estimate PM from AOD under consideration of ambient environmental conditions is a further option. Such models require a large amount of data to establish empirical relationship between the model inputs and PM concentrations, but showed promising results in the past (Stafoggia et al., 2017; Brokamp et al., 2018; Chen et al., 2018a; Stafoggia et al., 2020). Such an approach was taken in chapter 3 and will be discussed in the next section.

(2) Mapping and understanding PM10 patterns using AOD and environmental conditions

On the basis of the findings presented within chapter 2, it was hypothesized that PM10 can be estimated with higher accuracy from satellite-based AOD, if environmental conditions are explicitly considered. Therefore, links between PM10 concentrations and AOD under consideration of ambient environmental conditions were established using a statistical model. This model was set up to predict hourly PM10 concentrations measured in Germany and to analyse model sensitivities with regard to environmental conditions.

The meteorological parameters temperature, BLH, and east-west winds proved to be highly relevant to accurately predict PM10 (see Fig. 3.5 on page 62). Multiday inflow from eastern directions, for example, generally increases PM10 concentrations, while a decrease was found for situations dominated by inflow from western directions (see Fig. 3.6). The difference in PM concentrations that can be appointed to changes in east-west wind direction was on average $\sim$10 μg/m^3. This number agrees well with results by van Pinxteren et al. (2019), who determined the contribution of trans-boundary particle transport from countries east of Germany to be 13 μg/m^3 on average. That number was determined by comparing PM10 concentrations during western and eastern air mass inflow under similar meteorological conditions for sites in eastern Germany. In van Pinxteren et al. (2019), it is highlighted that trans-boundary particle transport from eastern European countries increases in winter due to higher emissions from domestic heating, which is in agreement with patterns shown in Fig. 3.6b. Furthermore, the confining effect of shallow BLHs, which causes particles to accumulate near ground (see Gupta and Christopher, 2009b; Wagner and Schäfer, 2017), is of major importance to PM10 predictions. An average increase of PM10 concentrations of $\sim$5–10 μg/m^3 was quantified for BLHs below 800 m (see Fig. 3.8 on page 66). This effect was found to be stronger in winter (up to $\sim$15 μg/m^3), while it is largely absent in summer. The influence of the boundary layer on PM concentrations stagnates between 800–2,000 m as the dilution effect of an increased boundary layer reaches its maximum. A similar pattern was found by Grange et al. (2018), who applied a RF-model on PM10 measurements and computed the partial dependence of BLH for a site in Zurich, Switzerland. Furthermore, in a study by Liu et al. (2018), extreme haze conditions over Bejing, China, were analysed, with peak PM2.5 concentrations of $\sim$400 μg/m^3. Still, the relationship between PM and the BLH seems to be comparable: the dilution of particles with increasing BLHs effectively reduces PM2.5 concentrations up to 800 m. Above this height, only a minor influence of BLH on PM2.5 concentrations is shown. In this thesis and in Grange et al. (2018), BLH is based on ECMWF reanalysis, while Liu et al. (2018) used a micro-pulse lidar to determine the BLH. Hence, the quantification of the influence of the BLH on PM10 concentrations appears to be robust and generalizable to other regions. An exception was pointed out by Grange et al. (2018):

if a measurement site is intermittently above the boundary layer due to topography, different mechanisms need to be considered.

Mean modelled PM10 concentrations increase slightly during conditions with very high BLHs (>2,000 m), which could be related to enhanced formation of secondary aerosols in summer (e.g., SO_4^{2-},) due strong solar irradiation and high temperatures (Dawson et al., 2007; Pay et al., 2012; Fuzzi et al., 2015; Grange et al., 2018). Indeed, the pattern is most prominent in summer, when these prerequisites are most likely to be met, but overall, the increase is weak, which suggests that the photochemistry-driven formation of new particles is not an important driver of PM10 concentrations. Still, this pattern is consistent with an observed increase of PM10 predictions (up to $12\,\mu g/m^3$) when de-seasonalised temperatures rise above average levels. High summer temperatures can also be related to enhanced biogenic activity (Laothawornkitkul et al., 2009) and increased dust resuspension due to dried-up soils (Hoffmann and Funk, 2015).

Precipitation has been identified as an important driver for PM concentrations in previous studies (Radke et al., 1980; Dawson et al., 2007; Rost et al., 2009; Li et al., 2015), but largely lacks influence on the statistical model's predictions. Possibly, this is related to the fact that only cloud-free days were considered in this study, as AOD is limited to these situations. Thus, the immediate effect of rainfall on particles in the atmosphere could not be investigated. A further analysis on the influence of precipitation on the fine-mode particle fraction is provided in chapter 4.

Results pertaining to the influence of AOD on PM10 predictions suggest that AOD is closely related to PM10 (see Fig. 3.14 on page 73), which highlights the general suitability of AOD for air quality studies when meteorological conditions are explicitly considered. AOD and PM10 are weaker correlated in summer when particles are more dispersed within a well-mixed boundary layer, and the AOD is largely determined by particles higher up in the atmosphere. The influence of BLH and the east-west wind component on the relationship between AOD and PM10 was quantified (see Figs 3.15 and 3.16 on page 74). Comparing PM10 predictions under conditions with low BLHs (200 m) with conditions with high BLHs (2,000 m) at a constant AOD, an average difference of $\sim$10 $\mu g/m^3$ is shown. The use of AOD without considering the BLH would not account for this difference. For the east-west wind flow, the average difference at a constant AOD between predictions under multi-day inflow from the west and predictions under multi-day inflow from the east was $\sim$8 $\mu g/m^3$.

The statistical model used in this study is more sensitive to meteorology than to land-surface parameters (see section 3.3.3). Of the analysed land cover types, only agricultural areas were found to positively influence PM10 predictions, using the partial dependence approach. The more pixels in the vicinity of a station are classified as agricultural area, the higher the PM10 predictions tend to be on average. Agriculture is known to be an important particle source by primary emissions of dust from arable lands and the application of fertilizers, which constitute important precursors for secondary particle formation (Hoffmann and Funk, 2015; Wagner et al., 2015). Only minor

influence of the NDVI data on the statistical model is found. MODIS-based NDVI data are frequently considered in comparable studies, but its importance was shown to be relatively low for different geographic settings, e.g., the U.S.A. (Hu et al., 2017), the UK (Yazdi et al., 2020), or Sweden (Stafoggia et al., 2020). The overall insensitivity of the model to land-surface parameters is in part related to the relatively coarse resolution of 250 m (for the CLC-classes and the MODIS NDVI), which might not be precise enough to adequately represent the conditions in in the vicinity of a measurement site. In addition, land surface parameters vary only in time but not in space and are thus of less use to the model to predict variations of PM10 concentrations. This may make it necessary to use different strategies for the determination of the importance of land use parameters in such a ML-framework, e.g., by comparing different sites during comparable meteorological conditions.

In general, results of the sensitivity analyses on the ML-framework with regard to environmental conditions agree well with other studies, also outside the field of ML (e.g., Gupta and Christopher, 2009b; Hoffmann and Funk, 2015; Wagner and Schäfer, 2017; Grange et al., 2018; Liu et al., 2018; van Pinxteren et al., 2019). This highlights the applicability of such a framework to derive physical processes and shows the robustness of the statistical analysis. Still, a direct comparison with ML-approaches used in other studies shows that results can somewhat vary with respect to differences in the model setups. In particular, studies may use similar parameters, but handle these differently, which complicates a comparison. For example, Grange et al. (2018) used instantaneous temperature as an input, while here, temperature was separated into instantaneous and de-seasonalised temperature, i.e., temperature anomalies. If this is taken into account, patterns pertaining to temperatures in Grange et al. (2018) agree well with those found here (see section 3.3.3). Different spatial or temporal resolutions of parameters can also lead to different results. Wind speed, for example, is considered here implicitly as instantaneous u and v components of the ERA-Interim data set, showing only minor relevance for the model. Grange et al. (2018) and also Yazdi et al. (2020), who conducted a statistical modelling study to predict PM2.5 over the greater London area, found a high relevance of wind speed. Wind speed data used in Grange et al. (2018) and Yazdi et al. (2020) are taken from nearby station measurements, which was possible due to a very dense network. These data more accurately describe local conditions and have a higher information content to the model.

Another important source of uncertainty when comparing the results of statistical sensitivity analyses is that methods to determine a feature's relevance to the model can differ. For example, Hu et al. (2017) predicted PM2.5 concentrations over the U.S.A. using a RF-model. Feature importances were approached by calculating the increase in mean squared error when a feature was permuted, and by the increase in node purities, i.e., the similarity of responses at the regression tree nodes. The methods provide very different results, which makes it difficult to decide which method is more reliable. Futhermore, Strobl et al. (2007) showed that the standard feature importance

provided by RF-models (used e.g., by Park et al., 2019; Stafoggia et al., 2020) is particularly unreliable because of biases in the subsampling process of the decision trees and biases introduced by categorical data. Hence, the comparison of ML-based sensitivity analyses should rely on in-depth analyses of the model's response to individual features, using e.g., partial dependence functions (Friedman, 2001) or SHAP regression values (Lundberg et al., 2020). These methods are preferable as they provide more insights and allow a robust analysis, which is not possible solely based on a number that represents a feature's importance.

The statistical modelling framework used here proved to be a powerful tool to infer on near-ground PM10 concentrations from satellite-based AOD if ambient environmental conditions, in particular meteorology, are considered. Furthermore, sensitivity analyses on the model produced relevant new insights regarding environmental drivers of PM10. With regard to the model's applicability to produce spatially coherent estimates of PM10, results of this study indicate that the performance of predictions on unseen locations is worse. This was tested by withholding a set of stations from the training data set, resulting in R^2 of 0.5–0.7, depending on the stations withheld, compared to R^2 of 0.77 if the training and test data were randomly split. This suggests that the statistical model is not always able to generalize to previously unseen locations. A more drastic reduction in R^2 was recently reported by Yazdi et al. (2020), who showed that their overall model is well able to capture the day-to-day variability of PM2.5 (R^2 of 0.882). For the model's ability to capture the spatial variability, however, a R^2 of 0.396 is reported. Similarly, in Hu et al. (2017), a reduction from 0.8 (overall R^2) to 0.7 (spatial R^2), and in Stafoggia et al. (2020), a reduction from 0.64 (overall R^2) to 0.5 (spatial R^2) was reported for ML-models predicting PM2.5. While this is not expected to affect the results of the sensitivity analysis presented here, it has implications for the reliability of the results of studies that train statistical models and extend predictions to areas where measurements are not available. The conventional performance metrics computed by randomly splitting train- and test-data sets might be over-optimistic in this regard (e.g., Hu et al., 2014; Park et al., 2019). The problem of spatial over-fitting of statistical models can be related to misleading geolocation variables, e.g., coordinates of the measurement sites or elevation (Meyer et al., 2018, 2019). These geolocation variables are highly autocorrelated in space and can be misinterpreted by statistical models. Furthermore, the unique combination of these geolocation variables provides a quasi-ID to the model, which improves the training score, but counteracts the model's ability to generalise (Meyer et al., 2018). Therefore, prediction frameworks that target spatially coherent maps of PM concentrations need to be engineered in a particular way, e.g., by allocating measurements from particular stations either to the train- or the test-data set. Incorporating such an approach also in the feature-selection-process would ensure that only features, which allow for predictions beyond the locations of the training samples, are used (Meyer et al., 2019).

As a next step, an in-depth, process-oriented study was conducted, focusing on how the interaction of meteorological conditions amplifies high concentrations of PM.

(3) Meteorology-driven, local-scale variability of PM1

Meteorology was identified as a major driver of PM10 concentrations in chapter 3. Meteorology-driven variations in day-to-day PM concentrations were targeted in chapter 4, with a particular emphasis on local-scale processes leading to high-pollution events. Focusing on the fine-mode particle fraction (PM1), particle formation and transformation processes of the most important chemical constituents, i.e., the particle species, were analysed. A statistical model was set up to predict daily variations of mean PM1 concentrations for the afternoon period (12–15 UTC), using meteorological parameters as input. The analysis of the model results is conducted with situation-based and individualised feature contributions using SHAP regression values (Lundberg and Lee, 2017; Lundberg et al., 2018b, 2020). These enable quantitative estimates of the contribution of each meteorological parameters to the final PM1 prediction.

The mean annual PM1 concentration at the measurement site were $8.84\,\mu g/m^3$. Shallow BLHs (below $500\,m$) drastically increase PM1 predictions by up to $5.5\,\mu g/m^3$ (see Fig. 4.6 on page 89), underscoring the validity of the results of chapter 3 pertaining to the influence of BLH on PM concentrations also for the fine particle fraction. The confining effect under shallow-BLH-conditions affects all species, which suggests the absence of major chemical feedbacks (see also Dawson et al., 2007). On the contrary, SO_4^{2-}, Org, NH_4^+, and BCff exhibit a positive contribution for BLHs above $\sim2,000\,m$. In the case of SO_4^{2-}, Org, and NH_4^+, this is likely related to the previously suggested enhanced oxidation rates due to high temperatures and strong solar irradiation under high-BLH-conditions, leading to increased formation rates of secondary particles (Dawson et al., 2007; Pay et al., 2012; Grange et al., 2018; Gen et al., 2019). Since BCff is not expected to be formed secondarily, the positive contribution of high BLHs on BCff concentrations must be due to different processes, e.g., increased vertical turbulence distributing BCff in the atmosphere.

Air masses from the northeast carry a higher particle burden to the site due to the high emissions in the Paris area (Bressi et al., 2013; Petetin et al., 2014; Petit et al., 2015; Srivastava et al., 2018). Here, the contribution of air masses from the northeastern wind sector to PM1 predictions was quantified to be ~5–$6\,\mu g/m^3$ on average. Wind speed increases PM1 predictions by ~2–$4\,\mu g/m^3$ under calm conditions (see Fig. 5.1). The relevance of wind speed supports the previous argumentation that information on wind speed is indeed an important driver of PM concentrations, but needs to represent local conditions (see Grange et al., 2018; Yazdi et al., 2020). Wind speed influences the contribution of other features to the model prediction, e.g., by increasing the contribution of shallow BLHs (see Fig. 4.8b on page 92). This empirical pattern in the model is likely related to the physical effect of reduced atmospheric ventilation due to the combination of low wind speeds and a shallow BLH, enhancing the accumulation of particles in the atmosphere near the ground. This is not a site-specific finding and is in agreement with results by Sujatha et al. (2016), who found atmospheric BC concentrations to negatively correlate with atmospheric

ventilation (defined as the product of BLH and wind speed) for measurements in southern India. Furthermore, Wang et al. (2018) pointed out the importance of stagnant conditions (defined based on 10-m wind speed, BLH, and the occurrence of precipitation) for recurring high-pollution events in China.

On site scarcity of rain increases PM1 predictions of the model. Results of chapter 4 suggest that the lack of rain is a prerequisite of peak PM1 concentrations, as the absence of wet deposition allows particles to accumulate in the atmosphere (Rost et al., 2009; van Pinxteren et al., 2019). This is also reflected in the model, since a lack of precipitation is shown to increase the contribution of low BLHs (see Fig. 4.8a on page 92), as the lack of wet deposition amplifies the accumulation effect of shallow BLHs. Petit et al. (2017) furthermore propose that particularly high PM concentrations are often associated with large-scale droughts, which enhance regional-scale formation of SIA. These large-scale phenomena are, however, not covered in this study and leave room for future analyses.
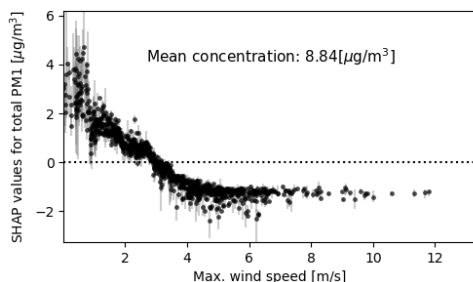

Figure 5.1: As Fig. 4.5 on page 87 for the contribution of maximum wind speed to the prediction of total PM1 (i.e., SHAP values) vs. absolute wind speed values [m/s].

Changes in instantaneous air temperature have shown a strong, nonlinear effect on PM1 species in this study (see Fig. 4.5 on page 87). The sensitivity of PM concentrations to temperature is already indicated in chapter 3, but using the information on the chemical constituents of PM, new insights are provided here. During low temperatures, temperature SHAP values, i.e., the contribution of temperatures to modelled PM1 concentrations, show a strong, positive influence on PM1 concentrations. Low temperatures stimulate new particle formation processes related to SO_4^{2-}, NH_4^+, and NO_3^- by shifting the gas-particle equilibrium (Clegg et al., 1998; Pay et al., 2012; Bressi et al., 2013; Petit et al., 2015). In addition, it is likely that the increasing signal for Org and BCwd is due to enhanced local anthropogenic biomass burning for residential heating (Favez et al., 2009; Sciare et al., 2010; Healy et al., 2012; Jiang et al., 2019). The effect of these processes on PM1 concentrations ceases at medium temperatures and PM1 concentrations become increasingly

insensitive to changes in temperature. At high temperatures, however, a temperature increases have a positive effect on SHAP values, leading to an increase in total PM1 predictions. This increase is driven by SO_4^{2-} and Org. Enhanced oxidation rates of SO_2 (Dawson et al., 2007) at higher temperatures lead to higher concentrations of SO_4^{2-}, while the increase in modelled concentrations of Org can be attributed to enhanced biogenic activity causing to a rise in biogenic emissions and secondary aerosol formation (Guenther et al., 1993; Megaritis et al., 2014; Churkina et al., 2017; Jiang et al., 2019). To enable an objective identification of important turning points in the relationship between air temperature and PM1 concentrations, a change-point analysis (CPA, Cermak et al., 2010; Andersen et al., 2016; Truong et al., 2020) was conducted on the data series of the temperature SHAP values sorted by absolute temperatures. The CPA pinpoints changes of statistical properties in the embedded signal by minimizing a radial basis cost function, and is able to identify discontinuities in a data series (Truong et al., 2020). Two turning points at $\sim$5 °C and $\sim$24 °C are identified, revealing three temperature-dependent PM1-regimes (see Fig. 5.2). For the low-temperature-regime (<5 °C), a linear regression reveals a trend of -1.36 µg/K. Between 24 °C–5 °C, only a very weak trend is found. The high-temperature-regime (>24 °C) shows an increase of 0.36 µg/K. The effect of temperature increases on PM concentrations has been previously indicated, e.g., in Dawson et al. (2007), who conducted a CTM-based study over the U.S.A. It was found that an increase of temperatures in winter effectively reduces PM2.5 concentrations due to the evaporation of NO_3^- and Org at higher temperatures. In summer, increases in SO_4^{2-} production counteract this effect. In a further CTM-based study by Megaritis et al. (2014), these findings were confirmed also for Europe. Megaritis et al. (2014) furthermore pointed out the role of higher biogenic emissions, leading to an increase in Org concentrations with increasing temperatures. The application of SHAP regression values allowed to expand this findings, and to present a coherent picture of how temperatures drive PM1 concentrations. Knowledge on the turning points in the relationship between temperature and PM1 concentrations as presented in Fig. 5.2 is essential to understand not only the development of high pollution episodes, but also how PM concentrations react with respect to climate change.

Overall, it is a complex combination of the described meteorological influences on PM1 that lead to the development of high pollution events, and processes vary substantially between and within seasons. While winter high pollution events were shown here to be frequently driven by temperatures below 5 °C and shallow BLHs, summer peak concentrations are generally lower and influenced by the high-temperature-regime above 24 °C and stagnant conditions with low wind speeds. The advection of particles from the highly polluted Paris area is a major driver of PM1 concentrations in all seasons. A detailed look on high pollution episodes is provided in Figs 4.11–4.14. A winter pollution episode in January 2016 with peak PM1 concentrations of 37 µg/m³ (see Fig. 4.11 on page 98) was tested for its main contributors on a daily basis. Model results show a strong contribution of temperature to the predicted PM1 concentrations during this
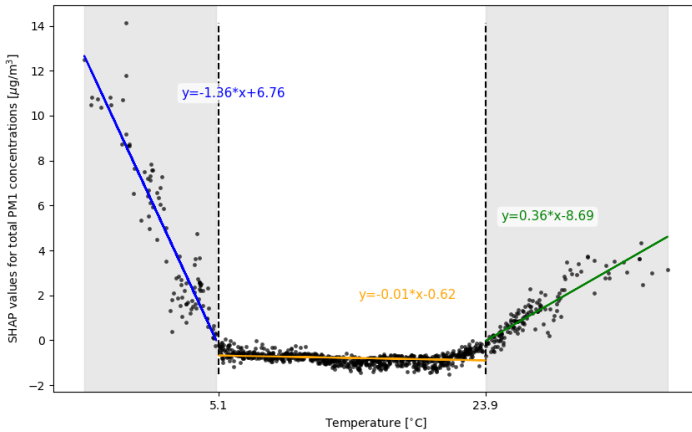
Figure 5.2: Regimes characterizing the influence of air temperatures [°C] on total PM1 concentrations. Absolute air temperatures [°C] are plotted against temperature SHAP values [µg/m³]. In addition, important turning points in the data series as identified by a CPA are indicated as vertical dotted lines. For each regime, the trends are shown, which are calculated using a linear regression.

episode (up to $11\,\mu g/m^3$), suggesting enhanced local, temperature-dependant particle formation as temperatures fell below freezing. In the case of another winter pollution episode in December 2016 with a main PM1 concentration peak $>40\,\mu g/m^3$ and several PM1 peaks $>27\,\mu g/m^3$ (see Fig. 4.12), temperatures remain relatively stable (5–10 °C) and the contribution of temperature to the PM1 predictions are low. BLHs (<500 m) and northeastern wind directions were quantified to be the major contributors to the PM1 peaks during this episode with contributions up to $6\,\mu g/m^3$. A further highly polluted episode happened in March 2013, with peak concentrations surpassing $50\,\mu g/m^3$ (see Fig. 4.14). These peaks are not as well reproduced by the model and are likely due to particle formation processes along the air mass trajectories, in particular of NO_3^- (Petit et al., 2017). The analysis of a summer pollution episode in June 2017 (see Fig. 4.14) with peak PM1 concentrations of $17\,\mu g/m^3$ suggests that temperatures above 30 °C contribute $\sim 3\,\mu g/m^3$ to the total PM1 concentration along with particle advection from Paris and continental Europe ($\sim 3\,\mu g/m^3$) and low wind speeds ($\sim 2\,\mu g/m^3$).

These findings relate to localized events, but the results are nevertheless highly relevant for other regions as well. For example, the importance of temperature-induced particle formation processes has been shown for the U.S.A. (Dawson et al., 2007), Europe (Megaritis et al., 2014), and China

(Wang et al., 2016). Hence, it is likely that the insights gained on the turning points governing the relationship between temperature and PM1 hold for other regions as well. Furthermore, the impact of BLH on PM1 concentrations found in this local study is similar to results presented in chapter 3 on a regional scale, which emphasizes that the dominant role of BLH on PM concentrations. The importance of wind direction highlights the role of advected pollution from highly polluted urban or industrial hotspots, indicating that urban surroundings can be severely affected by the urban pollution footprint. The diversity of processes leading to high pollution episodes emphasizes the need to develop a set of PM mitigation strategies that tackle both local and regional sources of particles and precursor gases.

The new insights on the highly variable relationship between AOD and PM10 provided in this thesis are a sound basis for improving the accuracy of future efforts to estimate PM concentrations based on satellite AOD retrievals, in particular for regions that lack an extensive network of in situ PM measurement sites. Accurate, spatially coherent PM estimates are beneficial as they allow for the identification of pollution hotspots and the long-term monitoring of air quality over a large spatial scale. This is an important step towards a reliable data basis on air pollution, which is also important for large-scale epidemiological studies. These studies depend on reliable air quality data to produce risk assessments for the population and to quantify health risks. Eventually, this would constitute a powerful tool for policymakers to develop the right measures towards better air quality. Furthermore, the results presented in this thesis advance the scientific understanding of environmental processes driving PM concentrations, contributing not only to the scientific progress, but are also important to the general public and policymakers. It is crucial to take atmospheric processes into account for the development of pollution mitigation strategies and to test the efficiency of these measures. The detailed knowledge on processes leading to high pollution also allows to identify pollution sources, e.g., by separating the contribution of advected particles from locally formed particles. The rate of locally-formed particles, for example, could be decreased by reducing the availability of precursor gases, e.g., SO_2 and NO_x, which are emitted in great quantities by anthropogenic sources (Pay et al., 2012). The results of this thesis, in particular the new insights gained on the influence of air temperature on PM concentrations, also have implications for the understanding of how PM concentrations might be affected by climate change. While this question has already been tackled by several modelling studies (Dawson et al., 2007; Megaritis et al., 2013, 2014; Sá et al., 2016; Doherty et al., 2017), the insights provided here regarding the response of PM concentrations to changes in air temperature, BLH, precipitation, wind, and RH advance the system understanding from an empirical perspective and might increase the confidence in PM estimations under climate change.

In the concluding outlook section, promising developments in the field of scientific air quality research are presented along with ideas on how the findings of this thesis could be put into practice.

5.2 Outlook

Future developments of PM estimates from satellite-based AOD

Estimates of PM concentrations from satellite-based AOD will very likely continue to play a major role in the field of air quality research. Future efforts may profit from the introduction of new data sets, e.g., the recently published ERA5 reanalysis data set, which provides more accurate information on meteorological conditions (Olauson, 2018; Tarek et al., 2019; Jiang et al., 2020). In addition, the applicability of geostationary satellites, e.g., the upcoming Meteosat Third Generation satellites for AOD retrievals is currently being discussed (Ceamanos et al., 2019). These have the potential to provide large spatial and temporal coverage, which would be of great value to further advance the scientific understanding of air quality patterns. Advances in ML techniques, e.g., deep learning approaches such as convolutional neural networks, have the potential to improve estimates of PM. These networks could, for example, leverage spatial information pertaining to pressure fields, wind fields, or terrain. In Soh et al. (2018), such an network was used to forecast PM concentrations, informing the public in Taiwan. A further promising development is the distribution of low-cost PM-sensors, which can be coupled with satellite-based AOD. These devices are easy to handle, but do not reach the accuracy of professional PM measurement devices. Still, they give some indication on PM concentrations and by distributing such devices to the public, a large coverage can be achieved. Malings et al. (2020), for example, suggest to link the PM measurements of those devices with satellite-based AOD, similar to the approach presented in chapter 3. This would be particularly beneficial for remote regions that lack conventional ground-based monitoring stations.

Putting the scientific understanding of PM into practice

The approach described in chapter 4 offers several conceivable practical applications. For example, the effect of measures towards better air quality could be analysed by setting up two statistical models: one model trained to predict PM concentrations before mitigation measures were taken and another model trained on PM concentrations after the measures come into effect. Both models would have to consider the same meteorological parameters as input. The difference of the PM predictions of both models would then provide a robust estimation of the effect of the mitigation measure under consideration of the meteorological conditions. Furthermore, the statistical approach could be coupled with weather forecasts to establish a short-term warning system. If high concentrations are to be expected due to the forecasted meteorological conditions,

a warning of air pollution could be issued to the public by local authorities. Furthermore, the importance of meteorological conditions for PM concentrations raises questions regarding air pollution thresholds such as the EU legal thresholds. These limits are fixed, but exceedances of PM values may not always be due to high emissions, but due to specific meteorological conditions. Since an exceedance of the EU thresholds has serious legal consequences for affected EU states, moving away from fixed limits to a dynamic perspective, which considers meteorological conditions during exceedances of air pollution thresholds might provide a level playing field for all member states.

In the light of recent events related to the global outbreak of SARS-CoV-2, air pollution could become even more the centre of the public attention. Preliminary studies suggest that air pollution might be a key factor in the virus' spreading. Although a thorough treatment of this question is still needed, the transmission by virus-laden aerosols is discussed and is considered as a possible spreading mechanism (Contini and Costabile, 2020). Further preliminary studies by Xiao et al. (2020) and Yongjian et al. (2020) indicate that long-term exposure to high levels of PM might increase mortality rates related to the COVID-19 disease that is triggered by SARS-CoV-2.

Looking forward, new ideas in the highly dynamic field of air quality research are very likely to emerge, which will help to pursue the continuous goal of ensuring good living conditions for all.

Scientific appendix of chapter 3

A   Information on UBA PM10 stations

The spatial distribution of PM10 measurement stations is shown in Fig. A.1. Overall, stations are distributed relatively homogeneously over the area of Germany. The number of stations is higher in urban agglomerations. The majority of stations are classified as "urban" or "suburban". "Rural" stations are relatively rare. Most stations are labelled as representative for background conditions by the data provider. Station altitudes [m.a.s.l.] are shown in Fig. A.2. Altitudes range from 0 to 970 m.a.s.l.

Furthermore, mean PM10 concentrations of all stations for the study period 2007–2015 are shown in Fig. A.3. Mean concentrations are highest in the Rhine-Ruhr area (north-west) and other urban areas such as Munich to the south, Berlin to the northeast and Hamburg to the north.
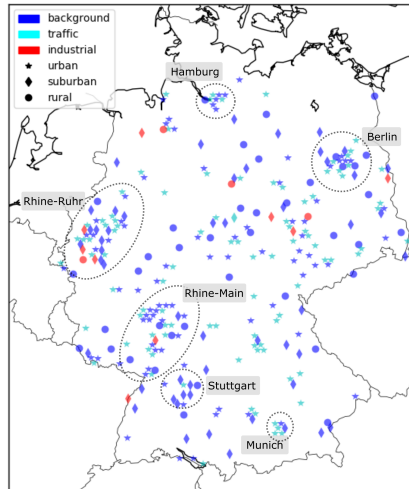

Figure A.1: Spatial distribution, type, and representativeness of UBA PM measurement stations in Germany
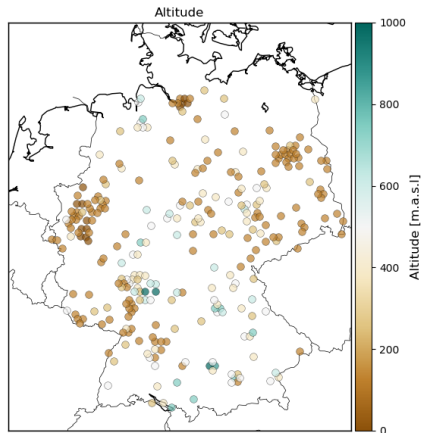
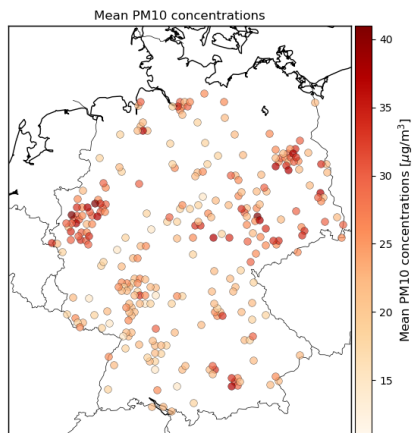*Figure A.2: Altitude [m.a.s.l.] of UBA PM measurement stations in Germany*

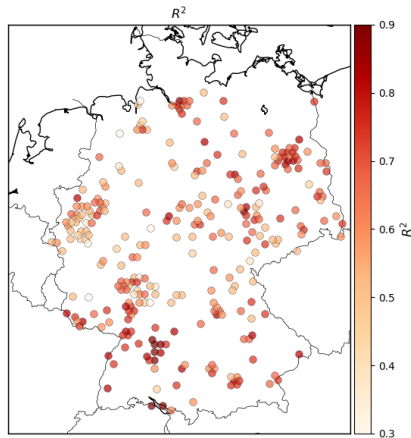*Figure A.3: Mean PM10 concentrations for measurement stations.*

Figure B.1: Spatial distribution of R^2 of the model set up.

B   Spatial distribution of model skill

Figs B.1 and B.2 show the spatial distribution of the coefficient of determination (R^2) and the RMSE, respectively. In the southwest and the northeast, R^2 tends to be higher and RMSE tends to be lower. In the Rhine-Ruhr region (northwest), performance seems to be generally worse. These stations generally have high mean PM10 concentrations (see Appendix A, Fig. A.3). However, it appears that other urban areas, which also have high mean PM10 concentrations can be modelled quite well (e.g. Berlin, in the northeast or Hamburg in the north).

C   Further Individual Conditional Expectation plots

Solar radiation (Fig. C.1) and time since the last precipitation (Fig. C.2) were analysed using the ICE method as described in chapter 2.6.3. Both input features show only minor influence on mean PM10 predictions. The ICE functions for time since last precipitation show particularly large variability of model responses of the individual data instances (shown by the shaded areas beneath the bold lines), indicating strong interactions with other features.
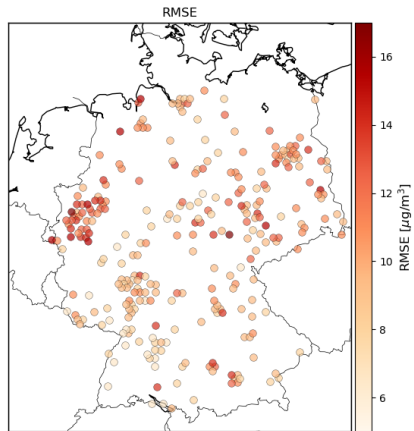
Figure B.2: Spatial distribution of the RMSE of the model set up.

D   Seasonal two-way partial dependence plots

In addition to Fig. 3.13 on page 72, which shows the full-year model two-way partial dependence of RH and BLH, Fig. D.1 additionally depicts the two-way partial dependence of the seasonal models. Similarly, Figs D.2 and D.3 show the seasonal two-way partial dependence of AOD and BLH and AOD and umean, respectively.
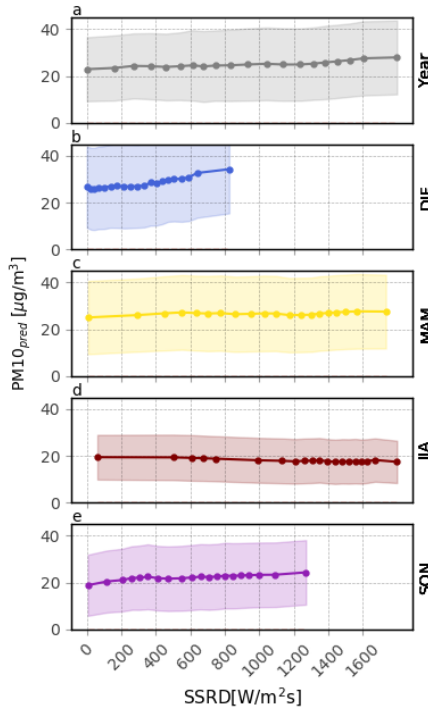
Figure C.1: PD plot showing the mean model response to changes in SSRD [W/m² s] for the full-year model (a) and each season separately (b-e, DJF, MAM, JJA, SON). Description as in Fig. 3.6
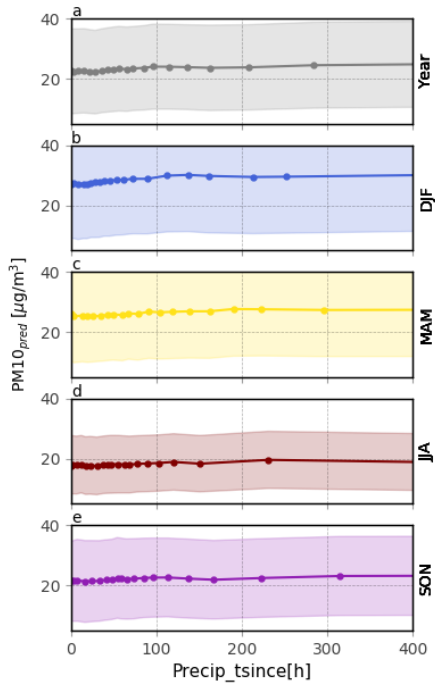
Figure C.2: PD plot showing the mean model response to changes in time since last precipitation [h] for the full-year model (a) and each season separately (b-e, DJF, MAM, JJA, SON). Description as in Fig. 3.6

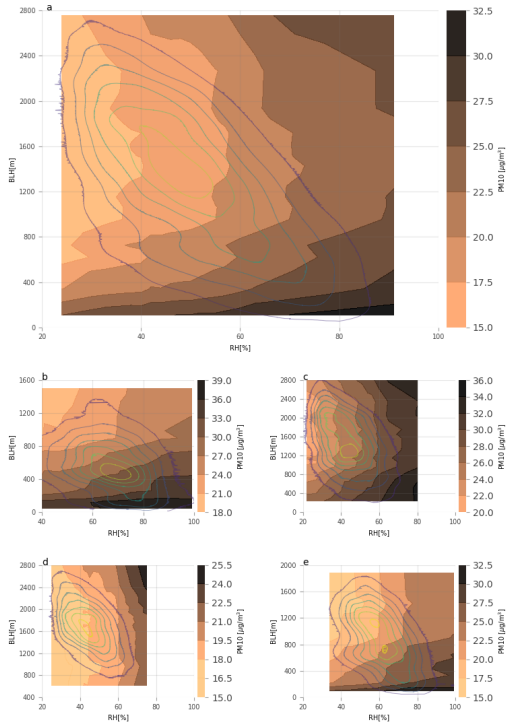Figure D.1: Two-way PD of RH and BLH, full-year model (a) and seasonal models (b-e, DJF, MAM, JJA, SON). Description as in Fig. 3.9
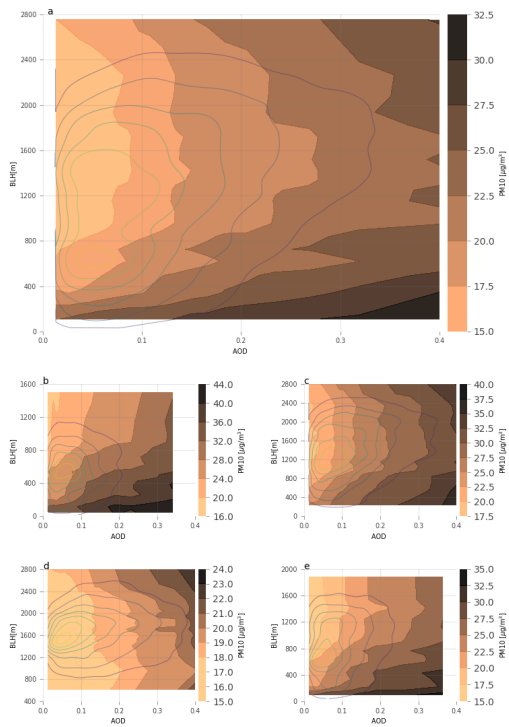
Figure D.2: Two-way PD of AOD and BLH, full-year model (a) and seasonal models (b-e, DJF, MAM, JJA, SON). Description as in Fig. 3.9
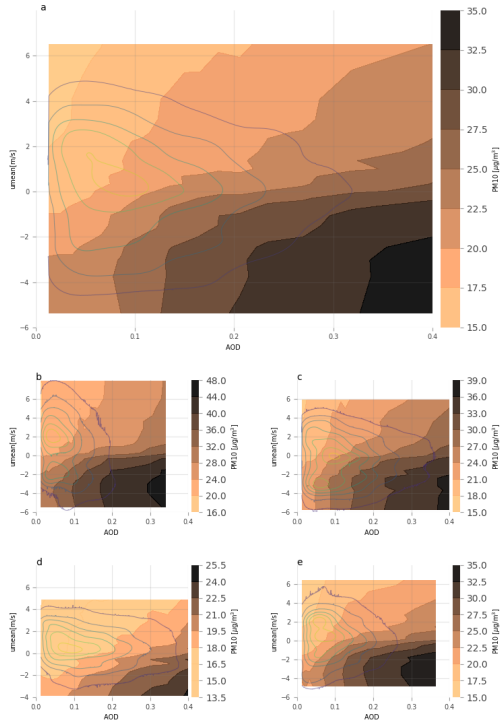
Figure D.3: Two-way PD of AOD and umean, full-year model (a) and seasonal models (b-e, DJF, MAM, JJA, SON). Description as in Fig. 3.9

Bibliography

Adamopoulos, A.D., Kambezidis, H.D., Kaskaoutis, D.G., Giavis, G., 2007. A study of aerosol particle sizes in the atmosphere of Athens, Greece, retrieved from solar spectral measurements. Atmos. Res. 86, 194–206. doi:10.1016/j.atmosres.2007.04.003.

Alanen, J., Simonen, P., Saarikoski, S., Timonen, H., Kangasniemi, O., Saukko, E., Hillamo, R., Lehtoranta, K., Murtonen, T., Vesala, H., Keskinen, J., Rönkkö, T., 2017. Comparison of primary and secondary particle formation from natural gas engine exhaust and of their volatility characteristics. Atmos. Chem. Phys. 17, 8739–8755. doi:10.5194/acp-17-8739-2017.

Andersen, H., Cermak, J., Fuchs, J., Knutti, R., Lohmann, U., 2017. Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks. Atmos. Chem. Phys. 17, 9535–9546. doi:10.5194/acp-17-9535-2017.

Andersen, H., Cermak, J., Fuchs, J., Schwarz, K., 2016. Global observations of cloud-sensitive aerosol loadings in low-level marine clouds. J. Geophys. Res. Atmos. 121, 12,936–12,946. doi:10.1002/2016JD025614.

Andreae, M., Rosenfeld, D., 2008. Aerosol–cloud–precipitation interactions. Part 1. The nature and sources of cloud-active aerosols. Earth-Science Rev. 89, 13–41. doi:10.1016/j.earscirev.2008.03.001.

Ansmann, A., Althausen, D., Wandinger, U., Franke, K., Müller, D., Wagner, F., Heintzenberg, J., 2000. Vertical profiling of the Indian aerosol plume with six-wavelength lidar during INDOEX: A first case study. Geophys. Res. Lett. 27, 963–966. doi:10.1029/1999GL010902.

Ansmann, A., Riebesell, M., Weitkamp, C., 1990. Measurement of atmospheric aerosol extinction profiles with a Raman lidar. Opt. Lett. 15, 746–748. doi:0146-9592/90/130746-0.

Ansmann, A., Wandinger, U., Riebesell, M., Weitkamp, C., 1992. Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar. Appl. Opt. 31, 7113–7131. doi:0003-6935/92/337113-19$05.00/0.

Arvani, B., Pierce, R.B., Lyapustin, A.I., Wang, Y., Ghermandi, G., Teggi, S., 2016. Seasonal monitoring and estimation of regional aerosol distribution over Po valley, northern Italy, using a high-resolution MAIAC product. Atmos. Environ. 141, 106–121. doi:10.1016/j.atmosenv.2016.06.037.

Atkinson, R., 2000. Atmospheric chemistry of VOCs and NOx. Atmos. Environ. 34, 2063–2101. doi:10.1016/S1352-2310(99)00460-4.

Atkinson, R.W., Ross Anderson, H., Sunyer, J., Ayres, J., Baccini, M., Vonk, J.M., Boumghar, A., Forastiere, F., Forsberg, B., Touloumi, G., Schartz, J., Katsouyanni, K., 2001. Acute Effects of Particulate Air Pollution on Respiratory Admissions. Am. J. Respir. Crit. Care Med. 164, 1860–1866. doi:10.1164/ajrccm.164.10.2010138.

Badyda, A., Gayer, A., Czechowski, P., Majewski, G., Dąbrowiecki, P., 2016. Pulmonary Function and Incidence of Selected Respiratory Diseases Depending on the Exposure to Ambient PM10. Int. J. Mol. Sci. 17, 1954. doi:10.3390/ijms17111954.

Baklanov, A., Molina, L.T., Gauss, M., 2016. Megacities, air quality and climate. Atmos. Environ. 126, 235–249. doi:10.1016/j.atmosenv.2015.11.059.

Barnaba, F., Putaud, J.P., Gruening, C., Dell'Acqua, A., Dos Santos, S., 2010. Annual cycle in co-located in situ, total-column, and height-resolved aerosol observations in the Po Valley (Italy): Implications for ground-level particulate matter mass concentration estimation from remote sensing. J. Geophys. Res. Atmos. 115, 1–22. doi:10.1029/2009JD013002.

Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., Gerlach, N., 2004. Projekt RADOLAN Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer) .

Bayat, R., Ashrafi, K., Shafiepour Motlagh, M., Hassanvand, M.S., Daroudi, R., Fink, G., Künzli, N., 2019. Health impact and related cost of ambient air pollution in Tehran. Environ. Res. 176, 108547. doi:10.1016/j.envres.2019.108547.

Belle, J.H., Chang, H.H., Wang, Y., Hu, X., Lyapustin, A., Liu, Y., 2017. The Potential Impact of Satellite-Retrieved Cloud Parameters on Ground-Level PM2.5 Mass and Composition. Int. J. Environ. Res. Public Health 14, 1244. doi:10.3390/ijerph14101244.

Beloconi, A., Chrysoulakis, N., Lyapustin, A., Utzinger, J., Vounatsou, P., 2018. Bayesian geostatistical modelling of PM10 and PM2.5 surface level concentrations in Europe using high-resolution satellite-derived products. Environ. Int. 121, 57–70. doi:10.1016/j.envint.2018.08.041.

Berlin Senate, 2013. Luftreinhalteplan 2011–2017. Available online: https://www.berlin.de/ (accessed on August 24th, 2018) .

Bertram, T.H., Heckel, A., Richter, A., Burrows, J.P., Cohen, R.C., 2005. Satellite measurements of daily variations in soil NOx emissions. Geophys. Res. Lett. 32, L24812. doi:10.1029/ 2005GL024640.

Birmili, W., Wiedensohler, A., 2000. New particle formation in the continental boundary layer: Meteorological and gas phase parameter influence. Geophys. Res. Lett. 27, 3325–3328. doi:10.1029/1999GL011221.

Blando, J.D., Turpin, B.J., 2000. Secondary organic aerosol formation in cloud and fog droplets: A literature evaluation of plausibility. Atmos. Environ. 34, 1623–1632. doi:10.1016/ S1352-2310(99)00392-1.

Bonn, B., Von Schneidemesser, E., Andrich, D., Quedenau, J., Gerwig, H., Lüdecke, A., Kura, J., Pietsch, A., Ehlers, C., Klemp, D., Kofahl, C., Nothard, R., Kerschbaumer, A., Junkermann, W., Grote, R., Pohl, T., Weber, K., Lode, B., Schönberger, P., Churkina, G., Butler, T.M., Lawrence, M.G., 2016. BAERLIN2014 -The influence of land surface types on and the horizontal heterogeneity of air pollutant levels in Berlin. Atmos. Chem. Phys. 16, 7785–7811. doi:10.5194/acp-16-7785-2016.

Bossard, M., Feranec, J., Otahel, J., 2000. CORINE land cover technical guide: Addendum 2000. Technical Report 40. European Environment Agency.

Bourgeois, Q., Ekman, A.M.L., Renard, J.B., Krejci, R., Devasthale, A., Bender, F.A.M., Riipinen, I., Berthet, G., Tackett, J.L., 2018. How much of the global aerosol optical depth is found in the boundary layer and free troposphere? Atmos. Chem. Phys. 18, 7709–7720. doi:10.5194/acp-18-7709-2018.

Boyouk, N., Léon, J.F., Delbarre, H., Podvin, T., Deroo, C., 2010. Impact of the mixing boundary layer on the relationship between PM2.5 and aerosol optical thickness. Atmos. Environ. 44, 271–277. doi:10.1016/j.atmosenv.2009.06.053.

Bressi, M., Sciare, J., Ghersi, V., Bonnaire, N., Nicolas, J.B., Petit, J.E., Moukhtar, S., Rosso, A., Mihalopoulos, N., Féron, A., 2013. A one-year comprehensive chemical characterisation of fine aerosol (PM2.5) at urban, suburban and rural background sites in the region of Paris (France). Atmos. Chem. Phys. 13, 7825–7844. doi:10.5194/acp-13-7825-2013.

Bressi, M., Sciare, J., Ghersi, V., Mihalopoulos, N., Petit, J.E., Nicolas, J.B., Moukhtar, S., Rosso, A., Féron, A., Bonnaire, N., Poulakis, E., Theodosi, C., 2014. Sources and geographical origins of fine aerosols in Paris (France). Atmos. Chem. Phys. 14, 8813–8839. doi:10.5194/acp-14-8813-2014.

Brokamp, C., Jandarov, R., Hossain, M., Ryan, P., 2018. Predicting Daily Urban Fine Particulate Matter Concentrations Using a Random Forest Model. Environ. Sci. Technol. 52, 4173–4179. doi:10.1021/acs.est.7b05381.

Brokamp, C., Jandarov, R., Rao, M.B., LeMasters, G., Ryan, P., 2017. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. Atmos. Environ. 151, 1–11. doi:10.1016/j.atmosenv.2016.11.066.

Bruegge, C.J., Diner, D.J., Kahn, R.A., Chrien, N., Helmlinger, M.C., Gaitley, B.J., Abdou, W.A., 2007. The MISR radiometric calibration process. Remote Sens. Environ. 107, 2–11. doi:10.1016/j.rse.2006.07.024.

Bundesministerium der Justiz und für Verbraucherschutz, 2010. Neununddreißigste Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes (Verordnung über Luftqualitätsstandards und Emissionshöchstmengen - 39. BImSchV). Anlage 1.

Carlton, A.G., Turpin, B.J., Altieri, K.E., Seitzinger, S., Reff, A., Lim, H.J., Ervens, B., 2007. Atmospheric oxalic acid and SOA production from glyoxal: Results of aqueous photooxidation experiments. Atmos. Environ. 41, 7588–7602. doi:10.1016/j.atmosenv.2007.05.035.

Ceamanos, X., Moparthy, S., Carrer, D., Seidel, F.C., 2019. Assessing the Potential of Geostationary Satellites for Aerosol Remote Sensing Based on Critical Surface Albedo. Remote Sens. 11, 2958. doi:10.3390/rs11242958.

Cermak, J., Knutti, R., 2009. Beijing Olympics as an aerosol field experiment. Geophys. Res. Lett. 36, L10806. doi:10.1029/2009GL038572.

Cermak, J., Wild, M., Knutti, R., Mishchenko, M.I., Heidinger, A.K., 2010. Consistency of global satellite-derived aerosol and cloud data sets with recent brightening observations. Geophys. Res. Lett. 37, L21704. doi:10.1029/2010GL044632.

Chafe, Z., Brauer, M., Héroux, M.E., Klimont, Z., Lanki, T., Salonen, R.O., Smith, K.R., 2015. Residential heating with wood and coal: health impacts and policy options in Europe and North America. Copenhagen.

Chen, G., Li, S., Knibbs, L.D., Hamm, N.A., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M.J., Guo, Y., 2018a. A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. Sci. Total Environ. 636, 52–60. doi:10.1016/j.scitotenv.2018.04.251.

Chen, G., Li, S., Zhang, Y., Zhang, W., Li, D., Wei, X., He, Y., Bell, M.L., Williams, G., Marks, G.B., Jalaludin, B., Abramson, M.J., Guo, Y., 2017. Effects of ambient PM 1 air pollution on daily emergency hospital visits in China: an epidemiological study. Lancet Planet. Heal. 1, e221–e229. doi:10.1016/S2542-5196(17)30100-6.

Chen, G.L., Guang, J., Li, Y., Che, Y.H., Gong, S.Q., 2018b. Retrieval of Atmospheric Particulate Matter Using Satellite Data Over Central and Eastern China. ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLII-3, 147–153. doi:10.5194/isprs-archives-XLII-3-147-2018.

Cheng, Y., Zheng, G., Wei, C., Mu, Q., Zheng, B., Wang, Z., Gao, M., Zhang, Q., He, K., Carmichael, G., Pöschl, U., Su, H., 2016. Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China. Sci. Adv. 2, e1601530. doi:10.1126/sciadv.1601530.

Chiriaco, M., Dupont, J.C., Bastin, S., Badosa, J., Lopez, J., Haeffelin, M., Chepfer, H., Guzman, R., 2018. ReOBS: a new approach to synthesize long-term multi-variable dataset and application to the SIRTA supersite. Earth Syst. Sci. Data 10, 919–940. doi:10.5194/essd-10-919-2018.

Chudnovsky, A., Koutrakis, P., Kloog, I., Melly, S., Nordio, F., Lyapustin, A., Wang, Y., Schwartz, J., 2014. Fine particulate matter predictions using high resolution Aerosol Optical Depth (AOD) retrievals. Atmos. Environ. 89, 189–198. doi:10.1016/j.atmosenv.2014.02.019.

Chudnovsky, A., Lyapustin, A., Wang, Y., Tang, C., Schwartz, J., Koutrakis, P., 2013a. High resolution aerosol data from MODIS satellite for urban air quality studies. Cent. Eur. J. Geosci. 6, 17–26. doi:10.2478/s13533-012-0145-4.

Chudnovsky, A., Tang, C., Lyapustin, A., Wang, Y., Schwartz, J., Koutrakis, P., 2013b. A critical assessment of high-resolution aerosol optical depth retrievals for fine particulate matter predictions. Atmos. Chem. Phys. 13, 10907–10917. doi:10.5194/acp-13-10907-2013.

Churkina, G., Kuik, F., Bonn, B., Lauer, A., Grote, R., Tomiak, K., Butler, T.M., 2017. Effect of VOC Emissions from Vegetation on Air Quality in Berlin during a Heatwave. Environ. Sci. Technol. 51, 6120–6130. doi:10.1021/acs.est.6b06514.

Clegg, S.L., Brimblecombe, P., Wexler, A.S., 1998. Thermodynamic Model of the System H^+ NH_4^+ Na^+ SO_4^{2-} NO_3^- Cl^- H_2O at 298.15 K. J. Phys. Chem. A 102, 2155–2171. doi:10.1021/jp973043j.

Clerbaux, N., Ipe, A., De Bock, V., Manon, U., Baudrez, E., Velazquez-Blazquez, A., Akkermans, T., Moreels, J., Hollmann, R., Selbach, N., Werscheck, M., 2017. CM SAF Aerosol Optical Depth (AOD) Data Record - Edition 1. doi:10.5676/EUM_SAF_CM/MSG_AOD/V001.

Conrad, V., 1946. Methods in Climatology. Cambridge, Massachusetts.

Contini, D., Costabile, F., 2020. Does Air Pollution Influence COVID-19 Outbreaks? Atmosphere (Basel). 11, 377. doi:10.3390/atmos11040377.

Crippa, M., DeCarlo, P.F., Slowik, J.G., Mohr, C., Heringa, M.F., Chirico, R., Poulain, L., Freutel, F., Sciare, J., Cozic, J., Di Marco, C.F., Elsasser, M., Nicolas, J.B., Marchand, N., Abidi, E., Wiedensohler, A., Drewnick, F., Schneider, J., Borrmann, S., Nemitz, E., Zimmermann, R., Jaffrezo, J.L., Prévôt, A.S.H., Baltensperger, U., 2013. Wintertime aerosol chemical composition and source apportionment of the organic fraction in the metropolitan area of Paris. Atmos. Chem. Phys. 13, 961–981. doi:10.5194/acp-13-961-2013.

Crumeyrolle, S., Chen, G., Ziemba, L., Beyersdorf, A., Thornhill, L., Winstead, E., Moore, R.H., Shook, M.A., Hudgins, C., Anderson, B.E., 2014. Factors that influence surface PM2.5 values inferred from satellite observations: Perspective gained for the US Baltimore-Washington metropolitan area during DISCOVER-AQ. Atmos. Chem. Phys. 14, 2139–2153. doi:10.5194/acp-14-2139-2014.

Curtius, J., 2006. Nucleation of atmospheric aerosol particles. Comptes Rendus Phys. 7, 1027–1045. doi:10.1016/j.crhy.2006.10.018.

Dawson, J.P., Adams, P.J., Pandis, S.N., 2007. Sensitivity of PM2.5 to climate in the Eastern US: a modeling case study. Atmos. Chem. Phys. 7, 4295–4309. doi:10.5194/acp-7-4295-2007.

De Meij, A., Bossioli, E., Penard, C., Vinuesa, J., Price, I., 2015. The effect of SRTM and Corine Land Cover data on calculated gas and PM10 concentrations in WRF-Chem. Atmos. Environ. 101, 177–193. doi:10.1016/j.atmosenv.2014.11.033.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A.P.,

Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. doi:10.1002/qj.828.

Dentener, F., Kinne, S., Bond, T., Boucher, O., Cofala, J., Generoso, S., Ginoux, P., Gong, S., Hoelzemann, J.J., Ito, A., Marelli, L., Penner, J.E., Putaud, J.P., Textor, C., Schulz, M., Van Der Werf, G.R., Wilson, J., 2006. Emissions of primary aerosol and precursor gases in the years 2000 and 1750 prescribed data-sets for AeroCom. Atmos. Chem. Phys. 6, 4321–4344. doi:10.5194/acp-6-4321-2006.

Deutscher Wetterdienst, 2020. Particle Size-Distribution. Available online: www.dwd.de/EN/research/observing_atmosphere/composition_atmosphere/aerosol/cont_nav/ particle_size_distribution_node (accessed on April 15th, 2020).

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. Environ. Sci. Technol. 50, 4712–4721. doi:10.1021/acs.est.5b06121.

Dinoi, A., Perrone, M.R., Burlizzi, P., 2010. Application of modis products for air quality studies over southeastern Italy. Remote Sens. 2, 1767–1796. doi:10.3390/rs2071767.

Doherty, R.M., Heal, M.R., O'Connor, F.M., 2017. Climate change impacts on human health over Europe through its effect on air quality. Environ. Heal. 16, 118. doi:10.1186/ s12940-017-0325-2.

van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C., Lyapustin, A., Sayer, A.M., Winker, D.M., 2016. Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environ. Sci. Technol. 50, 3762–3772. doi:10.1021/acs.est.5b05833.

van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. Environ. Health Perspect. 118, 847–855. doi:10.1289/ehp.0901623.

van Donkelaar, A., Martin, R.V., Park, R.J., 2006. Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. J. Geophys. Res. Atmos. 111, 1–10. doi:10.1029/2005JD006996.

Dubovik, O., Li, Z., Mishchenko, M.I., Tanré, D., Karol, Y., Bojkov, B., Cairns, B., Diner, D.J., Espinosa, W.R., Goloub, P., Gu, X., Hasekamp, O., Hong, J., Hou, W., Knobelspiesse, K.D., Landgraf, J., Li, L., Litvinov, P., Liu, Y., Lopatin, A., Marbach, T., Maring, H., Martins, V., Meijer, Y., Milinevsky, G., Mukai, S., Parol, F., Qiao, Y., Remer, L., Rietjens, J., Sano, I., Stammes, P., Stamnes, S., Sun, X., Tabary, P., Travis, L.D., Waquet, F., Xu, F., Yan, C., Yin, D., 2019. Polarimetric remote sensing of atmospheric aerosols: Instruments, methodologies, results, and perspectives. J. Quant. Spectrosc. Radiat. Transf. 224, 474–511. doi:10.1016/j.jqsrt.2018.11.024.

Dupont, J.C., Haeffelin, M., Badosa, J., Elias, T., Favez, O., Petit, J., Meleux, F., Sciare, J., Crenn, V., Bonne, J., 2016. Role of the boundary layer dynamics effects on an extreme air pollution event in Paris. Atmos. Environ. 141, 571–579. doi:10.1016/j.atmosenv.2016.06.061.

DWD Climate Data Center (CDC), 2017. Historical hourly station observations of 2m air temperature and humidity, version v004, 2016 .

DWD Climate Data Center (CDC), 2018. Grids of monthly averaged daily air temperature (2m) over Germany Version v1.0 .

Edwin, S.G., Mölders, N., 2020. Mesoscale Impacts on Cold Season PM2.5 in the Yukon Flats. J. Environ. Prot. (Irvine,. Calif). 11, 215–240. doi:10.4236/jep.2020.113013.

Egli, S., Thies, B., Bendix, J., 2018. A Hybrid Approach for Fog Retrieval Based on a Combination of Satellite and Ground Truth Data. Remote Sens. 10, 628. doi:10.3390/rs10040628.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813. doi:10.1111/j.1365-2656.2008.01390.x.

Ellison, R.B., Greaves, S.P., Hensher, D.A., 2013. Five years of London's low emission zone: Effects on vehicle fleet composition and air quality. Transp. Res. Part D Transp. Environ. 23, 25–33. doi:10.1016/j.trd.2013.03.010.

Emili, E., Lyapustin, A., Wang, Y., Popp, C., Korkin, S., Zebisch, M., Wunderle, S., Petitta, M., 2011a. High spatial resolution aerosol retrieval with MAIAC: Application to mountain regions. J. Geophys. Res. Atmos. 116, D23211. doi:10.1029/2011JD016297.

Emili, E., Popp, C., Petitta, M., Riffler, M., Wunderle, S., Zebisch, M., 2010. PM10 remote sensing from geostationary SEVIRI and polar-orbiting MODIS sensors over the complex terrain of the European Alpine region. Remote Sens. Environ. 114, 2485–2499. doi:10.1016/j.rse.2010.05.024.

Emili, E., Popp, C., Wunderle, S., Zebisch, M., Petitta, M., 2011b. Mapping particulate matter in alpine regions with satellite and ground-based measurements: An exploratory study for data assimilation. Atmos. Environ. 45, 4344–4353. doi:10.1016/j.atmosenv.2011.05.051.

Engel-Cox, J.A., Holloman, C.H., Coutant, B.W., Hoff, R.M., 2004. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. Atmos. Environ. 38, 2495–2509. doi:10.1016/j.atmosenv.2004.01.039.

Ervens, B., Turpin, B.J., Weber, R.J., 2011. Secondary organic aerosol formation in cloud droplets and aqueous particles (aqSOA): a review of laboratory, field and model studies. Atmos. Chem. Phys. 11, 11069–11102. doi:10.5194/acp-11-11069-2011.

EU, 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe.

Euopean Court of Justice, 2019. European Court of Justice: Case C-723/17 of 26 June 2019.

Favez, O., Cachier, H., Sciare, J., Sarda-Estève, R., Martinon, L., 2009. Evidence for a significant contribution of wood burning aerosols to PM2.5 during the winter season in Paris, France. Atmos. Environ. 43, 3640–3644. doi:10.1016/j.atmosenv.2009.04.035.

Fensterer, V., Küchenhoff, H., Maier, V., Wichmann, H.E., Breitner, S., Peters, A., Gu, J., Cyrys, J., 2014. Evaluation of the impact of low emission zone and heavy traffic ban in Munich (Germany) on the reduction of PM10 in ambient air. Int. J. Environ. Res. Public Health 11, 5094–5112. doi:10.3390/ijerph110505094.

Fowler, D., Pilegaard, K., Sutton, M., Ambus, P., Raivonen, M., Duyzer, J., Simpson, D., Fagerli, H., Fuzzi, S., Schjoerring, J., Granier, C., Neftel, A., Isaksen, I., Laj, P., Maione, M., Monks, P., Burkhardt, J., Daemmgen, U., Neirynck, J., Personne, E., Wichink-Kruit, R., Butterbach-Bahl, K., Flechard, C., Tuovinen, J., Coyle, M., Gerosa, G., Loubet, B., Altimir, N., Gruenhage, L., Ammann, C., Cieslik, S., Paoletti, E., Mikkelsen, T., Ro-Poulsen, H., Cellier, P., Cape, J., Horváth, L., Loreto, F., Niinemets, Ü., Palmer, P., Rinne, J., Misztal, P., Nemitz, E., Nilsson, D., Pryor, S., Gallagher, M., Vesala, T., Skiba, U., Brüggemann, N., Zechmeister-Boltenstern, S., Williams, J., O'Dowd, C., Facchini, M., de Leeuw, G., Flossman, A., Chaumerliac, N., Erisman, J., 2009. Atmospheric composition change: Ecosystems–Atmosphere interactions. Atmos. Environ. 43, 5193–5267. doi:10.1016/j.atmosenv.2009.07.068.

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29, 1189–1232. doi:10.1214/aos/1013203451.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378. doi:10.1016/S0167-9473(01)00065-2.

Fuchs, J., Cermak, J., Andersen, H., 2018. Building a cloud in the Southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning. Atmos. Chem. Phys. Discuss. 2, 1–23. doi:10.5194/acp-2018-593.

Funk, R., Reuter, H.I., Hoffmann, C., Engel, W., Öttl, D., 2008. Effect of moisture on fine dust emission from tillage operations on agricultural soils. Earth Surf. Process. Landforms 33, 1851–1863. doi:10.1002/esp.1737.

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M.C., Fowler, D., Koren, I., Langford, B., Lohmann, U., Nemitz, E., Pandis, S., Riipinen, I., Rudich, Y., Schaap, M., Slowik, J.G., Spracklen, D.V., Vignati, E., Wild, M., Williams, M., Gilardoni, S., 2015. Particulate matter, air quality and climate: lessons learned and future needs. Atmos. Chem. Phys. 15, 8217–8299. doi:10.5194/acp-15-8217-2015.

Garland, J., 1978. Dry and wet removal of sulphur from the atmosphere. Atmos. Environ. 12, 349–362. doi:10.1016/0004-6981(78)90217-2.

Geiß, A., Wiegner, M., Bonn, B., Schäfer, K., Forkel, R., Von Schneidemesser, E., Münkel, C., Lok Chan, K., Nothard, R., 2017. Mixing layer height as an indicator for urban air quality? Atmos. Meas. Tech. 10, 2969–2988. doi:10.5194/amt-10-2969-2017.

Gen, M., Zhang, R., Huang, D.D., Li, Y., Chan, C.K., 2019. Heterogeneous SO_2 Oxidation in Sulfate Formation by Photolysis of Particulate Nitrate. Environ. Sci. Technol. Lett. 6, 86–91. doi:10.1021/acs.estlett.8b00681.

Gentner, D.R., Isaacman, G., Worton, D.R., Chan, A.W., Dallmann, T.R., Davis, L., Liu, S., Day, D.A., Russell, L.M., Wilson, K.R., Weber, R., Guha, A., Harley, R.A., Goldstein, A.H., 2012. Elucidating secondary organic aerosol from diesel and gasoline vehicles through detailed characterization of organic carbon emissions. Proc. Natl. Acad. Sci. U. S. A. 109, 18318–18323. doi:10.1073/pnas.1212272109.

Ginoux, P., Prospero, J.M., Gill, T.E., Hsu, N.C., Zhao, M., 2012. Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. Rev. Geophys. 50, 1–36. doi:10.1029/2012RG000388.

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. J. Comput. Graph. Stat. 24, 44–65. doi:10.1080/10618600.2014.907095.

Goldstein, A.H., Galbally, I.E., 2007. Known and Unexplored Organic Constituents in the Earth's Atmosphere. Environ. Sci. Technol. 41, 1514–1521. doi:10.1021/es072476p.

Granados-Muñoz, M.J., Navas-Guzmán, F., Bravo-Aranda, J.A., Guerrero-Rascado, J.L., Lyamani, H., Valenzuela, A., Titos, G., Fernandez-Galvez, J., Alados-Arboledas, L., 2015. Hygroscopic growth of atmospheric aerosol particles based on active remote sensing and radiosounding measurements: Selected cases in southeastern Spain. Atmos. Meas. Tech. 8, 705–718. doi:10.5194/amt-8-705-2015.

Grange, S.K., Carslaw, D.C., Lewis, A.C., Boleti, E., Hueglin, C., 2018. Random forest meteorological normalisation models for Swiss PM10 trend analysis. Atmos. Chem. Phys. 18, 6223–6239. doi:10.5194/acp-18-6223-2018.

Größ, J., Hamed, A., Sonntag, A., Spindler, G., Elina Manninen, H., Nieminen, T., Kulmala, M., Hõrrak, U., Plass-Dülmer, C., Wiedensohler, A., Birmili, W., 2018. Atmospheric new particle formation at the research station Melpitz, Germany: Connection with gaseous precursors and meteorological parameters. Atmos. Chem. Phys. 18, 1835–1861. doi:10.5194/acp-18-1835-2018.

Guenther, A.B., Zimmerman, P.R., Harley, P.C., Monson, R.K., Fall, R., 1993. Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses. J. Geophys. Res. 98, 12609–12617. doi:10.1029/93JD00527.

Guo, J.P., Zhang, X.Y., Che, H.Z., Gong, S.L., An, X., Cao, C.X., Guang, J., Zhang, H., Wang, Y.Q., Zhang, X.C., Xue, M., Li, X.W., 2009. Correlation between PM concentrations and aerosol optical depth in eastern China. Atmos. Environ. 43, 5876–5886. doi:10.1016/j.atmosenv.2009.08.026.

Gupta, P., Christopher, S.A., 2009a. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. J. Geophys. Res. 114, D20205. doi:10.1029/2008JD011497.

Gupta, P., Christopher, S.A., 2009b. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. J. Geophys. Res. 114, D14205. doi:10.1029/2008JD011496.

Gupta, P., Christopher, S.A., Wang, J., Gehrig, R., Lee, Y., Kumar, N., 2006. Satellite remote sensing of particulate matter and air quality assessment over global cities. Atmos. Environ. 40, 5880–5892. doi:10.1016/j.atmosenv.2006.03.016.

Gupta, P., Levy, R.C., Mattoo, S., Remer, L.A., Munchak, L.A., 2016. A surface reflectance scheme for retrieving aerosol optical depth over urban surfaces in MODIS Dark Target retrieval algorithm. Atmos. Meas. Tech. 9, 3293–3308. doi:10.5194/amt-9-3293-2016.

Haeffelin, M., Bock, O., Boitel, C., Bony, S., Bouniol, D., Chepfer, H., Chiriaco, M., Cuesta, J., Drobinski, P., Flamant, C., Grall, M., Hodzic, A., Hourdin, F., Lapouge, F., Mathieu, A., Morille, Y., Naud, C., Pelon, J., Pietras, C., Protat, A., Romand, B., Scialom, G., Vautard, R., 2005. SIRTA, a ground-based atmospheric observatory for cloud and aerosol research. Ann. Geophys. 23, 253–275. <hal–00329353>.

Hall, S.J., Matson, P.A., Roth, P.M., 1996. NO_x Emissions From Soil: Implications for Air Quality Modeling in Agricultural Regions. Annu. Rev. Energy Environ. 21, 311–346. doi:10.1146/annurev.energy.21.1.311.

Hallquist, M., Wenger, J.C., Baltensperger, U., Rudich, Y., Simpson, D., Claeys, M., Dommen, J., Donahue, N.M., George, C., Goldstein, A.H., Hamilton, J.F., Herrmann, H., Hoffmann, T., Iinuma, Y., Jang, M., Jenkin, M.E., Jimenez, J.L., Kiendler-Scharr, A., Maenhaut, W., McFiggans, G., Mentel, T.F., Monod, A., Prévôt, A.S., Seinfeld, J.H., Surratt, J.D., Szmigielski, R., Wildt, J., 2009. The formation, properties and impact of secondary organic aerosol: Current and emerging issues. Atmos. Chem. Phys. 9, 5155–5236. doi:10.5194/acp-9-5155-2009.

Hande, L.B., Engler, C., Hoose, C., Tegen, I., 2015. Seasonal variability of Saharan desert dust and ice nucleating particles over Europe. Atmos. Chem. Phys. 15, 4389–4397. doi:10.5194/acp-15-4389-2015.

Harbula, J., Kopacková, V., 2011. Air pollution detection using MODIS data, in: Michel, U., Civco, D.L. (Eds.), Earth Resour. Environ. Remote Sensing/GIS Appl. II, p. 81811E. doi:10.1117/12.898107.

Healy, R.M., Sciare, J., Poulain, L., Kamili, K., Merkel, M., Müller, T., Wiedensohler, A., Eckhardt, S., Stohl, A., Sarda-Estève, R., McGillicuddy, E., O'Connor, I.P., Sodeau, J.R., Wenger, J.C., 2012. Sources and mixing state of size-resolved elemental carbon particles in a European megacity: Paris. Atmos. Chem. Phys. 12, 1681–1700. doi:10.5194/acp-12-1681-2012.

Hennig, F., Quass, U., Hellack, B., Küpper, M., Kuhlbusch, T.A.J., Stafoggia, M., Hoffmann, B., 2018. Ultrafine and Fine Particle Number and Surface Area Concentrations and Daily Cause-Specific Mortality in the Ruhr Area, Germany, 2009–2014. Environ. Health Perspect. 126, 027008. doi:10.1289/EHP2054.

Hoff, R.M., Christopher, S.A., Hidy, G., Hoff, R.M., Christopher, S.A., Sharma, P., Poulsen, T., Poulsen, T., Kalluri, P., Hoff, S., Bundy, D., Nelson, M., Zelle, B., Bundy, D., Nelson, M., Zelle, B., Jacobson, L., Heber, A., Ni, J., Ni, J., Zhang, Y., Koziel, J., Beasley, D., Joumard, R., Laurikko, J., Han, T., Geivanidis, S., Samaras, Z., Samaras, Z., Tei, T., Devaux, P., Andre´, J.M., Lacour, S., Lacour, S., Cornelis, E., Chang, V., Hildemann, L., Hildemann, L., Chang, C.h., Chang, V., Hildemann, L., Hildemann, L., Chang, C.h., Lee, J.Y., Keener, T., Yang, Y., Wang, S.W., Tang, X., Fan, Z.H., Wu, X., Lioy, P., Georgopoulos, P., Tang, X., Fan, Z.H., Wu, X., Lioy, P., Georgopoulos, P., Quek, A., Balasubramanian, R., Balasubramanian, R., Chen, Y.C., Chen, L.Y., Jeng, F.T., 2009. The AWMA 2009 Critical Review – Remote Sensing of Particulate Pollution from Space: Have We Reached the Promised Land? J. Air Waste Manage. Assoc. 59, 645–675. doi:10.3155/1047-3289.59.6.645.

Hoffmann, C., Funk, R., 2015. Diurnal changes of PM10-emission from arable soils in NE-Germany. Aeolian Res. 17, 117–127. doi:10.1016/j.aeolia.2015.03.002.

Hoffmann, C., Funk, R., Wieland, R., Li, Y., Sommer, M., 2008. Effects of grazing and topography on dust flux and deposition in the Xilingele grassland, Inner Mongolia. J. Arid Environ. 72, 792–807. doi:10.1016/j.jaridenv.2007.09.004.

Hoffmann, T., Odum, J., Bowman, F., Collins, D., Klockow, D., Flagan, R., Seinfeld, J.H., 1997. Formation of Organic Aerosols from the Oxidation of Biogenic Hydrocarbons. J. Geophys. Res. Atmos. 26, 189–222. doi:10.1023/A:1005734301837.

Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., Smirnov, A., 1998. AERONET — A Federated Instrument Network and Data Archive for Aerosol Characterization. Remote Sens. Environ. 66, 1–16. doi:10.1016/S0034-4257(98)00031-5.

Holzer-Popp, T., Schroedter, M., Gesell, G., 2002. Retrieving aerosol optical depth and type in the boundary layer over land and ocean from simultaneous GOME spectrometer and ATSR-2 radiometer measurements, 1, Method description. J. Geophys. Res. Atmos. 107, 4578. doi:10.1029/2001JD002013.

de Hoogh, K., Héritier, H., Stafoggia, M., Künzli, N., Kloog, I., 2018. Modelling daily PM2.5 concentrations at high spatio-temporal resolution across Switzerland. Environ. Pollut. 233, 1147–1154. doi:10.1016/j.envpol.2017.10.025.

Hsu, N.C., Jeong, M.J., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J., Tsay, S.C., 2013. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. J. Geophys. Res. Atmos. 118, 9296–9315. doi:10.1002/jgrd.50712.

Hsu, N.C., Lee, J., Sayer, A.M., Kim, W., Bettenhausen, C., Tsay, S.C., 2019. VIIRS Deep Blue Aerosol Products Over Land: Extending the EOS Long-Term Aerosol Data Records. J. Geophys. Res. Atmos. 124, 4026–4053. doi:10.1029/2018JD029688.

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. Environ. Sci. Technol. 51, 6936–6944. doi:10.1021/acs.est.7b01210.

Hu, X., Waller, L.A., Lyapustin, A., Wang, Y., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G., Estes, S.M., Quattrochi, D.A., Puttaswamy, S.J., Liu, Y., 2014. Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. Remote Sens. Environ. 140, 220–232. doi:10.1016/j.rse.2013.08.032.

Huang, W., Wang, G., Lu, S.E., Kipen, H., Wang, Y., Hu, M., Lin, W., Rich, D., Ohman-Strickland, P., Diehl, S.R., Zhu, P., Tong, J., Gong, J., Zhu, T., Zhang, J., 2012. Inflammatory and oxidative stress responses of healthy young adults to changes in air quality during the Beijing Olympics. Am. J. Respir. Crit. Care Med. 186, 1150–1159. doi:10.1164/rccm.201205-0850OC.

Hueglin, C., Gehrig, R., Baltensperger, U., Gysel, M., Monn, C., Vonmont, H., 2005. Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland. Atmos. Environ. 39, 637–651. doi:10.1016/j.atmosenv.2004.10.027.

Hughes, H.E., Morbey, R., Fouillet, A., Caserio-Schönemann, C., Dobney, A., Hughes, T.C., Smith, G.E., Elliot, A.J., 2018. Retrospective observational study of emergency department syndromic surveillance data during air pollution episodes across London and Paris in 2014. BMJ Open 8, 1–12. doi:10.1136/bmjopen-2017-018732.

Izquierdo, R., García Dos Santos, S., Borge, R., de la Paz, D., Sarigiannis, D., Gotti, A., Boldo, E., 2020. Health impact assessment by the implementation of Madrid City air-quality plan in 2020. Environ. Res. 183, 109021. doi:10.1016/j.envres.2019.109021.

Jenniches, S., Worrell, E., 2019. Regional economic and environmental impacts of renewable energy developments: Solar PV in the Aachen Region. Energy Sustain. Dev. 48, 11–24. doi:10.1016/j.esd.2018.10.004.

Jethva, H., Torres, O., Yoshida, Y., 2019. Accuracy assessment of MODIS land aerosol optical thickness algorithms using AERONET measurements over North America. Atmos. Meas. Tech. 12, 4291–4307. doi:10.5194/amt-12-4291-2019.

Jiang, H., Yang, Y., Bai, Y., Wang, H., 2020. Evaluation of the Total, Direct, and Diffuse Solar Radiations From the ERA5 Reanalysis Data in China. IEEE Geosci. Remote Sens. Lett. 17, 47–51. doi:10.1109/LGRS.2019.2916410.

Jiang, J., Aksoyoglu, S., El-Haddad, I., Ciarelli, G., Denier van der Gon, H.A.C., Canonaco, F., Gilardoni, S., Paglione, M., Minguillón, M.C., Favez, O., Zhang, Y., Marchand, N., Hao, L., Virtanen, A., Florou, K., Oapos;Dowd, C., Ovadnevaite, J., Baltensperger, U., Prévôt, A.S.H., 2019. Sources of organic aerosols in Europe: a modeling study using CAMx with modified volatility basis set scheme. Atmos. Chem. Phys. 19, 15247–15270. doi:10.5194/acp-19-15247-2019.

Just, A., De Carli, M., Shtein, A., Dorman, M., Lyapustin, A., Kloog, I., 2018. Correcting Measurement Error in Satellite Aerosol Optical Depth with Machine Learning for Modeling PM2.5 in the Northeastern USA. Remote Sens. 10, 803. doi:10.3390/rs10050803.

Kallweit, D., Wintermeyer, D., 2013. Berechnung der gesundheitlichen Belastung der Bevölkerung in Deutschland durch Feinstaub (PM 10). UMID Umwelt und Mensch - Informationsd. 4, 18–24.

Kanakidou, M., Seinfeld, J.H., Pandis, S.N., Barnes, I., Dentener, F.J., Facchini, M.C., Van Dingenen, R., Ervens, B., Nenes, A., Nielsen, C.J., Swietlicki, E., Putaud, J.P., Balkanski, Y., Fuzzi, S., Horth, J., Moortgat, G.K., Winterhalter, R., Myhre, C.E.L., Tsigaridis, K., Vignati, E., Stephanou, E.G., Wilson, J., 2005. Organic aerosol and global climate modelling: a review. Atmos. Chem. Phys. 5, 1053–1123. doi:10.5194/acp-5-1053-2005.

Kapadia, Z.Z., Spracklen, D.V., Arnold, S.R., Borman, D.J., Mann, G.W., Pringle, K.J., Monks, S.A., Reddington, C.L., Benduhn, F., Rap, A., Scott, C.E., Butt, E.W., Yoshioka, M., 2016. Impacts of aviation fuel sulfur content on climate and human health. Atmos. Chem. Phys. 16, 10521–10541. doi:10.5194/acp-16-10521-2016.

Kaufman, Y.J., Tanré, D., Remer, L.A., Vermote, E.F., Chu, A., Holben, B.N., 1997. Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. J. Geophys. Res. Atmos. 102, 17051–17067. doi:10.1029/96JD03988.

Kerschbaumer, A., 2007. On the aerosol budget over Berlin. Phd thesis. Free University, Berlin.

Keuken, M.P., Moerman, M., Voogt, M., Blom, M., Weijers, E.P., Röckmann, T., Dusek, U., 2013. Source contributions to PM2.5 and PM10 at an urban background and a street location. Atmos. Environ. 71, 26–35. doi:10.1016/j.atmosenv.2013.01.032.

Kiesewetter, G., Borken-Kleefeld, J., Schöpp, W., Heyes, C., Thunis, P., Bessagnet, B., Terrenoire, E., Fagerli, H., Nyiri, A., Amann, M., 2015. Modelling street level PM10 concentrations across Europe: source apportionment and possible futures. Atmos. Chem. Phys. 15, 1539–1553. doi:10.5194/acp-15-1539-2015.

Klimont, Z., Kupiainen, K., Heyes, C., Purohit, P., Cofala, J., Rafaj, P., Borken-Kleefeld, J., Schöpp, W., 2017. Global anthropogenic emissions of particulate matter including black carbon. Atmos. Chem. Phys. 17, 8681–8723. doi:10.5194/acp-17-8681-2017.

Klingner, M., Sähn, E., 2008. Prediction of PM10 concentration on the basis of high resolution weather forecasting. Meteorol. Zeitschrift 17, 263–272. doi:10.1127/0941-2948/2008/0288.

Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., Schwartz, J., 2011. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmos. Environ. 45, 6267–6275. doi:10.1016/j.atmosenv.2011.08.066.

Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2012. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the mid-atlantic states. Environ. Sci. Technol. 46, 11913–11921. doi:10.1021/es302673e.

Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A.C., Schwartz, J., Broday, D.M., 2015. Estimating daily PM 2.5 and PM 10 across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. Atmos. Environ. 122, 409–416. doi:10.1016/j.atmosenv.2015.10.004.

Knippertz, P., Todd, M.C., 2012. Mineral dust aerosols over the Sahara: Meteorological controls on emission and transport and implications for modeling. Rev. Geophys. 50, RG1007. doi:10.1029/2011RG000362.

Knüsel, B., Zumwald, M., Baumberger, C., Hirsch Hadorn, G., Fischer, E.M., Bresch, D.N., Knutti, R., 2019. Applying big data beyond small problems in climate research. Nat. Clim. Chang. 9, 196–202. doi:10.1038/s41558-019-0404-1.

Koelemeijer, R.B., Homan, C.D., Matthijsen, J., 2006. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. Atmos. Environ. 40, 5304–5315. doi:10.1016/j.atmosenv.2006.04.044.

Koren, I., Remer, L.A., Kaufman, Y.J., Rudich, Y., Martins, J.V., 2007. On the twilight zone between clouds and aerosols. Geophys. Res. Lett. 34, L08805. doi:10.1029/2007GL029253.

Kotthaus, S., Grimmond, C.S.B., 2018a. Atmospheric boundary-layer characteristics from ceilometer measurements. Part 1: A new method to track mixed layer height and classify clouds. Q. J. R. Meteorol. Soc. 144, 1525–1538. doi:10.1002/qj.3299.

Kotthaus, S., Grimmond, C.S.B., 2018b. Atmospheric boundary-layer characteristics from ceilometer measurements. Part 2: Application to London's urban boundary layer. Q. J. R. Meteorol. Soc. 144, 1511–1524. doi:10.1002/qj.3298.

Laborde, M., Crippa, M., Tritscher, T., Jurányi, Z., Decarlo, P.F., Temime-Roussel, B., Marchand, N., Eckhardt, S., Stohl, A., Baltensperger, U., Prévôt, A.S., Weingartner, E., Gysel, M., 2013. Black carbon physical properties and mixing state in the European megacity Paris. Atmos. Chem. Phys. 13, 5831–5856. doi:10.5194/acp-13-5831-2013.

Laothawornkitkul, J., Taylor, J.E., Paul, N.D., Hewitt, C.N., 2009. Biogenic volatile organic compounds in the Earth system. New Phytol. 183, 27–51. doi:10.1111/j.1469-8137.2009.02859.x.

Lee, H.J., 2019. Benefits of High Resolution PM2.5 Prediction using Satellite MAIAC AOD and Land Use Regression for Exposure Assessment: California Examples. Environ. Sci. Technol. 53, 12774–12783. doi:10.1021/acs.est.9b03799.

Lee, J., Kim, Y., 2007. Spectroscopic measurement of horizontal atmospheric extinction and its practical application. Atmos. Environ. 41, 3546–3555. doi:10.1016/j.atmosenv.2007.01.026.

de Leeuw, G., Andreas, E.L., Anguelova, M.D., Fairall, C.W., Lewis, E.R., O'Dowd, C., Schulz, M., Schwartz, S.E., 2011. Production flux of sea spray aerosol. Rev. Geophys. 49, RG2001. doi:10.1029/2010RG000349.

de Leeuw, J., Methven, J., Blackburn, M., 2015. Evaluation of ERA-Interim reanalysis precipitation products using England and Wales observations. Q. J. R. Meteorol. Soc. 141, 798–806. doi:10.1002/qj.2395.

Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525, 367–371. doi:10.1038/nature15371.

Lelieveld, J., Klingmüller, K., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., Münzel, T., 2019. Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. Eur. Heart J. 0, 1–7. doi:10.1093/eurheartj/ehz135.

Lenschow, P., Abraham, H.J., Kutzner, K., Lutz, M., Preuß, J.D., Reichenbächer, W., 2001. Some ideas about the sources of PM10. Atmos. Environ. 35, 23–33. doi:10.1016/s1352-2310(01)00122-4.

Levy, J.I., Hanna, S.R., 2011. Spatial and temporal variability in urban fine particulate matter concentrations. Environ. Pollut. 159, 2009–2015. doi:10.1016/j.envpol.2010.11.013.

Levy, R.C., Mattoo, S., Munchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., Hsu, N.C., 2013. The Collection 6 MODIS aerosol products over land and ocean. Atmos. Meas. Tech. 6, 2989–3034. doi:10.5194/amt-6-2989-2013.

Levy, R.C., Remer, L.A., Mattoo, S., Vermote, E.F., Kaufman, Y.J., 2007. Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. J. Geophys. Res. Atmos. 112, 1–21. doi:10.1029/2006JD007811.

Li, L., Yang, W., Xie, S., Wu, Y., 2020. Estimations and uncertainty of biogenic volatile organic compound emission inventory in China for 2008–2018. Sci. Total Environ. 733, 139301. doi:10.1016/j.scitotenv.2020.139301.

Li, R., Shinde, A., Liu, A., Glaser, S., Amini, A., 2019a. Identification of Nonlinear Interaction Effects in Prostate Cancer Survival Using Machine Learning-Based Modeling. Int. J. Radiat. Oncol. 105, S121. doi:10.1016/j.ijrobp.2019.06.091.

Li, Y., Chen, Q., Zhao, H., Wang, L., Tao, R., 2015. Variations in PM10, PM2.5 and PM1.0 in an Urban Area of the Sichuan Basin and Their Relation to Meteorological Factors. Atmosphere (Basel). 6, 150–163. doi:10.3390/atmos6010150.

Li, Y., Zhang, J., Sailor, D.J., Ban-Weiss, G.A., 2019b. Effects of urbanization on regional meteorology and air quality in Southern California. Atmos. Chem. Phys. 19, 4439–4457. doi:10.5194/acp-19-4439-2019.

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H., Zhu, B., 2017. Aerosol and boundary-layer interactions and impact on air quality. Natl. Sci. Rev. 4, 810–833. doi:10.1093/nsr/nwx117.

Liu, Q., Jia, X., Quan, J., Li, J., Li, X., Wu, Y., Chen, D., Wang, Z., Liu, Y., 2018. New positive feedback mechanism between boundary layer meteorology and secondary aerosol formation during severe haze events. Sci. Rep. 8, 1–8. doi:10.1038/s41598-018-24366-3.

148

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.i., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. doi:10.1038/s42256-019-0138-9.

Lundberg, S.M., Erion, G.G., Lee, S.I., 2018a. Consistent Individualized Feature Attribution for Tree Ensembles , 1–9. arXiv:1802.03888.

Lundberg, S.M., Lee, S.I., 2017. A Unified Approach to Interpreting Model Predictions , 1–10. arXiv: 1705.07874.

Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K.W., Newman, S.F., Kim, J., Lee, S.I., 2018b. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat. Biomed. Eng. 2, 749–760. doi:10.1038/s41551-018-0304-0.

Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., Korkin, S., 2011a. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. J. Geophys. Res. 116, D03210. doi:10.1029/2010JD014985.

Lyapustin, A., Wang, Y., Korkin, S., Huang, D., 2018. MODIS Collection 6 MAIAC algorithm. Atmos. Meas. Tech. 11, 5741–5765. doi:10.5194/amt-11-5741-2018.

Lyapustin, A., Wang, Y., Laszlo, I., Hilker, T., G.Hall, F., Sellers, P.J., Tucker, C.J., Korkin, S.V., 2012. Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. Atmospheric correction. Remote Sens. Environ. 127, 385–393. doi:10.1016/j.rse.2012.09.002.

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., Reid, J.S., 2011b. Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. J. Geophys. Res. 116, D03211. doi:10.1029/2010JD014986.

Malings, C., Westervelt, D.M., Hauryliuk, A., Presto, A., Subramanian, R., 2020. Application of Low-Cost Fine Particulate Mass Monitors to Convert Satellite Aerosol Optical Depth Measurements to Surface Concentrations in North America and Africa. Atmos. Meas. Tech. Discuss. submitted, 1–30. doi:10.5194/AMT-2020-67.

McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G.E., Elmore, K.L., Homeyer, C.R., Smith, T., 2019. Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. Bull. Am. Meteorol. Soc. 100, 2175–2199. doi:10.1175/BAMS-D-18-0195.1.

Megaritis, A.G., Fountoukis, C., Charalampidis, P.E., Denier Van Der Gon, H.A., Pilinis, C., Pandis, S.N., 2014. Linking climate and air quality over Europe: Effects of meteorology on PM2.5 concentrations. Atmos. Chem. Phys. 14, 10283–10298. doi:10.5194/acp-14-10283-2014.

Megaritis, A.G., Fountoukis, C., Charalampidis, P.E., Pilinis, C., Pandis, S.N., 2013. Response of fine particulate matter concentrations to changes of emissions and temperature in Europe. Atmos. Chem. Phys. 13, 3423–3443. doi:10.5194/acp-13-3423-2013.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. doi:10.1016/j.envsoft.2017.12.001.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecol. Modell. 411, 108815. doi:10.1016/j.ecolmodel.2019.108815.

Möhler, O., DeMott, P.J., Vali, G., Levin, Z., 2007. Microbiology and atmospheric processes: The role of biological particles in cloud physics. Biogeosciences 4, 1059–1071. doi:10.5194/bg-4-1059-2007.

Morfeld, P., Erren, T., 2019. Warum ist die „Anzahl vorzeitiger Todesfälle durch Umweltexpositionen" nicht angemessen quantifizierbar? Das Gesundheitswes. 81, 144–149. doi:10.1055/a-0832-2038.

NASA, 2020. NASA Worldview application - part of the NASA Earth Observing System Data and Information System (EOSDIS). Available at: https://worldview.earthdata.nasa.gov (accessed on April 15th, 2020).

Ng, N.L., Herndon, S.C., Trimborn, A., Canagaratna, M.R., Croteau, P.L., Onasch, T.B., Sueper, D., Worsnop, D.R., Zhang, Q., Sun, Y.L., Jayne, J.T., 2011. An Aerosol Chemical Speciation Monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol. Aerosol Sci. Technol. 45, 770–784. doi:10.1080/02786826.2011.560211.

Nordio, F., Kloog, I., Coull, B.A., Chudnovsky, A., Grillo, P., Bertazzi, P.A., Baccarelli, A.A., Schwartz, J., 2013. Estimating spatio-temporal resolved PM10 aerosol mass concentrations using MODIS satellite data and land use regression over Lombardy, Italy. Atmos. Environ. 74, 227–236. doi:10.1016/j.atmosenv.2013.03.043.

Olauson, J., 2018. ERA5: The new champion of wind power modelling? Renew. Energy 126, 322–331. doi:10.1016/j.renene.2018.03.056.

Park, S., Shin, M., Im, J., Song, C.K., Choi, M., Kim, J., Lee, S., Park, R., Kim, J., Lee, D.W., Kim, S.K., 2019. Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. Atmos. Chem. Phys. 19, 1097–1113. doi:10.5194/acp-19-1097-2019.

Pay, M.T., Jiménez-Guerrero, P., Baldasano, J.M., 2012. Assessing sensitivity regimes of secondary inorganic aerosol formation in Europe with the CALIOPE-EU modeling system. Atmos. Environ. 51, 146–164. doi:10.1016/j.atmosenv.2012.01.027.

Pearce, J.L., Beringer, J., Nicholls, N., Hyndman, R.J., Uotila, P., Tapper, N.J., 2011. Investigating the influence of synoptic-scale meteorology on air quality using self-organizing maps and generalized additive modelling. Atmos. Environ. 45, 128–136. doi:10.1016/j.atmosenv.2010.09.032.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2012. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. , 2825–2830. arXiv: 1201.0490.

Perraud, V., Bruns, E.A., Ezell, M.J., Johnson, S.N., Yu, Y., Alexander, M.L., Zelenyuk, A., Imre, D., Chang, W.L., Dabdub, D., Pankow, J.F., Finlayson-Pitts, B.J., 2012. Nonequilibrium atmospheric secondary organic aerosol formation and growth. Proc. Natl. Acad. Sci. 109, 2836–2841. doi:10.1073/pnas.1119909109.

Petäjä, T., Järvi, L., Kerminen, V.M., Ding, A., Sun, J., Nie, W., Kujansuu, J., Virkkula, A., Yang, X., Fu, C., Zilitinkevich, S., Kulmala, M., 2016. Enhanced air pollution via aerosol-boundary layer feedback in China. Sci. Rep. 6, 18998. doi:10.1038/srep18998.

Petetin, H., Beekmann, M., Sciare, J., Bressi, M., Rosso, A., Sanchez, O., Ghersi, V., 2014. A novel model evaluation approach focusing on local and advected contributions to urban PM2.5 levels - Application to Paris, France. Geosci. Model Dev. 7, 1483–1505. doi:10.5194/gmd-7-1483-2014.

Petit, J.E., Amodeo, T., Meleux, F., Bessagnet, B., Menut, L., Grenier, D., Pellan, Y., Ockler, A., Rocq, B., Gros, V., Sciare, J., Favez, O., 2017. Characterising an intense PM pollution episode in March 2015 in France from multi-site approach and near real time data: Climatology,

variabilities, geographical origins and model evaluation. Atmos. Environ. 155, 68–84. doi:10.1016/j.atmosenv.2017.02.012.

Petit, J.E., Favez, O., Sciare, J., Canonaco, F., Croteau, P., Močnik, G., Jayne, J., Worsnop, D., Leoz-Garziandia, E., 2014. Submicron aerosol source apportionment of wintertime pollution in Paris, France by double positive matrix factorization (PMF2) using an aerosol chemical speciation monitor (ACSM) and a multi-wavelength Aethalometer. Atmos. Chem. Phys. 14, 13773–13787. doi:10.5194/acp-14-13773-2014.

Petit, J.E., Favez, O., Sciare, J., Crenn, V., Sarda-Estève, R., Bonnaire, N., Močnik, G., Dupont, J.C., Haeffelin, M., Leoz-Garziandia, E., 2015. Two years of near real-time chemical composition of submicron aerosols in the region of Paris using an Aerosol Chemical Speciation Monitor (ACSM) and a multi-wavelength Aethalometer. Atmos. Chem. Phys. 15, 2985–3005. doi:10.5194/acp-15-2985-2015.

van Pinxteren, D., Mothes, F., Spindler, G., Fomba, K.W., Herrmann, H., 2019. Trans-boundary PM10: Quantifying impact and sources during winter 2016/17 in eastern Germany. Atmos. Environ. 200, 119–130. doi:10.1016/j.atmosenv.2018.11.061.

Pope, C.A., Burnett, R.T., Thun, M.J., E., C.E., Krewski, D., Ito, K., Thurston, G.D., 2002. Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. JAMA 287, 1132–1141. doi:10.1001/jama.287.9.1132.

Putaud, J.P., Raes, F., Van Dingenen, R., Brüggemann, E., Facchini, M.C., Decesari, S., Fuzzi, S., Gehrig, R., Hüglin, C., Laj, P., Lorbeer, G., Maenhaut, W., Mihalopoulos, N., Müller, K., Querol, X., Rodriguez, S., Schneider, J., Spindler, G., ten Brink, H., Tørseth, K., Wiedensohler, A., 2004. A European aerosol phenomenology—2: chemical characteristics of particulate matter at kerbside, urban, rural and background sites in Europe. Atmos. Environ. 38, 2579–2595. doi:10.1016/j.atmosenv.2004.01.041.

Putaud, J.P., Van Dingenen, R., Alastuey, A., Bauer, H., Birmili, W., Cyrys, J., Flentje, H., Fuzzi, S., Gehrig, R., Hansson, H., Harrison, R., Herrmann, H., Hitzenberger, R., Hüglin, C., Jones, A., Kasper-Giebl, A., Kiss, G., Kousa, A., Kuhlbusch, T., Löschau, G., Maenhaut, W., Molnar, A., Moreno, T., Pekkanen, J., Perrino, C., Pitz, M., Puxbaum, H., Querol, X., Rodriguez, S., Salma, I., Schwarz, J., Smolik, J., Schneider, J., Spindler, G., ten Brink, H., Tursic, J., Viana, M., Wiedensohler, A., Raes, F., 2010. A European aerosol phenomenology – 3: Physical and chemical characteristics of particulate matter from 60 rural, urban, and kerbside sites across Europe. Atmos. Environ. 44, 1308–1320. doi:10.1016/j.atmosenv.2009.12.011.

Qadir, R.M., Abbaszade, G., Schnelle-Kreis, J., Chow, J.C., Zimmermann, R., 2013. Concentrations and source contributions of particulate organic matter before and after implementation of a low emission zone in Munich, Germany. Environ. Pollut. 175, 158–167. doi:10.1016/j.envpol.2013.01.002.

Radke, L.F., Hobbs, P.V., Eltgroth, M.W., 1980. Scavenging of aerosol particles by precipitation. J. Appl. Meteorol. 19, 715–722. doi:10.1175/1520-0450(1980)019<0715:SOAPBP>2.0.CO;2.

Reizer, M., Juda-Rezler, K., 2016. Explaining the high PM10concentrations observed in Polish urban areas. Air Qual. Atmos. Heal. 9, 517–531. doi:10.1007/s11869-015-0358-z.

Rengarajan, R., Sudheer, A., Sarin, M., 2011. Wintertime PM2.5 and PM10 carbonaceous and inorganic constituents from urban site in western India. Atmos. Res. 102, 420–431. doi:10.1016/j.atmosres.2011.09.005.

Román, M.O., Wang, Z., Sun, Q., Kalb, V., Miller, S.D., Molthan, A., Schultz, L., Bell, J., Stokes, E.C., Pandey, B., Seto, K.C., Hall, D., Oda, T., Wolfe, R.E., Lin, G., Golpayegani, N., Devadiga, S., Davidson, C., Sarkar, S., Praderas, C., Schmaltz, J., Boller, R., Stevens, J., Ramos González, O.M., Padilla, E., Alonso, J., Detrés, Y., Armstrong, R., Miranda, I., Conte, Y., Marrero, N., MacManus, K., Esch, T., Masuoka, E.J., 2018. NASA's Black Marble nighttime lights product suite. Remote Sens. Environ. 210, 113–143. doi:10.1016/j.rse.2018.03.017.

Rost, J., Holst, T., Sahn, E., Klingner, M., Anke, K., Ahrens, D., Mayer, H., 2009. Variability of PM10 concentrations dependent on meteorological conditions. Int. J. Environ. Pollut. 36, 3–18. doi:10.1504/IJEP.2009.021813.

Rybarczyk, Y., Zalakeviciute, R., 2018. Machine learning approaches for outdoor air quality modelling: A systematic review. doi:10.3390/app8122570.

Sá, E., Martins, H., Ferreira, J., Marta-Almeida, M., Rocha, A., Carvalho, A., Freitas, S., Borrego, C., 2016. Climate change and pollutant emissions impacts on air quality in 2050 over Portugal. Atmos. Environ. 131, 209–224. doi:10.1016/j.atmosenv.2016.01.040.

Sayer, A.M., Munchak, L.A., Hsu, N.C., Levy, R.C., Bettenhausen, C., Jeong, M.J., 2014. MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and "merged" data sets, and usage recommendations. J. Geophys. Res. Atmos. 119, 13965–13989. doi:10.1002/2014JD022453.

Schäfer, K., Harbusch, A., Emeis, S., Koepke, P., Wiegner, M., 2008. Correlation of aerosol mass near the ground with aerosol optical depth during two seasons in Munich. Atmos. Environ. 42, 4036–4046. doi:10.1016/j.atmosenv.2008.01.060.

Schäfer, K., Wagner, P., Emeis, S., Jahn, C., Münkel, C., Suppan, P., Ecology, L., Gmbh, V., 2012. Mixing layer height and air pollution levels in urban area. Proc. SPIE 8534, 1–10. doi:10.1117/12.974328.

Schwarz, K., Cermak, J., Fuchs, J., Andersen, H., 2017. Mapping the twilight zone-What we are missing between clouds and aerosols. Remote Sens. 9, 1–10. doi:10.3390/rs9060577.

Sciare, J., D'Argouges, O., Zhang, Q.J., Sarda-Estève, R., Gaimoz, C., Gros, V., Beekmann, M., Sanchez, O., 2010. Comparison between simulated and observed chemical composition of fine aerosols in Paris (France) during springtime: Contribution of regional versus continental emissions. Atmos. Chem. Phys. 10, 11987–12004. doi:10.5194/acp-10-11987-2010.

Shapley, L., 1953. A Value for n-Person Games, in: Kuhn, H.W., Tucker, A.W. (Eds.), Contrib. to Theory Games (AM-28), Vol. II. Princeton University Press, Princeton. volume 28, pp. 307–318. doi:10.1515/9781400881970-018.

Shell, J.R.J., 2004. Bidirectional Reflectance : An Overview with Remote Sensing Applications Measurement Recommendations. R.I.T. , 1–80.

Shin, M., Kang, Y., Park, S., Im, J., Yoo, C., Quackenbush, L.J., 2020. Estimating ground-level particulate matter concentrations using satellite-based data: a review. GIScience Remote Sens. 57, 174–189. doi:10.1080/15481603.2019.1703288.

Sinharay, R., Gong, J., Barratt, B., Ohman-Strickland, P., Ernst, S., Kelly, F.J., Zhang, J.J., Collins, P., Cullinan, P., Chung, K.F., 2018. Respiratory and cardiovascular responses to walking down a traffic-polluted road compared with walking in a traffic-free area in participants aged 60 years and older with chronic lung or heart disease and age-matched healthy controls: a randomised, crosso. Lancet 391, 339–349. doi:10.1016/S0140-6736(17)32643-0.

Sogacheva, L., Popp, T., Sayer, A.M., Dubovik, O., Garay, M.J., Heckel, A., Christina Hsu, N., Jethva, H., Kahn, R.A., Kolmonen, P., Kosmale, M., De Leeuw, G., Levy, R.C., Litvinov, P., Lyapustin, A., North, P., Torres, O., Arola, A., 2020. Merging regional and global aerosol optical depth records from major available satellite products. Atmos. Chem. Phys. 20, 2031–2056. doi:10.5194/acp-20-2031-2020.

Soh, P.W., Chang, J.W., Huang, J.W., 2018. Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations. IEEE Access 6, 38186–38199. doi:10.1109/ACCESS.2018.2849820.

Sorek-Hamer, M., Broday, D.M., Chatfield, R., Esswein, R., Stafoggia, M., Lepeule, J., Lyapustin, A., Kloog, I., 2017. Monthly analysis of PM ratio characteristics and its relation to AOD. J. Air Waste Manage. Assoc. 67, 27–38. doi:10.1080/10962247.2016.1208121.

Srivastava, D., Favez, O., Bonnaire, N., Lucarelli, F., Haeffelin, M., Perraudin, E., Gros, V., Villenave, E., Albinet, A., 2018. Speciation of organic fractions does matter for aerosol source apportionment. Part 2: Intensive short-term campaign in the Paris area (France). Sci. Total Environ. 634, 267–278. doi:10.1016/j.scitotenv.2018.03.296.

Stafoggia, M., Johansson, C., Glantz, P., Renzi, M., Shtein, A., de Hoogh, K., Kloog, I., Davoli, M., Michelozzi, P., Bellander, T., 2020. A Random Forest Approach to Estimate Daily Particulate Matter, Nitrogen Dioxide, and Ozone at Fine Spatial Resolution in Sweden. Atmosphere (Basel). 11, 239. doi:10.3390/atmos11030239.

Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., de' Donato, F., Gaeta, A., Leone, G., Lyapustin, A., Sorek-Hamer, M., de Hoogh, K., Di, Q., Forastiere, F., Kloog, I., 2017. Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. Environ. Int. 99, 234–244. doi:10.1016/j.envint.2016.11.024.

Stieb, D.M., Shutt, R., Kauri, L., Mason, S., Chen, L., Szyszkowicz, M., Dobbin, N.A., Rigden, M., Jovic, B., Mulholland, M., Green, M.S., Liu, L., Pelletier, G., Weichenthal, S.A., Dales, R.E., Luginaah, I., 2017. Cardio-Respiratory Effects of Air Pollution in a Panel Study of Outdoor Physical Activity and Health in Rural Older Adults. J. Occup. Environ. Med. 59, 356–364. doi:10.1097/JOM.0000000000000954.

Stirnberg, R., Cermak, J., Andersen, H., 2018. An Analysis of Factors Influencing the Relationship between Satellite-Derived AOD and Ground-Level PM10. Remote Sens. 10, 1353. doi:10.3390/rs10091353.

Stirnberg, R., Cermak, J., Fuchs, J., Andersen, H., 2020. Mapping and Understanding Patterns of Air Quality Using Satellite Data and Machine Learning. J. Geophys. Res. Atmos. 125, e2019JD031380. doi:10.1029/2019JD031380.

Stolwijk, A.M., Straatman, H., Zielhuis, G.A., 1999. Studying seasonality by using sine and cosine functions in regression analysis. J. Epidemiol. Community Heal. 53, 235–238. doi:10.1136/jech.53.4.235.

Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8, 1–25. doi:10.1186/1471-2105-8-25.

Sujatha, P., Mahalakshmi, D., Ramiz, A., Rao, P., Naidu, C., 2016. Ventilation coefficient and boundary layer height impact on urban air quality. Cogent Environ. Sci. 2, 1–9. doi:10.1080/23311843.2015.1125284.

Superczynski, S.D., Kondragunta, S., Lyapustin, A.I., 2017. Evaluation of the multi-angle implementation of atmospheric correction (MAIAC) aerosol algorithm through intercomparison with VIIRS aerosol products and AERONET. J. Geophys. Res. Atmos. 122, 3005–3022. doi:10.1002/2016JD025720.

Tanré, D., Kaufman, Y.J., Herman, M., Mattoo, S., 1997. Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances. J. Geophys. Res. Atmos. 102, 16971–16988. doi:10.1029/96jd03437.

Tarek, M., Brisette, F.P., Arsenault, R., 2019. Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modeling over North-America. Hydrol. Earth Syst. Sci. Discuss. submitted, 1–35. doi:10.5194/hess-2019-316.

Theloke, J., Thiruchittampalam, B., Orlikova, S., Uzbasich, M., Gauger, T., 2009. Methodology development for the spatial distribution of the diffuse emissions in Europe. Technical Report. European Commission.

Titos, G., Jefferson, A., Sheridan, P.J., Andrews, E., Lyamani, H., Alados-Arboledas, L., Ogren, J.A., 2014. Aerosol light-scattering enhancement due to water uptake during the TCAP campaign. Atmos. Chem. Phys. 14, 7031–7043. doi:10.5194/acp-14-7031-2014.

Truong, C., Oudre, L., Vayatis, N., 2020. Selective review of offline change point detection methods. Signal Processing 167. doi:10.1016/j.sigpro.2019.107299.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. doi:10.1016/0034-4257(79)90013-0.

Turnock, S.T., Butt, E.W., Richardson, T.B., Mann, G.W., Reddington, C.L., Forster, P.M., Haywood, J., Crippa, M., Janssens-Maenhout, G., Johnson, C.E., Bellouin, N., Carslaw, K.S., Spracklen, D.V., 2016. The impact of European legislative and technology measures to reduce air pollutants on air quality, human health and climate. Environ. Res. Lett. 11, 024010. doi:10.1088/1748-9326/11/2/024010.

TÜV Rheinland, 2012. Report on the suitability test of the ambient air quality measuring system TEOM 1405-DF Ambient Particulate Monitor with PM10 pre-separator and virtual impactor of the company Thermo Fisher Scientific for the components PM10 and PM2.5. Technical Report. TÜV Rheinland Energie und Umwelt GmbH.

Umweltbundesamt, 2004. Qualitätssicherungshandbuch des UBA-Messnetzes. Technical Report. Federal Environment Agency. Berlin.

Umweltbundesamt, 2019. Air Quality 2018: Preliminary Evaluation. Technical Report. Umweltbundesamt. Dessau-Roßlau. doi:ISSN2363-829X.

Várnai, T., Marshak, A., Yang, W., 2013. Multi-satellite aerosol observations in the vicinity of clouds. Atmos. Chem. Phys. 13, 3899–39808. doi:10.5194/acp-13-3899-2013.

VDI, 2002. VDI-Richtlinien: Minimum requirements for suitability tests of automated ambient air quality measuring systems Point-related measurement methods of gaseous and particulate pollutants.

Von Engeln, A., Teixeira, J., 2013. A planetary boundary layer height climatology derived from ECMWF reanalysis data. J. Clim. 26, 6575–6590. doi:10.1175/JCLI-D-12-00385.1.

Wagner, P., Schäfer, K., 2017. Influence of mixing layer height on air pollutant concentrations in an urban street canyon. Urban Clim. 22, 64–79. doi:10.1016/j.uclim.2015.11.001.

Wagner, S., Angenendt, E., Beletskaya, O., Zeddies, J., 2015. Costs and benefits of ammonia and particulate matter abatement in German agriculture including interactions with greenhouse gas emissions. Agric. Syst. 141, 58–68. doi:10.1016/j.agsy.2015.09.003.

Wang, G., Zhang, R., Gomez, M.E., Yang, L., Zamora, M.L., Hu, M., Lin, Y., Peng, J., Guo, S., Meng, J., Li, J., Cheng, C., Hu, T., Ren, Y., Wang, Y., Gao, J., Cao, J., An, Z., Zhou, W., Li, G., Wang, J., Tian, P., Marrero-Ortiz, W., Secrest, J., Du, Z., Zheng, J., Shang, D., Zeng, L., Shao, M., Wang, W., Huang, Y., Wang, Y., Zhu, Y., Li, Y., Hu, J., Pan, B., Cai, L., Cheng, Y., Ji, Y., Zhang, F., Rosenfeld, D., Liss, P.S., Duce, R.A., Kolb, C.E., Molina, M.J., 2016. Persistent sulfate formation from London Fog to Chinese haze. Proc. Natl. Acad. Sci. U. S. A. 113, 13630–13635. doi:10.1073/pnas.1616540113.

Wang, J., 2003. Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies. Geophys. Res. Lett. 30, 2095. doi:10.1029/2003GL018174.

Wang, J., Martin, S.T., 2007. Satellite characterization of urban aerosols: Importance of including hygroscopicity and mixing state in the retrieval algorithms. J. Geophys. Res. Atmos. 112, 1–18. doi:10.1029/2006JD008078.

Wang, J., Ogawa, S., 2015. Effects of meteorological conditions on PM2.5 concentrations in Nagasaki, Japan. Int. J. Environ. Res. Public Health 12, 9089–9101. doi:10.3390/ijerph120809089.

Wang, X., Dickinson, R.E., Su, L., Zhou, C., Wang, K., 2018. PM2.5 Pollution in China and How It Has Been Exacerbated by Terrain and Meteorological Conditions. Bull. Am. Meteorol. Soc. 99, 105–119. doi:10.1175/BAMS-D-16-0301.1.

Wang, X., Wang, W., Yang, L., Gao, X., Nie, W., Yu, Y., Xu, P., Zhou, Y., Wang, Z., 2012. The secondary formation of inorganic aerosols in the droplet mode through heterogeneous aqueous reactions under haze conditions. Atmos. Environ. 63, 68–76. doi:10.1016/j.atmosenv.2012.09.029.

Weaver, A.M., McGuinn, L., Neas, L., Mirowsky, J., Devlin, R.B., Dhingra, R., Ward-Caviness, C., Cascio, W.E., Kraus, W.E., Hauser, E.R., Di, Q., Schwartz, J., Diaz-Sanchez, D., 2019. Neighborhood Sociodemographic Effects on the Associations Between Long-term PM2.5 Exposure and Cardiovascular Outcomes and Diabetes Mellitus. Environ. Epidemiol. 3, e038. doi:10.1097/EE9.0000000000000038.

Weigl, E., 2017. RADOLAN: Radar Online Adjustment. Radar based quantitative precipitation estimation products. Available at: www.dwd.de/RADOLAN (accessed on April 15th, 2020) .

Wiegner, M., Madonna, F., Binietoglou, I., Forkel, R., Gasteiger, J., Geiß, A., Pappalardo, G., Schäfer, K., Thomas, W., 2014. What is the benefit of ceilometers for aerosol remote sensing? An answer from EARLINET. Atmos. Meas. Tech. 7, 1979–1997. doi:10.5194/amt-7-1979-2014.

Winker, D.M., Vaughan, M.A., Omar, A., Hu, Y., Powell, K., Liu, Z., Hunt, W.H., Young, S.A., 2009. Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms. J. Atmos. Ocean. Technol. 26, 2310–2323. doi:0.1175/2009JTECHA1281.1.

World Health Organization, 2006. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: Global update 2005. Technical Report. World Health Organization. Geneva. doi:10.1016/0004-6981(88)90109-6.

Xiao, W., Nethery, R.C., Sabath, M.B., Braun, D., Dominici, F., 2020. Exposure to air pollution and COVID-19 mortality in the United States. medRxiv 04.05., 20054502. doi:10.1101/2020.04.05.20054502.

Yahi, H., Marticorena, B., Thiria, S., Chatenet, B., Schmechtig, C., Rajot, J.L., Crepon, M., 2013. Statistical relationship between surface PM10 concentration and aerosol optical depth over the Sahel as a function of weather type, using neural network methodology. J. Geophys. Res. Atmos. 118, 13265–13281. doi:10.1002/2013JD019465.

Yang, M., Chu, C., Bloom, M.S., Li, S., Chen, G., Heinrich, J., Markevych, I., Knibbs, L.D., Bowatte, G., Dharmage, S.C., Komppula, M., Leskinen, A., Hirvonen, M.R., Roponen, M., Jalava, P., Wang, S.Q., Lin, S., Zeng, X.W., Hu, L.W., Liu, K.K., Yang, B.Y., Chen, W., Guo, Y., Dong, G.H., 2018. Is smaller worse? New insights about associations of PM1 and respiratory health in children and adolescents. Environ. Int. 120, 516–524. doi:10.1016/j.envint.2018.08.027.

Yang, Y.Q., Wang, J.Z., Gong, S.L., Zhang, X.Y., Wang, H., Wang, Y.Q., Wang, J., Li, D., Guo, J.P., 2016. PLAM - A meteorological pollution index for air quality and its applications in fog-haze forecasts in North China. Atmos. Chem. Phys. 16, 1353–1364. doi:10.5194/acp-16-1353-2016.

Yazdi, D.M., Kuang, Z., Dimakopoulou, K., Barratt, B., Suel, E., Amini, H., Lyapustin, A., Katsouyanni, K., Schwartz, J., 2020. Predicting Fine Particulate Matter (PM2.5) in the Greater London Area: An Ensemble Approach using Machine Learning Methods. Remote Sens. 12, 914. doi:10.3390/rs12060914.

Yongjian, Z., Jingu, X., Fengming, H., Liqing, C., 2020. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. Sci. Total Environ. 138704. doi:10.1016/j.scitotenv.2020.138704.

Zhang, H., Hoff, R.M., Kondragunta, S., Laszlo, I., Lyapustin, A., 2013. Aerosol optical depth (AOD) retrieval using simultaneous GOES-East and GOES-West reflected radiances over the western United States. Atmos. Meas. Tech. 6, 471–486. doi:10.5194/amt-6-471-2013.

Zhang, J., Reid, J.S., 2010. A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products. Atmos. Chem. Phys. 10, 10949–10963. doi:10.5194/acp-10-10949-2010.

Zhang, X., Chu, Y., Wang, Y., Zhang, K., 2018. Predicting daily PM2.5 concentrations in Texas using high-resolution satellite aerosol optical depth. Sci. Total Environ. 631-632, 904–911. doi:10.1016/j.scitotenv.2018.02.255.

Zhao, C., Huang, M., Fast, J.D., Berg, L.K., Qian, Y., Guenther, A., Gu, D., Shrivastava, M., Liu, Y., Walters, S., Pfister, G., Jin, J., Shilling, J.E., Warneke, C., 2016. Sensitivity of biogenic volatile organic compounds to land surface parameterizations and vegetation distributions in California. Geosci. Model Dev. 9, 1959–1976. doi:10.5194/gmd-9-1959-2016.

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., Chen, T., Wu, F., Qiu, Y., 2017. Analysis of influential factors for the relationship between PM 2.5 and AOD in Beijing. Atmos. Chem. Phys. 17, 13473–13489. doi:10.5194/acp-17-13473-2017.

Zieger, P., Fierz-Schmidhauser, R., Poulain, L., Müller, T., Birmili, W., Spindler, G., Wiedensohler, A., Baltensperger, U., Weingartner, E., 2014. Influence of water uptake on the aerosol particle light scattering coefficients of the Central European aerosol. Tellus B Chem. Phys. Meteorol. 66, 22716. doi:10.3402/tellusb.v66.22716.

Zieger, P., Fierz-Schmidhauser, R., Weingartner, E., Baltensperger, U., 2013. Effects of relative humidity on aerosol light scattering: Results from different European sites. Atmos. Chem. Phys. 13, 10609–10631. doi:10.5194/acp-13-10609-2013.

Further appendices

Original publication: Stirnberg et al. (2018)

Stirnberg, R., Cermak, J. and Andersen, H.: An Analysis of Factors Influencing the Relationship between Satellite-Derived AOD and Ground-Level PM10. Remote Sensing 2018, 10, 1353. doi: 10.3390/rs10091353

Contributions: Roland Stirnberg conceived the study design, conducted the investigation and developed the methodology including coding. Roland Stirnberg visualised and interpreted the results. Roland Stirnberg wrote the original draft and incorporated the reviewer comments.

Peer-Review: The manuscript was submitted to Remote Sensing on July 27th, 2018. After a peer-review with two anonymous reviewers, the manuscript was accepted on August 22nd, 2018 and published on August 25th, 2018.

Usage of text passages and modifications: The publication is included in abbreviated form in chapter 2. Parts of the method section and of the introduction of the original manuscript were moved to chapter 1.3. The numbering of figures was changed and section titles were renamed.

remote sensing

MDPI

Article

An Analysis of Factors Influencing the Relationship between Satellite-Derived AOD and Ground-Level PM10

Roland Stirnberg [1,2,*] ![ORCID], **Jan Cermak** [1,2] ![ORCID] **and Hendrik Andersen** [1,2] ![ORCID]

1   Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research,
    H.-v.-Helmholtz-Platz 1, 76344 Leopoldshafen, Germany; jan.cermak@kit.edu (J.C.);
    hendrik.andersen@kit.edu (H.A.)
2   Karlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing,
    Englerstr. 7, 76131 Karlsruhe, Germany
*   Correspondence: roland.stirnberg@kit.edu; Tel.: +49-721-608-24854

☑ check for updates

Abstract: Air pollution can endanger human health, especially in urban areas. Assessment of air quality primarily relies on ground-based measurements, but these provide only limited information on the spatial distribution of pollutants. In recent years, satellite derived Aerosol Optical Depth (AOD) has been used to approximate particulate matter (PM) with varying success. In this study, the relationship between hourly mean concentrations of particulate matter with a diameter of 10 micrometers or less (PM10) and instantaneous AOD measurements is investigated for Berlin, Germany, for 2001–2015. It is found that the relationship between AOD and PM10 is rarely linear and strongly influenced by ambient relative humidity (RH), boundary layer height (BLH), wind direction and wind speed. Generally, when a moderately dry atmosphere (30% < RH ≤ 50%) coincides with a medium BLH (600–1200 m), AOD and PM10 are in the same range on a semi-quantitative scale. AOD increases with ambient RH, leading to an overestimation of the dry particle concentration near ground. However, this effect can be compensated if a low boundary layer (<600 m) is present, which in turn significantly increases PM10, eventually leading to satellite AOD and PM10 measurements of similar magnitude. Insights of this study potentially influence future efforts to estimate near-ground PM concentrations based on satellite AOD.

Keywords: Aerosol Optical Depth; PM10; MAIAC; MODIS; air quality; boundary layer height; ambient relative humidity

1. Introduction

The adverse health effects of Particulate Matter (PM) on the human respiratory and cardiovascular system are well known and include asthma, emphysema, and lung cancer [1–3]. Although measures to reduce ambient PM concentrations are mandatory in many countries once legal thresholds are exceeded, the most effective approach to reduce PM concentrations is still a matter of debate. Low emission zones (LEZ) pertaining to automotive traffic or other driving restrictions have been set up in cities worldwide, for example in Berlin [4,5], Munich [6], London [7], Tokyo or Mexico City [8]. One fact that makes it hard for policy makers to choose the right measures is that the processes leading to excessive particulate concentrations in urban regions are not yet fully understood. For example, vegetation can both enhance [9] and decrease PM concentrations [4], depending on the type of vegetation, the primary pollutant, and the size of green spaces.

To better constrain the processes driving PM distribution, extensive and spatially continuous data are required; however, this information is not readily available. In general, networks of PM monitoring

stations are not dense enough to reflect interactions between urban areas and their surroundings. Moreover, measurement stations are much more common in urban areas than in rural areas. Satellite data have the potential to fill in the blanks.

Satellite-based AOD observations have been used as a tool for air quality assessment with varying success [10–12]. However, relying on satellite AOD observations as proxy for near-ground air pollution can be misleading in some situations. AOD is defined as the extinction of radiation in an atmospheric column at a certain wavelength, while particulate matter concentrations reflect a dry mass concentration of particles of a certain size distribution at a specific measurement site [10]. The relationship between these two entities is influenced by numerous factors and is therefore highly variable and often non-linear [10,13–15]. On a seasonal scale, the relationship can even be inverse, because AOD tends to peak in summer, while PM usually peaks during the winter months [13,16,17]. The main reason therefore is the varying vertical distribution of aerosols within the atmosphere: in situations with a high boundary layer, the number of particles near ground decreases as they dissipate in the atmosphere, whereas a low BLH tends to increase near-ground particle numbers as the particles are confined in the lower atmosphere. BLH tends to be higher in summer due to pronounced turbulence [12,17,18]. Information about the vertical distribution of aerosols, approximated for example by the BLH, is therefore necessary for understanding variations in the relationship between AOD and PM [19–21]. Furthermore, moisture in the atmosphere, in this study approximated by ambient RH measurements, exerts influence on the AOD. Increasing RH will usually lead to hygroscopic growth of particles (aerosol swelling), enhancing the scattering coefficient and thus increasing the AOD [10,12,20]. The magnitude of hygroscopic growth depends on size and chemical species of particles [22–24]. In situ PM measurements, on the other hand, are usually carried out under dry conditions [25] and are not directly affected by ambient relative humidity. Additional meteorological factors influencing the relationship between AOD and PM include wind speed and wind direction. Generally, it is known that ambient weather conditions can exert strong influence on this relationship, superimposing small-scale effects of PM measurement sites [12,26]. Wind speed and direction govern the amount and type of particles in the atmosphere [27–30].

With regard to determinants of the relationship between AOD and PM, previous studies have identified favorable meteorological boundary conditions to derive near-ground PM from satellite AOD, often with varying conclusions. Although frameworks differ, some of these are presented here briefly. Gupta et al. (2006) [20] report a strong linear correlation between AOD and particulate matter with a diameter of 2.5 micrometer or less (PM2.5) for situations with low BLH. They conclude that when a low BLH is present, the satellite observes a very similar amount of particles as detected by ground-based measurements. In another study, Chudnovsky et al. (2013) [28] argue that in winter conditions in the Boston area, U.S.A., a shallow BLH would rather lead to low correlations between AOD and PM2.5. In these situations, particle concentrations are low in the upper parts of the atmosphere as particles are mostly confined within the shallow boundary layer. This might limit the path length for satellite measurements and decrease the sensitivity of AOD [31]. The role of RH has been controversially discussed in recent studies. While some do not find a significant influence of RH on the correlation of AOD and PM2.5 [28], others report that AOD is more representative of PM2.5 mass concentration under dry conditions, i.e., RH < 50% [12]. Results of a field campaign conducted in northeastern U.S. suggest that water uptake by particles can indeed play a major role in the relationship between AOD and PM2.5 [32].

While AOD has sometimes been used as a proxy for PM, a thorough and detailed treatment of the specific roles of meteorological conditions in determining the link between both parameters is not available. Accordingly, the aim of this study is to identify a range of meteorological conditions under which AOD can be reliably used to approximate PM, and to constrain under which conditions satellite AOD observations underestimate or overestimate ground PM concentrations. Knowledge of these conditions is a prerequisite for generating reliable spatial estimates of PM. Based on the literature reviewed above, the guiding hypothesis is that for a continental mid-latitude location BLH, RH and wind are of the greatest importance for inferring PM10 from satellite-based AOD.

The data basis of this study is comprised of almost 15 years of data including satellite observations from the Moderate Resolution Imaging Spectroradiometer (MODIS), model output from the European Centre for Medium-Range Weather Forecasts (ECMWF) and station data from the German Meteorological Service (Deutscher Wetterdienst, DWD). Using the recent, high-resolution Multi-Angle Implementation of Atmospheric Correction (MAIAC) AOD algorithm and ground-based PM10 measurements, numerous collocated data pairs were analyzed in an area around Berlin, Germany. Berlin was chosen as a study site because of its continental location in flat terrain and excellent data availability. A novel approach is introduced, which includes a monthly ranking of AOD and PM10 observations and allows for a direct comparison of the magnitude of paired values. In this way, it was possible to characterize the relationship between AOD and PM10 for each data pair individually and to perform in-depth analyses. Ranked AOD/PM10 data pairs were linked to surrounding meteorological parameters. The approach presented in this study can be generalized for the use in other study areas.

2. Materials and Methods

2.1. Study Domain

The study focuses on a rectangular area ($\sim$12,000 km^2) around the cities of Berlin and Potsdam in northeastern Germany. A relatively small area was chosen to properly quantify and interpret influencing factors of the relationship between AOD and PM10. The area is dominated by the cities in its center and is surrounded by agricultural land and large vegetated areas to the north and south (see Figure 1). A great number of PM10 measurements in the city area and its surroundings are available. The area is mainly characterized by continental climate with only sparse, episodic maritime influence when advection of air masses from northern and northwestern directions occurs. Thus, the area is suitable for studying the influence of wind direction to discriminate between these regimes.

Local PM10 sources dominate in the region, with traffic the major contributor [4]. Vegetated areas surrounding Berlin influence the mixture of particles in and around the urban area. The emission of Biogenic Volatile Organic Compounds (BVOCs, e.g., isoprene and monoterpenes) leads to secondary aerosol formation. Depending on the vegetation type, this effect can vary significantly; agricultural areas, for example, can have pollution levels similar to the urban background. It has been argued that particle transport from the east can markedly contribute to overall pollution levels [4,33]. In general, two different air mass regimes can be discriminated. Western air masses tend to transport lower particle concentrations, whereas continental, eastern air masses accumulate particles and thus carry elevated particle concentrations [34].
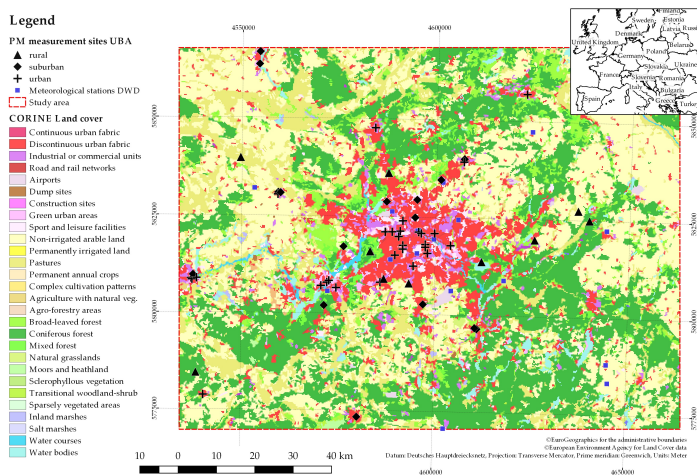
164

Figure 1. Geographical extent of the study area. Different land cover classes are colored, positions of meteorological stations and PM10-measurement sites are indicated.

2.2. Materials

2.2.1. PM-Measurements

PM10 is analyzed in this study instead of PM2.5 due to a greater availability of data for the study region. PM10 measurements might be more adequate for the analysis, as larger particles are less easily transported and more representative of local conditions. Additionally, PM2.5 excludes the fraction of larger particles, which are nevertheless accountable for light extinction and are thus measured by the satellite. The stations are run by the German Federal Environmental Agency (UBA) and consist of hourly mean concentrations [μg /m^3]. PM concentrations are determined by measuring the attenuation of β-radiation by a dust-coated filter. Measurements are largely uninfluenced by ambient temperature and relative humidity. To avoid condensation, the particle inlet is heated permanently [25].

The locations of PM measurement sites and station types are shown in Figure 1. Instructions by the European Union prescribe measurement sites to be sited in such a way that micro-environment effects can be avoided. Stations can be of type background or traffic and be representative for urban, rural or suburban areas. Traffic sites are generally close to main roads or busy intersections [35]. These sites are excluded from the analysis as they are by design not representative for conditions beyond the specific traffic situation and thus the area covered by a satellite pixel. PM concentrations within the city area reach their maximum in the morning hours ($\sim$ 7–8 a.m.) and in the evening ($\sim$ 9–11 p.m.) [36]. Urban background sites capture the integrated contribution of all sources near the site without one particular source dominating. These stations are representative for some km^2. Suburban stations need to be placed downwind (referring to the main wind direction) of emission sources and ensure a representativeness of some dozends km^2. Rural background stations may not be influenced by agglomerations or industrial sites closer than five kilometers. These stations may be located in small settlements, given a representativeness of some hundred km^2 [35].

165

2.2.2. Satellite AOD Data

High-resolution (1 × 1 km) AOD computed by application of the MAIAC technique is used [37–40]. The product is based on data from MODIS, Collection 6, aboard the Terra and Aqua satellites. The MAIAC algorithm makes use of look-up tables, explicitly taking into account surface bidirectional reflectance factors (BRF). The calculation of AOD relies on the assumption that surface BRFs remain largely constant over time, considering a time series of 16 consecutive days. This assumption might not be true in case of rapid surface changes, e.g., green-up events over bright surfaces [37,38]. To avoid these pixels, uncertainty estimates as included in the product are applied for filtering here. Pixels with an uncertainty larger 0.4 were discarded. For valid daytime AOD data used in this study, Aqua and Terra satellite overpass times range from ∼9 a.m. UTC to ∼1 p.m. UTC.

2.2.3. Meteorological Data

To represent regional-scale meteorological conditions, ERA-Interim reanalysis data generated by the ECMWF are used with a resolution of 0.125° [41]. These include BLH [m] and wind speed [m/s] in u and v direction (10 m height). Wind speed and wind direction [°] are calculated using u and v wind components. BLH data are employed as a proxy for the vertical concentration of aerosols in the lower troposphere, assuming particles are well-mixed within the boundary layer [12,42] and AOD mostly represents attenuation in the boundary layer [43]. Depending on the closest satellite overpass time, ERA-Interim data at 9 a.m. UTC, 12 noon UTC or 3 p.m. UTC are used.

Ground measurements were obtained from the DWD. One hour mean values of RH [%] measured in 2 m height are used [44]. The locations of DWD measurement sites are given in Figure 1 as blue dots. This study uses near-ground RH measurements, as these are more widely available than measurements of the vertical distribution of water vapour. Under dry conditions (RH < 55%), near-ground RH measurements show reasonable agreement with vertically resolved measurements. This might not be the case at higher RH [32].

2.3. Methods

UBA PM station coordinates are used as spatial reference, i.e., AOD pixels, ERA-Interim pixels and meteorological station data coordinates are collocated with the position of these stations. Only the nearest pixels below a threshold of 0.015° (∼1 km) distance from the PM measurement were collocated with the corresponding closest UBA station. The relatively small distance threshold was chosen to ensure that the MAIAC AOD reflects the situation in close vicinity of the UBA station. Terra/Aqua satellite overpass time is used as temporal reference for the PM10 and meteorological data, i.e., values closest to the overpass time are used. The work flow of this study is summarized in Figure 2.

For the analysis of factors influencing the relationship between AOD and PM10, this study introduces the Air Quality Rank Difference Index (AiRDI). The AiRDI allows for a direct, semi-quantitative assessment of whether AOD and PM10 agree, or whether the satellite qualitatively over- or underestimates PM10. Accordingly, situations in which these parameters are of similar magnitude can be identified systematically. The AiRDI is dimensionless and is computed for each spatially and temporally collocated AOD-PM10 data pair. First, AOD and PM10 values are sorted from lowest to highest on a monthly basis. The sorting is performed for each month individually to eliminate seasonal effects as outlined in the introduction. The sorted AOD (PM10) values are then assigned ranks corresponding to their positions on a scale from 0 to the number of AOD (PM10) observations per month (*n*), yielding AOD_{ranked} and $PM10_{ranked}$. The rank of AOD (PM10) within each month is then divided by n and multiplied by 100, yielding normalized ranks greater than 0 and up to 100. Finally, the AiRDI is computed by subtracting $PM10_{rank}$ from AOD_{rank} (see Equation (3)).

$$AOD_{rank}(month) = \frac{AOD_{ranked}}{n(month)} * 100 \qquad (1)$$

$$PM10_{rank}(month) = \frac{PM10_{ranked}}{n(month)} * 100 \qquad (2)$$

$$AiRDI(month) = AOD_{rank}(month) - PM10_{rank}(month) \qquad (3)$$

Positive AiRDI means that AOD_{rank} is higher than $PM10_{rank}$. In other words, the satellite observation is relatively higher than the PM10 concentration, indicating a relative overestimation of PM10 by the MAIAC AOD. The opposite is true for a negative AiRDI; in this case, lower AOD_{rank} coincide with higher $PM10_{rank}$, indicating an underestimation of PM10 by the MAIAC AOD. In cases where the AiRDI is close to zero, AOD and PM10 are in the same range.
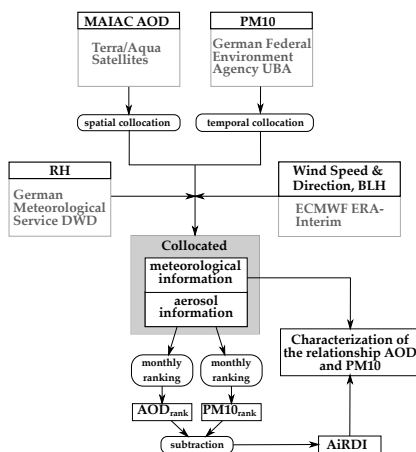
Figure 2. Overview of the work flow of this study. Methods are shown in curved boxes, inputs and outputs in rectangular boxes.

3. Results

3.1. Nonlinearity of the relationship between AOD and PM10

Several studies report a robust linear correlation between daily means of AOD and PM, e.g., for the southern U.S. [10], the Po Valley in Italy [16] or selected cities worldwide [20]. In the Berlin study area, the relationship between AOD and PM10 is highly nonlinear. This is suggested by patterns shown in Figures 3a,b, which depict AiRDI in relation to AOD and PM10. A strong positive linearity would result in a slope of zero, as increasing AOD would coincide with increasing PM10, resulting in a constant AiRDI. In fact, the figures show an almost linear increase (decrease) of AiRDI for increasing AOD_{rank} ($PM10_{rank}$). Higher (lower) AOD is likely to overestimate (underestimate) PM10 concentrations; AiRDI is more likely to be positive for high AOD, but more likely to be negative for low AOD. The pattern is inverse for the comparison of PM10 and AiRDI.
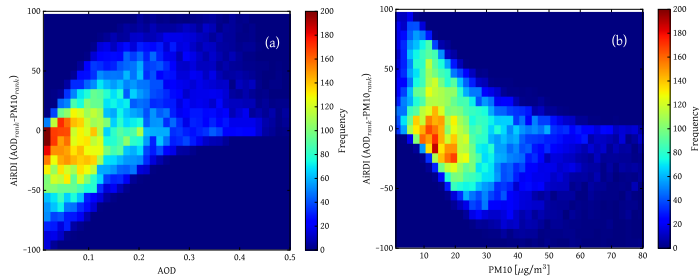
Figure 3. 2D-histograms of AiRDI, AOD$_{rank}$ (**a**) and AiRDI, PM10$_{rank}$ (**b**). The frequency of occurrence of associated data pairs is color-coded.

3.2. Meteorological Conditions of AOD-PM10 Agreement and Divergence

In this section, meteorological factors governing the relationship between AOD and PM10 are identified and connected to the AiRDI. An in-depth analysis concerning the role of RH and BLH is conducted. Wind speed and direction is investigated to determine the role of weather conditions and the prevailing inflow direction. AiRDI patterns related to wind are partly driven by the redistribution of RH and BLH.

To identify possible factors influencing AiRDI, data was split in positive (AiRDI > 25), negative values (AiRDI < −25) and AiRDI values close to zero (25 ≥ AiRDI ≥ −25). Subsequently, the median of each parameter for each season was calculated. Then, the relative deviation of each data point from this median for each AiRDI-class was determined, resulting in the distribution shown in Figure 4.

The class AiRDI > 25 is characterized by strong positive deviations of BLH and slightly positive RH deviations. This is consistent with what was introduced above: a higher BLH tends to decrease PM10 concentrations near ground because particles disperse within the atmospheric column, while elevated ambient RH tends to initiate hygroscopic growth, increasing AOD [10,12]. Both effects combined result in a positive AiRDI. Additionally, a high BLH might increment the number of particles distributed in the lower troposphere due to enhanced turbulence. Subsequently, AOD would also increase [13]. BLH does not show a marked deviation in summer, pointing to a decreased importance of BLH during the summer months.

Patterns of the class AiRDI < −25 display almost inverse characteristics. Here, BLH and RH show negative deviations. Negative AiRDI values result from low BLHs, which increase PM10 concentrations near ground, whereas low ambient RHs decrease the influence of hygroscopic growth. PM10 concentrations appear to have very high positive deviations in winter months, when particles are often confined to layers close to the ground [13]. The frequent occurrence of stable stratification situations along with low turbulent atmospheric exchange during stagnant synoptic situations could lead to very high particle concentrations [26].

Parameters in the class 25 ≥ AiRDI ≥ −25 do not show strong deviations of any parameter. The best agreement between AOD and PM10 is achieved when RH and BLH are near their seasonal median values.
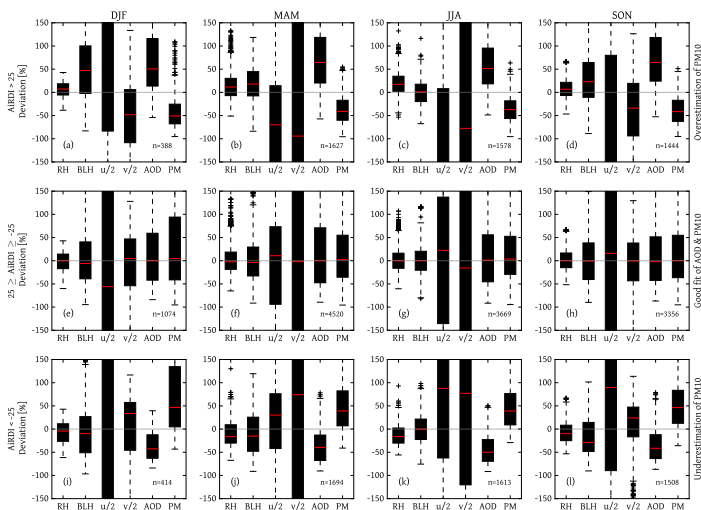
Figure 4. Determination of factors influencing AiRDI including relative humidity (RH), boundary layer height (BLH), wind velocity in u and v direction (u/2,v/2), aerosol optical depth (AOD) and PM10 (PM) . Data are split on a seasonal basis and divided in positive (**a–d**), negative values (**e–h**) and AiRDI values close to zero (**i–l**). Winter months are denoted as DJF, spring months as MAM, summer months as JJA, fall months as SON. Box plots show the relative deviation of each parameter observation to the seasonal mean. Deviations for u and v were multiplied by a factor of 0.5 to prevent a distortion of the y-axes. The box plots extend from the lower to the upper quartile value, with a red line at the median. Whiskers show the range with outliers marked as short, horizontal black lines. The number of observations for each class is denoted as n.

In summary, AOD observations coinciding with high (low) ambient RH and high (low) BLH have a higher probability to cause an overestimation (underestimation) of PM10. Patterns are largely constant throughout the year. Wind speed and direction also play an important role. But like RH and BLH, the relationship is complex and is treated in more detail in the following sections.

3.3. The Role of RH and BLH

This section quantifies the interconnected effects of RH and BLH on the relationship between AOD and PM10. To this end, a multivariate analysis was conducted using ranges of RH and BLH. These were related to AOD_{rank}, $PM10_{rank}$ and AiRDI.

Three AiRDI-regimes can be discriminated in Figure 5a. Low RH coinciding with low BLH leads to negative AiRDI values, relatively high RH coinciding with high BLH leads to positive AiRDI values. A curved area of near-zero AiRDI values separates negative and positive AiRDI patterns, indicating meteorological conditions where the magnitudes of AOD and PM10 match closely. The shape of this patch emphasizes the dependency of the AiRDI on both RH and BLH. With decreasing BLH, the near-zero values are shifted towards higher RH. Obviously, the effects on AiRDI of both parameters can compensate each other, leading to a AiRDI close to zero.

Figure 5b highlights the effect of hygroscopic growth on AOD, with increased RH leading to higher AOD_{rank}. Furthermore, the data suggest that AOD_{rank} is positively correlated not only to RH, but also to BLH (Figure 5b). Relatively high AOD_{rank} coincides with relatively high BLH. This effect might be due to convection processes, which increase the BLH and cause particles to be lifted up into the atmosphere more easily. In these situations, the deposition of particles would be prevented and particles would remain in the atmosphere, causing the AOD to increase. Low BLH, as an indicator of low turbulence, would lead to reduced particle release and thus relatively low AOD. Additionally, when particles are confined in a shallow boundary layer, a reduced path length for satellite measurements might reduce the sensitivity to AOD [31].

$PM10_{rank}$ is elevated for BLH < 600 m (Figure 5c) owing to limited dissipation of particles into higher atmospheric layers and subsequent accumulation of particles near the ground. At very high BLH (>2400 m), $PM10_{rank}$ is also slightly elevated. Around 3% of all observations are related to this pattern. $PM10_{rank}$ at these high BLH values should be interpreted with care: For mid-latitudes, BLH > 2400 m appears to be very high [45] and might not properly reflect real conditions. A weak dependence of $PM10_{rank}$ on RH exists: $PM10_{rank}$ decreases with increasing RH. This might be due to the fact that RH is elevated when precipitation influenced the atmosphere prior to the measurement, leading to increased RH and decreased particle concentrations due to wet deposition of particles [26]. Another possible reason might be that the influence of maritime air is overrepresented in situations with high RH. These air masses are humid and clean, i.e., carry fewer particles. Finally, this pattern could be due to measurement errors of the UBA PM sites caused by condensation on the particle inlet of the PM sensors, despite the heating of the sensor inlet. This way, fewer particles would reach the sensor, decreasing reported particle concentrations.

When BLH is greater than 600 m, the AiRDI is mostly driven by RH. In a very moist atmosphere (RH > 90%), the median AiRDI is near zero only at very low BLH. The influence of hygroscopic growth on AOD is apparent when AiRDI approaches positive values. In these situations, AOD is increased while PM10 remains unaffected. This effect can be seen from about RH > 40%, depending on BLH. Previous studies report that hygroscopic growth starts at RH ~ 50% [46].

Results presented in this section partly contradict earlier studies, which identify situations with low BLH and low RH as favorable [20]. Figure 5a suggests that a very low BLH tends to cause an underestimation of PM10 concentrations by the satellite (negative AiRDI). RH and BLH clearly have an interconnected influence on the relationship between AOD and PM10, which emphasizes the need to examine both parameters simultaneously. Generally, AOD and PM10 match best when the atmosphere is relatively dry (40% < RH ≤ 70%) and BLH is at medium levels (600–1200 m). High ambient RH, on the other hand, can be compensated by a low BLH, leading to a good agreement of AOD and PM10. RH measurements used in this study only capture near-ground ambient RH. Nevertheless, Figure 5a suggests that near-ground RH has a large influence on the relationship between AOD and PM10.
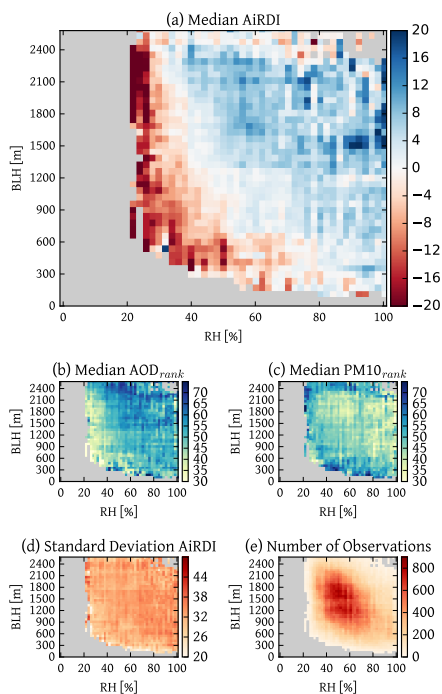
Figure 5. (**a**) Relationship between AiRDI (AOD$_{rank}$ − PM10$_{rank}$), RH and BLH. BLH is gridded in steps of 60 m, RH in steps of 2%, each grid cell represents the median AiRDI. (**b**) Relationship between AOD$_{rank}$, RH and BLH, same grouping of RH and BLH as (**a**). (**c**) Relationship between PM10$_{rank}$, RH and BLH, same grouping of RH and BLH as (**a**). (**d**) Standard deviation of AiRDI for each class, same grouping of RH and BLH as (**a**). (**e**) Number of observations for each grid cell, same grouping of RH and BLH as (**a**).

3.4. The Role of Wind Direction and Wind Speed

In this section, patterns of AiRDI, AOD$_{rank}$ and PM10$_{rank}$ in relation to large scale patterns of wind speed and wind direction are analyzed and connected to results of previous chapters. Patterns of wind speed and direction give hints on particle origins. Knowledge on the direction of origin of air masses, for example, helps characterize particle species and concentrations [23,24]. Wind speeds help to distinguish between the dominance of local sources or particles originating further away.

To distinguish between the effects of wind speed and direction, wind speed data are split in four classes 0–3 m/s, >3–5 m/s, >5–7 m/s and >7 m/s. The resulting polar plots are comprised of 16 wind direction classes with a range of 22.5°, starting from 0°. Each class is colored with the corresponding median AiRDI value, computed of all associated AOD$_{rank}$ and PM10$_{rank}$ that appear in the corresponding class. Radii of the polar bars indicate the frequency of occurrence.

A clear AiRDI pattern can be identified in Figure 6. Air masses from northwestern or western direction tend to cause a positive AiRDI, while air masses from the east and south-east tend to coincide with negative AiRDI values. This pattern is most distinctive for wind speed classes >3–5 m/s and >5–7 m/s. For lower wind speeds (0–3 m/s), the pattern is less clearly defined. This is likely due to the dominance of local particle sources, superimposing the effects of large-scale air mass transports. Results suggest that satellite AOD tends to overestimate particle concentrations near the ground when strong northwestern or western winds prevail, whereas an underestimation occurs when eastern and southeastern winds dominate. There is a higher chance of AiRDI close to zero at lower wind speeds, as local emissions can be captured more easily by both ground stations and the satellite.

Figure 6. Relationship between the AiRDI ($AOD_{rank} - PM10_{rank}$), wind direction and wind speed for the study area. Figures (**a–d**) represent a wind speed class as shown on the upper left of each panel. Polar bars enclose a wind direction class, each 22.5° in range. Colors indicate the mean AiRDI for each polar bar, the radius of each bar indicates the frequency of occurrence.

$PM10_{rank}$ (Figure 7a–d) is generally highest during conditions with low wind speeds (0–3 m/s). This is not surprising, as transport of pollutants takes place less effectively and particles from local sources inside the urban area of Berlin accumulate in the atmosphere. Additionally, dry deposition of particles might take place more effectively, as particles reach the surface more easily due to decreased vertical winds. This might lead to increased PM10 concentrations at the measurement sites. Winds from eastern directions show elevated $PM10_{rank}$, indicating advection of continental air masses with high particle numbers near the ground. Air masses from the west, possibly of maritime origin, tend to be cleaner than continental air, because particles are scavenged by wet deposition [23,26]. At higher wind speeds, this pattern is most distinct.

Generally, patterns for AOD_{rank} are less pronounced (Figure 8a–d). In contrast to $PM10_{rank}$, AOD_{rank} is slightly elevated when high wind speeds occur from the west . The inverse pattern of AOD_{rank} versus $PM10_{rank}$ in terms of wind speed and wind direction demonstrates that satellite derived AOD is not suited for directly deriving ground-based PM10 in all situations.
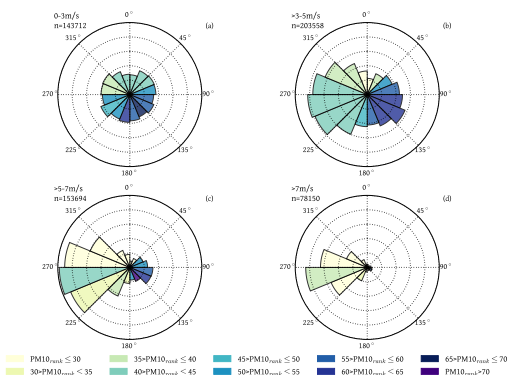
Figure 7. Relationship between PM10$_{rank}$, wind direction and wind speed. Setup of figures (**a–d**) is identical to previous plot. PM10$_{rank}$ is generally higher when continental, i.e., easterly air masses dominate.
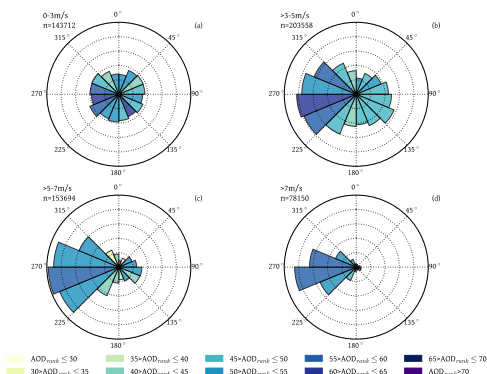

Figure 8. Relationship between AOD$_{rank}$, wind direction and wind speed. Setup of figures (**a–d**) is identical to previous plot.

Several explanations for these observations are possible. Advection of western air masses transports higher amounts of moisture to the Berlin study area, indicated by higher values of RH in Figure 9. Increased availability of moisture in the atmosphere could promote particle water uptake and cause hygroscopic swelling of particles. Increased particle sizes would lead to rising AOD$_{rank}$ but not affect PM10$_{rank}$, which ultimately leads to positive AiRDI. Eastern winds, on the other hand, transport continental air masses with a potentially larger particle loading [4,33]. This might increase PM10 concentrations.
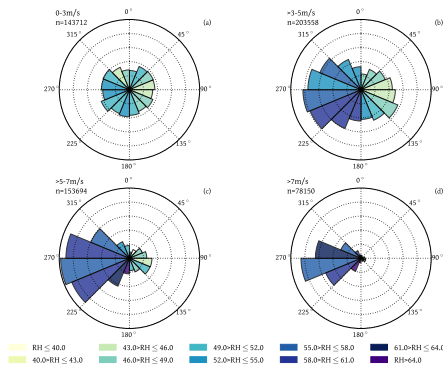
Figure 9. Relationship between RH [%], wind direction and speed. Setup of figures (**a**–**d**) is identical to previous plot.

A further possible explanation might be the varying hygroscopicity of particles carried by different air masses. Western air masses are thought to carry a higher amount of hydrophilic particles (e.g., sea salts) which tend to be more prone to hygroscopic growth due to more effective water uptake [23,24,47]. The opposite might be true for air masses from eastern and northeastern origins carrying anthropogenic and industrial aerosols. For the measurement station Melpitz, located around 130 km south of Berlin, it was found that the scattering enhancement of particles was lowest when air masses were advected from the east and northeast [23,34]. Thus, these particles are less likely to increase the AOD by hygroscopic growth [23,24]. Finally, eastern wind directions might indicate a stable atmospheric situation with lower BLH (see Figure 10), causing pollution levels to increase near the ground [12,18]. As can be seen in Figure 10, advection of air masses from the west at higher wind speeds tends to increase BLH.
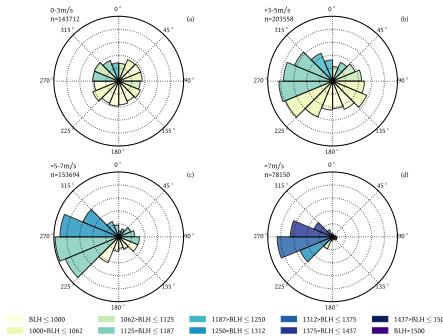
Figure 10. Relationship between BLH [m], wind direction and speed. Setup of figures (**a**–**d**) is identical to previous plot.

In summary, advection of western air masses tends to increase RH and BLH, coinciding with slightly elevated AOD_{rank} and low $PM10_{rank}$. Eastern and southeastern air masses tend to show lower RH and BLH, coinciding with slightly decreased AOD_{rank} and elevated $PM10_{rank}$. These patterns are generally more pronounced for higher wind speeds, whereas dependence on wind direction decreases with lower wind speeds. To conclude, the following statements summarize the findings of this chapter:

- Western air masses with higher wind speeds establish a relatively high probability that the satellite AOD overestimates PM10.
- The opposite is true for eastern direction of origin. In these cases, a higher probability of the satellite AOD underestimating PM10 levels exists.
- Differences in RH and BLH patterns seem to be the main drivers behind the dependence of AiRDI on wind direction besides the transport of pollutants and varying hygroscopicity of particles.
- Agreement between AOD and PM10 is more likely to occur when wind speeds are low.

4. Conclusions

Information on RH, BLH, wind direction and speed proved to be of importance, driving the relationship between AOD and PM10. Generally, AOD serves as a good proxy for PM10 concentrations near ground when a very dry atmosphere (30% < RH ≤ 50%) coincides with a high BLH (>1200 m), when moderate ambient RH (50–80%) coincides with moderate BLH (600–1200 m) or when humid conditions (RH > 80%) coincide with a shallow BLH (<600 m). In case of a dry atmosphere, AOD is not influenced by hygroscopic growth and thus relatively low. This can be compensated by low PM10 levels, which appear more frequently in situations with a high BLH. When ambient RH is high, satellite-derived AOD tends to overestimate street-level PM10 concentrations. This effect can be compensated by a low BLH, which tends to increase PM10 levels near the ground. In combination, these effects can lead to an AiRDI close to zero, indicating a good match of satellite-retrieved AOD and ground-based PM10 measurements. When moderate RH and BLH prevail, AOD is only marginally affected by hygroscopic growth due to limited availability of moisture, and particle concentrations in the boundary layer seem to agree well with particle concentrations near the ground. These results suggest that it might be misleading to examine RH and BLH separately.

In terms of wind speed and wind direction, it was found that situations with prevailing western direction of origin have a higher probability to cause a positive AiRDI, whereas winds from the east tend to cause a negative AiRDI. The AiRDI is more likely to be near-zero in situations with low wind speeds.

The AiRDI proved to be a useful approach to a perform semi-quantitative comparison of the parameters AOD and PM10 and to better understand their interactions with surrounding meteorological conditions. Knowledge of the effect of these conditions can help future approaches to derive high-quality PM estimates based on AOD. The AiRDI approach can be generalized for application in other study areas. Still, this study did not account for some sources of uncertainties, for example effects of cloud contamination or aerosol layers above the BLH [32,48]. Future analyses could further focus on the role of different aerosol species or the role of precipitation events on the relationship between AOD and PM.

Author Contributions: Conceptualization, R.S.; Investigation, R.S.; Methodology, R.S., J.C. and H.A.; Visualization, R.S.; Writing—original draft, R.S.; Writing—review & editing, R.S., J.C. and H.A.

Conflicts of Interest: The authors declare no conflict of interest.

Further appendices

Abbreviations

The following abbreviations are used in this manuscript:

AiRDI	Air Quality Rank Difference Index
AOD	Aerosol Optical Depth
BLH	Boundary Layer Height
BRF	Bidirectional Reflectance Factor
BVOC	Biogenic Volatile Organic Compounds
DWD	German Meteorological Service
ECMWF	European Center for Medium-Range Weather Forecasts
LEZ	Low Emission Zones
MAIAC	Multi-Angle Implementation of Atmospheric Correction
MODIS	Moderate Resolution Imaging Spectroradiometer
PM	Particulate Matter
RH	Relative Humidity

References

1.　Wichmann, H.; Spix, C.; Tuch, T.; Wölke, G.; Peters, A.; Heinrich, J.; Kreyling, W.; Heyder, J. Daily mortality and fine and ultrafine particles in Erfurt, Germany part I: role of particle number and particle mass. *Res. Rep. (Health Eff. Inst.)* **2000**, *98*, 5–86.

2.　Lim, S.S.; Vos, T.; Flaxman, A.D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; Amann, M.; Anderson, H.R.; Andrews, K.G.; Aryee, M.; et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **2012**, *380*, 2224–2260. [CrossRef]

3.　Samoli, E.; Stafoggia, M.; Rodopoulou, S.; Ostro, B.; Declercq, C.; Alessandrini, E.; Díaz, J.; Karanasiou, A.; Kelessis, A.G.; Tertre, A.L.; et al. Associations between fine and coarse particles and mortality in Mediterranean cities: Results from the MED-PARTICLES project. *Environ. Health Perspect.* **2013**, *121*, 932–938. [CrossRef] [PubMed]

4.　Bonn, B.; Von Schneidemesser, E.; Andrich, D.; Quedenau, J.; Gerwig, H.; Lüdecke, A.; Kura, J.; Pietsch, A.; Ehlers, C.; Klemp, D.; et al. BAERLIN2014—The influence of land surface types on and the horizontal heterogeneity of air pollutant levels in Berlin. *Atmos. Chem. Phys.* **2016**, *16*, 7785–7811. [CrossRef]

5.　Berlin Senate. Luftreinhalteplan 2011–2017. Available online: https://www.berlin.de/ (accessed on 24 August 2018).

6.　Qadir, R.M.; Abbaszade, G.; Schnelle-Kreis, J.; Chow, J.C.; Zimmermann, R. Concentrations and source contributions of particulate organic matter before and after implementation of a low emission zone in Munich, Germany. *Environ. Pollut.* **2013**, *175*, 158–167. [CrossRef] [PubMed]

7.　Ellison, R.B.; Greaves, S.P.; Hensher, D.A. Five years of London's low emission zone: Effects on vehicle fleet composition and air quality. *Transp. Res. Part D Transp. Environ.* **2013**, *23*, 25–33. [CrossRef]

8.　Wolff, H. Keep your clunker in the suburb: Low-emission zones and adoption of green vehicles. *Econ. J.* **2014**, *124*, F481–F512. [CrossRef]

9.　Churkina, G.; Kuik, F.; Bonn, B.; Lauer, A.; Grote, R.; Tomiak, K.; Butler, T.M. Effect of VOC Emissions from Vegetation on Air Quality in Berlin during a Heatwave. *Environ. Sci. Technol.* **2017**, *51*, 6120–6130. [CrossRef] [PubMed]

10.　Wang, J.; Christopher, S. Intercomparison between satellite-derived aerosol optical thickness and PM 2.5 mass: Implications for air quality studies. *Geophys. Res. Lett.* **2003**, *30*, 2095. [CrossRef]

11.　Engel-Cox, J.A.; Holloman, C.H.; Coutant, B.W.; Hoff, R.M. Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality. *Atmos. Environ.* **2004**, *38*, 2495–2509. [CrossRef]

12.　Gupta, P.; Christopher, S.A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *J. Geophys. Res.* **2009**, *114*, D14205. [CrossRef]

13. Barnaba, F.; Putaud, J.P.; Gruening, C.; Dell'Acqua, A.; Dos Santos, S. Annual cycle in co-located in situ, total-column, and height-resolved aerosol observations in the Po Valley (Italy): Implications for ground-level particulate matter mass concentration estimation from remote sensing. *J. Geophys. Res. Atmos.* **2010**, *115*, 1–22. [CrossRef]

14. Kloog, I.; Koutrakis, P.; Coull, B.A.; Lee, H.J.; Schwartz, J. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ.* **2011**, *45*, 6267–6275. [CrossRef]

15. Chudnovsky, A.; Tang, C.; Lyapustin, A.; Wang, Y.; Schwartz, J.; Koutrakis, P. A critical assessment of high-resolution aerosol optical depth retrievals for fine particulate matter predictions. *Atmos. Chem. Phys.* **2013**, *13*, 10907–10917. [CrossRef]

16. Arvani, B.; Pierce, R.B.; Lyapustin, A.I.; Wang, Y.; Ghermandi, G.; Teggi, S. Seasonal monitoring and estimation of regional aerosol distribution over Po valley, northern Italy, using a high-resolution MAIAC product. *Atmos. Environ.* **2016**, *141*, 106–121. [CrossRef]

17. Koelemeijer, R.B.; Homan, C.D.; Matthijsen, J. Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmos. Environ.* **2006**, *40*, 5304–5315. [CrossRef]

18. Geiß, A.; Wiegner, M.; Bonn, B.; Schäfer, K.; Forkel, R.; Von Schneidemesser, E.; Münkel, C.; Lok Chan, K.; Nothard, R. Mixing layer height as an indicator for urban air quality? *Atmos. Meas. Tech.* **2017**, *10*, 2969–2988. [CrossRef]

19. Van Donkelaar, A.; Martin, R.V.; Park, R.J. Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing. *J. Geophys. Res. Atmos.* **2006**, *111*, 1–10. [CrossRef]

20. Gupta, P.; Christopher, S.; Wang, J.; Gehrig, R.; Lee, Y.; Kumar, N. Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmos. Environ.* **2006**, *40*, 5880–5892. [CrossRef]

21. Boyouk, N.; Léon, J.F.; Delbarre, H.; Podvin, T.; Deroo, C. Impact of the mixing boundary layer on the relationship between PM2.5 and aerosol optical thickness. *Atmos. Environ.* **2010**, *44*, 271–277. [CrossRef]

22. Wang, J.; Martin, S.T. Satellite characterization of urban aerosols: Importance of including hygroscopicity and mixing state in the retrieval algorithms. *J. Geophys. Res. Atmos.* **2007**, *112*, 1–18. [CrossRef]

23. Zieger, P.; Fierz-Schmidhauser, R.; Weingartner, E.; Baltensperger, U. Effects of relative humidity on aerosol light scattering: Results from different European sites. *Atmos. Chem. Phys.* **2013**, *13*, 10609–10631. [CrossRef]

24. Titos, G.; Jefferson, A.; Sheridan, P.J.; Andrews, E.; Lyamani, H.; Alados-Arboledas, L.; Ogren, J.A. Aerosol light-scattering enhancement due to water uptake during the TCAP campaign. *Atmos. Chem. Phys.* **2014**, *14*, 7031–7043. [CrossRef]

25. Umweltbundesamt. *Qualitätssicherungshandbuch des UBA-Messnetzes*; Technical Report; Federal Environment Agency: Berlin, Germany, 2004.

26. Rost, J.; Holst, T.; Sahn, E.; Klingner, M.; Anke, K.; Ahrens, D.; Mayer, H. Variability of PM10 concentrations dependent on meteorological conditions. *Int. J. Environ. Pollut.* **2009**, *36*, 3–18. [CrossRef]

27. Levy, J.I.; Hanna, S.R. Spatial and temporal variability in urban fine particulate matter concentrations. *Environ. Pollut.* **2011**, *159*, 2009–2015. [CrossRef] [PubMed]

28. Chudnovsky, A.; Lyapustin, A.; Wang, Y.; Tang, C.; Schwartz, J.; Koutrakis, P. High resolution aerosol data from MODIS satellite for urban air quality studies. *Cent. Eur. J. Geosci.* **2013**, *6*, 17–26. [CrossRef]

29. Stafoggia, M.; Schwartz, J.; Badaloni, C.; Bellander, T.; Alessandrini, E.; Cattani, G.; de' Donato, F.; Gaeta, A.; Leone, G.; Lyapustin, A.; et al. Estimation of daily PM10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environ. Int.* **2017**, *99*, 234–244. [CrossRef] [PubMed]

30. Zheng, C.; Zhao, C.; Zhu, Y.; Wang, Y.; Shi, X.; Wu, X.; Chen, T.; Wu, F.; Qiu, Y. Analysis of Influential Factors for the Relationship between PM2.5 and AOD in Beijing. *Atmos. Chem. Phys. Discuss.* **2017**, 1–57. [CrossRef]

31. Gupta, P.; Christopher, S.A. Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *J. Geophys. Res.* **2009**, *114*, D20205. [CrossRef]

32. Crumeyrolle, S.; Chen, G.; Ziemba, L.; Beyersdorf, A.; Thornhill, L.; Winstead, E.; Moore, R.H.; Shook, M.A.; Hudgins, C.; Anderson, B.E. Factors that influence surface PM2.5values inferred from satellite observations: Perspective gained for the US Baltimore-Washington metropolitan area during DISCOVER-AQ. *Atmos. Chem. Phys.* **2014**, *14*, 2139–2153. [CrossRef]

33. Kerschbaumer, A. On the Aerosol Budget over Berlin. Ph.D. Thesis, Free University, Berlin, Germany, 2007.

34. Zieger, P.; Fierz-Schmidhauser, R.; Poulain, L.; Müller, T.; Birmili, W.; Spindler, G.; Wiedensohler, A.; Baltensperger, U.; Weingartner, E. Influence of water uptake on the aerosol particle light scattering coefficients of the Central European aerosol. *Tellus B Chem. Phys. Meteorol.* **2014**, *66*. [CrossRef]

35. EU. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. *Off. J. Eur. Communities* **2008**, *152*, 1–43. Available online: https://www.eea.europa.eu/policy-documents/directive-2008-50-ec-of (accessed on 24 August 2018).

36. Von Schneidemesser, E.; Bonn, B.; Butler, T.M.; Ehlers, C.; Gerwig, H.; Hakola, H.; Hellén, H.; Kerschbaumer, A.; Klemp, D.; Kofahl, C.; et al. BAERLIN2014-stationary measurements and source apportionment at an urban background station in Berlin, Germany. *Atmos. Chem. Phys.* **2018**, *18*, 8621–8645. [CrossRef]

37. Lyapustin, A.; Martonchik, J.; Wang, Y.; Laszlo, I.; Korkin, S. Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *J. Geophys. Res. Atmos.* **2011**, *116*. [CrossRef]

38. Lyapustin, A.; Wang, Y.; Laszlo, I.; Kahn, R.; Korkin, S.; Remer, L.; Levy, R.; Reid, J.S. Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *J. Geophys. Res. Atmos.* **2011**, *116*, 1–15. [CrossRef]

39. Lyapustin, A.I.; Wang, Y.; Laszlo, I.; Hilker, T.; G.Hall, F.; Sellers, P.J.; Tucker, C.J.; Korkin, S.V. Multi-angle implementation of atmospheric correction for MODIS (MAIAC): 3. Atmospheric correction. *Remote Sens. Environ.* **2012**, *127*, 385–393. [CrossRef]

40. Lyapustin, A.; Wang, Y.; Korkin, S.; Huang, D. MODIS Collection 6 MAIAC Algorithm. *Atmos. Meas. Tech. Discuss.* **2018**, 1–50. [CrossRef]

41. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [CrossRef]

42. Ansmann, A.; Althausen, D.; Wandinger, U.; Franke, K.; Müller, D.; Wagner, F.; Heintzenberg, J. Vertical profiling of the Indian aerosol plume with six-wavelength lidar during INDOEX: A first case study. *Geophys. Res. Lett.* **2000**, *27*, 963–966. [CrossRef]

43. Schäfer, K.; Harbusch, A.; Emeis, S.; Koepke, P.; Wiegner, M. Correlation of aerosol mass near the ground with aerosol optical depth during two seasons in Munich. *Atmos. Environ.* **2008**, *42*, 4036–4046. [CrossRef]

44. DWD Climate Data Center (CDC). Historical Hourly Station Observations of 2m Air Temperature and Humidity. 2016. Available online: ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/hourly/(accessed on 24 August 2018)

45. Von Engeln, A.; Teixeira, J. A planetary boundary layer height climatology derived from ECMWF reanalysis data. *J. Clim.* **2013**, *26*, 6575–6590. [CrossRef]

46. Lee, J.; Kim, Y. Spectroscopic measurement of horizontal atmospheric extinction and its practical application. *Atmos. Environ.* **2007**, *41*, 3546–3555. [CrossRef]

47. Granados-Muñoz, M.J.; Navas-Guzmán, F.; Bravo-Aranda, J.A.; Guerrero-Rascado, J.L.; Lyamani, H.; Valenzuela, A.; Titos, G.; Fernandez-Galvez, J.; Alados-Arboledas, L. Hygroscopic growth of atmospheric aerosol particles based on active remote sensing and radiosounding measurements: Selected cases in southeastern Spain. *Atmos. Meas. Tech.* **2015**, *8*, 705–718. [CrossRef]

48. Bourgeois, Q.; Ekman, A.M.L.; Renard, J.b.; Krejci, R.; Devasthale, A.; Bender, F.A.; Riipinen, I.; Berthet, G.; Tackett, J.L. How much of the global aerosol optical depth is found in the boundary layer and free troposphere? *Atmos. Chem. Phys.* **2018**, *18*, 7709–7720. [CrossRef]

Original publication: Stirnberg et al. (2020)

Contributions: Roland Stirnberg developed the study concept and the methodology. Roland Stirnberg conducted the analysis including coding. Roland Stirnberg visualised and interpreted the results. Roland Stirnberg wrote the original draft and incorporated reviewer comments.

Peer-Review: The manuscript was submitted to Journal of Geophysical Research: Atmospheres on July 19th, 2019. After a peer-review with three anonymous reviewers, the manuscript was accepted on February 4th, 2020 and published on February 6th, 2020.

Usage of text passages and modification: The publication is included in abbreviated form in chapter 3. Parts of the method section of the original manuscript were moved to chapter 1.3 and parts of the introduction of the original manuscript were moved to chapters 1.2 and 1.3. Section titles were renamed. The numbering of figures was changed and section titles were renamed.

JGR Atmospheres

Mapping and Understanding Patterns of Air Quality Using Satellite Data and Machine Learning

Roland Stirnberg[1,2] (iD), Jan Cermak[1,2] (iD), Julia Fuchs[1,2] (iD), and Hendrik Andersen[1,2] (iD)

[1]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, [2]Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract The quantification of factors leading to harmfully high levels of particulate matter (PM) remains challenging. This study presents a novel approach using a statistical model that is trained to predict hourly concentrations of particles smaller than 10 μm (PM10) by combining satellite-borne aerosol optical depth (AOD) with meteorological and land-use parameters. The model is shown to accurately predict PM10 (overall $R^2 = 0.77$, RMSE = 7.44 $\mu g/m^3$) for measurement sites in Germany. The capability of satellite observations to map and monitor surface air pollution is assessed by investigating the relationship between AOD and PM10 in the same modeling setup. Sensitivity analyses show that important drivers of modeled PM10 include multiday mean wind flow, boundary layer height (BLH), day of year (DOY), and temperature. Different mechanisms associated with elevated PM10 concentrations are identified in winter and summer. In winter, mean predictions of PM10 concentrations >35 $\mu g/m^3$ occur when BLH is below ~500 m. Paired with multiday easterly wind flow, mean model predictions surpass 40 $\mu g/m^3$ of PM10. In summer, PM10 concentrations seemingly are less driven by meteorology, but by emission or chemical particle formation processes, which are not included in the model. The relationship between AOD and predicted PM10 concentrations depends to a large extent on ambient meteorological conditions. Results suggest that AOD can be used to assess air quality at ground level in a machine learning approach linking it with meteorological conditions.

Plain Language Summary In this study, factors leading to severe air pollution are determined using machine learning. In addition, it is tested, to what extent, that the use of satellite data is adequate to derive information on air quality near ground. It is shown that besides human emissions, concentrations of particles in the air are to a large extent driven by meteorological factors such as wind direction, state of the atmospheric boundary layer, and season.

1. Motivation and Research Questions

Extensive research has been conducted in recent years on the adverse health effects of particulate matter (PM) on the human cardiovascular system and the lungs. Cohort studies show that negative effects include emphysema, lung cancer, diabetes, and hypertension (Lelieveld et al., 2019; Lim et al., 2012; Pope et al., 2002; Wichmann et al., 2000), which cause a large number of premature deaths (Lelieveld et al., 2019, 2015). Although these risks are largely known and confirmed, discussions on effective measures to reduce exposure to air pollution are ongoing. Suggested measures range from traffic bans for certain vehicle types (Ellison et al., 2013; Qadir et al., 2013) over a reduced or more efficient use of solid fuel-based residential heating (Chafe et al., 2015) to the expansion of urban vegetation (Bonn et al., 2016). However, the actual effects of those measures are not always evident as not only local emissions but also ambient conditions such as meteorology, vegetation, and other land cover can play a substantial role in determining local PM concentrations. Therefore, understanding and quantifying drivers of PM concentrations is necessary to improve the efficiency of measures toward better air quality. The influence of meteorological conditions on particulate matter concentrations are diverse (see Figure 1) and have been described, for example, by Rost et al. (2009), Reizer and Juda-Rezler (2016), Dupont et al. (2016), Li et al. (2017), and Fuzzi et al. (2015). Boundary layer height (BLH) determines the height up to which particles are distributed in the atmosphere (Gupta & Christopher, 2009a; Schäfer et al., 2012). Precipitation leads to a substantial reduction in PM concentrations by wet scavenging (Li et al., 2015; Rost et al., 2009). Wind regulates particle transport and turbulent mixing away from the surface (Li et al., 2017). Depending on the location of the measuring point, air masses from
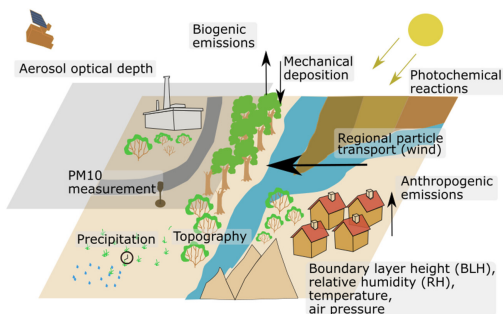
Figure 1. Conceptual design of potential influences on variations in hourly PM10 concentrations. The machine learning model is set up to reproduce these variations.

certain wind directions transport particles and lead to elevated concentrations (Lenschow et al., 2001; Li et al., 2015). Temperature can regulate the particle number in the atmosphere by stimulating photochemical reactions, which transform precursor gases to secondary aerosols (Gupta & Christopher, 2009a; Wang & Martin, 2007) or by causing partitioning of condensed precursor gases (Petit et al., 2014). Various land cover types can act as sinks or sources for particles (Beloconi et al., 2018; Bonn et al., 2016; Churkina et al., 2017). Topography can be of importance for PM concentrations, for example, as particles accumulate in valleys (Emili et al., 2011; Emili et al., 2011).

Satellite AOD provides information on atmospheric particle concentrations and can expand information to areas where station measurements are sparse (Emili et al., 2011), revealing hotspots and spatiotemporal variations of pollution (Cermak & Knutti, 2009; Gupta et al., 2006). However, relying on satellite AOD as a proxy for near-ground air pollution can be misleading, as AOD reflects the extinction of radiation in an atmospheric column, while particulate matter concentrations reflect a highly localized dry mass concentration of particles of a certain size distribution typically measured near ground (Wang & Christopher, 2003). Several studies have trained statistical models on the relationship between AOD and PM, accounting for a range of additional parameters, and mostly with a focus on applications (see review by Rybarczyk, 2018). Methods include linear regression models (Arvani et al., 2016), multiple-additive regression models (Gupta & Christopher, 2009a; Zhang et al., 2018), land-use models (Kloog et al., 2011; Nordio et al., 2013), or a combination of the latter two (Chudnovsky et al., 2014; Hu et al., 2014; Kloog et al., 2012). With increasing data availability and computational power, machine learning methods, for example, artificial neural networks (Gupta & Christopher, 2009b; Di et al., 2016) and random forests (RF) (Brokamp et al., 2017; Chen et al., 2018; Grange et al., 2018) have been applied frequently in recent years. These machine learning models are beneficial as they efficiently reproduce nonlinear relationships and interactions of input features (Brokamp et al., 2017; Elith et al., 2008). In contrast to physical models, machine learning approaches do not require extensive prior process knowledge and thus have the potential to reveal or quantify processes that are as yet undetermined (Knüsel et al., 2019). Multivariate processes can be investigated by isolating certain variables and studying inputs and responses for dominant patterns (Andersen et al., 2017; Cermak & Knutti, 2009; Fuchs et al., 2018).

While numerous air pollution studies apply statistical models mainly to accurately predict PM concentrations (Hu et al., 2017; Kloog et al., 2015; Stafoggia et al., 2017; van Donkelaar et al., 2010; Zhang et al., 2018), recent studies additionally analyzed feature importances and the information content of explanatory variables in the statistical models to infer processes. For example, Park et al. (2019) used a RF model to predict PM10 in South Korea and found large influence of AOD, day of the year (DOY), wind speed, and solar radiation on the modeled PM concentrations. Similarly, Grange et al. (2018) obtained high feature importances for wind speed, back trajectory cluster, DOY, and air temperature, using the RF approach for PM10

181

Table 1
List of Input Features Stating Data Origin, Data Representation, and Abbreviation

Data set	Variable	Abbreviation
	Input features	
MODIS MAIAC	Aerosol optical depth	AOD
MODIS	Normalized difference	NDVI
	Vegetation index	
NASA Earth at night	Lights at night	Nightlights
EU-DEM, v1.1	Elevation (m),	Elev,
	Topographic position index (m)	TPI
CORINE land cover	Land cover types	1CLC, 2CLC,
		3CLC, 4CLC
		5CLC, 6CLC
DWD meteorological	Air pressure (sea level) (hPa)	AirPres
measurements	Relative humidity (%)	RH
	Continentality	Conti
RADOLAN	Time since last precip. (hr),	Precip_tsince,
	Magnitude of last precip. (mm/hr),	Precip_magn,
	Cumulative precip. last 24h (mm),	Precip_acc
ERA-Interim	wind vectors u,v (m/s),	u,v
	mean wind vectors of	umean,
	previous hours (m/s)	vmean
	Boundary layer height (m)	BLH
	Temperature (°),	T
	temperature anomalies (K),	Tan
	Convective available	CAPE
	potential energy (J/kg)	
	Surface solar radiation	SSRD
	downwards (J/m^2)	
EEA emission data base	Annual emission of NH_3,	NH3, PM10em,
	PM10, SO_2, NO_x (t/year)	SO2, NOx
Other	Day of year	DOY
	Day of week	Weekday
	Model outcome	
UBA air quality measurements	PM10 concentrations ($\mu g/m^3$)	PM10

measurements in Switzerland. The present study builds upon the approaches applied in these studies but provides a more in-depth analysis of model-inherent relationships. To this end, gradient boosted regression trees (GBRT) are used to understand and quantify the conditions driving air quality, as well as determinants of the relationship between AOD and PM10. GBRT have been successfully applied to study sensitivities of aerosol processes before (cf. Fuchs et al., 2018). Thus, a basis is set for targeted satellite-based analyses of spatial patterns of air quality.

2. Data and Methods

The data basis of this study is comprised of 8 years (2007–2015) including satellite observations from the moderate resolution imaging spectroradiometer (MODIS) and others, model output from the European Centre for Medium-Range Weather Forecasts (ECMWF), and station data from the German Meteorological Service (Deutscher Wetterdienst, DWD) and the German Federal Environmental Agency (Umweltbundesamt, UBA). AOD observations are based on the multi-angle implementation of atmospheric correction (MAIAC) algorithm. A statistical model is set up to predict hourly PM10 concentrations based on a variety of input features, which are summarized in Table 1. Data with high temporal resolution are selected to

182

reflect the atmospheric situation close to the satellite overpass times. Generally, station measurements were preferred for parameters, which have adequate coverage and which are not expected to vary on small scales (e.g., air pressure).

2.1. Satellite Data
2.1.1. MAIAC AOD

Satellite-borne, high-resolution (1*1 km) MAIAC Collection 6 AOD is used (Lyapustin et al., 2011a; Lyapustin, Martonchik, et al. 2011; Lyapustin et al. 2011b; Lyapustin et al., 2018). The product is based on data from the MODIS sensor aboard the Terra and Aqua satellites. The MAIAC algorithm makes use of look-up tables, explicitly taking into account surface bidirectional reflectance factors (BRF). The calculation of AOD relies on the assumption that surface BRFs remain largely constant over time, considering a time series of 16 consecutive days (Lyapustin et al., 2011a; Lyapustin, Martonchik, et al. 2011). Quality flags are applied for filtering. Pixels were only incorporated when clear conditions are reported, that is, no contamination of the data due to clouds is to be expected (Lyapustin et al., 2018). An additional filter was set up to exclude AOD close to clouds to avoid increased AOD near cloud fringes due to aerosol swelling effects (Schwarz et al., 2017; Várnai et al., 2013). Therefore, the distance to the nearest cloud as classified by the MAIAC algorithm (Lyapustin et al., 2018) was calculated and a threshold of 0.1° was set, which corresponds to ~7 km. This is in the range of what previous studies used as threshold (Emili et al., 2011; Koren et al., 2007). Terra or Aqua satellite overpass times were used as temporal reference for data collocation, that is, other input feature values closest to the overpass times were used. Valid daytime AOD acquisition times as used in this study range from ~9 to ~1 UTC. AOD is an important input to the statistical model as it provides implicit information on the total aerosol loading in the atmosphere, reflecting natural as well as anthropogenic sources.

2.1.2. MODIS NDVI

Sixteen-day NDVI averages obtained from the MODIS MOD13Q1 Version 6 product are used (DOI: 10.5067/MODIS/MOD13Q1.006) to approximate photosynthetically active vegetation (Tucker C. J., 1979). The NDVI was found to influence PM concentrations in several previous works (Beloconi et al., 2018; Chudnovsky et al., 2014; Stafoggia et al., 2017). Vegetation acts as a sink by increasing the aerodynamic roughness and available surface for mechanical deposition (Fuzzi et al., 2015). On the other hand, vegetation can increase background particle concentrations by emitting pollen in spring (Fuzzi et al., 2015) and by enhancing the creation of secondary organic aerosols (SOA, Churkina et al., 2017). Around each PM station coordinate, a window with an edge size of 20 km is established to reflect the local contribution of vegetation to the total particle concentration. An edge size of 20 km is comparable to the mean representativeness of the PM stations (EU, 2008). The mean NDVI of each window was used as a predictor.

2.1.3. NASA Night Lights

The NASA night lights data set is included as a surrogate for population density. Based on the data from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi National Polar-Orbiting Platform (SNPP), the night-time lights product is available at 500 m resolution (Román et al., 2018). Population density is an important factor contributing to PM_{10} concentrations as it reflects human activity (Beloconi et al., 2018; Park et al., 2019). The mean night light intensity of a 20 km window around each PM station was used as a predictor.

2.1.4. EEA DEM

Topography can have a marked influence on the accumulation of particles (Emili et al., 2011). Here, the v1.1 EU-DEM is used, which is a hybrid product produced by the European Environment Agency (EEA) based on SRTM and ASTER GDEM fused by a weighted averaging approach. It has a spatial resolution of 25*25 m (more information and download at https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata). Station altitude and the dominant topography around each PM station are incorporated as predictors. A topographic position index (TPI) is computed to provide information on the topography of a pixel relative to its surrounding pixels. It employs the station altitude and subtracts the mean altitude of surrounding pixels in its vicinity. Positive values indicate a summit position, whereas negative values indicate a valley position (Egli et al., 2018). Again, a window size of 20 km was chosen.

2.1.5. EEA Corine Land Cover

Data from the CORINE land cover (CLC) inventory for the years 2006 and 2012 are used (Bossard et al., 2000), accessed via https://land.copernicus.eu/pan-european/corine-land-cover/view. In its finest thematic accuracy, the CLC data set consists of 44 classes. For this study, a more simplistic classification in six land cover types was chosen, which represent the most relevant land cover types influencing air quality (see

Table 2
Original Labels and Aggregation of Corine Land Cover Classes Used in This Study

Code	Label	Original labels included
1	Artificial surfaces	Continuous urban fabric, discontinuous urban fabric, industrial or commercial units, road and rail networks and associated land, port areas, airports, mineral extraction sites, dump sites, construction sites, green urban areas
2	Agricultural areas	Nonirrigated arable land, permanently irrigated land, rice fields, vineyards, fruit trees and berry plantations, olive groves, pastures, annual crops associated with permanent crops, complex cultivation patterns, land principally occupied by agriculture with significant areas of natural vegetation, agro-forestry areas
3	Forest and seminatural areas	Broad-leaved forest, coniferous forest, mixed forest, natural grasslands, moors and heathland, sclerophyllous vegetation, transitional woodland-shrub, beaches & dunes & sands, bare rocks, sparsely vegetated areas, burnt areas, glaciers and perpetual snow
4	Wetlands	Inland marshes, peat bogs, salt marshes, salines, intertidal flats
5	Water bodies	Water courses, water bodies, coastal lagoons, estuaries, sea and ocean
6	No data, unclassified	No data, unclassified land surface, unclassified water bodies, unclassified

Table 2). The data have a spatial resolution of 250 m. Because the CLC data are categorical, a window of 20 km edge size was set up, and the number of occurrences of each type in that window was calculated. The number of occurrences of each feature is used as predictor for the model.

2.2. Reanalysis Data

The Era-Interim reanalysis data by the ECMWF provides full spatial coverage for the whole study period with an interpolated spatial resolution of 0.125 ° (Dee et al., 2011). The data set has been successfully used in previous air quality studies (cf. Chen et al., 2018; Stafoggia et al., 2017; Zheng et al., 2017). To capture regional transport of particles, ERA-Interim reanalysis wind components (m/s) in east-west and north-south direction (10 m height) are used. Wind direction and speed can influence both particle concentrations (Beloconi et al., 2018; Chudnovsky et al., 2013; Li et al., 2015) and the relationship between AOD and PM10 (Stirnberg et al., 2018; Zheng et al., 2017). Wind direction and speed are included as instantaneous values and as the mean of precedent days (72 hr). BLH data are employed to approximate the vertical distribution of aerosols in the lower troposphere, implying that particles are well mixed within the boundary layer (Ansmann et al., 2000; Gupta & Christopher, 2009a). Dispersion of particles within a high and well-mixed boundary layer leads to reduced particle numbers near ground. If the BLH is low, particles accumulate and increase PM concentrations near ground as they are constrained to a smaller volume (Gupta & Christopher, 2009a; Wagner & Schäfer, 2017). Era-Interim temperature in 2 m height is included as instantaneous values and as temperature anomaly. Temperature anomalies are determined as the deviation of the expected value for each day of the year. Expected values are calculated by averaging daily values over 30 years, then calculating a running mean over 30 days to achieve a smooth sequence of the expected temperature. In addition to temperature, downward surface solar radiation (SSRD) and convective potential energy (CAPE) are included to capture potential secondary aerosol formation based on photochemical transformation processes (Wang & Martin, 2007) and to capture convective mixing in the atmosphere (Chudnovsky et al., 2013).
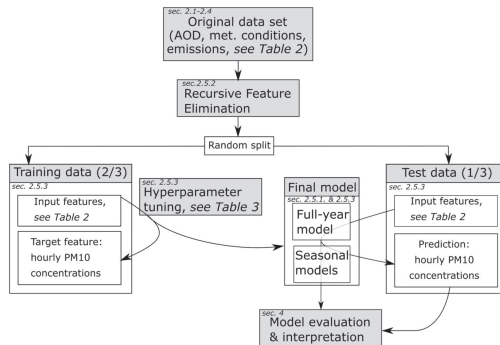
184

Figure 2. Framework of this study: from input feature selection (i.e., separating important features from noise), hyperparameter tuning, model training, and model evaluation to the interpretation of the results.

2.3. Ground-Based Data
2.3.1. UBA PM10

The focus here is on PM10 rather than PM2.5 because the latter excludes the fraction of larger particles, which are nevertheless accountable for light extinction and thus contribute to AOD measured by the satellite (Emili et al., 2011). In addition, PM10 measurements are more widely available. The PM stations are maintained by the UBA and measure the hourly mean PM10 concentration. PM concentrations are determined by measuring the attenuation of β-radiation by a dust-coated filter (Umweltbundesamt, 2004) or by the principle of oscillating micro weighing (TÜVRheinland, 2012). To avoid condensation, the particle inlet is heated permanently. Thus, measurements are largely uninfluenced by temperature and humidity (Umweltbundesamt, 2004). The uncertainty of the continuous measurements is prescribed to be below 25% (Bundesministerium der Justiz und für Verbraucherschutz, 2010; VDI, 2002). Instructions by the European Union prescribe site locations that avoid microenvironmental effects. Stations are classified as background, industrial, or traffic and are designed to be representative for urban, rural, or suburban areas. Traffic sites are generally situated relatively close to main roads or intersections (EU, 2008). Urban background sites are assumed to capture the contribution of all sources near the site without one particularly dominating source. Suburban background stations need to be placed downwind (referring to the main wind direction) of emission sources. Rural background stations should not be influenced by agglomerations or industrial sites closer than 5 km (EU, 2008). The spatial distribution of station types and their altitudes are shown in Figures A1 and A2. UBA PM station coordinates are used as spatial reference for data collocation, that is, pixels from continuous data grids are collocated with the position of these stations if below a distance threshold of 0.01 ° (~0.7 km). Urban industrial stations are not considered in this study, as they are primarily influenced by local emissions from point sources that cannot be adequately represented with the available data. PM10 concentrations were checked for sudden peaks, which could be due to localized events. These are filtered out by eliminating situations, where the PM10 concentration was more than double the mean of the previous and following hours.

2.3.2. DWD Meteorological Data

Air pressure and relative humidity (RH) data were obtained from the German Meteorological Service (DWD) (DWD Climate Data Center [CDC], 2017). Previous studies found a positive correlation of PM10 concentrations and air pressure. Higher air pressure indicates stable synoptic conditions, which favors the accumulation of particles (Li et al., 2015). RH potentially enhances particle numbers by stimulating the formation of aqueous SOA, forming in cloud or aerosol water (Ervens et al., 2011). In addition, RH can influence the relationship between AOD and PM10, as higher levels of humidity lead to hygroscopic particle growth. Moisture on particles increases their diameter, eventually causing a rise in AOD without affecting PM10 measurements (Schwarz et al., 2017; Titos et al., 2014; Stirnberg et al., 2018).

2.3.3. Continentality Factor

The relationship between AOD and PM10 as well as driving factors of PM10 concentrations may vary in different climatic regions (Di et al., 2016). This is accounted for by the inclusion of a dimensionless continentality factor k, which is calculated following the formula by Conrad (1946)

$$k = (1.7 * A/sin(\phi + 10°)) - 14, \tag{1}$$

where A corresponds to the difference between the hottest and coldest mean monthly temperature and ϕ to the latitude in decimal degree. Here, k was calculated using the mean temperature of July (1980–2010) and the mean temperature of January (1980–2010). Temperature data are provided by the DWD (DWD Climate Data Center [CDC], 2018).

2.3.4. DWD Radolan Precipitation

The influence of precipitation on PM10 concentrations includes wash-out effects and the limitation of movement of particles after precipitation events (Fuzzi et al., 2015; Li et al., 2015; Rost et al., 2009). In this study, data from the Radar Online Adjustment project (RADOLAN) are used. The data set is produced by the DWD and merges radar measurements with rain gauge data, also including orographic correction (Bartels et al., 2004; Weigl, 2017). The data are available with high spatial coverage and are expected to be better suited for the present analysis than the Era-Interim precipitation product, which has some known biases (de Leeuw et al., 2015). In the statistical model, the time since the last precipitation event (hr) and its magnitude (mm/hr) as well as the accumulated precipitation of the last 24 hr (mm) are included. Hourly means of precipitation around each PM station are averaged within a window of edge size 5 km. The chosen window size is smaller than the previous ones, as precipitation effects typically vary on small scales.

2.4. Other Input Data
2.4.1. EEA Emission Database

Annual emissions of NO_x, PM10, SO_2, and NH_3 (in tonnes, based on the year 2008) are included to approximate background pollution levels. The data are gathered by the European Environment Agency (EEA) and described in Theloke et al. (2009). Diffuse air releases from traffic, agricultural, industrial, and residential sources are covered. Strong emitters that fall under the European Pollutant Release and Transfer Register (E-PRTR) are not included in this data set.

2.4.2. Spatiotemporal Information

Seasonality was shown to be an important factor in previous studies (Grange et al., 2018). Here, DOY is used as seasonal proxy. To mirror the seasonal cycle, DOY was converted to a sine curve with +1 representing summer solstice and −1 representing winter solstice (Park et al., 2019; Stolwijk et al., 1999). To further approximate variability in emission strengths based on human activity, day of the week is included.

2.5. Gradient Boosted Regression Trees
2.5.1. Model Specifications

GBRT as implemented in python's scikit-learn module are used (Hastie et al., 2009; Pedregosa et al., 2012). GBRT merge several statistical approaches found in machine learning applications: decision trees (1) and boosting (2) with gradient descent (3).

1. Decision trees use decision nodes to split the predictor space in subsets, which provide the most homogeneous distribution, that is, the subsets' variance is minimized. For each subset, regression trees fit the mean response of the observations that go into the model (Elith et al., 2008).
2. Similar to the RF method, GBRT consist of an ensemble of decision trees. In GBRT models, however, the construction of the ensemble is different, as decision tree regressors are sequentially added to the ensemble (Elith et al., 2008; Hastie et al., 2009; Rybarczyk, 2018). Each new tree that is added to the ensemble boosts its predecessor with the goal to minimize a loss function, and existing trees are not changed when new trees are added. The trees are fitted on a subset of the complete data set, which induces a random component to the model to reduce overfitting (Elith et al., 2008; Hastie et al., 2009).
3. Each new predictor is fitted to the predecessor's previous residual error using gradient descent (Hastie et al., 2009; Elith et al., 2008).

GBRT capture complex interactions and interactive effects between individual predictors (Brokamp et al., 2017; Elith et al., 2008), which nevertheless need to be considered when interpreting the model outcome. An advantage of tree-based methods such as RF or GBRT is that, compared to deep learning methods,

Table 3
List of Hyperparameters and Parameter Grid That is Applied During the Grid Search

Hyperparameter	Description	Parameter grid
loss	Loss function to be optimized	fixed: least squares regression
learning_rate	Contribution of trees to ensemble	[0.08,0.05,0.01]
n_estimators	Number of boosting iterations	[800, 1,800, 2,500, 4,000]
subsample	Fraction of samples used for individual tree	[0.6, 0.8]
min_samples_split	Minimum number of samples used to split a decision node	[6, 10, 14]
min_samples_leaf	Minimum number of samples required for a leaf node	[14, 18, 20]
max_depth	Maximum depth for individual tree, i.e., maximum number of node layers	[6, 10, 14]
max_features	Fraction of features to be considered when searching for best split	[0.5, 0.8]

Note. Hyperparameters determine the architecture of the GBRT model.

model decisions can be retraced and dependencies of the model outcome to input features can be quantified, allowing for conclusions regarding physical processes. This makes GBRT an interpretable machine learning method. GBRT theoretically produce results more effectively than the RF method, as trees are built systematically and less iterations are required (Elith et al., 2008). GBRT have shown to have good predictive power in previous studies (Elith et al., 2008; Fuchs et al., 2018; Just et al., 2018). The general framework of setting up the model is shown in Figure 2 and includes input feature selection, hyperparameter tuning, model training, and model validation.

2.5.2. Feature Selection: Recursive Feature Elimination

Redundant input features lack information and potentially degrade model performance by inducing misleading information, thus weakening the target orientation of the model (Meyer et al., 2018). Therefore, a feature selection is conducted to eliminate redundant predictors. A base model is initialized with a fixed number of trees (500), a fixed learning rate (0.1), and all other hyperparameters on default settings. One feature is then excluded, and the data set is randomly split into a training data set (2/3) and a test data set (1/3) 50-fold. After 50 repetitions, the decrease in model performance on the test data set due to the exclusion of the feature was determined using R^2 as indicator. In the final model, the magnitude of the last precipitation event, the accumulated rainfall of the last 24 hr, and the annual mean emission of SO_2 are excluded, as their exclusion did not lead to a decrease in model performance.

2.5.3. Hyperparameter Tuning, Model Training, and Model Evaluation

Hyperparameters refer to the model architecture, for example, the number of trees or the number of decision nodes. The determination of adequate model hyperparameters is essential to avoid overfitting of the model but at the same time ensures that the model is able to generalize. A grid search is executed, where several parameter combinations are tested. A list of tested parameters is provided in Table 3.

There are trade-offs between *max_depth*, *n_estimators*, and *learning_rate*. A lower *learning_rate* requires a higher number of trees, as the contribution of each tree is decreased. The contribution of each tree is effectively the step size of the gradient descent. Thus, if *learning_rate* is too large, the model cannot adapt to the training data. The number of decision nodes in a tree (*max_depth*) also affects *n_estimators*: Increasing *max_depth* reduces the number of necessary trees but increases the risk of overfitting.

The model penalizes erroneous predictions to improve accuracy. The calculation of the penalty value is determined by the loss function. In this study, a least squares loss function is chosen. The least squares loss function is sensitive to very high (low) values as it strongly penalizes large deviations between predictions and observations and will adjust the model accordingly. This is desirable, as the model should be able to reproduce high concentrations of PM10. Model performance is validated using two kinds of validation strategies. The integrated scikit-learn split function creates a random training (2/3) and test data set (1/3), ensuring a comparable distribution of both data sets. To test the spatial generalizability of the model,
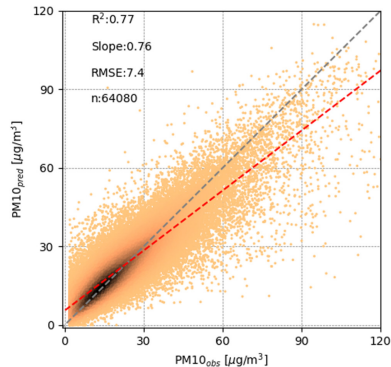
187

Figure 3. Scatter plot showing the full-year model predictions for hourly PM10 concentrations. Also shown are coefficient of determination (R^2), slope (red dotted line), and root mean square error (RMSE). The color range from black (high) to orange (low) indicates the frequency of occurrence. The relatively high R^2 shows that the model covers the majority of occurring variance. However, an underestimation of higher PM10 concentrations leads to a lower slope.

a leave-location-out (LLO) approach is conducted. Therefore, one third of randomly chosen stations are restrained for validation. If the model performance is lower compared to the random approach, the spatial generalizability of the model is limited (Meyer et al., 2018).

For further analysis, four seasonal models are trained on seasonal subsets of the data (see also Figure 2) using the hyperparameters determined in the full-year grid search. One input feature set representative for all seasonal models and the full-year model is used to ensure their comparability. The only exception is the EEA emission data set, which is only included in the yearly model, since it represents yearly emissions.

2.6. Model Interpretation: Isolation of Feature Contributions
2.6.1. Feature Importance

The relative feature importance reflects the explanatory power a feature provides to the model. Assuming that the model is able to capture physical processes well, the feature importance represents a valuable qualitative measure to determine the relative magnitude of the influence of input features to predicted PM10 concentrations. The feature importance is calculated by repeated permutation of one feature (Strobl et al., 2007).

2.6.2. Partial Dependence

To quantify the influence of input features on the model, the partial dependence (PD) of modeled PM10 concentrations on input features is calculated. PDs express the average effect of one input feature on the modeled PM10 outcome while accounting for average effects of complement input features (Elith et al., 2008; Hastie et al., 2009; Goldstein et al., 2015). The investigated input feature is gridded and the corresponding average PM10 prediction is calculated with respect to complement features, which are varied over their marginalized distributions (e.g., 1st–99th percentile). Thus, PD plots reflect the mean change of average predicted PM10 concentrations based on one input feature. The isolated effects of input features on the model response can be evaluated and put into context regarding their significance to physical and chemical processes determining PM10 concentrations (Grange et al., 2018).

2.6.3. Individual Conditional Expectation

PD plots reflect the mean model response and neglect model heterogeneity. Model responses to single data instances (i.e., one set of input features related to one PM10 observation) can be unwrapped and bundled in one plot using individual conditional expectation (ICE). ICE plots reflect individual predicted responses as a function of one data instance depending on correspondent feature observations. Model responses are computed by keeping the complement features constant while the investigated feature varies, thus creating new
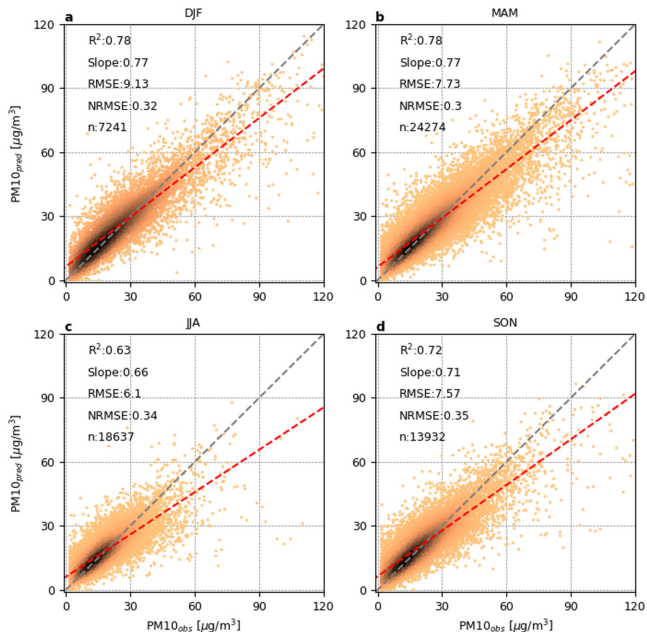
188

Figure 4. Scatter plot showing seasonal model predictions for hourly PM10 concentrations. Also shown are seasonal R^2, slope (red dotted line), normalized RMSE (NRMSE), and RMSE. The NRMSE was calculated by dividing the RMSE by the mean of the corresponding seasonal subgroup of PM10 observations. Colors as in Figure 3.

data instances and predictions from the model. The average over all ICE lines yields the PD plot, allowing model heterogeneity and mean model response to be depicted simultaneously (Goldstein et al., 2015).

3. Results and Discussion

3.1. Model Performance

The overall model performance is shown in Figure 3, depicting observed PM10 versus predicted PM10 for the validation data. The full-year model explains 77% of the variance. The slope (0.76) shows a slight overestimation of low PM10 concentrations as well as an underestimation of high PM10 concentrations. Presumably, the underestimation is due to processes not captured by the input features, that is, street-scale processes not covered by AOD observations but still influencing PM10 observations, such as increased PM10 emissions due to traffic jams or localized dust resuspension. In addition, the model possibly tends to underestimate higher PM10 observations, because most valid data points are available for medium to low PM10 concentrations. Thus, the model is optimized to best reproduce these observations. This tendency was reduced by the choice of the least squares loss function as describes in chapter 2.5.3 but likely still continues to affect the model accuracy. The model performance is comparable to similar studies, also in its underestimation of PM (Hu et al., 2017; Grange et al., 2018; Stafoggia et al., 2017; Zhang et al., 2018). Tenfold random train/test splits were conducted, resulting in 10 models. Validation of these models revealed very similar performances, which is why only one model is shown and used for subsequent model analysis. Applying the LLO split
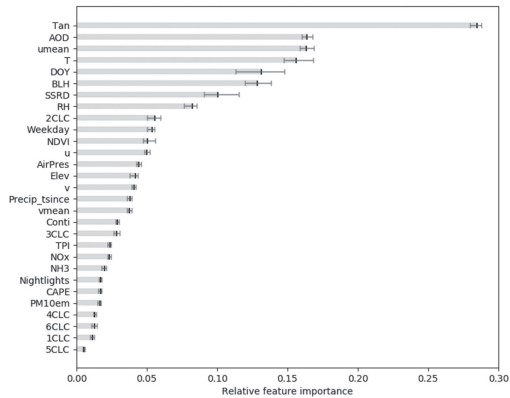
189

Figure 5. Relative importance of input features based on repeated permutation (see section 2.6). The range represents the standard deviation. Abbreviations correspond to those shown in Table 1.

approach slightly deteriorated model performance. Depending on the restrained stations, R^2 ranged from 0.5 to 0.7. The model is considered as adequate to be used for further investigations. Nevertheless, note that high PM10 values tend to be underestimated by the model (see Figure 3). The spatial distribution of the model skill is shown in Figures B1 and B2.

Model performances vary seasonally (see Figure 4). The model performs best in winter and spring with high R^2 values (0.77) and low NRMSEs (0.32 and 0.3), while R^2 is lowest in summer with an R^2 of 0.63 and a slightly increased NRMSE of 0.34. PM10 concentrations generally show less variance in summer, which reduces the RMSE but possibly provides the model less variance to learn from and deteriorates its skill (i.e., R^2). Obviously, processes governing PM10 concentrations in summer are not as well captured by the model. This will be further addressed in the following chapters.

3.2. Information Content of Input Features

Temperature (anomaly and absolute), AOD, 3-day mean east-west wind component, DOY, and BLH are of high importance to the model (see Figure 5). The importance of the DOY suggests that the model captures the seasonality of PM10 concentrations, which are higher in winter and lower in summer. The relatively high importance of AOD and the good model performance emphasize the suitability of AOD to infer on PM10 concentrations when additional parameters are taken into account. A comparison to similar studies, for example, by Grange et al. (2018) and Park et al. (2019), reveals comparable relative feature importances of AOD, solar radiation (Park et al., 2019), DOY, and BLH (Grange et al., 2018; Park et al., 2019). The high importance of the 3-day mean of the east-west wind component found in this study aligns with the high importance of the back trajectory clusters in the study by Grange et al. (2018). Both parameters reflect regional particle transport. However, there is a discrepancy regarding the importance of wind speed and temperature. Wind speed is considered as instantaneous wind components in this study, which have limited importance. While the importance of absolute air temperature is comparable, the high importance of temperature anomalies found here shows that the approach of splitting temperature information into absolute values and anomalies as pursued in this study provides additional information to the model. Since comparing feature importance values can provide only limited insights into processes behind air pollution patterns, further quantitative analyses are presented in the following chapters using the ICE approach.
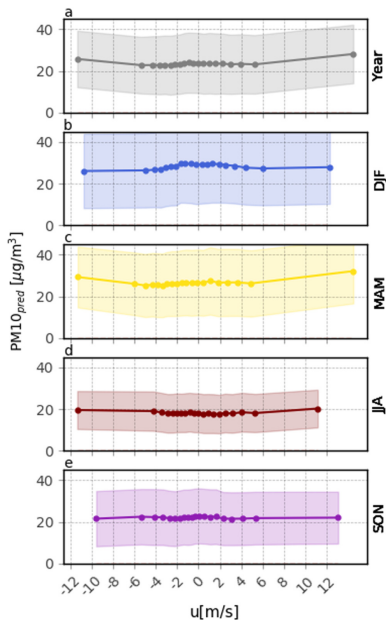
190

Figure 6. Partial dependence plot showing the mean model response to changes in 3-day mean east-west wind component (m/s) as bold lines for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Shaded areas show the range of the individual conditional expectation (ICE) lines (10th to 90th percentiles). The horizontal distance of dots on the bold line indicates the distribution of valid data points. Negative (positive) values represent dominant inflow of eastern (western) air masses.

3.3. Model Sensitivity

3.3.1. Mesoscale Wind Information

The PD of the 3-day mean east-west wind component shows a consistent pattern throughout all seasons. Positive values (i.e., prevailing western direction of inflow) are associated with reduced concentrations of PM10, whereas a negative east-west wind component (i.e., winds from the east) is associated with increased model PM10 concentrations (Figure 6). For the full-year model the maximum difference in mean PM10 predictions is ~10 $\mu g/m^3$, while in winter, the maximum difference is ~20 $\mu g/m^3$. These numbers agree well with results from van Pinxteren et al. (2019), who quantified the influence of eastern air masses on eastern Germany. They found the contribution of trans-boundary transport from eastern European countries to be 13 $\mu g/m^3$ on average, depending on meteorological conditions. Air masses from continental eastern Europe tend to transport higher amounts of particles, whereas western, more maritime air tends to be cleaner due to precipitation along the trajectories of air masses. Source regions of PM10 include industrial and residential areas in Poland and the Czech Republic with heavy industries or extensive usage of solid fuels for residential heating (Beloconi et al., 2018; Kiesewetter et al., 2015; Reizer & Juda-Rezler, 2016; van Pinxteren et al., 2019). Results by Grange et al. (2018) also show increased values of PM10 for northern and northeastern wind directions, although to a lesser extent. In winter, the effect of particle transportation is strongest, presumably due to increased emissions from domestic heating in eastern Europe (Reizer & Juda-Rezler, 2016;
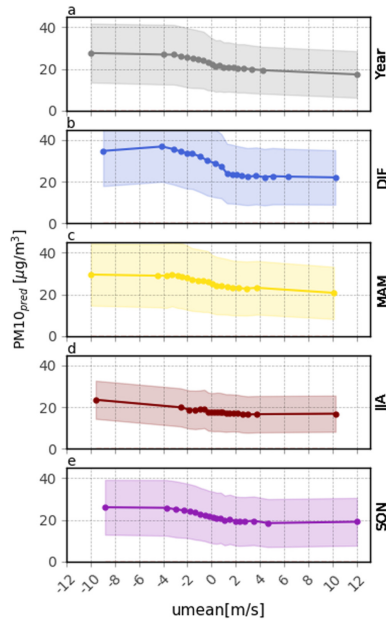
Figure 7. PD plot showing the mean model response to changes in east-west wind component (m/s) as bold lines for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Negative (positive) values represent dominant inflow of eastern (western) air masses. Description as in Figure 6.

van Pinxteren et al., 2019). A slight increase in modeled PM10 at low values of instantaneous east-west wind components is visible (see Figure 7). This could indicate insufficient mixing of the atmosphere, which would lead to an accumulation of particles near ground (Chudnovsky et al., 2013). Overall, instantaneous wind information has little influence on the model, causing the PD to remain relatively constant. The relatively constant PD of instantaneous wind information implies little influence on PM10 predictions. This suggests that wind information needs to be extended to a longer time scale to influence PM predictions. As shown in Grange et al. (2018), wind speed aggregated for a daily period can substantially influence PM10 concentrations, with lower speeds causing higher concentrations. Park et al. (2019) also found the maximum wind speed of previous 3 hr to be of importance for their statistical predictions of PM10. The north-south wind component PDs (instantaneous and 3 days) do not provide clear trends.

3.3.2. BLH and CAPE

The PD of BLH shows that the model is able to reproduce the pattern of decreasing particle concentrations with increasing BLH (Gupta & Christopher, 2009a; Wagner & Schäfer, 2017). The shape of the full-year PD of BLH shown in Figure 8a is similar to that provided by Grange et al. (2018) for observations in Switzerland. They found a reduction of ∼8 $\mu g/m^3$ for daily PM10 predictions. A reduction in mean PM10 predictions of ∼10 $\mu g/m^3$ for situations with higher BLH can be seen in Figure 8a. This similarity is encouraging and proves the robustness of the modeling approach, since both studies use ERA-Interim BLH in a comparable geographic setting.
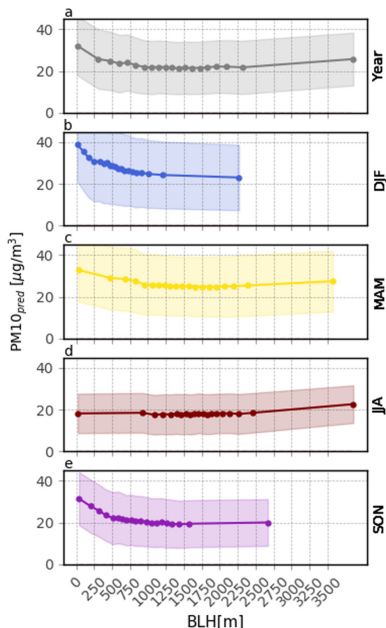
Figure 8. PD plot showing the mean model response to changes in instantaneous BLH (m) for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Description as in Figure 6.

For BLH values above ~800 m, the PD remains constant, that is, the influence of the boundary layer on PM concentrations stagnates (Liu et al., 2018). Mean modeled PM10 concentrations increase slightly in conditions with very high BLH (>2,000 m). This pattern could be related to the formation of a deep, convective boundary layer coinciding with high temperatures, enhancing the formation of secondary aerosols (Grange et al., 2018). The abundance of radiation, high temperatures, and precursor gases at excess concentrations would be needed therefore (Fuzzi et al., 2015). Indeed, the pattern is most prominent in summer, when these prerequisites are most likely to be met. The PD of BLH in summer is almost constant, that is, little information is provided to the model. Figure 8d reflects a shift of the frequency of occurrence of data points toward higher BLH: During summer months, medium to high BLHs are more likely to occur due to enhanced convection. As mentioned before, a BLH above 800 m provides only little information to the model, thus reducing its predictive ability. The accumulation effect of a lower BLH is most pronounced in winter with increased mean PM10 predictions of almost 20 μg/m³. In addition, PM10 emissions are expected to be higher in wintertime due to combustion of solid fuels for domestic heating. This has been shown for Eastern European countries (Reizer & Juda-Rezler, 2016; van Pinxteren et al., 2019) and likely influences PM10 concentrations in Germany. Thus, with higher locally produced or advected PM10 emissions, the accumulation effect of a low BLH is more distinct in winter than in summer, where reduced emissions are expected and an accumulation of particles is not expected (Wagner & Schäfer, 2017). A dependence of model PM10 on CAPE could not be identified using the PD approach.
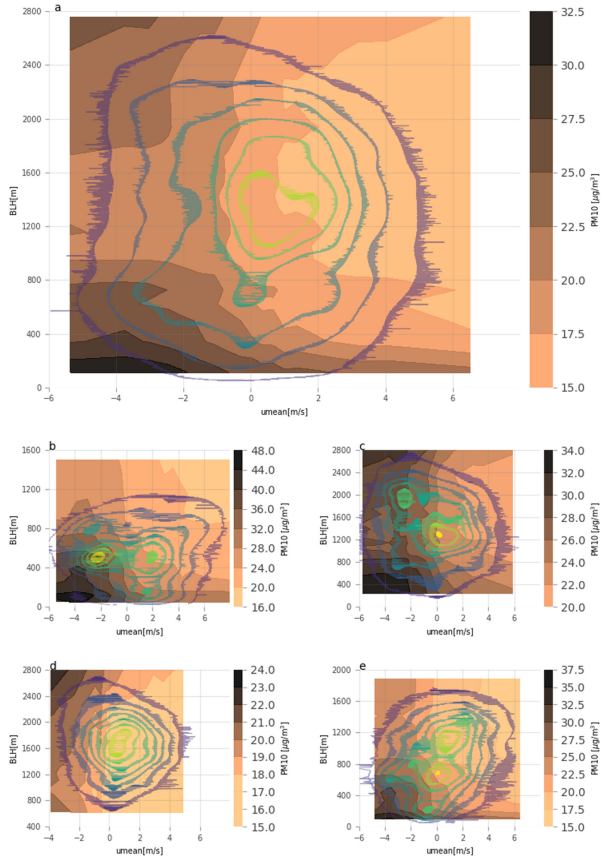
193

Figure 9. Two-way PD of umean and BLH, full-year model (a) and seasonal models (b–e, DJF, MAM, JJA, and SON). Similar to the previously shown PD, the two-way PD shows the mean modeled PM10 (color coded from light orange to black) to the isolated effects of umean and BLH. On top of the PD isolines, a probability density function (PDF) using Gaussian kernel density estimates is added to indicate the qualitative frequency of occurrence of values from high (yellow) to low (dark blue).
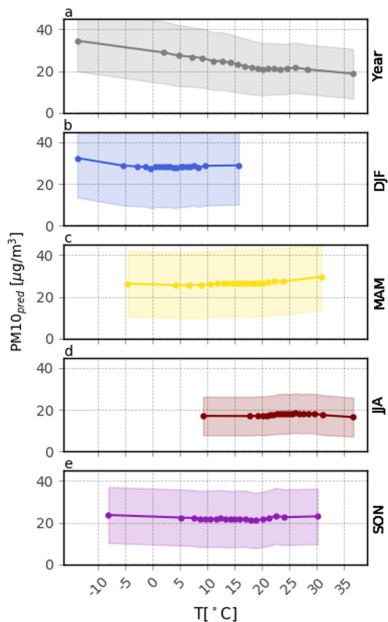
194

Figure 10. PD plot showing the mean model response to changes in temperature anomalies (K) for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Description as in Figure 6.

3.3.3. Two-Way PD: umean and BLH

Depicting the PD of BLH and the 3-day mean east-west wind component simultaneously allows for the quantification of particularly high-polluted situations, considering combined effects of both features. A probability density function (PDF) using Gaussian kernel density estimates is added to provide a qualitative estimate of the frequency of occurrence of PD values. Values outside the PDF estimation as shown in Figure 9 are extrapolated based on the trained model and do not represent observed data.

The two-way PD suggests mean modeled PM10 concentrations double due to changes in BLH and wind flow, referring to the full-year model. Highest mean predictions ($\sim$35 $\mu g/m^3$) are modeled when eastern winds coincide with shallow boundary layers, whereas lowest mean predictions occur during medium BLH and western winds ($\sim$17 $\mu g/m^3$). Note that due to the tendency of the model to underestimate high PM10 levels, concentrations could be higher in reality. Patterns differ seasonally. In winter, highest mean PM10 predictions surpass 45 $\mu g/m^3$ during shallow BLH conditions and wind flow from the east. In summer, this is not the case. As mentioned in chapter 3.3.2, there is indication of elevated PM10 concentrations during very high BLH ($\sim$>2,000 m) conditions, coinciding with eastern wind flow. Highest mean predictions in summer do not surpass $\sim$22 $\mu g/m^3$ (within the limits of the PDF estimation).

3.3.4. Thermal Influence on PM10

In spring, summer, and autumn, positive temperature anomalies cause a marked increase of mean model PM10 predictions (see Figure 10). Presumably, higher temperatures in spring and summer (Figures 10c and 10d) reflect enhanced biogenic activity, as vegetation is generally more active at higher temperatures.
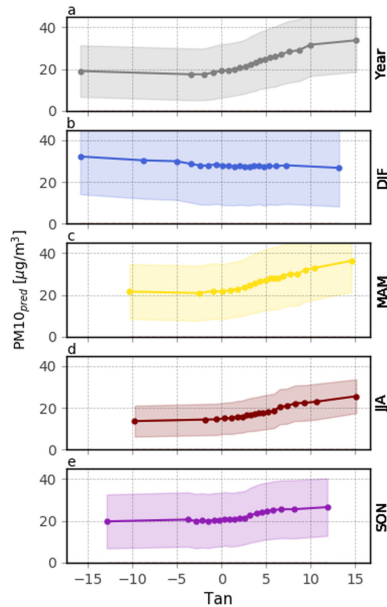
195

Figure 11. PD plot showing the mean model response to changes in instantaneous temperature (° C) for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Description as in Figure 6.

Consequently, emissions of primary particles such as debris or pollen and the emission of biogenic volatile organic compounds (BVOCs) are stimulated (Laothawornkitkul et al., 2009). With higher BVOC emission, an enhancement of secondary SOA formations is leveraged (Churkina et al., 2017; Megaritis et al., 2013). In addition to increased biogenic activities, higher temperatures cause the soil to dry up more quickly, thus increasing dust emissions (Hoffmann & Funk, 2015). In winter, positive anomalies have very little effect on predicted PM10 concentrations (Figure 10b). This supports these hypotheses, since neither increased biogenic activity nor dried-up soils are to be expected in winter. Higher temperatures could reflect increased photochemical oxidation processes, which trigger photochemical reactions leading to new particle formation processes (Birmili & Wiedensohler, 2000; Größ et al., 2018; Wiedensohler, 2000). However, there was no trend in the partial dependence of SSRD, which shows a weak influence of SSRD on PM10 predictions (see Figure C1). Note however that the lacking influence of SSRD could also be due to the fact that this study is confined to cloud-free situations due to the availability of AOD. The variation of SSRD would be higher when including cloudy days, which would likely improve the information content provided to the model by including SSRD. Another possible explanation for increased PM10 concentrations at higher temperatures could be that these situations are associated with stable synoptic conditions (at least in spring, summer, and autumn), causing particles to accumulate in the atmosphere. PD trends for instantaneous temperature (Figure 11) are inverse to those described for temperature anomalies. The full-year model PD shows a decrease in predicted PM10 concentrations for higher temperatures. Presumably, instantaneous temperature reflects the annual cycle of PM10 (similar to DOY), which is why the seasonal PDs show no trends.
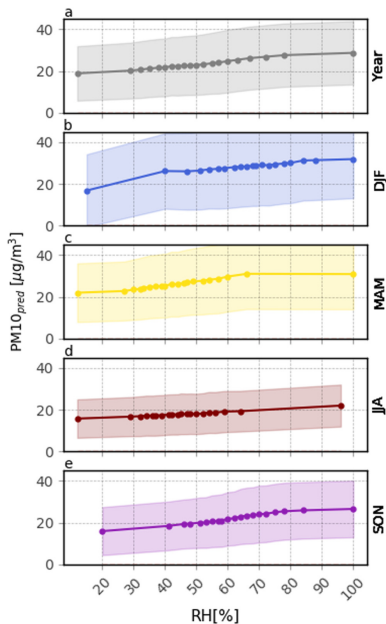
Figure 12. PD plot showing the mean model response to changes in RH (%) for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Description as in Figure 6.

PD of air temperature as presented in Grange et al. (2018) reveal a different pattern. In their study, temperatures below freezing are associated with high PM10 concentrations, medium temperature in the range of 0–15 °C with low PM10 concentrations, and temperature above 15 °C with high PM10 concentrations. However, when combining the effects of temperature anomalies and temperature presented in Figures 10a and 11a, the emerging pattern would be similar. This suggests that the model presented in this study is able to discriminate between the seasonal component of temperature and the immediate effect of temperature on particle emissions, for example, due to new particle formation (cf. Birmili & Wiedensohler, 2000; Bressi et al., 2013; Größ et al., 2018; Petetin et al., 2014; Wiedensohler, 2000)

3.3.5. RH and Precipitation

Increased RH is associated with higher PM10 predictions (see Figure 12). This is likely related to an increase in AOD due to aerosol swelling in humid conditions (Crumeyrolle et al., 2014; Wang & Christopher, 2003).

Other than increasing AOD, the importance of RH could also be related to a correlation between BLH and RH (see PDF estimate in Figure 13). Higher RH reduces the magnitude of turbulent vertical flux and subsequently reduces the BLH (Adamopoulos et al., 2007; Petäjä et al., 2016), which in turn could increase PM10 predictions. On the other hand, vertical transport of water vapor is impeded by a low BLH, which increases RH and stimulates formation of aqueous secondary aerosols (Liu et al., 2018). To analyze the influences of BLH and RH on predicted PM10, a two-way PD of RH and BLH is calculated (see Figure 13). It shows a
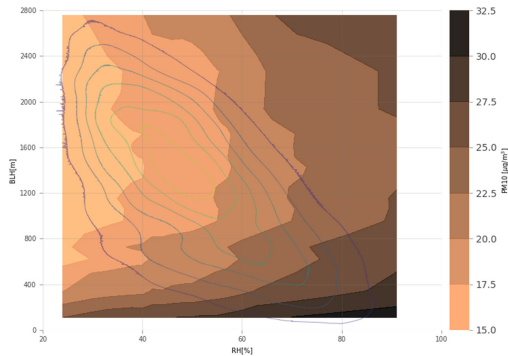
197

Figure 13. Two-way PD of RH and BLH, full-year model. Description as in Figure 9.

changing pattern with increasing RH, visible in a change of line structure orientation from vertical to more horizontal in the two-way PD plot. While BLH dominates during shallow boundary layer conditions (vertical lines), the influence of RH is more prominent at higher BLH values (horizontal lines). This pattern points to an influence of RH on PM10 predictions, which is decoupled from the BLH, as the influence of BLH above 800 m is marginal. A similar pattern is found for all seasons except for summer. In summer, the model outcome does not show any response to changes in BLH, as there are rarely any BLH values below 800 m (see Figure D1). A study by Belle et al. (2017) conducted in the United States found RH to have positive impact on PM2.5 concentrations during cloud-free conditions due to an increase in sulfate and nitrate masses. However, they found PM2.5 to decrease with increasing RH during cloudy conditions.

The more time passed since the last precipitation event, the higher the PM10 prediction tends to be, reflecting the accumulation of local emissions in the atmosphere. The influence of this effect on the model is not pronounced and stagnates from about 100 hr (see Figure C2). The magnitude of the last precipitation event and accumulated precipitation of last 24 hr were not included in the model due to lacking importance as determined in the feature selection (see chapter 2.5.2). Note however that the low importance of precipitation could be related to the consideration of only cloud-free days in this study. Thus, the immediate effect of rainfall on particles in the atmosphere cannot be investigated. In addition, possible effects of precipitation along the trajectories of advected air masses are not covered.

3.3.6. NDVI, Corine Land Cover, and Spatiotemporal Factors

The NDVI was of minor importance for the prediction of PM10 concentrations. No trends or seasonal differences were found by application of the PD approach. However, with increasing number of pixels in the vicinity of a PM station classified as agricultural areas (2CLC), PM10 concentrations tend to be higher. Likely, this is related to primary emission of dust from arable lands and the application of fertilizers (NH3), which constitute important precursors for secondary particle formation (Hoffmann & Funk, 2015; Wagner et al., 2015). For the other land cover classes, no trends were found using the PD approach. Lower PM10 concentrations are predicted on Saturdays and Sundays, indicating that reduced anthropogenic activity (less traffic, reduced industrial production) has an immediate effect on PM10 concentrations. Increasing altitude slightly reduces the mean PM10 prediction, possibly due to lower population density and more effective pollution dispersion processes (Beloconi et al., 2018; Hu et al., 2014). The PDs of NH3, continentality, PM10 mean annual emissions, and of the TPI do not show a distinct trend. SO$_2$ was excluded from the model during the feature selection process.
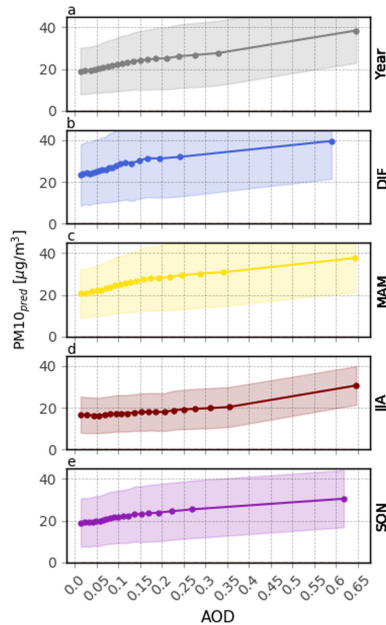
198

Figure 14. PD plot showing the mean model response to changes in AOD for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Description as in Figure 6.

3.4. Determinants of the Relationship Between AOD and PM10

The full-year model and the seasonal models associate increasing AOD with increasing PM10 (see Figure 14). This pattern is less distinct in summer (except for very high AOD) when particles are generally more dispersed within a well-mixed boundary layer, and the AOD is largely determined by particles higher up in the atmosphere, thus weakening the relation between AOD and PM10.

The relationship between AOD and PM10 is not bivariate and can be modified by ambient meteorology (Gupta & Christopher, 2009a; Guo et al., 2009; Sorek-Hamer et al., 2017; Stirnberg et al., 2018). A quantification of this effect is approached here by using the two-way PD method.

The two-way PD of AOD and BLH reveals a dependence of the model on both AOD and BLH (see Figure 15). The importance of interactive effects of these features can be illustrated by the following example: assume an AOD of 0.2 and BLH of 2,000 m versus an AOD of 0.2 and BLH of 200 m. In the latter case, the mean predicted PM10 concentration is ~10 $\mu g/m^3$ higher as the aerosol content determining AOD is closer to the ground and thus more relevant for the PM10 prediction. In other words, a prediction based on AOD (assuming that AOD is largely determined by attenuation in the boundary layer Schäfer et al., 2008) alone would lead to erroneous PM10 predictions, as AOD does not fully capture the particle accumulation effect of a shallow boundary layer (cf. Stirnberg et al., 2018). Similar effects can be observed for the two-way PD of AOD and the 3-day mean east-west wind component (see Figure 15) and for the two-way PD of AOD and
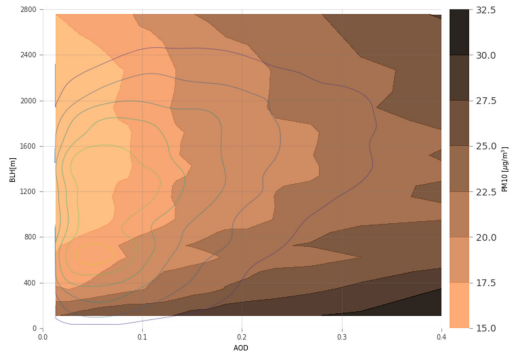
199

Figure 15. Two-way PD of AOD and BLH, full-year model. Description as in Figure 9.

temperature anomalies (plot not shown). The two-way PD of AOD and BLH shows only a minor seasonal pattern, which is mostly driven by BLH (see Figure D2)

Westerly wind flow (positive umean) leads to substantially lower PM10 predictions when compared to similar AOD values in situations dominated by easterly wind flow (negative umean). The two-way PD suggests this effect to be as large as ∼8 $\mu g/m^3$ (see Figure 16). Air masses from the east possibly carry a higher amount of near-ground particles (Beloconi et al., 2018; Bonn et al., 2016; Reizer & Juda-Rezler, 2016), affecting PM10 observations more strongly than AOD. Another reason for the observed effect could be that western air masses carry a relatively large amount of sea salts with a high hygroscopic growth factor. By effectively taking up water, these constituents enhance light scattering, thus increasing AOD without increasing PM10 measurements (Stirnberg et al., 2018; Zieger et al., 2014; Zieger et al., 2013). The seasonality for the two-way PD of AOD and umean is weak (see Figure D3).
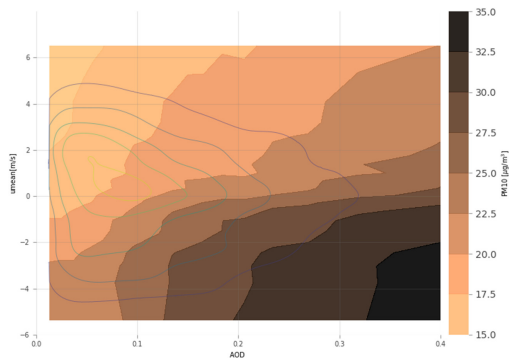

Figure 16. Two-way PD of AOD and umean, full-year model. Description as in Figure 9.
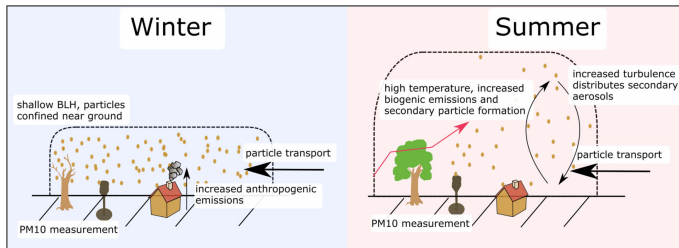
200

Figure 17. Schematic representation of different processes driving high pollution situations in winter (left) and summer (right).

4. Summary, Conclusions, and Outlook

A machine learning model is used to advance the understanding of drivers of near-ground PM10 and the capability to use satellite AOD to infer on PM10. Parameters pertaining to meteorology, land cover, and satellite-based AOD are considered and related to hourly PM10 concentrations. The model performs well (overall R^2 of 0.77, RMSE $= 7.44\,\mu g/m^3$) and provides a basis to assess sensitivities. These allow for the isolation and quantification of effects of ambient conditions on PM10. Overall, the model is more sensitive to meteorological conditions than to land cover parameters. BLH, east-west winds, DOY, temperature, and RH are identified as the important driving factors of PM10 variations. Representing regional particle transport, the 3-day mean of the east-west wind component substantially modifies PM10 concentrations,
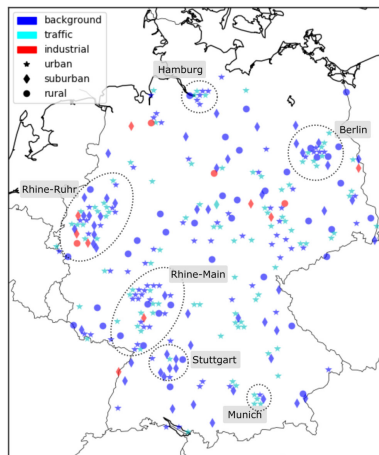

Figure A1. Spatial distribution, type, and representativeness of UBA PM measurement stations in Germany.
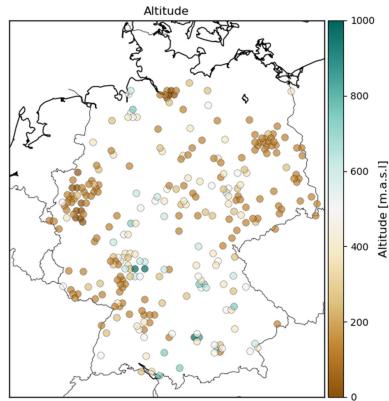
Figure A2. Altitude (m.a.s.l.) of UBA PM measurement stations in Germany.

depending on the direction of inflow. Eastern inflow generally increases PM10 concentrations. Modeled PM10 concentrations were also increased during higher than average temperatures. Possibly, this is due to stimulated vegetation activity, increasing primary particle and precursor gas emissions. The influence of BLH is most prominent at very low ($\sim$<500 m) values. However, there is indication that very high BLH values ($\sim$<2,500 m) influence PM10 concentrations as well. While the former threshold marks the effects of
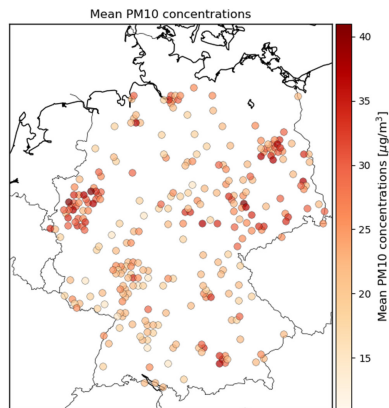

Figure A3. Mean PM10 concentrations for measurement stations used in this study. Time period is the time frame analyzed in this study (2007–2015).
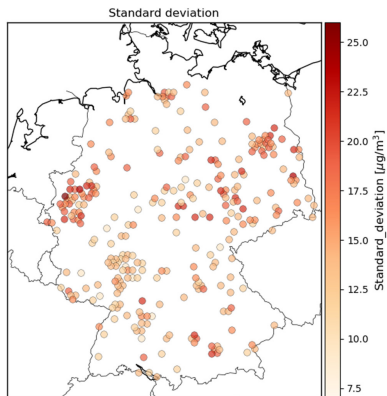
Figure A4. Standard deviation for PM10 concentrations for measurement stations used in this study. Time period is the time frame analyzed in this study (2007–2015).

particle accumulation within a shallow boundary layer, the latter threshold could indicate the formation of a deep boundary layer with stimulated formation of secondary aerosols as suggested by Grange et al. (2018). If BLH is between these thresholds, its explanatory power is limited. In these situations, other processes determine PM10 concentrations. Overall, the model outcome suggests that there are different meteorological boundary conditions that potentially cause elevated PM10 concentrations in winter and summer (Figure 17).

Figure B1. Spatial distribution of R^2.

203

Figure B2. Spatial distribution of the RMSE.

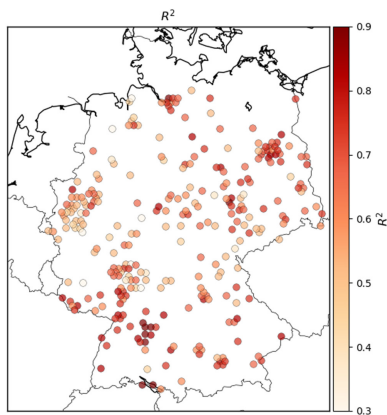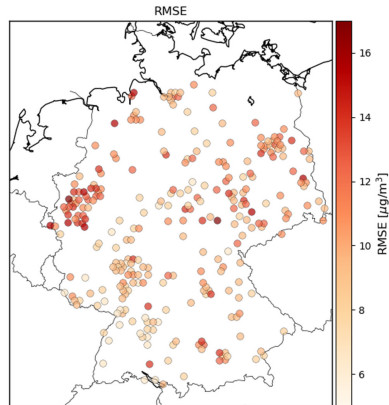In winter, very shallow boundary layers, coinciding with multiday easterly wind flow cause highest mean PM10 predictions ($>30\ \mu g/m^3$). Mean PM10 predictions for these conditions are as high as $40\ \mu g/m^3$ (see Figure 9). This is probably related to higher anthropogenic emissions in winter and frequently low BLH.

In summer, higher temperatures associated with increased formation of secondary aerosols and coinciding with multiday easterly wind flow lead to mean predicted PM10 concentrations $>27\ \mu g/m^3$ (predicted figure not shown). High PM10 concentrations in summer appear to be largely uncoupled from changes in BLH (see Figure 8d). The model R^2 decreases in summer, suggesting that the statistical model does not as well resolve these processes. In addition, the relationship between AOD and PM10 is weaker in summer.

Results presented in this study suggest that meteorology plays a substantial role in the development of high pollution situations. This has potential implications for plans toward better air quality in high-polluted areas, as meteorological conditions need to be taken into account, for example, for temporary traffic bans. In addition, there is a need to introduce measures to reduce air pollution on a regional scale. Measures limited to city scales can only decrease pollution levels associated with local emission sources, which can be superimposed by transported particles.

The importance of AOD for the statistical model highlights the suitability of AOD for air quality studies. However, potential implications and limitations for the use of satellite AOD for air quality studies are described. This study has shown that satellite-derived AOD can be used to infer street-level PM10 concentrations, if ambient meteorological conditions are taken into account explicitly. In particular, temperature anomalies, the east-west regional wind component, and BLH modify the relationship between PM10 and AOD. A drawback of including AOD is the restriction to cloud-free situations, which potentially introduces a bias due to nonrandom data gaps (Belle et al., 2017). Depending on the situation and location, both an overestimation or underestimation of PM10 could be the consequence (Belle et al., 2017). In addition, the influence of certain meteorological variables could be underestimated due to important processes under cloudy conditions, which are not covered (Belle et al., 2017; Brokamp et al., 2018). The use of GBRT proved fruitful to understand interconnected processes and the approach presented here can be potentially expanded to other research questions focusing on the understanding multivariate processes. Future efforts will further address the determination of mechanisms leading to high pollution events using machine learning not only for total PM10 concentrations but for individual aerosol species.
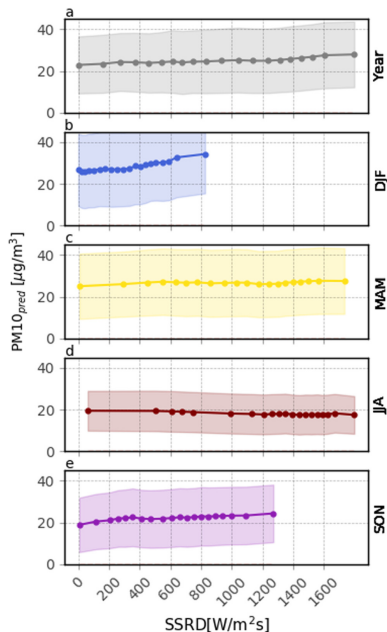
204

Figure C1. PD plot showing the mean model response to changes in SSRD (W/m^{2}s) for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Description as in Figure 6.

Appendix A: Distribution of UBA PM10 Station Types, Altitude, Mean Concentrations, and Standard Deviations of PM10 Concentrations

The spatial distribution of PM10 measurement stations is shown in Figure A1. Overall, stations are distributed relatively homogeneously over the area of Germany. The number of stations is higher in urban agglomerations. The majority of stations are classified as "urban" or "suburban." "Rural" stations are relatively rare. Most stations are labelled as representative for background conditions by the data provider. Station altitudes (m.a.s.l.) are shown in Figure A2. Altitudes range from 0 to 970 m.a.s.l.

Furthermore, mean PM10 concentrations and standard deviations of all stations for the study period 2007–2015 are shown in Figures A3 and A4, respectively. Mean concentrations are highest in the Rhine-Ruhr area (north-west) and other urban areas such as Munich to the south, Berlin to the northeast, and Hamburg to the north.

Appendix B: Spatial Distribution of Model Skill

Figures B1 and B2 show the spatial distribution of the coefficient of determination and the RMSE, respectively. In the southwest and the northeast, R^2 tends to be higher, and RMSE tends to be lower. In the
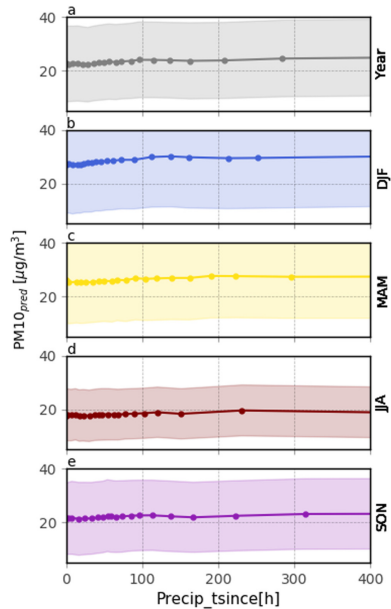
Figure C2. PD plot showing the mean model response to changes in time since last precipitation (hr) for the full-year model (a) and each season separately (b–e, DJF, MAM, JJA, and SON). Description as in Figure 6.

Rhine-Ruhr region (northwest), performance seems to be generally worse. These stations generally have high mean PM10 concentrations (see Figure A3). However, it appears that other urban areas, which also have high mean PM10 concentrations can be modeled quite well (e.g., Berlin in the northeast or Hamburg in the north).

Appendix C: Further Individual Conditional Expectation Plots

Solar radiation (Figure C1) and time since the last precipitation (Figure C2) were analyzed using the ICE method as described in chapter 2.6.3. Both input features show only minor influence on mean PM10 predictions. The ICE functions for time since last precipitation show particularly large variability of model responses of the individual data instances (shown by the shaded areas beneath the bold lines), indicating strong interactions with other features.

Appendix D: Seasonal Two-Way Partial Dependence Plots

In addition to Figure 13, which shows the full-year model two-way partial dependence of RH and BLH, Figure D1 additionally depicts the two-way partial dependence of the seasonal models. Similarly, Figures D2 and D3 show the seasonal two-way partial dependence of AOD and BLH and AOD and umean, respectively.

206

Figure D1. Two-way PD of RH and BLH, full-year model (a) and seasonal models (b–e, DJF, MAM, JJA, and SON). Description as in Figure 9.

Acronyms

AOD aerosol optical depth
BLH boundary layer height
BRF bidirectional reflectance factor
CAPE convective available potential energy
CLC Corine land cover
DEM digital elevation model

207

Further appendices

Figure D2. Two-way PD of AOD and BLH, full-year model (a) and seasonal models (b–e, DJF, MAM, JJA, and SON). Description as in Figure 9.

DOY day of year
DWD German Meteorological Service
EEA European Environmental Agency
ECMWF Centre for Medium-Range Weather Forecasts
GBRT gradient boosted regression trees
LLO leave location out
MAIAC multi-angle implementation of atmospheric correction

208

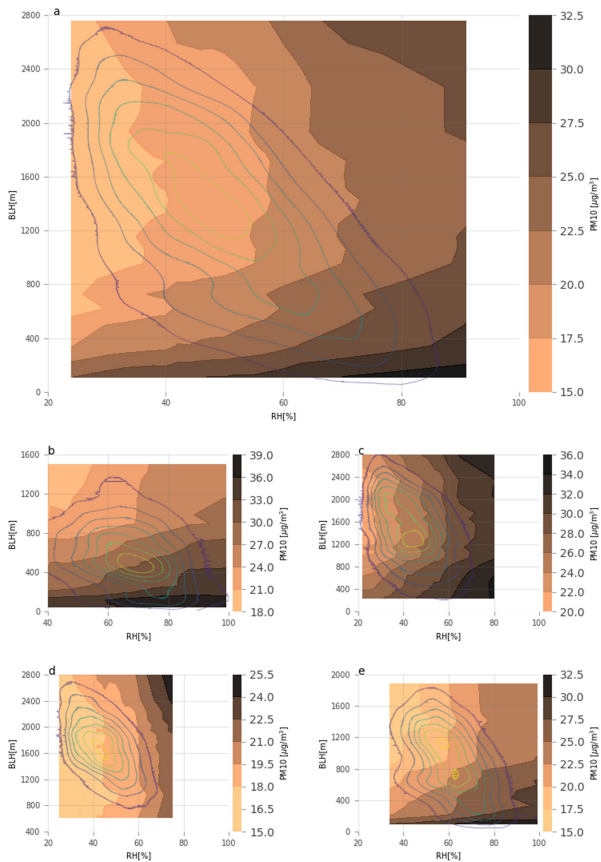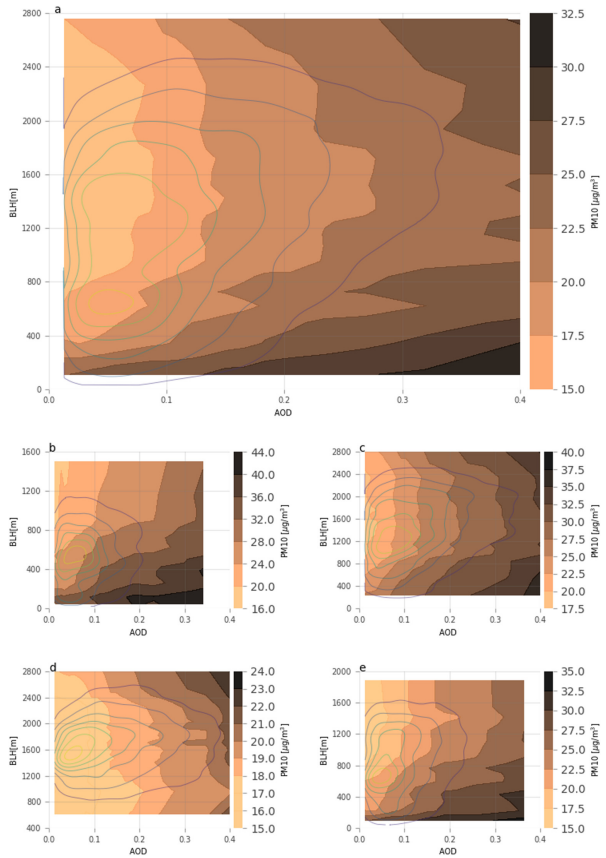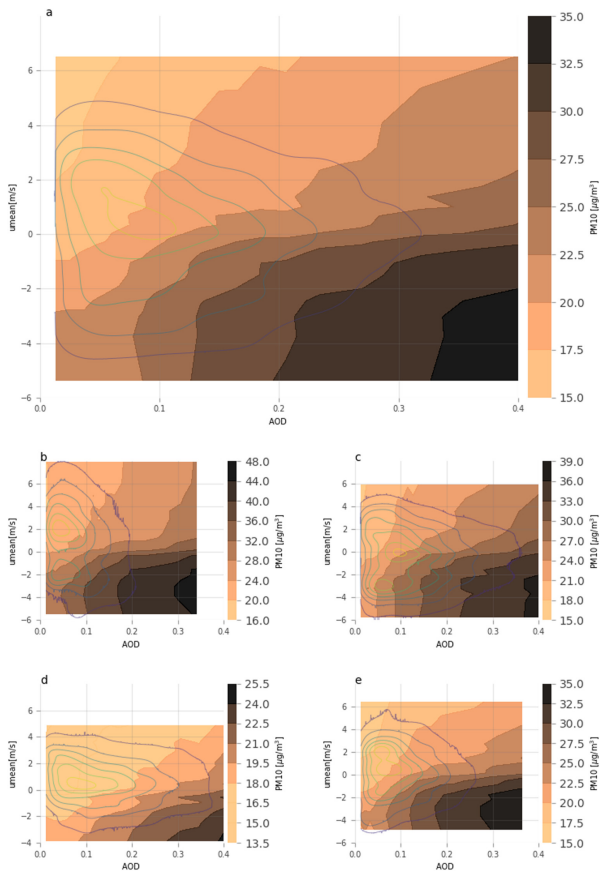**Figure D3.** Two-way PD of AOD and umean, full-year model (a) and seasonal models (b–e, DJF, MAM, JJA, and SON). Description as in Figure 9.

MODIS moderate resolution imaging spectroradiometer
NDVI normalized difference vegetation index
PDF probability density function
PM particulate matter
RADOLAN Radar-Online-Aneichung
RF random forest

RH relative humidity
RMSE root mean square error
SOA secondary organic aerosols
SSRD surface solar radiation downwards
TPI topographic position index
UBA German Environmental Agency
VIIRS Visible Infrared Imaging Radiometer Suite
SNPP Suomi National Polar-Orbiting Platform,

Conflict of Interest

The authors declare no conflict of interest.

References

Adamopoulos, A. D., Kambezidis, H. D., Kaskaoutis, D. G., & Giavis, G. (2007). A study of aerosol particle sizes in the atmosphere of Athens, Greece, retrieved from solar spectral measurements. *Atmospheric Research*, 86(3-4), 194–206. https://doi.org/10.1016/j.atmosres.2007.04.003

Andersen, H., Cermak, J., Fuchs, J., Knutti, R., & Lohmann, U. (2017). Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks. *Atmospheric Chemistry and Physics*, 17(15), 9535–9546. https://doi.org/10.5194/acp-17-9535-2017

Ansmann, A., Althausen, D., Wandinger, U., Franke, K., Müller, D., Wagner, F., & Heintzenberg, J. (2000). Vertical profiling of the Indian aerosol plume with six-wavelength lidar during INDOEX: A first case study. *Geophysical Research Letters*, 27(7), 963–966. https://doi.org/10.1029/1999GL010902

Arvani, B., Pierce, R. B., Lyapustin, A. I., Wang, Y., Ghermandi, G., & Teggi, S. (2016). Seasonal monitoring and estimation of regional aerosol distribution over Po valley, northern Italy, using a high-resolution MAIAC product. *Atmospheric Environment*, 141(June), 106–121. https://doi.org/10.1016/j.atmosenv.2016.06.037

Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., & Gerlach, N. (2004). Projekt RADOLAN Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer). Deutscher Wetterdienst. Retrieved from https://www.dwd.de/DE/leistungen/radolan/radolan_info/

Belle, J. H., Chang, H. H., Wang, Y., Hu, X., Lyapustin, A., & Liu, Y. (2017). The potential impact of satellite-retrieved cloud parameters on ground-level PM2.5mass and composition. *International Journal of Environmental Research and Public Health*, 14(10). https://doi.org/10.3390/ijerph14101244

Beloconi, A., Chrysoulakis, N., Lyapustin, A., Utzinger, J., & Vounatsou, P. (2018). Bayesian geostatistical modelling of PM10 and PM2.5 surface level concentrations in Europe using high-resolution satellite-derived products. *Environment International*, 121(August), 57–70. https://doi.org/10.1016/j.envint.2018.08.041

Birmili, W., & Wiedensohler, A. (2000). New particle formation in the continental boundary layer: Meteorological and gas phase parameter influence. *Geophysical Research Letters*, 27(20), 3325–3328. https://doi.org/10.1029/1999GL011221

Bonn, B., Von Schneidemesser, E., Andrich, D., Quedenau, J., Gerwig, H., Lüdecke, A., et al. (2016). BAERLIN2014—The influence of land surface types on the horizontal heterogeneity of air pollutant levels in Berlin. *Atmospheric Chemistry and Physics*, 16(12), 7785–7811. https://doi.org/10.5194/acp-16-7785-2016

Bossard, M., Feranec, J., & Otahel, J. (2000). CORINE land cover technical guide: Addendum 2000 (Tech. Rep. 40). European Environment Agency.

Bressi, M., Sciare, J., Ghersi, V., Bonnaire, N., Nicolas, J. B., Petit, J. E., et al. (2013). A one-year comprehensive chemical characterisation of fine aerosol (PM2.5) at urban, suburban and rural background sites in the region of Paris (France). *Atmospheric Chemistry and Physics*, 13(15), 7825–7844. https://doi.org/10.5194/acp-13-7825-2013

Brokamp, C., Jandarov, R., Hossain, M., & Ryan, P. (2018). Predicting daily urban fine particulate matter concentrations using a random forest model. *Environmental Science & Technology*, 52, 4173–4179. https://doi.org/10.1021/acs.est.7b05381

Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., & Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environment*, 151, 1–11. https://doi.org/10.1016/j.atmosenv.2016.11.066

Bundesministerium der Justiz und für Verbraucherschutz (2010). Neununddreißigste Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes (Verordnung über Luftqualitätsstandards und Emissionshöchstmengen - 39. BImSchV). Anlage 1. Bundesministerium der Justiz und für Verbraucherschutz. Retrieved from https://www.gesetze-im-internet.de/bimschv_39/anlage_1.html

Cermak, J., & Knutti, R. (2009). Beijing Olympics as an aerosol field experiment. *Geophysical Research Letters*, 36, L10806. https://doi.org/10.1029/2009GL038572

Chafe, Z., Brauer, M., Héroux, M.-E., Klimont, Z., Lanki, T., Salonen, R. O., & Smith, K. R. (2015). Residential heating with wood and coal: Health impacts and policy options in Europe and North America. Copenhagen: World Health Organization.

Chen, G. L., Guang, J., Li, Y., Che, Y. H., & Gong, S. Q. (2018). Retrieval of atmospheric particulate matter using satellite data over central and eastern China. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3, 147–153. https://doi.org/10.5194/isprs-archives-XLII-3-147-2018

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T., et al. (2018). A machine learning method to estimate PM2.5concentrations across China with remote sensing, meteorological and land use information. *Science of the Total Environment*, 636, 52–60. https://doi.org/10.1016/j.scitotenv.2018.04.251

Chudnovsky, A., Koutrakis, P., Kloog, I., Melly, S., Nordio, F., Lyapustin, A., et al. (2014). Fine particulate matter predictions using high resolution aerosol optical depth (AOD) retrievals. *Atmospheric Environment*, 89, 189–198. https://doi.org/10.1016/j.atmosenv.2014.02.019

Chudnovsky, A., Tang, C., Lyapustin, A., Wang, Y., Schwartz, J., & Koutrakis, P. (2013). A critical assessment of high-resolution aerosol optical depth retrievals for fine particulate matter predictions. *Atmospheric Chemistry and Physics, 13*(21), 10,907–10,917. https://doi.org/10.5194/acp-13-10907-2013

Churkina, G., Kuik, F., Bonn, B., Lauer, A., Grote, R., Tomiak, K., & Butler, T. M. (2017). Effect of VOC emissions from vegetation on air quality in Berlin during a heatwave. *Environmental Science & Technology, 51*, 6120–6130. https://doi.org/10.1021/acs.est.6b06514

Conrad, V. (1946). *Methods in Climatology* (pp. 296–300). Cambridge, Massachusetts: Harvard University Press.

Crumeyrolle, S., Chen, G., Ziemba, L., Beyersdorf, A., Thornhill, L., Winstead, E., et al. (2014). Factors that influence surface PM2.5 values inferred from satellite observations: Perspective gained for the US Baltimore-Washington metropolitan area during DISCOVER-AQ. *Atmospheric Chemistry and Physics, 14*(4), 2139–2153.

DWD Climate Data Center (CDC) (2017). Historical hourly station observations of 2m air temperature and humidity, version v004, 2016. (Tech. Rep.). Offenbach. Retrieved from ftp://ftp-cdc.dwd.de/pub/DWD/CDC/observations_germany/climate/hourly/air_temperature/historical/.

DWD Climate Data Center (CDC) (2018). Grids of monthly averaged daily air temperature (2m) over Germany Version v1.0. Offenbach. Retrieved from ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/hourly/air_temperature/historical/.

de Leeuw, J., Methven, J., & Blackburn, M. (2015). Evaluation of ERA-Interim reanalysis precipitation products using England and Wales observations. *Quarterly Journal of the Royal Meteorological Society, 141*(688), 798–806. https://doi.org/10.1002/qj.2395

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 137*(656), 553–597. https://doi.org/10.1002/qj.828

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. *Environmental Science & Technology, 50*(9), 4712–4721. https://doi.org/10.1021/acs.est.5b06121

Dupont, J.-C., Haeffelin, M., Badosa, J., Elias, T., Favez, O., Petit, J. E., et al. (2016). Role of the boundary layer dynamics effects on an extreme air pollution event in Paris. *Atmospheric Environment, 141*, 571–579. https://doi.org/10.1016/j.atmosenv.2016.06.061

EU (2008). Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Vol. 152). European Parliament, European Council. doi: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:152:0001:0044:EN:PDF

Egli, S., Thies, B., & Bendix, J. (2018). A hybrid approach for fog retrieval based on a combination of satellite and ground truth data. *Remote Sensing, 10*(4), 628. https://doi.org/10.3390/rs10040628

Elith, J., Leathwick, JR., & Hastie, T. (2008). A working guide to boosted regression trees, 77, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Ellison, R. B., Greaves, S. P., & Hensher, D. A. (2013). Five years of London's low emission zone: Effects on vehicle fleet composition and air quality. *Transportation Research Part D: Transport and Environment, 23*, 25–33. https://doi.org/10.1016/j.trd.2013.03.010

Emili, E., Lyapustin, A., Wang, Y., Popp, C., Korkin, S., Zebisch, M., et al. (2011). High spatial resolution aerosol retrieval with MAIAC: Application to mountain regions. *Journal of Geophysical Research, 116*, D23211. https://doi.org/10.1029/2011JD016297

Emili, E., Popp, C., Wunderle, S., Zebisch, M., & Petitta, M. (2011). Mapping particulate matter in alpine regions with satellite and ground-based measurements: An exploratory study for data assimilation. *Atmospheric Environment, 45*(26), 4344–4353. https://doi.org/10.1016/j.atmosenv.2011.05.051

Ervens, B., Turpin, B. J., Weber, R. J., Brunswick, N., & Sciences, A. (2011). Physics secondary organic aerosol formation in cloud droplets and aqueous particles (aqSOA): A review of laboratory, field and model studies, 11, 11,069–11,102. https://doi.org/10.5194/acp-11-11069-2011

Fuchs, J., Cermak, J., & Andersen, H. (2018). Building a cloud in the Southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning. *Atmospheric Chemistry and Physics, 18*, 16,537–16,552. https://doi.org/10.5194/acp-18-16537-2018

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., et al. (2015). Particulate matter, air quality and climate: lessons learned and future needs. *Atmospheric Chemistry and Physics, 15*(14), 8217–8299. https://doi.org/10.5194/acp-15-8217-2015

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics, 24*, 44–65.

Größ, J., Hamed, A., Sonntag, A., Spindler, G., & Manninen, H. E. (2018). Atmospheric new particle formation at the research station Melpitz, Germany : Connection with gaseous precursors and meteorological parameters. *Atmospheric Chemistry and Physics, 18*, 1835–1861.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., & Hueglin, C. (2018). Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmospheric Chemistry and Physics, 18*(9), 6223–6239. https://doi.org/10.5194/acp-18-6223-2018

Guo, J. P., Zhang, X. Y., Che, H. Z., Gong, S. L., An, X., Cao, C. X., et al. (2009). Correlation between PM concentrations and aerosol optical depth in eastern China. *Atmospheric Environment, 43*(37), 5876–5886. https://doi.org/10.1016/j.atmosenv.2009.08.026

Gupta, P., & Christopher, S. A. (2009a). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach. *Journal of Geophysical Research, 114*, D14205. https://doi.org/10.1029/2008JD011496

Gupta, P., & Christopher, S. A. (2009b). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *Journal of Geophysical Research, 114*, D20205. https://doi.org/10.1029/2008JD011497

Gupta, P., Christopher, S., Wang, J., Gehrig, R., Lee, Y., & Kumar, N. (2006). Satellite remote sensing of particulate matter and air quality assessment over global cities. *Atmospheric Environment, 40*(30), 5880–5892. https://doi.org/10.1016/j.atmosenv.2006.03.016

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*, Springer Series in Statistics, vol. 26. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hoffmann, C., & Funk, R. (2015). Diurnal changes of PM10-emission from arable soils in NE-Germany. *Aeolian Research, 17*, 117–127. https://doi.org/10.1016/j.aeolia.2015.03.002

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., & Liu, Y. (2017). Estimating PM2.5 concentrations in the conterminous United States using the random forest approach, 51, 6936–6944. https://doi.org/10.1021/acs.est.7b01210

Hu, X., Waller, L. A., Lyapustin, A., Wang, Y., Al-Hamdan, M. Z., Crosson, W. L., et al. (2014). Estimating ground-level PM2.5 concentrations in the Southeastern United States using MAIAC AOD retrievals and a two-stage model. *Remote Sensing of Environment, 140*, 220–232. https://doi.org/10.1016/j.rse.2013.08.032

Just, A. C., De Carli, M. M., Shtein, A., Dorman, M., Lyapustin, A., & Kloog, I. (2018). Correcting measurement error in satellite aerosol optical depth with machine learning for modeling PM2.5 in the Northeastern USA. *Remote Sensing, 10*(5), 803. https://doi.org/10.3390/rs10050803

Kiesewetter, G., Borken-Kleefeld, J., Schöpp, W., Heyes, C., Thunis, P., Bessagnet, B., et al. (2015). Modelling street level PM10 concentrations across Europe: Source apportionment and possible futures. *Atmospheric Chemistry and Physics*, 15(3), 1539–1553. https://doi.org/10.5194/acp-15-1539-2015

Kloog, I., Koutrakis, P., Coull, B. A., Lee, H. J., & Schwartz, J. (2011). Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmospheric Environment*, 45(35), 6267–6275. https://doi.org/10.1016/j.atmosenv.2011.08.066

Kloog, I., Nordio, F., Coull, B. A., & Schwartz, J. (2012). Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the mid-atlantic states. *Environmental Science & Technology*, 46(21), 11,913–11,921. https://doi.org/10.1016/j.atmosenv.2011.08.066

Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A. C., et al. (2015). Estimating daily PM 2.5 and PM 10 across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. *Atmospheric Environment*, 122(March 2016), 409–416. https://doi.org/10.1016/j.atmosenv.2015.10.004

Knüsel, B., Zumwald, M., Baumberger, C., Hirsch Hadorn, G., Fischer, E. M., Bresch, D. N., & Knutti, R. (2019). Applying big data beyond small problems in climate research. *Nature Climate Change*, 9(3), 196–202. https://doi.org/10.1038/s41558-019-0404-1

Koren, I., Remer, L. A., Kaufman, Y. J., Rudich, Y., & Martins, J. V. (2007). On the twilight zone between clouds and aerosols. *Geophysical Research Letters*, 34, L08805. https://doi.org/10.1029/2007GL029253

Laothawornkitkul, J., Taylor, J. E., Paul, N. D., & Hewitt, C. N. (2009). Biogenic volatile organic compounds in the Earth system (vol 183, pg 27, 2009). *New Phytologist*, 184(1), 276.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525(7569), 367–371. https://doi.org/10.1038/nature15371

Lelieveld, J., Klingmüller, K., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., & Münzel, T. (2019). Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions. *European Heart Journal*, 40, 1590–1596. https://doi.org/10.1093/eurheartj/ehz135

Lenschow, P., Abraham, H. J., Kutzner, K., Lutz, M., Preu, J. D., & Reichenbach, W. (2001). Some ideas about the sources of PM10. *Atmospheric Environment*, 35, S23–S33.

Li, Y., Chen, Q., Zhao, H., Wang, L., & Tao, R. (2015). Variations in PM10, PM2.5 and PM1.0 in an urban area of the Sichuan Basin and their relation to meteorological factors. *Atmosphere (Basel)*, 6(1), 150–163. https://doi.org/10.3390/atmos6010150

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., et al. (2017). Aerosol and boundary-layer interactions and impact on air quality. *National Science Review*, 4(6), 810–833. https://doi.org/10.1093/nsr/nwx117

Lim, S. S., Vos, T., Flaxman, A. D., Danaei, G., Shibuya, K., Adair-Rohani, H., et al. (2012). A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380(9859), 2224–2260. https://doi.org/10.1016/S0140-6736(12)61766-8

Liu, Q., Jia, X., Quan, J., Li, J., Li, X., Wu, Y., et al. (2018). New positive feedback mechanism between boundary layer meteorology and secondary aerosol formation during severe haze events. *Scientific Reports*, 8(1), 1–8. https://doi.org/10.1038/s41598-018-24366-3

Lyapustin, A., Martonchik, J., Wang, Y., Laszlo, I., & Korkin, S. (2011). Multiangle implementation of atmospheric correction (MAIAC): 1. Radiative transfer basis and look-up tables. *Journal of Geophysical Research*, 116, D03210. https://doi.org/10.1029/2010JD014985

Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS Collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques Discussions*, 11, 5741–5765. https://doi.org/10.5194/amt-2018-141

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., et al. (2011a). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research*, 116, D03211. https://doi.org/10.1029/2010JD014986

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., et al. (2011b). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research*, 116, D03211. https://doi.org/10.1029/2010JD014986

Megaritis, A. G., Fountoukis, C., Charalampidis, P. E., Pilinis, C., & Pandis, S. N. (2013). Response of fine particulate matter concentrations to changes of emissions and temperature in Europe. *Atmospheric Chemistry and Physics*, 13(6), 3423–3443. https://doi.org/10.5194/acp-13-3423-2013

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling & Software*, 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001

Nordio, F., Kloog, I., Coull, B. A., Chudnovsky, A., Grillo, P., Bertazzi, P. A., et al. (2013). Estimating spatio-temporal resolved PM10 aerosol mass concentrations using MODIS satellite data and land use regression over Lombardy, Italy. *Atmospheric Environment*, 74, 227–236. https://doi.org/10.1016/j.atmosenv.2013.03.043

Park, S., Shin, M., Im, J., Song, C.-k., Choi, M., Kim, J., et al. (2019). Estimation of ground-level particulate matter concentrations through the synergistic use of satellite observations and process-based models over South Korea. *Atmospheric Chemistry and Physics*, 19(2), 1097–1113. https://doi.org/10.5194/acp-19-1097-2019

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2012). Scikit-learn: Machine Learning in Python. Retrieved from http://arxiv.org/abs/1201.0490

Petäjä, T., Järvi, L., Kerminen, V.-M., Ding, A. J., Sun, J. N., Nie, W., et al. (2016). Enhanced air pollution via aerosol-boundary layer feedback in China. *Scientific Reports*, 6(1), 18898. https://doi.org/10.1038/srep18898

Petetin, H., Beekmann, M., Sciare, J., Bressi, M., Rosso, A., Sanchez, O., & Ghersi, V. (2014). A novel model evaluation approach focusing on local and advected contributions to urban PM2.5 levels—Application to Paris, France. *Geoscientific Model Development*, 7(4), 1483–1505. https://doi.org/10.5194/gmd-7-1483-2014

Petit, J. E., Favez, O., Sciare, J., Canonaco, F., Croteau, P., Močnik, G., et al. (2014). Submicron aerosol source apportionment of wintertime pollution in Paris, France by double positive matrix factorization (PMF2) using an aerosol chemical speciation monitor (ACSM) and a multi-wavelength aethalometer. *Atmospheric Chemistry and Physics*, 14(23), 13,773–13,787. https://doi.org/10.5194/acp-14-13773-2014

Pope, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, 287, 1132–1141. https://doi.org/10.1001/jama.287.9.1132

Qadir, R. M., Abbaszade, G., Schnelle-Kreis, J., Chow, J. C., & Zimmermann, R. (2013). Concentrations and source contributions of particulate organic matter before and after implementation of a low emission zone in Munich, Germany. *Environmental Pollution*, 175(2), 158–167. https://doi.org/10.1016/j.envpol.2013.01.002

Reizer, M., & Juda-Rezler, K. (2016). Explaining the high PM10 concentrations observed in Polish urban areas. *Air Quality, Atmosphere and Health*, 9(5), 517–531. https://doi.org/10.1007/s11869-015-0358-z

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., et al. (2018). NASA's Black Marble nighttime lights product suite. *Remote Sensing of Environment*, 210(November 2017), 113–143. https://doi.org/10.1016/j.rse.2018.03.017

212

Rost, J., Holst, T., Sahn, E., Klingner, M., Anke, K., Ahrens, B., & Mayer, H. (2009). Variability of PM10 concentrations dependent on meteorological conditions. *International Journal of Environment and Pollution*, *36*(March 2014), 3–18. https://doi.org/10.1504/IJEP.2009.021813

Rybarczyk, Y. (2018). Machine learning approaches for outdoor air quality modelling: A systematic review. *Applied Sciences*, *8*(12), 2570. https://doi.org/10.3390/app8122570

Schäfer, K., Harbusch, A., Emeis, S., Koepke, P., & Wiegner, M. (2008). Correlation of aerosol mass near the ground with aerosol optical depth during two seasons in Munich. *Atmospheric Environment*, *42*(18), 4036–4046. https://doi.org/10.1016/j.atmosenv.2008.01.060

Schäfer, K., Wagner, P., Emeis, S., Jahn, C., Münkel, C., Suppan, P., et al. (2012). Mixing layer height and air pollution levels in urban area. *Proceedings of SPIE*, *8534*, 1–10. https://doi.org/10.1117/12.974328

Schwarz, K., Cermak, J., Fuchs, J., & Andersen, H. (2017). Mapping the twilight zone—What we are missing between clouds and aerosols. *Remote Sensing*, *9*(6), 1–10. https://doi.org/10.3390/rs9060577

Sorek-Hamer, M., Broday, D. M., Chatfield, R., Esswein, R., Stafoggia, M., Lepeule, J., et al. (2017). Monthly analysis of PM ratio characteristics and its relation to AOD. *Journal of the Air & Waste Management Association*, *67*(1), 27–38. https://doi.org/10.1080/10962247.2016.1208121

Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., et al. (2017). Estimation of daily PM10 concentrations in Italy (2006-2012) using finely resolved satellite data, land use variables and meteorology. *Environment International*, *99*, 234–244. https://doi.org/10.1016/j.envint.2016.11.024

Stirnberg, R., Cermak, J., & Andersen, H. (2018). An analysis of factors influencing the relationship between satellite-derived AOD and ground-level PM10. *Remote Sensing*, *10*, 1353. https://doi.org/10.3390/rs10091353

Stolwijk, A. M., Straatman, H., & Zielhuis, G. A. (1999). Studying seasonality by using sine and cosine functions in regression analysis. *Journal of Epidemiology and Community Health*, *53*(4), 235–238. https://doi.org/10.1136/jech.53.4.235

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. https://doi.org/10.1186/1471-2105-8-25

TÜVRheinland (2012). Report on the suitability test of the ambient air quality measuring system TEOM 1405-DF ambient particulate monitor with PM10 pre-separator and virtual impactor of the company Thermo Fisher Scientific for the components PM10 and PM2.5: TÜV Rheinland Energie und Umwelt GmbH.

Theloke, J., Thiruchittampalam, B., Orlikova, S., Uzbasich, M., & Gauger, T. (2009). Methodology development for the spatial distribution of the diffuse emissions in Europe.

Titos, G., Jefferson, A., Sheridan, P. J., Andrews, E., Lyamani, H., Alados-Arboledas, L., & Ogren, J. A. (2014). Aerosol light-scattering enhancement due to water uptake during the TCAP campaign. *Atmospheric Chemistry and Physics*, *14*(13), 7031–7043. https://doi.org/10.5194/acp-14-7031-2014

Tucker C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150.

Umweltbundesamt (2004). Qualitätssicherungshandbuch des UBA-Messnetzes. Berlin: Federal Environment Agency. Retrieved from http://www.umweltbundesamt.de/uba-info-daten/daten/mbm

Várnai, T., Marshak, A., & Yang, W. (2013). Multi-satellite aerosol observations in the vicinity of clouds. *Atmospheric Chemistry and Physics*, *13*(8), 3899–3908. https://doi.org/10.5194/acp-13-3899-2013

VDI (2002). VDI-RICHTLINIEN: Minimum requirements for suitability tests of automated ambient air quality measuring systems Point-related measurement methods of gaseous and particulate pollutants. Retrieved from https://www.umweltbundesamt.de/sites/default/files/medien/1/dokumente/vdi_4202_1_de.pdf

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environmental Health Perspectives*, *118*(6), 847–855. https://doi.org/10.1289/ehp.0901623

van Pinxteren, D., Mothes, F., Spindler, G., Fomba, K. W., & Herrmann, H. (2019). Trans-boundary PM10: Quantifying impact and sources during winter 2016/17 in eastern Germany. *Atmospheric Environment*, *200*(March 2019), 119–130. https://doi.org/10.1016/j.atmosenv.2018.11.061

Wagner, S., Angenendt, E., Beletskaya, O., & Zeddies, J. (2015). Costs and benefits of ammonia and particulate matter abatement in German agriculture including interactions with greenhouse gas emissions. *Agricultural Systems*, *141*, 58–68. https://doi.org/10.1016/j.agsy.2015.09.003

Wagner, P., & Schäfer, K. (2017). Influence of mixing layer height on air pollutant concentrations in an urban street canyon. *Urban Climate*, *22*(May 2018), 64–79. https://doi.org/10.1016/j.uclim.2015.11.001

Wang, J., & Christopher, S. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM 2.5 mass: Implications for air quality studies. *Geophysical Research Letters*, *30*(21), 2095. https://doi.org/10.1029/2003GL018174

Wang, J., & Martin, S. T. (2007). Satellite characterization of urban aerosols: Importance of including hygroscopicity and mixing state in the retrieval algorithms. *Journal of Geophysical Research*, *112*, D17203. https://doi.org/10.1029/2006JD008078

Weigl, E. (2017). RADOLAN: Radar online adjustment. Radar based quantitative precipitation estimation products. 63067 Retrieved from https://www.dwd.de/DE/leistungen/radolan/radolan_info/radolan_poster_201711_en_pdf.pdf?__blob=publicationFile&v=2

Wichmann, H., Spix, C., Tuch, T., Wölke, G, Peters, A., Heinrich, J., et al. (2000). Daily mortality and fine and ultrafine particles in Erfurt, Germany part I: Role of particle number and particle mass. *Research report (Health Effects Institute)*, *98*(December 2000), 5–86.

Wiedensohler, A. (2000). New particle formation in the continental boundary layer´ meteorological and gas phase parameter. *Geophysical Research Letters*, *27*(20), 3325–3328.

Zhang, X., Chu, Y., Wang, Y., & Zhang, K. (2018). Predicting daily PM2.5concentrations in Texas using high-resolution satellite aerosol optical depth. *Science of the Total Environment*, *631-632*, 904–911. https://doi.org/10.1016/j.scitotenv.2018.02.255

Zheng, C., Zhao, C., Zhu, Y., Wang, Y., Shi, X., Wu, X., et al. (2017). Analysis of influential factors for the relationship between PM2.5 and AOD in Beijing. *Atmospheric chemistry and physics, Discussions*, *17*, 13,473–13,489. https://doi.org/10.5194/acp-2016-1170

Zieger, P., Fierz-Schmidhauser, R., Poulain, L., Müller, T., Birmili, W., Spindler, G., et al. (2014). Influence of water uptake on the aerosol particle light scattering coefficients of the Central European aerosol. *Tellus B: Chemical and Physical Meteorology*, *66*(1), 105462. https://doi.org/10.3402/tellusb.v66.22716

Zieger, P., Fierz-Schmidhauser, R., Weingartner, E., & Baltensperger, U. (2013). Effects of relative humidity on aerosol light scattering: Results from different European sites. *Atmospheric Chemistry and Physics*, *13*(21), 10,609–10,631. https://doi.org/10.5194/acp-13-10609-2013

213

Original manuscript: Stirnberg et al. (ready for peer review)

Stirnberg, R., Cermak, J., Kotthaus, S., Haeffelin, M., Andersen, H., Fuchs, J. Kim, M. (ready for peer review). Meteorology-driven variability of air pollution (PM1) revealed with explainable machine learning. To be submitted to Atmospheric Chemistry and Physics.

Contributions: Roland Stirnberg conceived the study concept and developed the methodology. Roland Stirnberg conducted the formal analysis including coding. Roland Stirnberg visualised and interpreted the results. Roland Stirnberg wrote the original draft.

Peer-Review: -

Usage of text passages and modifications: The manuscript is included in abbreviated form in chapter 4. Parts of the introduction of the original manuscript were moved to the chapter 1.2 and arts of the method section of the original manuscript were moved to chapter 1.3. The numbering of figures was changed and section titles were renamed. The abbreviation MLH was changed to BLH and the abbreviation Tprec was changed to Precip_tsince for consistency.

Meteorology-driven variability of air pollution (PM1) revealed with explainable machine learning

Roland Stirnberg[1,2], Jan Cermak[1,2], Simone Kotthaus[3], Martial Haeffelin[3], Hendrik Andersen[1,2], Julia Fuchs[1,2], and Miae Kim[1,2,]

[1]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Germany
[2]Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Germany
[3]Institut Pierre Simon Laplace, École Polytechnique, CNRS, Institut Polytechnique de Paris, Palaiseau, France

Correspondence: Roland Stirnberg (Roland.Stirnberg@kit.edu)

Abstract. Air pollution, in particular high concentrations of particulate matter smaller than $1\,\mu$m in diameter (PM1), continues to be a major health problem, and meteorology is known to substantially impact atmospheric PM concentrations. However, the scientific understanding of the complex mechanisms leading to high pollution episodes is inconclusive, as the effects of meteorological variables are not easy to separate and quantify. In this study, a novel, data-driven approach based on empirical

5 relationships is used to quantify drivers of atmospheric concentrations of PM1 during high pollution episodes. A tree-based machine learning model is set up to predict concentrations of speciated PM1 at a suburban site southwest of Paris, France, using meteorological variables as input features. The contributions of each meteorological feature to modeled PM1 concentrations are quantified using SHapley Additive exPlanation (SHAP) regression values. Meteorological contributions to PM1 concentrations are analysed in seasonal and high-resolution case studies. Model results suggest that winter pollution episodes

10 are mainly driven by shallow mixed layer heights (MLH), low temperatures, low wind speeds and inflow from northeastern wind directions. Contributions of MLHs to the winter pollution episodes are quantified to be on average $\sim$5 μg/m^3 for MLHs below <500 m agl, and temperatures below freezing contribute as much as $\sim$9 μg/m^3. Northeasterly winds are found to advect particles, contributing $\sim$5 μg/m^3 to total PM1 concentrations (combined effects of u- and v-wind components). However, in calm conditions (i.e. wind speeds < $\sim$2 m/s), the lack of dispersion leads to increased PM1 concentrations by $\sim$3 μg/m^3.

15 Summer peak PM1 concentrations are generally lower than in winter, and are characterised by a higher content of organics. Drivers of summer peak PM1 concentrations are temperatures above $\sim$25 °C (contributions of up to $\sim$2.5 μg/m^3), dry spells of several days (maximum contributions of $\sim$1.5 μg/m^3) and wind speeds below $\sim$2 m/s (maximum contributions of $\sim$3 μg/m^3). High-resolution case studies show a large variability of processes, which together lead to high PM1 concentrations. Processes vary even within seasons. A high pollution episode in January 2016 is shown to be driven by a drop in temperature (maximum

20 contributions of 11 μg/m^3), which enhances formation of secondary inorganic aerosols (SIA) and likely causes an increase in local wood-burning emissions. In contrast, during December 2016, high PM1 concentrations are caused mainly by a shallow MLH and particle advection. It is shown that an observed decrease in pollution levels is linked to a change in wind direction, advecting cleaner, maritime air to the PM measurement site (combined contributions of u- and v-wind-components of $\sim$-4 μg/m^3). The application of SHAP regression values within a machine learning framework presents a novel and promising

25 way of analysing observational data sets in environmental sciences.

1

Copyright statement. TEXT

1 Introduction

Air pollution has serious implications on human well-being, including damaging effects on the cardiovascular system and the lungs (Hennig et al., 2018; Lelieveld et al., 2019), and an increased number of asthma seizures (Hughes et al., 2018). This
30   includes particles smaller than 1 μm in diameter (PM1), which are associated with fits of coughing (Yang et al., 2018) and an increase in emergency hospital visits (Chen et al., 2017). The adverse health effect lead to an increase in mortality of people exposed to high particle concentrations (Samoli et al., 2008, 2013; Lelieveld et al., 2015). People living in urban areas are particularly affected by poor air quality and with ongoing urbanization, their number is projected to grow (Baklanov et al., 2016; Li et al., 2019b). These developments have motivated several countermeasures to improve air quality. Proposed efforts
35   to reduce anthropogenic particle emissions include partial traffic bans (Su et al., 2015; Dey et al., 2018) and the reduction of solid fuel use for domestic heating (Chafe et al., 2014). Although anthropogenic emissions play an important role for PM concentrations in the atmosphere, meteorological conditions related to large-scale circulation patterns as well as local-scale boundary layer processes and interactions with the land surface are major drivers of PM concentrations as well (Cermak and Knutti, 2009; Bressi et al., 2013; Megaritis et al., 2014; Dupont et al., 2016; Petäjä et al., 2016; Yang et al., 2016; Li et al.,
40   2017). It is therefore crucial to take atmospheric and environmental processes into account during the development of efficient pollution mitigation strategies. Wind speed and direction generally have a strong influence on air quality as they determine the advection of pollutants (Petetin et al., 2014; Petit et al., 2015; Srivastava et al., 2018). Limiting the vertical exchange of air masses, the mixed layer height (MLH) governs the volume of air in which particles typically disperse. Although some authors indicate that mixed layer height cannot be related directly to concentrations of pollutants and that other meteorological
45   parameters and local sources need to be considered (Geiß et al., 2017), a lower MLH can increase PM concentrations as particles are not mixed into higher atmospheric levels and accumulate near the ground (Gupta and Christopher, 2009; Schäfer et al., 2012; Stirnberg et al., 2020).

Higher MLHs in combination with high wind speeds increase atmospheric ventilation processes, thus decreasing near-surface particle concentrations (Sujatha et al., 2016; Wang et al., 2018). Air temperature can influence PM concentrations in multiple
50   ways, e.g. by modifying the emission of secondary PM precursors such as volatile organic compounds (VOCs) during summer (Fowler et al., 2009; Megaritis et al., 2013; Churkina et al., 2017), and by condensating high saturation vapour pressure compounds such as nitric acid and sulfuric acid (Hueglin et al., 2005; Pay et al., 2012; Bressi et al., 2013; Megaritis et al., 2014). The wet removal of particles by precipitation is known to be an efficient atmospheric aerosol sink (Radke et al., 1980; Bressi et al., 2013). Although all these atmospheric conditions and processes have been identified as drivers of local air quality,
55   it is usually a complex combination of meteorological and chemical processes that lead to the formation of high-pollution events (cf. Petit et al., 2015; Dupont et al., 2016; Stirnberg et al., 2020). The metropolitan area of Paris is one of the most densely populated and industrialised areas in Europe. Thus, air quality is a recurring issue and has been at the focus of many studies in the past years (Bressi et al., 2014; Petetin et al., 2014; Petit et al., 2015; Dupont et al., 2016; Petit et al., 2017;

2

Srivastava et al., 2018). Results indicate that the Paris metropolitan region is often affected by mid- to long-range transport of pollutants, as due to the city's flat terrain, an efficient horizontal exchange of air masses is frequent (cf. Bressi et al., 2013; Petit et al., 2015). High-pollution events often occur in late autumn, winter, and early spring. Often, these episodes are characterised by stagnant atmospheric conditions and a combination of local contributions, e.g. traffic emissions, residential emissions, or regionally transported particles, e.g. nitrates from manure spreading, or sulfates from point sources (Petetin et al., 2014; Petit et al., 2014, 2015; Srivastava et al., 2018). Air masses originating from continental Europe (Belgium, Netherlands, West Germany) generally enhance particle concentrations, especially of secondary inorganic aerosols (SIA, Bressi et al. (2013); Srivastava et al. (2018). Advected particle contribution has been found to be in the range of 70 % for background concentrations in Paris of particles with a diameter smaller 2.5 μm (Petetin et al., 2014). Hence the variability between high-pollution episodes in terms of timing, sources and meteorological boundary conditions is considerable (Petit et al., 2017). Previous approaches to determine meteorological drivers of air pollution included, for example, the use of chemical transport models (CTMs), which, however, require comprehensive knowledge on emission sources and secondary particle formation pathways and are associated with considerable uncertainties (Sciare et al., 2010; Petetin et al., 2014; Kiesewetter et al., 2015). Further methods rely on data exploration, e.g. the statistical analysis of time-series (cf. Dupont et al., 2016), which can be coupled with positive matrix factorization to derive PM sources (Petit et al., 2014; Srivastava et al., 2018). To take into account the interconnected nature of PM drivers, multivariate statistical approaches such as principal component analysis (PCA) have been applied (Chen et al., 2014; Leung et al., 2017). In recent years, machine learning techniques have been increasingly used to expand the analysis of PM concentrations with respect to meteorology, allowing to retrace general patterns (Hu et al., 2017; Grange et al., 2018).

Here, the multivariate and highly interconnected nature of the processes determining local PM1 concentrations is captured and analysed in a data-driven way, capturing empirical relationships between PM1 concentrations and meteorological parameters. A state-of-the-art explainable machine learning model is set up to predict PM1 concentrations, and is then used to advance our process understanding of the complex mechanisms that lead to high pollution events in Paris. Localised and individualised attributions of feature contributions are performed using SHapley Additive exPlanation regression (SHAP) values (Lundberg and Lee, 2017; Lundberg et al., 2018a, 2020). With the use of SHAP values, a detailed insight to the decisions of the statistical model can be provided, hence allowing an advancement of previous ML approaches (cf. Hastie et al., 2009; Lundberg et al., 2018a). The attribution of local (i.e. situation-based) statistical feature contributions enables quantitative estimates of drivers of PM concentrations and allows to infer on processes driving PM concentrations at high temporal resolution. Meteorological effects on speciated PM1 concentrations are quantified and the roles of the most critical atmospheric variables for driving peak particle concentrations are highlighted. The improved scientific understanding of processes is crucial for the assessment of the effectiveness of measures towards better air quality, and therefore of high value to political decision makers.

2 Data sets

Seven years (2012-2018) of meteorological and air quality data from the Site Instrumental de Recherche par Télédétection Atmosphérique (SIRTA, Haeffelin et al., 2005) supersite are the basis of this study. The SIRTA Atmospheric Observatory is
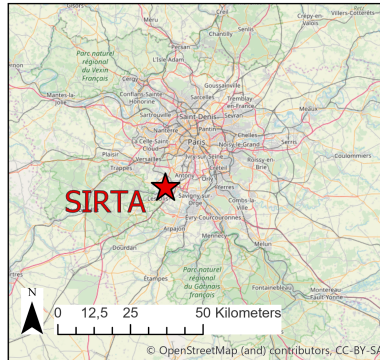
3

Figure 1. Location of the SIRTA supersite southwest of Paris. © OpenStreetMap contributors 2020. Distributed under a Creative Commons BY-SA License.

located about 25km southwest of Paris (48.713°N and 2.208°E, Fig. 1). This study focuses on day-to-day variations of total and speciated PM1. To separate diurnal effects e.g. the development of the boundary layer during morning hours (Petit et al., 2014; Dupont et al., 2016; Kotthaus and Grimmond, 2018a) from day-to-day variations of PM1, mean concentrations of total and
95 speciated PM1 for the afternoon period 12-15 UTC are considered. The diurnal cycle of PM concentrations is thus not included in the data. In sections 2.1 and 2.2, the PM1 and meteorological data and preprocessing steps before setting up the machine learning model are described. The applied machine learning model and data analysis techniques are presented in sections 3.1 and 3.2.

2.1 Submicron particle measurements

100 Aerosol chemical speciation monitorm (ACSM, Ng et al., 2011) measurements are conducted at SIRTA in the framework of the ACTRIS project. The ACSM provides continuous, near real-time and artifact-free measurements of the major chemical composition of non-refractory submicron aerosols. A detailed description of its functionality can be found in (Ng et al., 2011; Petit et al., 2014, 2015). Species measured include organics (Org), ammonium (NH4), sulfate (SO4), nitrate (NO3) and chloride (Cl). In addition, black carbon (BC) is monitored by a seven-wavelength Magee Scientific Aethalometer (model AE31, Petit
105 et al., 2015), which measures light extinction on a collecting filter. Using the multispectral information, a differentiation into fossil fuel-based BC (BCff) and BC from wood burning (BCwd) is achieved (Sciare et al., 2010; Healy et al., 2012; Petit et al., 2014). Here, the sum of all measured species is assumed to represent the total PM1 content (cf. Petit et al., 2014, 2015). The consistency of ACSM and Aethalometer measurements is checked by comparing the sum of all monitored species

4

with measurements of a nearby Tapered Element Oscillating Microbalance equipped with a Filter Dynamic Measurement System (TEOM-FDMS). While there is some inconsistency, in particular for low particle concentrations, the measurements are considered to be robust (Petit et al., 2014). As an additional input to the machine learning model, the fraction of NO3 of the previous day is added (NO3_frac). Pollution events dominated by NO3 are often linked to regional-scale events, which depend on anthropogenically-influenced processes in the source regions of NO3 precursors (Petit et al., 2017). This is approximated by the inclusion of the fraction of NO3 of the previous day.

2.2 Meteorological data

Meteorological variables included in this study are air temperature (in 2 m height, T), relative humidity (RH), air pressure (AirPres), precipitation, wind speed (ws), wind direction (u, v) and total incoming solar radiation (TISR). Data are taken from the quality-controlled and 1h averaged re-analysed observation (ReObs) dataset. Further information on the instrumentation used for the acquisition of these variables is provided in Chiriaco et al. (2018). MLH is derived from automatic lidar and ceilometer (ALC) measurements of a Vaisala CL31 ceilometer using the CABAM algorithm Characterising the Atmospheric Boundary layer based on ALC Measurements (Characterising the Atmospheric Boundary layer based on ALC Measurements, Kotthaus and Grimmond, 2018a, b). Due to an instrument failure, during the period July to mid-November 2016, SIRTA ALC measurements had to be replaced with measurements conducted at the Paris Charle de Gaulle Airport, which is located northeast of Paris.

Meteorological factors are chosen as input features for the statistical model based on findings of previous studies (cf. section 1). Meteorological observations are converted to suitable input information for the statistical model (cf. section 3.1). Wind speed (ws) is calculated from the ReObs u and v components [m/s] and the maximum wind speed of the afternoon period is included in the model. U and v wind components are then normalised to values between 0 and 1, thus only depicting the direction information. To reduce the impact of short-term fluctuation in wind direction, the 3-day mean wind direction is calculated based on the normalised u and v wind components (umean and vmean). Hours since the last precipitation event (Tprec) are counted and used as input to capture the particle accumulation effect between precipitation events (Rost et al., 2009; Petit et al., 2017).

3 Methods

3.1 Machine learning model: technique and application

Gradient Boosted Regression Trees (GBRT, used here in a python 3.6.4 environment with the scikit-learn module, Friedman, 2002; Pedregosa et al., 2012) are applied to predict daily total and speciated PM1 concentrations. As a tree-based method, GBRTs use a tree regressor, which sets up decision trees based on a training data set. The trees split the training data along decision nodes, creating homogeneous subsamples of the data by minimizing the variance of each subsample. For each sub-sample, regression trees fit the mean response of the model to the observations (Elith et al., 2008). To increase confidence in

5

140   the prediction, decision trees are combined to form an ensemble prediction. Trees are sequentially added to the ensemble (Elith et al., 2008; Rybarczyk and Zalakeviciute, 2018) and each new tree is fitted to the predecessor's previous residual error using gradient descent (Friedman, 2002). This is an advantage of GBRT over standard ensemble tree methods (e.g. Random Forests (RF), Just et al., 2018) as trees are built systematically and fewer iterations are required (Elith et al., 2008). Characteristics of the meteorological training data set with respect to observed total and speciated PM1 concentrations are conveyed to the

145   statistical model. The learned relationships are then used for model interpretation and to produce estimates of PM1 based on unseen meteorological data to test the model. The architecture of the statistical model is determined by the hyperparameters, e.g. the number of trees, the maximum depth of each tree (i.e., the number of split nodes on each tree) and the learning rate (i.e., the magnitude of the contribution of each tree to the model outcome, which is basically the step size of the gradient descent). The hyperparameters are tuned by executing a grid search, systematically validating testing previously defined hy-

150   perparameter combinations and determining the best combination via a three-fold cross validation. Note that PM1 data is not normally distributed, i.e. there is more data available for mid-range concentrations. To avoid that the model primarily optimizes its predictions on these values, a least-squares loss function was chosen. This loss function is more sensitive to higher PM1 values (i.e. outliers of the PM1 data distribution), as it strongly penalises high absolute differences between predictions and observations. Accordingly, the model is adjusted to reproduce higher concentrations as well.

155   For each PM species, a specific GBRT model is set up and used for the analysis of meteorological influences on individual PM1 species (see section 4.2). Additionally, a quasi-total PM1 model is used to predict the sum of all species at once, which is used for an analysis of meteorological drivers of high-pollution events (see sections 4.3 and 4.4). Train and test data sets to evaluate each model are created by randomly splitting the full data set. These splits, however, are the same for the species models and the full PM1 model to ensure comparability between the models. Three quarters of the data are used for training

160   and hyperparameter tuning with cross-validation (n=1086), and one quarter for testing (n=363). In addition, the robustness of the model results is tested by repeating this process ten times, resulting in ten models with different train-/ test-splits and different hyperparameters.

3.2   Explaining model decisions: SHapley Additive exPlanation (SHAP) values

While being powerful predictive models, tree-based machine learning methods also have a high interpretability (Lundberg et al.,

165   2020). In order to understand physical mechanisms on the basis of model decisions, the contributions of the meteorological input features to the model outcome are analysed. Feature contributions are attributed using SHAP values, which allow for an individualised, unique feature attribution for every prediction (Shapley, 1953; Lundberg and Lee, 2017; Lundberg et al., 2018a, 2020). SHAP values provide a deeper understanding of model decisions than the relatively widely used partial dependence plots (Hastie et al., 2009; Goldstein et al., 2015; Fuchs et al., 2018; Lundberg et al., 2018a; McGovern et al., 2019; Stirnberg et al.,

170   2020). Partial dependence plots show the global mean effect of an input feature to the model outcome, while SHAP values quantify the feature contribution to each single model prediction , accounting for multicollinearity. Feature contributions are calculated from the difference in model predictions with that feature present, versus predictions for a retrained model, without the feature. Since the effect of withholding a feature depends on other features in the model due to interactive effects between
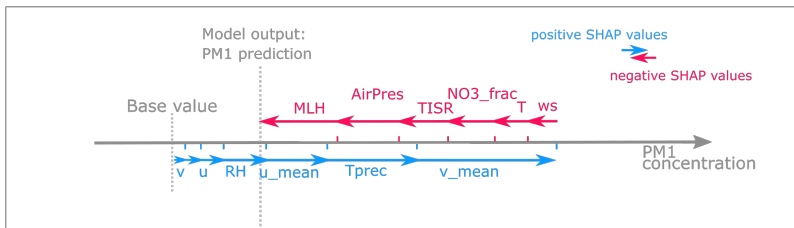
6

220

Figure 2. Conceptual figure illustrating the interaction of SHAP values and model output. Starting with a base value, which is the mean prediction if all data points are considered, positive SHAP values (blue) increase the final prediction of total and speciated PM1 concentrations, while negative SHAP values (red) decrease the prediction. The sum of all SHAP values for each input feature yields the final prediction. Depending on whether positive or negative SHAP values dominate, the prediction is higher or lower than the base value (Lundberg et al., 2018b). Adapted from https://github.com/slundberg/shap.

the features, differences are computed for all possible feature subset combinations of each data instance (Lundberg and Lee, 2017).

175

Summing up SHAP values for each input feature at a single time step yields the final model prediction. SHAP values can be negative since SHAP values are added to the base value, which is the mean prediction when taking into account all possible input feature combinations. Negative (positive) SHAP values reduce (raise) the prediction below (above) the base value. The higher the absolute SHAP value of a feature, the more distinct is the influence of that feature on the model predictions. The

180 sum of all SHAP values at one time step yields the final prediction of PM1 concentrations. An example of breaking down a model prediction into feature contributions using SHAP values is shown schematically in Fig. 2. The computation of traditional Shapley Regression values is time consuming, since a large number of all possible feature combinations have to be included. The SHAP framework for tree-based models allows a faster computation compared to full shapley regression values while maintaining a high accuracy (Lundberg and Lee, 2017; Lundberg et al., 2018a) and is therefore used here. The shap python

185 implementation is used for the computation of SHAP values (https://github.com/slundberg/shap).

Pairwise interactive effects between input features can be estimated using the SHAP approach. Interactive effects are defined as the difference between the SHAP values for one feature when a second feature is present and the SHAP values for the one feature when the other feature is absent (Lundberg et al., 2018a). SHAP values are a novel tool to better understand multivariate natural systems, in particular when applied in state-of-the-art machine learning models as GBRT. So far, SHAP values have

190 been used in the fields of computer science (cf. Antwarg et al., 2019) and medical science (Lundberg et al., 2018b; Li et al., 2019a; Lundberg et al., 2020), but have yet to be applied to study environmental systems.
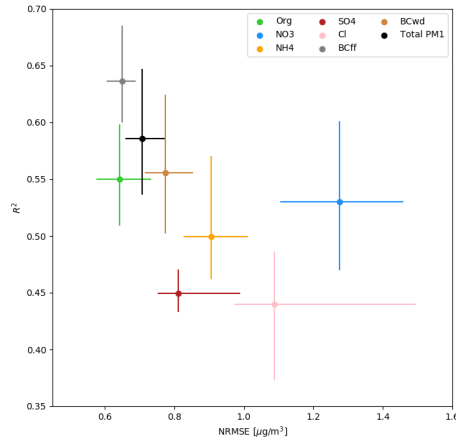
7

Figure 3. Performance indicators for ten model iterations: coefficient of determination R^2 against normalised Root Mean Squared Error (NRMSE) for the separate species models and the total PM1 model. Vertical and horizontal lines indicate the maximum spread in R^2 and NRMSE, respectively.

4 Results and discussion

4.1 Model performance

The performance of the ten model iterations is assessed by comparing the coefficient of determination (R^2) and normalised root
195   mean square error (NRSME) for independent test data (Fig. 3). While the models for BCwd, BCff and total PM1 show small spread, Cl and NO3 exhibit larger variations between model runs (indicated by horizontal and vertical lines in Fig. 3). This suggests that while drivers of BCff concentration variations are well covered by the model, this is less so in the case of Cl and NO3. Possible reasons for this are that no explicit information on anthropogenic emissions or chemical formation pathways are included in the models. Still, the model performance indicators highlight that a large fraction of the variations in particle
200   concentrations are explained by the meteorological variables used as model inputs. Performances of model iterations of the species-specific and total PM1 are generally similar, suggesting a robust model outcome.
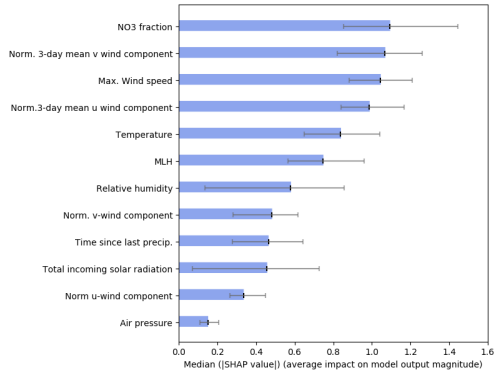
8

Figure 4. Ranked median SHAP values of the model input features, i.e. the average absolute value that a feature adds to the final model outcome, referring to the total PM1 model [$\mu g/m^3$] (Lundberg et al., 2018b). Horizontal lines indicate the variability between model runs.

The mean input feature importance, ordered from high to low, of the total PM1 model run by means of the SHAP feature attribution values is shown in Fig. 4, The NO3 fraction of the previous day has the highest impact on the model, followed by temperature, wind direction information, and MLH. Hence, the impact of these meteorological variables on model decisions is

205   analysed in more detail in the following.

4.2   Contribution of meteorological input features on modelled particle species and total PM1 concentrations

To gain insights on relevant processes governing particle concentrations in the Paris region, the contribution of input features on species and total PM1 concentration predictions from the statistical model, i.e. the SHAP values, are plotted as a function of absolute feature values (Figs 5-7). The contribution of an input feature to each (local) prediction of the species or total PM1

210   concentrations is shown while taking into account intra-model variability. Intra-model variability of SHAP values, i.e. different SHAP value attributions for the same feature value within one model, is shown by the vertical distribution of dots for absolute input feature values. Intra-model variability is caused by interactions of the different model input features.

9

4.2.1 Contribution of temperature

The impact of ambient air temperature on modelled species concentrations is highly non-linear (Fig. 5). All species show
215 increased contributions to predictions at temperatures below ∼4 °C while the contribution of high temperatures on model pre-
dictions differs substantially between species. The statistical model is able to reproduce well-known characteristics of species
concentration variations related to temperature. For example, sulfate formation is enhanced with increasing temperatures (Fig.
5d) due to an increased oxidation rate of SO_2 (cf. Dawson et al., 2007; Li et al., 2017) and strong solar irradiation due to
photochemical oxidation (Gen et al., 2019). Dawson et al. (2007) reported an increase of 34 ng/m^{3}K for PM2.5 concentrations
220 using a CTM. The increase in sulfate at low ambient temperatures is not reported in this study. It is likely linked to increased
aqueous phase particle formation in cold and foggy situations (Rengarajan et al., 2011; Petetin et al., 2014; Cheng et al., 2016).
Considerable local formation of nitrate at low temperatures (Fig. 5b) is consistent with results from a previous study in the area
by Petetin et al. (2014). Enhanced formation of ammonium nitrate at lower temperatures (Fig. 5c) by the shifting gas-particle
equilibrium has been previously reported by Clegg et al. (1998); Pay et al. (2012); Bressi et al. (2013); Petit et al. (2015); **?**.
225 The increase in organic matter and BCwd concentrations at low temperatures (Fig. 5g) is likely related to human behaviour,
as biomass burning is often used for domestic heating in the study area (Favez et al., 2009; Sciare et al., 2010; Healy et al.,
2012; Jiang et al., 2019). In addition, organic matter concentrations are linked to the condensation of semi-volatile organic
species at low temperatures (Putaud et al., 2004; Bressi et al., 2013). The sharp increase in modelled concentrations of organics
above 25°C (Fig 5a) could be due to enhanced biogenic activity leading to a rise in biogenic emissions and secondary aerosol
230 formation (Guenther et al., 1993; Churkina et al., 2017; Jiang et al., 2019).

The contribution of temperature on modelled total PM1 concentrations (Fig. 6h) is consistent with the response patterns to
changes in temperatures described for the individual species in panels 6a-6g, with positive contributions at both low (<4 °C)
and high air temperatures (>25 °C). For temperatures below freezing, the model allocates maximum contributions to modelled
total PM1 concentrations of up to 12 μg/m^3. The spread of SHAP values between model iterations is generally higher for low
235 temperatures (vertical grey bars in Fig. 6), where absolute values are of greater magnitude, but in all cases the signal contained
in the feature contributions far exceeds the spread between model runs. The inter-model variability of the feature contribution
depends to some extent on the distribution of the different training data sets. This effect is particularly pronounced for BCff,
which shows very high variability at low temperatures. As BCff concentrations depend on anthropogenic emissions, BCff
predictions highly depend on the diversity of emissions included in the training data.

4.2.2 Contribution of MLH
240

Variations in MLH can have a substantial impact on near-surface particle concentrations, as the mixed layer is the atmospheric
volume in which the particles are dispersed (cf. Klingner and Sähn, 2008; Dupont et al., 2016; Wagner and Schäfer, 2017). The
effect of MLH variations on modelled particle concentrations is highly nonlinear for all species (Fig. 6), with the magnitude
of the contribution varying by species. Possible reasons for this will be discussed in the following. Similar to the pattern
245 observed for temperature SHAP values, the inter-model variation of predictions is highest for low MLHs since predicted
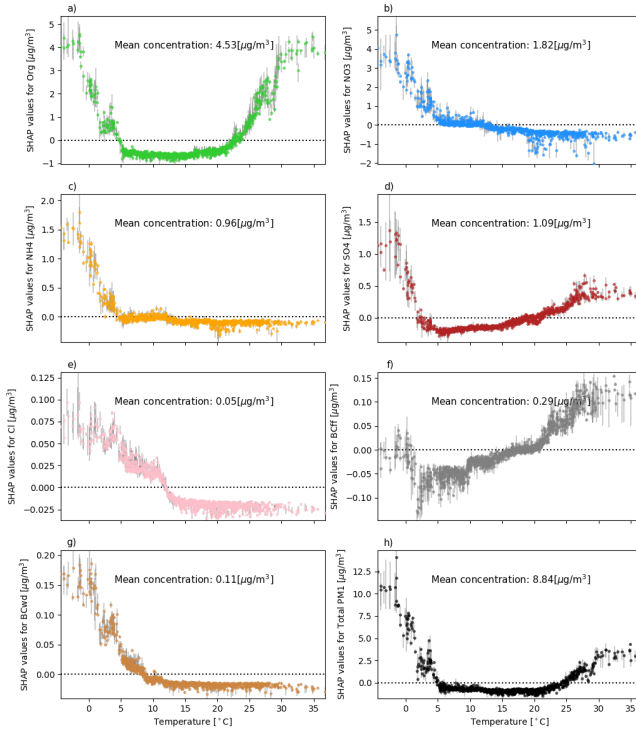
10

Figure 5. Air temperature SHAP values (contribution of temperature to the prediction of species and total PM1 concentrations [$\mu g/m^3$] for each data instance) vs. absolute air temperature [$^\circ C$]. Inter-model variability of allocated SHAP values is shown as the variance of predicted values between the ten model iterations and plotted as vertical grey bars. The dotted horizontal line indicates the transition from positive to negative SHAP values.

particle concentrations then have the highest variation. Contributions of MLH to predicted particle concentrations are highest for very shallow mixed layers due to the accumulation of particles close to the ground under shallow MLH conditions (Dupont

et al., 2016; Wagner and Schäfer, 2017). A relatively distinct transition from positive contributions during shallow boundary layer conditions ($\sim$0-800 m) towards negative contributions at high MLHs is evident for all species except SO4. Modelled SO4 concentrations show a less distinct response to changes in MLH as they are largely driven by point sources and thus susceptible to particle advection, independent of MLH (Pay et al., 2012; Petit et al., 2014, 2015). The accumulation effect under low MLH conditions is hence less important. Furthermore, an increase of SO4 concentrations with higher MLHs ($> \sim 1500$ m agl) could be due to a more effective transport of SO4 and its precursor SO2 under high MLH conditions (cf. Pay et al., 2012).

In agreement with results from previous studies focusing on PM10 (Grange et al., 2018; Stirnberg et al., 2020) or PM2.5 (Liu et al., 2018), SHAP values do not change much for MLH above $\sim$800-900 m, i.e. boundary layer height variations above this level do not strongly influence submicron particle concentrations. Positive contributions for MLHs above $\sim$800-900 m have been previously reported by Grange et al. (2018), who relate this pattern to enhanced secondary aerosol formation in a very deep and dry boundary layer. The positive contributions of high MLHs on SO4 and organics could be explained following this argumentation. Processes driving the positive contribution of high MLHs on BCff, however, remain inconclusive. For predicted total PM1 concentrations, the maximum positive contribution of the MLH is as high as 5.5 μg/m^3 while negative contributions can amount to -2 μg/m^3. While the maximum contribution by MLH is lower than the maximum contribution determined for air temperature, the frequency of shallow MLH is far greater than that of the minimum temperatures that have the largest effect (Figs 5d & 6d). In addition to causing particles to accumulate near the surface, low MLH can also provide effective pathways for local new particle formation. Secondary pollutants, such as nitrate and ammonium, are increased at low MLHs as conditions favorable to the formation of nitrate usually coincide with shallow MLH conditions (i.e., low temperatures, often in combination with high RH, Pay et al., 2012; Petetin et al., 2014; Dupont et al., 2016; Wang et al., 2016). BC concentrations, however, are mainly associated with primary emissions, as is a substantial fraction of organic matter (Petit et al., 2015). Hence, the accumulation of these particles during low turbulence conditions can explain the strong influence of MLH on BCwd and BCff.

4.2.3 Contribution of wind direction

To analyse the contribution of wind direction to predicted particle concentrations, SHAP values of normalised 3-day mean u and v wind components were added up and transformed to degrees (Fig. 7). Generally, contributions of wind are highest when winds from the northern to northeastern sectors prevail, while negative values are evident for southwesterly directions. Given the location of the measurement site, this pattern undoubtedly reflects the advection of particles emitted in the urban area of Paris and from continental Europe versus cleaner marine air masses during southwestern inflow (Bressi et al., 2013; Petetin et al., 2014; Petit et al., 2015; Srivastava et al., 2018). Increased concentrations of organic matter are predicted for northerly, northeasterly and easterly winds. These patterns suggest a significant contribution of advected organic particles from a specific wind sector. This is in agreement with the findings of Petetin et al. (2014) who estimated that 69 % of the PM25 organic matter fraction is advected by northeasterly winds, which is related to advected particles from wood burning sources in the Paris area and SOA formation along the transport trajectories. While Petit et al. (2015) did not find a wind direction dependence of organic matter measured at SIRTA using wind regression, they reported the regional background of organic matter to be of
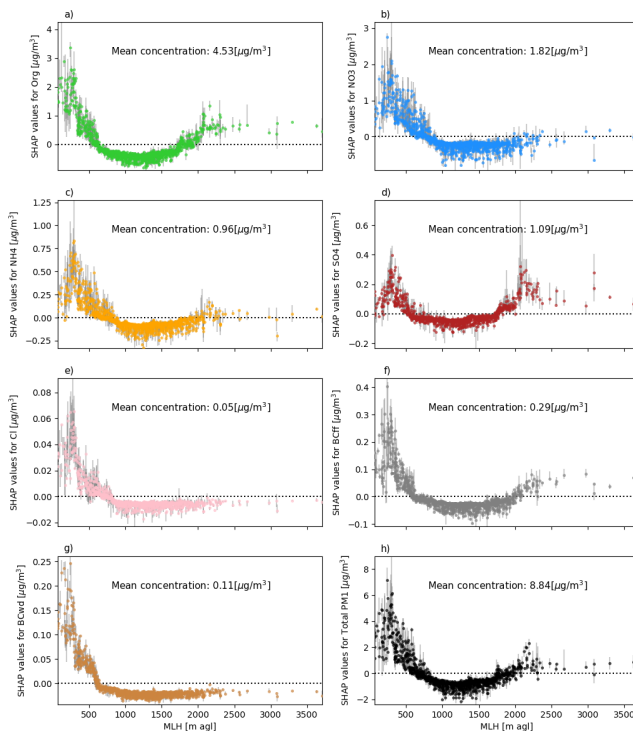
<div align="center">12</div>

Figure 6. As Fig. 5 for MLH SHAP values (contribution of MLH to the prediction of species and total PM1 for each data instance) vs. absolute MLH values [m agl].

importance. Comparing upwind rural stations to urban sites, Bressi et al. (2013) concluded organic matter is largely driven by mid- to long-range transport. Contributions to the SO4-model are highest for northeastern and eastern wind direction, which aligns with previous findings by Pay et al. (2012); Bressi et al. (2014); Petit et al. (2017), who identified the Benelux region

285 and western Germany as strong emitters of sulfur dioxide (SO2). SO2 can be transformed to particulate SO4 (cf. Pay et al.,

13

227

2012) while being transported towards the measurement site. A further possible source is the Orly Airport to the east of SIRTA as jet exhausts also contain SO2 (cf. Kapadia et al., 2016). Nitrate concentrations are affected by long-range transport from continental Europe (Benelux, western Germany), which are imported towards SIRTA from northeastern directions (Petetin et al., 2014; Petit et al., 2014). It is to be expected that the influence of mid- to long-range transport on the particle observations

290 at SIRTA is rather substantial, with most high pollution days affected by particle advection from continental Europe (Bressi et al., 2013). As shown by Petetin et al. (2014), SIA species like NO3, SO4 or NH4 are rarely formed within the Paris city area, but advected from further away, at least during the spring and summer months. Concerning BCff and BCwd, model results suggest a dependence on wind direction during northwestern to northeastern inflow. Although BC concentrations are expected to be largely determined by local emissions (Bressi et al., 2013), e.g. from local residential areas, a substantial contribution of

295 imported particles from wood burning and traffic emissions from the Paris region (Laborde et al., 2013; Petetin et al., 2014) and continental sources is likely (Petetin et al., 2014).

4.2.4 Contribution of feature interactions

Pairwise interactive effects between MLH vs. time since last precipitation and MLH vs. maximum wind speed are investigated in Figs 8a and 8b. SHAP interaction effects between MLH and time since last precipitation are most pronounced for MLHs

300 below ~ 500 m agl (Fig. 8a). Interaction values are negative for low MLHs paired with time since last precipitation close to zero hours. With increasing time since last precipitation, interaction effects become positive, thus increasing the contribution of Tprec and MLH to the model outcome. An explanation of this pattern concerning underlying processes could be that due to the lack of precipitation, a higher number of particles is available in the atmosphere for accumulation, hence increasing the accumulation effect of a shallow MLH. In case of recent precipitation, the accumulation effect of a shallow MLH is weakened.

305 For higher MLHs, interactive effects with time since the last precipitation event are marginal. Interactive effects between MLH and wind speed are shown in Fig. 8b. Positive SHAP values for maximum wind speeds below ~ 2 m/s reflect stable situations, favoring the accumulation of particles, whereas high wind speeds enhance the ventilation of particles (Sujatha et al., 2016). This can also be deduced from Fig. 8b, which shows increased SHAP values for low wind speeds in combination with a low MLH. Low wind speeds combined with a high MLH ($> \sim 1000$ m agl), on the other hand, result in decreased SHAP values.

310 Similarly, low MLHs combined with higher wind speeds ($> \sim 2$ m/s) also decrease predictions of total PM1 concentrations.

4.3 Meteorological conditions of high-pollution events

To further identify conditions that favor high pollution episodes, the data set is split into situations with exceptionally high total PM1 concentrations (>95th percentile) and situations with typical concentrations of total PM1 (interquartile range, IQR) for the meteorological summer and winter seasons. Mean SHAP values refer to the total PM1 model, corresponding input feature

315 distributions and species fractions for the two subgroups are aggregated seasonally. This allows for a quantification of seasonal feature contributions to average or polluted situations. Figs 9 & 10 show mean SHAP values for typical (left) and high-pollution (right) situations in the upper panel. The distribution of SHAP values are shown as box plots for each feature. Absolute feature value distributions are given in the bottom of the figure. In the lowest subpanel, the chemical composition of the total PM1
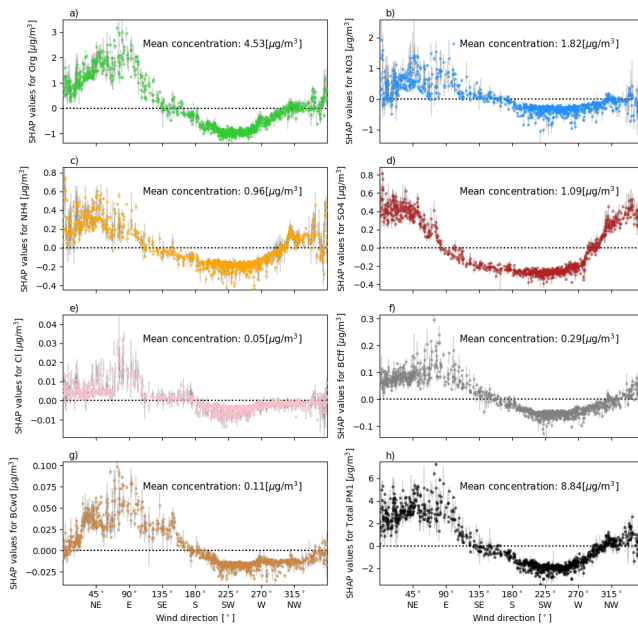
14

Figure 7. As Fig. 5 for wind direction SHAP values (contribution of 3-day mean wind direction to the prediction of species and total PM1 for each data instance) vs. absolute wind direction [°].

concentration for each subgroup is shown. The largest contributor to high pollution situations in winter is air temperature
320   (Fig. 9). SHAP values for temperature are substantially increased during high pollution situations, when temperatures are
systematically lower. Further contributing factors to high pollution situations are the MLH, wind speed, NO3 fraction of the
previous day and the wind direction components. In winter, the PM1 composition shows a relatively large fraction of nitrates,
which is increased during high pollution situations (Fig. 9, lower panel). High concentrations of nitrate in winter can be linked
to advection or to enhanced formation due to the temperature-dependent low volatility of ammonium nitrate (Petetin et al.,
325   2014). The organic matter fraction is slightly decreased during high pollution situations. MLH and maximum wind speed
contributions to high pollution situations are linked to low ventilation conditions which are very frequent in winter (cf. (cf.
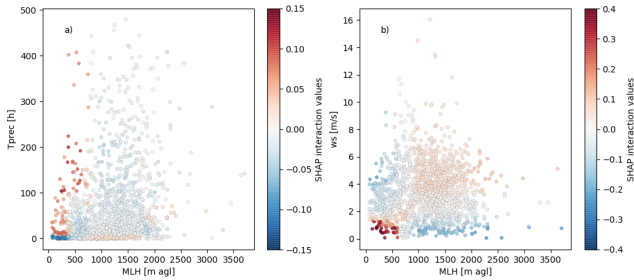
15

Figure 8. MLH vs. a) time since last precipitation and MLH vs. b) maximum wind speed, respectively, colored by the SHAP interaction values for the respective features.

Dupont et al., 2016). Positive contributions of wind direction for inflow from the northern and eastern sectors are dominant during high pollution situations while inflow from the southern and western sectors prevails during average pollution situations (see Fig. 7, Bressi et al., 2013; Petetin et al., 2014; Srivastava et al., 2018). Note that the time since the last precipitation is increased during high pollution situations, but the effects on the model outcome is weak. This suggests that lacking precipitation is not a direct driver of modelled total PM1 concentrations, but increases the contribution of other input features (see Fig. 8a) or is a meaningful factor in only some situations.

Summer total PM1 composition (Fig. 10) is characterised by a larger fraction of organics compared to the winter season (Fig. 9). As a considerable fraction of organic matter is formed locally (Petetin et al., 2014), the increased proportion of organics could be due to more frequent stagnant synoptic situations that may limit the advection of transported SIA particles. In addition, the positive SHAP values of solar irradiation and temperature highlight that the solar irradiation stimulates transformation processes and increases the number of biogenic SOA particles (Guenther et al., 1993; Petetin et al., 2014). As mean temperatures are highest in summer, positive temperature SHAP values are associated with increased organic matter concentrations (Fig. 5). The higher importance (i.e. higher SHAP values) of time since the last precipitation event during high pollution situations points to an accumulation of particles in the atmosphere. Dry situations can also enhance the emission of dust over dry soils (Hoffmann and Funk, 2015). The negative contribution of MLH during both typical and high pollution situations reflects seasonality, as afternoon MLHs in summer are usually too high to have a substantial positive impact on total PM1 concentrations (see Fig. 6). MLH is thus not expected to be a driver of day-to-day variations of summer total PM1 concentrations. Note that the average MLH is higher during high pollution situations, which likely points to increased formation of SO4 (see Fig. 6).
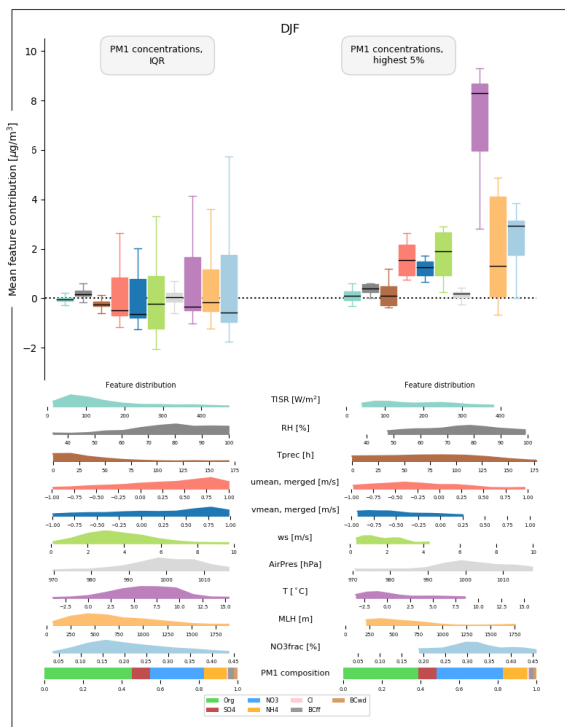
16

230

Figure 9. Mean feature contributions (i.e., SHAP values) for situations with low total PM1 concentrations (left) and situations with high pollution (right), respectively, during winter (December, January, February). Respective range of SHAP values by species are shown as box plots, with median (bold line), 25-75th percentile range (boxes), and 10-90th percentile range (whiskers). Both training and test data are included. Absolute feature value distributions (given as normalised frequencies) as well as the chemical composition of the total PM1 concentration are shown in the subpanels. Colors of the box plots correspond to colors in the feature distribution subpanels. SHAP values of the input features u_norm_3d and u_norm as well v_norm_3d and v_norm were merged to u_norm, merged and v_norm, merged to achieve better transparency.
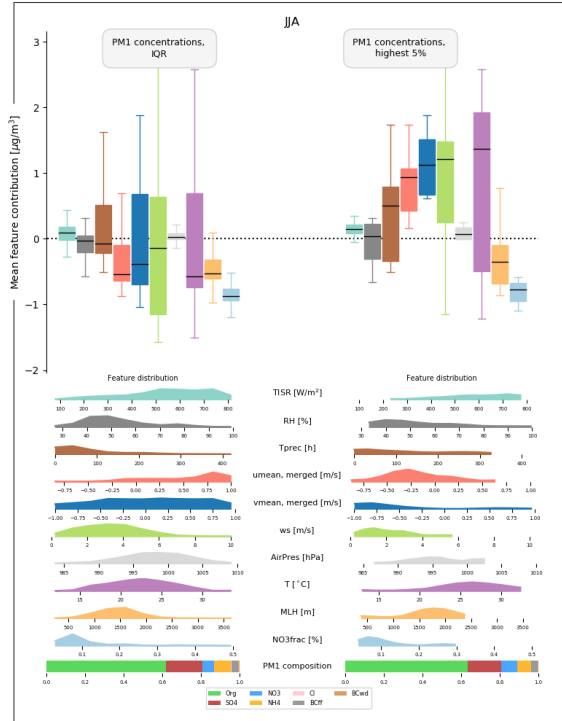
17

Figure 10. As Fig. 9 for mean feature contributions (i.e., SHAP values) for situations with low total PM1 concentrations (left) and situations with high pollution (right), respectively, during summer (July, June, August).

345   **4.4   Day-to-day variability of selected high pollution events**

Analysing the combination of SHAP values of the various input features on a daily basis allows for direct attribution of the respective implications for modelled total PM1 concentrations (cf. Lundberg et al., 2020). Here, four particular pollution episodes are selected to analyse the model outcome with respect to physical processes (Figs 11-14). The examples highlight the

18

advantages but also the limitations of the interpretation of the statistical model results. The high pollution episodes took place
in winter 2016 (10th - 30th January and 25th November - 25th December), spring 2015 (11th - 31st March) and summer 2017
(8th - 28th June). The upper panels in Figs 13-16 indicate the total PM1 prediction as horizontal black line with vertical black
lines denoting the range of predictions of all 10 models. The observed species concentrations are shown in the corresponding
colors. The subsequent panels show absolute values and SHAP values for the most relevant meteorological input features.

4.4.1 January 2016

Prior to the onset of the high-pollution episode in January 2016 (Fig. 11), the situation is characterised by MLHs in the range
of 1000m, temperatures above freezing ($\sim$5-10°C), frequent precipitation and winds from the southwest. The organic matter
fraction dominates the particle speciation. The episode itself is reproduced well by the model. According to the model results,
the event is largely temperature-driven, i.e. SHAP values of temperature explain a large fraction of the total PM1 concentra-
tion variation (note the adjusted y-axis of the temperature SHAP values). On 18th January, temperatures drop below freezing,
coupled with a decrease in MLH. As a consequence, both modelled and observed PM1 concentrations start to rise. A further
increase in total PM1 concentrations is driven by a sharp transition from stronger southwestern to weaker northeastern winds
(strong negative u component, weak negative v component) on January 19th. The combined effects of these changes lead to
a marked increase in total PM1 concentrations, peaking at $\sim$37 μg/m^3 on 20th January. On the following days, temperatures
increase steadily, thus the contribution of temperature decreases. At the same time, although values of MLH remain almost
constant, the contribution of MLH drops substantially from $\sim$5 μg/m^3 to $\sim$2 μg/m^3. This is due to interactive effects between
MLH and the features wind speed, time since last precipitation and normalised v-wind-component. All of these features in-
crease the contribution of MLH on 20th January, but decrease its contribution on 21st-23rd January. The high pollution episode
ceases after a shift to southeastern winds and the increasing temperatures. The pollution episode is characterised by a relatively
large fraction of NO3 and NH4, which explains the strong feature contribution of temperature to the modeled total PM1 con-
centration, as the abundance of these species is temperature dependent (see Fig. 5) and points to a large contribution of locally
formed inorganic particles. Still, the contribution of wind direction and speed also suggests that advected secondary particles
and their build-up in the boundary layer are relevant factors during the development of the high pollution episode (Petetin et al.,
2014; Petit et al., 2014; Srivastava et al., 2018).

4.4.2 December 2016

A high-pollution episode with several peaks of total PM1 is observed in November and December 2016. The first peak on 26th
December is followed by an abrupt minimum in total PM1 concentrations on 28th November, and a build-up of pollution in a
shallow boundary layer towards the second peak on 2nd December with total PM1 concentrations exceeding 40 μg/m^3. In the
following days, total PM1 concentrations continuously decrease, eventually reaching a second minimum on 11th December.
A gradual increase in total PM1 concentrations follows, resulting in a third (double-)peak total PM1 concentration on 17th
December. Total PM1 concentrations drop to lower levels afterwards. Throughout the 3.5 week-long episode, high pollution is
largely driven by shallow MLH (<$\sim$500m), and weak north-northeasterly winds, i.e. a regime of low ventilation and advection
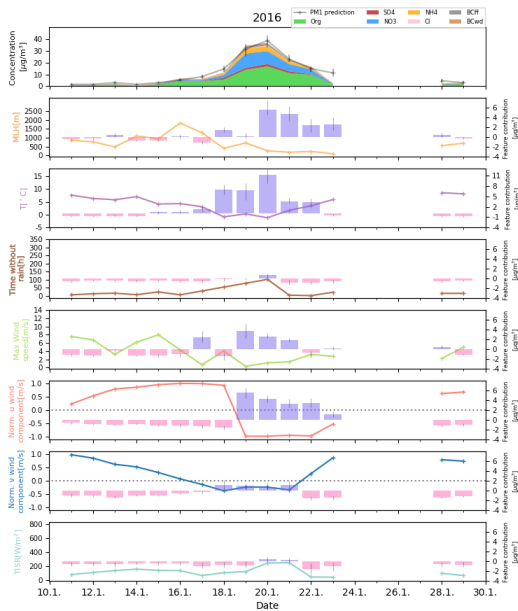
19

Figure 11. Winter pollution episode in January 2016. (a): predicted total PM1 and observed PM1 species concentrations, with absolute input feature values and corresponding SHAP values of (b) MLH, (c) temperature, (d) hours after rain, (e) maximum wind speed (f) normalised u wind and (g) normalised v wind component.

of continental air. During the brief periods with lower total PM1 concentrations, these conditions are disrupted by a higher MLH (∼28th November), or a change in prevailing winds (∼11th December). In contrast to the pollution episode in January 2016, this December 2016 episode is not driven by temperature changes. Temperatures range between ∼5-12°C and have a 385 minor contribution to predicted total PM1 concentrations (see also Fig. 5), emphasizing the different processes causing air
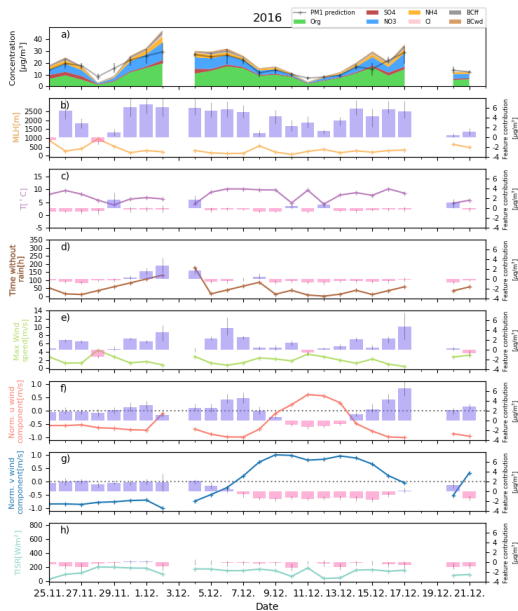
20

Figure 12. As Fig. 11 for a further winter pollution episode in December 2016.

pollution in the Paris region. The large contributions of wind direction and speed to the modeled total PM1 concentration highlight the importance of advected particles during this period that continuously build-up in the shallow boundary layer to produce these highly polluted conditions. Note that the model is not able to fully reproduce the pollution peak on December 2nd, which may be indicative of missing input features in the model. Judging from the PM1 species composition during this time (relatively high fraction of NO3 and BC), it seems likely that missing information on particle emissions may be the reason for the difference between modeled and observed total PM1 concentration.

4.4.3 June 2017

A period of above average total PM1 concentrations occurred in June 2017. The episode is very well reproduced by the model, suggesting a strong dependence of the observed total PM1 concentration to meteorological drivers. Although absolute total
395 PM1 concentrations are substantially lower than during the previously described winter pollution episodes, the event is still above average for summer pollution levels. Organic matter particles dominate the PM1 fraction throughout the episode, with a relatively high SO4 fraction. Conditions during this episode are characterised by strong solar irradiation (positive SHAP values) and high MLHs (mostly negative SHAP values), which show low day-to-day variability and reflect characteristic summer conditions. A lack of precipitation (no rain for a period of more than 2 weeks) and high temperatures also contribute
400 to the total PM1 concentrations during this episode. While solar irradiation and time since last precipitation are associated with positive SHAP values throughout this period, air temperature only has a positive contribution when exceeding $\sim$25 °C. This aligns with patterns shown in Fig. 5, where increased concentrations of organic matter and SO4 are identified for high temperatures. Peak total PM1 concentrations of $\sim$17 μg/m^3 are observed on June 20th and 21st. A change in the east-west wind component from western to eastern inflow directions in conjunction with an increase in temperatures to above 30 °C
405 are the drivers of the modeled peak in total PM1 concentrations. MLH is also increased with values $\sim$2000 m agl, which are associated with slightly positive SHAP values. This observation fits with findings described in section 4.2.2 and is likely linked to enhanced secondary particle formation (Megaritis et al., 2014; Jiang et al., 2019). As suggested by response patterns of species to changes in MLH shown in Fig. 7, this effect is linked to an increase in SO4 concentrations. The main fraction of the peak total PM1 values, however, is linked to an increase in organic matter concentrations due to the warm temperatures (cf.
410 Fig. 5).

4.4.4 March 2015

High particle concentrations are measured in early March 2015 with high day-to-day variability. This modelled course of the pollution episode is chosen to compare results to previous studies focusing on the evolution of this episode (Petit et al., 2017; Srivastava et al., 2018). The episode is of particular interest, since it also shows some limitations of the statistical model setup.
415 The episode is characterised by high fractions of SIA particles, in particular SO4, NH4 and NO3 (Fig. 14, upper panel) and similar concentrations observed at multiple measurement sites in France (Petit et al., 2017). Contributions of local sources are low and much of the episode is characterised by winds blowing in from the northwest, advecting aged SIA particles towards SIRTA (Petit et al., 2017; Srivastava et al., 2018). A widespread scarcity of rain probably enhanced the large-scale formation of secondary pollution across western Europe (in particular western Germany, The Netherlands, Luxemburg, Petit et al., 2017),
420 which were then transported towards SIRTA. This is reflected by the SHAP values of the u and v wind components, which are positive throughout the episode (see Fig. 14g & 14h). Concentration peaks of total PM1 are measured on 18th and 20th March. Both peaks are characterised by a rapid development of total PM1 concentrations. As described in Petit et al. (2017), these strong daily variations of total PM1, which are mainly driven by the SIA fraction, could be due to varying synoptic cycles, especially the passage of cold fronts. The contribution of MLH and temperature is relatively small, which is consistent with the
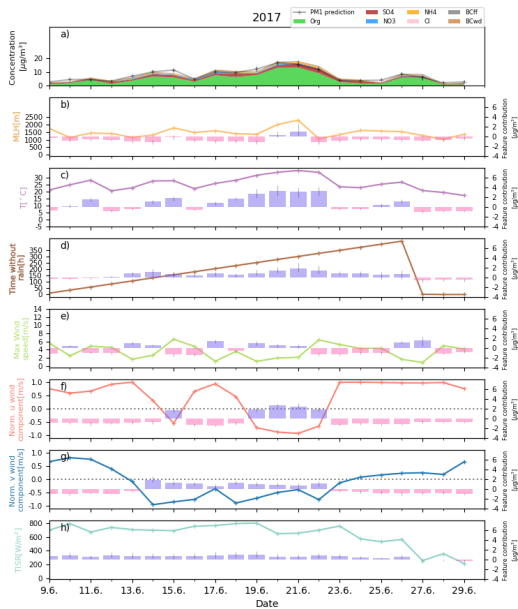
<div align="center">22</div>

236

Figure 13. As Fig. 11 for an exemplary summer pollution episode in June 2017.

high influence of advection on total PM1 concentrations during the episode. The exceptional character of the episode (cf. Petit
et al., 2017) partly explains the bad performance of the model in capturing total PM1 variability during the event. Unusual rain
shortage is observed in large areas of Western Europe prior to the episode (Petit et al., 2017). While time since precipitation at
the SIRTA-site is a large positive contributor to the model predictions (see Fig. 14d), it is not driving the day-to-day variations.
The unusual nature of this event and lacking information on emission in the source regions and formation processes along air
mass trajectories in the model likely explain why the model has difficulties in reproducing this pollution episode. While this
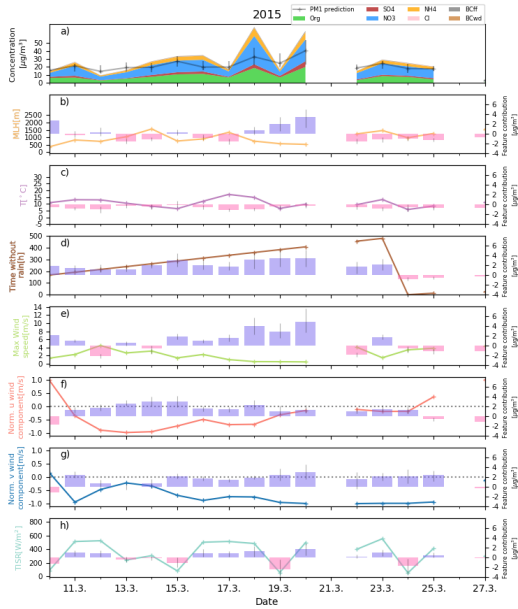
Figure 14. As Fig. 11 for an exemplary spring pollution episode in March 2015.

has implications for the application of explainable machine learning models for rare events, this is not expected to be an issue for the other cases and seasonal results presented here.

24

5   Conclusions and outlook

In this study, dominant patterns of drivers of PM1 species and total PM1 concentrations are identified and analysed using a novel, data-driven statistical approach. A machine learning model is set up to predict measured speciated and total PM1 concentrations based on meteorological measurements from the SIRTA supersite, southwest of Paris. The statistical model is able to reproduce daily variability of particle concentrations well, and is used to analyse and quantify the atmospheric processes causing high-pollution episodes during different seasons using a SHAP-value framework. Comparison of the results based on the statistical model with findings of previous studies on air quality patterns in the Paris region shows good agreement. Peak concentrations of total PM1 in winter are mainly driven by shallow MLHs, low temperatures and low wind speeds. These conditions are often amplified by northeastern wind inflow. A detailed analysis revealed different drivers of high-pollution episodes in winter. For an episode in January 2016, model results show a strong contribution of temperature to the elevated PM1 concentrations during this episode (up to 11 μg/m^3 are attributed to temperature), suggesting enhanced local, temperature-dependant particle formation. During a different, prolonged pollution episode in December 2016, temperature levels were relatively stable and had no influence. Here, contributions of MLHs ($<$500 m asl) were quantified to be the main driver of PM1 peak concentrations with contributions up to 6 μg/m^3, along with wind direction contributions of up to $\sim$6 μg/m^3. Total PM1 concentrations in spring can be as high as 50 μg/m^3. These peaks in spring are not as well reproduced by the model as winter episodes and are likely due to new particle formation processes along the air mass trajectories, in particular of nitrate. Summer PM1 concentrations are lower than in other seasons. Model results suggest that summer peak concentrations are largely driven by high temperatures, particle advection from Paris and continental Europe with low wind speeds and prolonged periods without precipitation. For an example episode in June 2017, temperatures above 30°C contribute $\sim$3 μg/m^3 to the total PM1 concentration. On site scarcity of rain increases air pollution, but does not appear to be a main driver of strong day-to-day variations in particle concentrations. Main drivers of day-to-day variability of predicted PM1 concentrations are changes in wind direction, air temperature and MLH. These changes often superimpose the influence of time without precipitation. Individual PM1 species are shown to respond differently to changes in temperature. While SO4 and organic matter concentrations are increased during both high and low temperature situations, NH4 and NO3 are substantially increased only at low temperatures. Model results indicate that SIA particle formation is enhanced during shallow MLH conditions.

Many of the results presented here hold true for regions other than suburban Paris and are thus beneficial for the general understanding of drivers of air pollution. This includes the importance of formation processes of secondary pollutants as well as the dominant role of the MLH for PM1 concentrations. The contribution of wind direction furthermore highlights the role of advected pollution and emphasizes the need for large-scale measures against air pollution. The GBRT approach in combination with the SHAP regression values presented here provides an intuitive tool to assess meteorological drivers of air pollution and to advance the understanding of high pollution events by uncovering different physical mechanisms leading to high-pollution episodes. To our knowledge, this is the first time that the SHAP-framework for explainable machine learning is applied in atmospheric sciences. The results of this study are highly relevant for policy makers, e.g. by providing a basis for future clean air programs or by providing the potential of a statistically-based early warning system for high pollution episodes. Statistical

25

models such as the one developed in the present study can be used to attribute changes in air quality to policy measures (cf. Cermak and Knutti, 2009), and may help political decision makers develop and implement effective clean-air policies. This study could be extended in the future, e.g. by including information on anthropogenic emissions or further stations down- and 470 upwind of SIRTA, which would allow further analysis of dominant advection patterns. Future efforts could also focus on the inclusion of short-time weather forecasts, thus providing an air quality forecast based on the statistical model.

26

240

480 References

Antwarg, L., Shapira, B., and Rokach, L.: Explaining Anomalies Detected by Autoencoders Using SHAP, SIGIR 2019 - Proc. 42nd Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., pp. 1281–1284, https://doi.org/10.1145/nnnnnnn.nnnnnnn, http://arxiv.org/abs/1903.02407, 2019.

Baklanov, A., Molina, L. T., and Gauss, M.: Megacities, air quality and climate, Atmos. Environ., 126, 235–249, https://doi.org/10.1016/j.atmosenv.2015.11.059, 2016.

485 Bressi, M., Sciare, J., Ghersi, V., Bonnaire, N., Nicolas, J. B., Petit, J. E., Moukhtar, S., Rosso, A., Mihalopoulos, N., and Féron, A.: A one-year comprehensive chemical characterisation of fine aerosol (PM2.5) at urban, suburban and rural background sites in the region of Paris (France), Atmos. Chem. Phys., 13, 7825–7844, https://doi.org/10.5194/acp-13-7825-2013, 2013.

Bressi, M., Sciare, J., Ghersi, V., Mihalopoulos, N., Petit, J. E., Nicolas, J. B., Moukhtar, S., Rosso, A., Féron, A., Bonnaire, N., Poulakis, E., and Theodosi, C.: Sources and geographical origins of fine aerosols in Paris (France), Atmos. Chem. Phys., 14, 8813–8839, 490 https://doi.org/10.5194/acp-14-8813-2014, 2014.

Cermak, J. and Knutti, R.: Beijing Olympics as an aerosol field experiment, Geophys. Res. Lett., 36, 1–5, https://doi.org/10.1029/2009GL038572, 2009.

Chafe, Z. A., Brauer, M., Klimont, Z., Van Dingenen, R., Mehta, S., Rao, S., Riahi, K., Dentener, F., and Smith, K. R.: Household Cooking with Solid Fuels Contributes to Ambient PM 2.5 Air Pollution and the Burden of Disease, Environ. Health Perspect., 122, 1314–1320, 495 https://doi.org/10.1289/ehp.1206340, https://ehp.niehs.nih.gov/doi/10.1289/ehp.1206340, 2014.

Chen, G., Li, S., Zhang, Y., Zhang, W., Li, D., Wei, X., He, Y., Bell, M. L., Williams, G., Marks, G. B., Jalaludin, B., Abramson, M. J., and Guo, Y.: Effects of ambient PM 1 air pollution on daily emergency hospital visits in China: an epidemiological study, Lancet Planet. Heal., 1, e221–e229, https://doi.org/10.1016/S2542-5196(17)30100-6, http://dx.doi.org/10.1016/S2542-5196(17)30100-6https://linkinghub.elsevier.com/retrieve/pii/S2542519617301006, 2017.

500 Chen, Y., Schleicher, N., Chen, Y., Chai, F., and Norra, S.: The influence of governmental mitigation measures on contamination characteristics of PM2.5 in Beijing, Sci. Total Environ., 490, 647–658, https://doi.org/10.1016/j.scitotenv.2014.05.049, http://dx.doi.org/10.1016/j.scitotenv.2014.05.049https://linkinghub.elsevier.com/retrieve/pii/S0048969714007268, 2014.

Cheng, Y., Zheng, G., Wei, C., Mu, Q., Zheng, B., Wang, Z., Gao, M., Zhang, Q., He, K., Carmichael, G., Pöschl, U., and Su, H.: Reactive nitrogen chemistry in aerosol water as a source of sulfate during haze events in China, Sci. Adv., 2, e1601 530, 505 https://doi.org/10.1126/sciadv.1601530, http://advances.sciencemag.org/lookup/doi/10.1126/sciadv.1601530, 2016.

Chiriaco, M., Dupont, J.-C., Bastin, S., Badosa, J., Lopez, J., Haeffelin, M., Chepfer, H., and Guzman, R.: ReOBS: a new approach to synthesize long-term multi-variable dataset and application to the SIRTA supersite, Earth Syst. Sci. Data, 10, 919–940, https://doi.org/10.5194/essd-10-919-2018, https://www.earth-syst-sci-data.net/10/919/2018/, 2018.

Churkina, G., Kuik, F., Bonn, B., Lauer, A., Grote, R., Tomiak, K., and Butler, T. M.: Effect of VOC Emissions from Vegetation on Air 510 Quality in Berlin during a Heatwave, Environ. Sci. Technol., p. acs.est.6b06514, https://doi.org/10.1021/acs.est.6b06514, http://pubs.acs.org/doi/abs/10.1021/acs.est.6b06514, 2017.

Clegg, S. L., Brimblecombe, P., and Wexler, A. S.: Thermodynamic Model of the System H + -NH 4 + -Na + -SO 4 2- -NO 3 - -Cl - -H 2 O at 298.15 K, J. Phys. Chem. A, 102, 2155–2171, https://doi.org/10.1021/jp973043j, https://pubs.acs.org/doi/10.1021/jp973043j, 1998.

Dawson, J. P., Adams, P. J., and Pandis, S. N.: Sensitivity of PM2.5 to climate in the Eastern US: a modeling case study, Atmos. 515 Chem. Phys., 7, 4295–4309, https://doi.org/10.5194/acp-7-4295-2007, http://www.atmos-chem-phys.net/7/4295/2007/, 2007.

27

Dey, S., Caulfield, B., and Ghosh, B.: Potential health and economic benefits of banning diesel traffic in Dublin, Ireland, J. Transp. Heal., 10, 156–166, https://doi.org/10.1016/j.jth.2018.04.006, https://linkinghub.elsevier.com/retrieve/pii/S2214140517307740, 2018.

Dupont, J.-C., Haeffelin, M., Badosa, J., Elias, T., Favez, O., Petit, J., Meleux, F., Sciare, J., Crenn, V., and Bonne, J.: Role of the boundary layer dynamics effects on an extreme air pollution event in Paris, Atmos. Environ., https://doi.org/10.1016/j.atmosenv.2016.06.061, http://linkinghub.elsevier.com/retrieve/pii/S1352231016304940, 2016.

Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, pp. 802–813, https://doi.org/10.1111/j.1365-2656.2008.01390.x, 2008.

Favez, O., Cachier, H., Sciare, J., Sarda-Estève, R., and Martinon, L.: Evidence for a significant contribution of wood burning aerosols to PM2.5 during the winter season in Paris, France, Atmos. Environ., 43, 3640–3644, https://doi.org/10.1016/j.atmosenv.2009.04.035, 2009.

Fowler, D., Pilegaard, K., Sutton, M., Ambus, P., Raivonen, M., Duyzer, J., Simpson, D., Fagerli, H., Fuzzi, S., Schjoerring, J., Granier, C., Neftel, A., Isaksen, I., Laj, P., Maione, M., Monks, P., Burkhardt, J., Daemmgen, U., Neirynck, J., Personne, E., Wichink-Kruit, R., Butterbach-Bahl, K., Flechard, C., Tuovinen, J., Coyle, M., Gerosa, G., Loubet, B., Altimir, N., Gruenhage, L., Ammann, C., Cieslik, S., Paoletti, E., Mikkelsen, T., Ro-Poulsen, H., Cellier, P., Cape, J., Horváth, L., Loreto, F., Niinemets, Ü., Palmer, P., Rinne, J., Misztal, P., Nemitz, E., Nilsson, D., Pryor, S., Gallagher, M., Vesala, T., Skiba, U., Brüggemann, N., Zechmeister-Boltenstern, S., Williams, J., O'Dowd, C., Facchini, M., de Leeuw, G., Flossman, A., Chaumerliac, N., and Erisman, J.: Atmospheric composition change: Ecosystems–Atmosphere interactions, Atmos. Environ., 43, 5193–5267, https://doi.org/10.1016/j.atmosenv.2009.07.068, https://linkinghub.elsevier.com/retrieve/pii/S1352231009006633, 2009.

Friedman, J. H.: Stochastic gradient boosting, Comput. Stat. Data Anal., 38, 367–378, https://doi.org/10.1016/S0167-9473(01)00065-2, http://linkinghub.elsevier.com/retrieve/pii/S0167947301000652, 2002.

Fuchs, J., Cermak, J., and Andersen, H.: Building a cloud in the Southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning, Atmos. Chem. Phys. Discuss., 2, 1–23, https://doi.org/10.5194/acp-2018-593, https://www.atmos-chem-phys-discuss.net/acp-2018-593/, 2018.

Geiß, A., Wiegner, M., Bonn, B., Schäfer, K., Forkel, R., Von Schneidemesser, E., Münkel, C., Lok Chan, K., and Nothard, R.: Mixing layer height as an indicator for urban air quality?, Atmos. Meas. Tech., 10, 2969–2988, https://doi.org/10.5194/amt-10-2969-2017, 2017.

Gen, M., Zhang, R., Huang, D. D., Li, Y., and Chan, C. K.: Heterogeneous SO 2 Oxidation in Sulfate Formation by Photolysis of Particulate Nitrate, Environ. Sci. Technol. Lett., 6, 86–91, https://doi.org/10.1021/acs.estlett.8b00681, https://pubs.acs.org/doi/10.1021/acs.estlett.8b00681, 2019.

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E.: Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation, J. Comput. Graph. Stat., 24, 44–65, https://doi.org/10.1080/10618600.2014.907095, http://www.tandfonline.com/doi/full/10.1080/10618600.2014.907095, 2015.

Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E., and Hueglin, C.: Random forest meteorological normalisation models for Swiss PM 10 trend analysis, Atmos. Chem. Phys., 18, 6223–6239, https://doi.org/10.5194/acp-18-6223-2018, https://www.atmos-chem-phys.net/18/6223/2018/, 2018.

Guenther, A. B., Zimmerman, P. R., Harley, P. C., Monson, R. K., and Fall, R.: Isoprene and monoterpene emission rate variability: Model evaluations and sensitivity analyses, J. Geophys. Res., 98, 12 609, https://doi.org/10.1029/93JD00527, http://doi.wiley.com/10.1029/93JD00527, 1993.

28

Gupta, P. and Christopher, S. A.: Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach, J. Geophys. Res., 114, D14 205, https://doi.org/10.1029/2008JD011496, http://doi.wiley.com/10.1029/2008JD011496, 2009.

Haeffelin, M., Bock, O., Boitel, C., Bony, S., Bouniol, D., Chepfer, H., Chiriaco, M., Cuesta, J., Drobinski, P., Flamant, C., Grall, M., Hodzic, A., Hourdin, F., Lapouge, F., Mathieu, A., Morille, Y., Naud, C., Pelon, J., Pietras, C., Protat, A., Romand, B., Scialom, G., and Vautard, R.: SIRTA, a ground-based atmospheric observatory for cloud and aerosol research, Ann. Geophys., 23, 253–275, 2005.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, vol. 26 of *Springer Series in Statistics*, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-84858-7, http://link.springer.com/10.1007/978-0-387-84858-7, 2009.

Healy, R. M., Sciare, J., Poulain, L., Kamili, K., Merkel, M., Müller, T., Wiedensohler, A., Eckhardt, S., Stohl, A., Sarda-Estève, R., McGillicuddy, E., O'Connor, I. P., Sodeau, J. R., and Wenger, J. C.: Sources and mixing state of size-resolved elemental carbon particles in a European megacity: Paris, Atmos. Chem. Phys., 12, 1681–1700, https://doi.org/10.5194/acp-12-1681-2012, https://www.atmos-chem-phys.net/12/1681/2012/, 2012.

Hennig, F., Quass, U., Hellack, B., Küpper, M., Kuhlbusch, T. A. J., Stafoggia, M., and Hoffmann, B.: Ultrafine and Fine Particle Number and Surface Area Concentrations and Daily Cause-Specific Mortality in the Ruhr Area, Germany, 2009–2014, Environ. Health Perspect., 126, 027 008, https://doi.org/10.1289/EHP2054, https://ehp.niehs.nih.gov/doi/10.1289/EHP2054, 2018.

Hoffmann, C. and Funk, R.: Diurnal changes of PM10-emission from arable soils in NE-Germany, Aeolian Res., 17, 117–127, https://doi.org/10.1016/j.aeolia.2015.03.002, http://dx.doi.org/10.1016/j.aeolia.2015.03.002, 2015.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., and Liu, Y.: Estimating PM 2 . 5 Concentrations in the Conterminous United States Using the Random Forest Approach, https://doi.org/10.1021/acs.est.7b01210, 2017.

Hueglin, C., Gehrig, R., Baltensperger, U., Gysel, M., Monn, C., and Vonmont, H.: Chemical characterisation of PM2.5, PM10 and coarse particles at urban, near-city and rural sites in Switzerland, Atmos. Environ., 39, 637–651, https://doi.org/10.1016/j.atmosenv.2004.10.027, 2005.

Hughes, H. E., Morbey, R., Fouillet, A., Caserio-Schönemann, C., Dobney, A., Hughes, T. C., Smith, G. E., and Elliot, A. J.: Retrospective observational study of emergency department syndromic surveillance data during air pollution episodes across London and Paris in 2014, BMJ Open, 8, 1–12, https://doi.org/10.1136/bmjopen-2017-018732, 2018.

Jiang, J., Aksoyoglu, S., El-Haddad, I., Ciarelli, G., Denier van der Gon, H. A., Canonaco, F., Gilardoni, S., Paglione, M., Minguillón, M. C., Favez, O., Zhang, Y., Marchand, N., Hao, L., Virtanen, A., Florou, K., O'Dowd, C., Ovadnevaite, J., Baltensperger, U., and Prévôt, A. S.: Sources of organic aerosols in Europe: A modelling study using CAMx with modified volatility basis set scheme, Atmos. Chem. Phys. Discuss., pp. 1–35, https://doi.org/10.5194/acp-2019-468, 2019.

Just, A., De Carli, M., Shtein, A., Dorman, M., Lyapustin, A., and Kloog, I.: Correcting Measurement Error in Satellite Aerosol Optical Depth with Machine Learning for Modeling PM2.5 in the Northeastern USA, Remote Sens., 10, 803, https://doi.org/10.3390/rs10050803, http://www.mdpi.com/2072-4292/10/5/803, 2018.

Kapadia, Z. Z., Spracklen, D. V., Arnold, S. R., Borman, D. J., Mann, G. W., Pringle, K. J., Monks, S. A., Reddington, C. L., Benduhn, F., Rap, A., Scott, C. E., Butt, E. W., and Yoshioka, M.: Impacts of aviation fuel sulfur content on climate and human health, Atmos. Chem. Phys., 16, 10 521–10 541, https://doi.org/10.5194/acp-16-10521-2016, 2016.

Kiesewetter, G., Borken-Kleefeld, J., Schöpp, W., Heyes, C., Thunis, P., Bessagnet, B., Terrenoire, E., Fagerli, H., Nyiri, A., and Amann, M.: Modelling street level PM10 concentrations across Europe: Source apportionment and possible futures, Atmos. Chem. Phys., 15, 1539–1553, https://doi.org/10.5194/acp-15-1539-2015, 2015.

29

Klingner, M. and Sähn, E.: Prediction of PM10 concentration on the basis of high resolution weather forecasting, Meteorol. Zeitschrift, 17, 263–272, https://doi.org/10.1127/0941-2948/2008/0288, http://www.schweizerbart.de/papers/metz/detail/17/56689/Prediction{_}of{_}PM10{_}concentration{_}on{_}the{_}basis{_}of{_}h?af=crossref, 2008.

Kotthaus, S. and Grimmond, C. S. B.: Atmospheric boundary-layer characteristics from ceilometer measurements. Part 1: A new method to
595   track mixed layer height and classify clouds, Q. J. R. Meteorol. Soc., 144, 1525–1538, https://doi.org/10.1002/qj.3299, 2018a.

Kotthaus, S. and Grimmond, C. S. B.: Atmospheric boundary-layer characteristics from ceilometer measurements. Part 2: Application to
London's urban boundary layer, Q. J. R. Meteorol. Soc., 144, 1511–1524, https://doi.org/10.1002/qj.3298, 2018b.

Laborde, M., Crippa, M., Tritscher, T., Jurányi, Z., Decarlo, P. F., Temime-Roussel, B., Marchand, N., Eckhardt, S., Stohl, A., Baltensperger,
U., Prévôt, A. S., Weingartner, E., and Gysel, M.: Black carbon physical properties and mixing state in the European megacity Paris,
600   Atmos. Chem. Phys., 13, 5831–5856, https://doi.org/10.5194/acp-13-5831-2013, 2013.

Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and Pozzer, A.: The contribution of outdoor air pollution sources to premature mortality
on a global scale, Nature, 525, 367–371, https://doi.org/10.1038/nature15371, 2015.

Lelieveld, J., Klingmüller, K., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., and Münzel, T.: Cardiovascular disease burden from ambient
air pollution in Europe reassessed using novel hazard ratio functions, Eur. Heart J., pp. 1–7, https://doi.org/10.1093/eurheartj/ehz135,
605   https://academic.oup.com/eurheartj/advance-article/doi/10.1093/eurheartj/ehz135/5372326, 2019.

Leung, D. M., Tai, A. P. K., Mickley, L. J., Moch, J. M., Van Donkelaar, A., Shen, L., and Martin, R. V.: Synoptic meteorological modes of
variability for fine particulate matter (PM 2.5) air quality in major metropolitan regions of China, Atmos. Chem. Phys. Discuss. Chem.
Phys. Discuss., started, 1–29, https://doi.org/10.5194/acp-2017-916, 2017.

Li, R., Shinde, A., Liu, A., Glaser, S., and Amini, A.: Identification of Nonlinear Interaction Effects in Prostate Cancer Survival Using
610   Machine Learning-Based Modeling, Int. J. Radiat. Oncol., 105, S121, https://doi.org/10.1016/j.ijrobp.2019.06.091, https://linkinghub.
elsevier.com/retrieve/pii/S0360301619309265, 2019a.

Li, Y., Zhang, J., Sailor, D. J., and Ban-Weiss, G. A.: Effects of urbanization on regional meteorology and air quality in Southern California,
Atmos. Chem. Phys., 19, 4439–4457, https://doi.org/10.5194/acp-19-4439-2019, 2019b.

Li, Z., Guo, J., Ding, A., Liao, H., Liu, J., Sun, Y., Wang, T., Xue, H., Zhang, H., and Zhu, B.: Aerosol and boundary-layer interactions
615   and impact on air quality, Natl. Sci. Rev., 4, 810–833, https://doi.org/10.1093/nsr/nwx117, https://academic.oup.com/nsr/article/4/6/810/
4191281, 2017.

Liu, Q., Jia, X., Quan, J., Li, J., Li, X., Wu, Y., Chen, D., Wang, Z., and Liu, Y.: New positive feedback mechanism between boundary layer
meteorology and secondary aerosol formation during severe haze events, Sci. Rep., 8, 1–8, https://doi.org/10.1038/s41598-018-24366-3,
http://dx.doi.org/10.1038/s41598-018-24366-3, 2018.

620   Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, pp. 1–10, http://arxiv.org/abs/1705.07874, 2017.

Lundberg, S. M., Erion, G. G., and Lee, S.-I.: Consistent Individualized Feature Attribution for Tree Ensembles, https://doi.org/10.1007/978-
0-387-84858-7, http://arxiv.org/abs/1802.03888, 2018a.

Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K.-W., Newman, S.-F., Kim, J., and
Lee, S.-I.: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery, Nat. Biomed. Eng., 2, 749–760,
625   https://doi.org/10.1038/s41551-018-0304-0, http://www.nature.com/articles/s41551-018-0304-0, 2018b.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-i.: From local
explanations to global understanding with explainable AI for trees, Nat. Mach. Intell., 2, 56–67, https://doi.org/10.1038/s42256-019-0138-
9, http://dx.doi.org/10.1038/s42256-019-0138-9http://www.nature.com/articles/s42256-019-0138-9, 2020.

30

McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., and Smith, T.: Making the Black
630   Box More Transparent: Understanding the Physical Implications of Machine Learning, Bull. Am. Meteorol. Soc., 100, 2175–2199,
       https://doi.org/10.1175/BAMS-D-18-0195.1, http://journals.ametsoc.org/doi/10.1175/BAMS-D-18-0195.1, 2019.

Megaritis, A. G., Fountoukis, C., Charalampidis, P. E., Pilinis, C., and Pandis, S. N.: Response of fine particulate matter concentrations to
       changes of emissions and temperature in Europe, Atmos. Chem. Phys., 13, 3423–3443, https://doi.org/10.5194/acp-13-3423-2013, 2013.

Megaritis, A. G., Fountoukis, C., Charalampidis, P. E., Denier Van Der Gon, H. A., Pilinis, C., and Pandis, S. N.: Linking climate and air qual-
635   ity over Europe: Effects of meteorology on PM2.5 concentrations, Atmos. Chem. Phys., 14, 10 283–10 298, https://doi.org/10.5194/acp-
       14-10283-2014, 2014.

Ng, N. L., Herndon, S. C., Trimborn, A., Canagaratna, M. R., Croteau, P. L., Onasch, T. B., Sueper, D., Worsnop, D. R., Zhang, Q., Sun, Y. L.,
       and Jayne, J. T.: An Aerosol Chemical Speciation Monitor (ACSM) for routine monitoring of the composition and mass concentrations of
       ambient aerosol, Aerosol Sci. Technol., 45, 770–784, https://doi.org/10.1080/02786826.2011.560211, 2011.

640   Pay, M. T., Jiménez-Guerrero, P., and Baldasano, J. M.: Assessing sensitivity regimes of secondary inorganic aerosol formation in Europe
       with the CALIOPE-EU modeling system, Atmos. Environ., 51, 146–164, https://doi.org/10.1016/j.atmosenv.2012.01.027, http://dx.doi.
       org/10.1016/j.atmosenv.2012.01.027, 2012.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Pretten-
       hofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn:
645   Machine Learning in Python, http://arxiv.org/abs/1201.0490, 2012.

Petäjä, T., Järvi, L., Kerminen, V.-M., Ding, A., Sun, J., Nie, W., Kujansuu, J., Virkkula, A., Yang, X., Fu, C., Zilitinkevich, S., and Kulmala,
       M.: Enhanced air pollution via aerosol-boundary layer feedback in China, Sci. Rep., 6, 18 998, https://doi.org/10.1038/srep18998, http:
       //www.nature.com/articles/srep18998, 2016.

Petetin, H., Beekmann, M., Sciare, J., Bressi, M., Rosso, A., Sanchez, O., and Ghersi, V.: A novel model evaluation approach focus-
650   ing on local and advected contributions to urban PM2.5 levels - Application to Paris, France, Geosci. Model Dev., 7, 1483–1505,
       https://doi.org/10.5194/gmd-7-1483-2014, 2014.

Petit, J. E., Favez, O., Sciare, J., Canonaco, F., Croteau, P., Močnik, G., Jayne, J., Worsnop, D., and Leoz-Garziandia, E.: Submicron aerosol
       source apportionment of wintertime pollution in Paris, France by double positive matrix factorization (PMF2) using an aerosol chemical
       speciation monitor (ACSM) and a multi-wavelength Aethalometer, Atmos. Chem. Phys., 14, 13 773–13 787, https://doi.org/10.5194/acp-
655   14-13773-2014, 2014.

Petit, J. E., Favez, O., Sciare, J., Crenn, V., Sarda-Estève, R., Bonnaire, N., Močnik, G., Dupont, J. C., Haeffelin, M., and Leoz-Garziandia,
       E.: Two years of near real-time chemical composition of submicron aerosols in the region of Paris using an Aerosol Chemical Speciation
       Monitor (ACSM) and a multi-wavelength Aethalometer, Atmos. Chem. Phys., 15, 2985–3005, https://doi.org/v, 2015.

Petit, J. E., Amodeo, T., Meleux, F., Bessagnet, B., Menut, L., Grenier, D., Pellan, Y., Ockler, A., Rocq, B., Gros, V., Sciare, J., and Favez, O.:
660   Characterising an intense PM pollution episode in March 2015 in France from multi-site approach and near real time data: Climatology,
       variabilities, geographical origins and model evaluation, Atmos. Environ., 155, 68–84, https://doi.org/10.1016/j.atmosenv.2017.02.012,
       http://dx.doi.org/10.1016/j.atmosenv.2017.02.012, 2017.

Putaud, J.-P., Raes, F., Van Dingenen, R., Brüggemann, E., Facchini, M.-C., Decesari, S., Fuzzi, S., Gehrig, R., Hüglin, C., Laj, P., Lorbeer,
       G., Maenhaut, W., Mihalopoulos, N., Müller, K., Querol, X., Rodriguez, S., Schneider, J., Spindler, G., ten Brink, H., Tørseth, K., and
665   Wiedensohler, A.: A European aerosol phenomenology—2: chemical characteristics of particulate matter at kerbside, urban, rural and

31

245

background sites in Europe, Atmos. Environ., 38, 2579–2595, https://doi.org/10.1016/j.atmosenv.2004.01.041, https://linkinghub.elsevier.com/retrieve/pii/S1352231004000949, 2004.

Radke, L. F., Hobbs, P. V., and Eltgroth, M. W.: Scavenging of aerosol particles by precipitation., J. Appl. Meteorol., 19, 715–722, https://doi.org/10.1175/1520-0450(1980)019<0715:SOAPBP>2.0.CO;2, 1980.

670 Rengarajan, R., Sudheer, A., and Sarin, M.: Wintertime PM2.5 and PM10 carbonaceous and inorganic constituents from urban site in western India, Atmos. Res., 102, 420–431, https://doi.org/10.1016/j.atmosres.2011.09.005, http://dx.doi.org/10.1016/j.atmosres.2011.09.005https://linkinghub.elsevier.com/retrieve/pii/S0169809511002900, 2011.

Rost, J., Holst, T., Sahn, E., Klingner, M., Anke, K., Ahrens, D., and Mayer, H.: Variability of PM10 concentrations dependent on meteorological conditions, Int. J. Environ. Pollut., 36, 3–18, https://doi.org/10.1504/IJEP.2009.021813, 2009.

675 Rybarczyk, Y. and Zalakeviciute, R.: Machine learning approaches for outdoor air quality modelling: A systematic review, https://doi.org/10.3390/app8122570, http://www.mdpi.com/2076-3417/8/12/2570, 2018.

Samoli, E., Peng, R., Ramsay, T., Pipikou, M., Touloumi, G., Dominici, F., Burnett, R., Cohen, A., Krewski, D., Samet, J., and Katsouyanni, K.: Acute effects of ambient particulate matter on mortality in Europe and North America: Results from the APHENA study, Environ. Health Perspect., 116, 1480–1486, https://doi.org/10.1289/ehp.11345, 2008.

680 Samoli, E., Stafoggia, M., Rodopoulou, S., Ostro, B., Declercq, C., Alessandrini, E., Díaz, J., Karanasiou, A., Kelessis, A. G., Le Tertre, A., Pandolfi, P., Randi, G., Scarinzi, C., Zauli-Sajani, S., Katsouyanni, K., and Forastiere, F.: Associations between Fine and Coarse Particles and Mortality in Mediterranean Cities: Results from the MED-PARTICLES Project, Environ. Health Perspect., 121, 932–938, https://doi.org/10.1289/ehp.1206124, https://ehp.niehs.nih.gov/doi/10.1289/ehp.1206124, 2013.

Schäfer, K., Wagner, P., Emeis, S., Jahn, C., Münkel, C., Suppan, P., Ecology, L., and Gmbh, V.: Mixing layer height and air pollution levels
685 in urban area, Proc. SPIE, 8534, 1–10, https://doi.org/10.1117/12.974328, 2012.

Sciare, J., D'Argouges, O., Zhang, Q. J., Sarda-Estève, R., Gaimoz, C., Gros, V., Beekmann, M., and Sanchez, O.: Comparison between simulated and observed chemical composition of fine aerosols in Paris (France) during springtime: Contribution of regional versus continental emissions, Atmos. Chem. Phys., 10, 11 987–12 004, https://doi.org/10.5194/acp-10-11987-2010, 2010.

Shapley, L.: A Value for n-Person Games, in: Contrib. to Theory Games (AM-28), Vol. II, edited by Kuhn, H. W. and Tucker, A. W., vol. 28,
690 pp. 307–318, Princeton University Press, Princeton, https://doi.org/10.1515/9781400881970-018, http://www.degruyter.com/view/books/9781400881970/9781400881970-018/9781400881970-018.xml, 1953.

Srivastava, D., Favez, O., Bonnaire, N., Lucarelli, F., Haeffelin, M., Perraudin, E., Gros, V., Villenave, E., and Albinet, A.: Speciation of organic fractions does matter for aerosol source apportionment. Part 2: Intensive short-term campaign in the Paris area (France), Sci. Total Environ., 634, 267–278, https://doi.org/10.1016/j.scitotenv.2018.03.296, https://doi.org/10.1016/j.scitotenv.2018.03.296https://linkinghub.elsevier.com/retrieve/pii/S0048969718310568, 2018.

Stirnberg, R., Cermak, J., Fuchs, J., and Andersen, H.: Mapping and Understanding Patterns of Air Quality Using Satellite Data and Machine Learning, J. Geophys. Res. Atmos., 125, https://doi.org/10.1029/2019JD031380, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031380?af=R{&}utm{_}source=researcher{_}app{&}utm{_}medium=referral{&}utm{_}campaign=RESR{_}MRKT{_}Researcher{_}inbound{&}sid=researcherhttps://onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031380, 2020.

700 Su, J. G., Apte, J. S., Lipsitt, J., Garcia-Gonzales, D. A., Beckerman, B. S., de Nazelle, A., Texcalac-Sangrador, J. L., and Jerrett, M.: Populations potentially exposed to traffic-related air pollution in seven world cities, Environ. Int., 78, 82–89, https://doi.org/10.1016/j.envint.2014.12.007, http://dx.doi.org/10.1016/j.envint.2014.12.007, 2015.

32

Sujatha, P., Mahalakshmi, D., Ramiz, A., Rao, P., and Naidu, C.: Ventilation coefficient and boundary layer height impact on urban air quality, Cogent Environ. Sci., 2, 1–9, https://doi.org/10.1080/23311843.2015.1125284, http://dx.doi.org/10.1080/23311843.2015.1125284https: //www.cogentoa.com/article/10.1080/23311843.2015.1125284, 2016.

Wagner, P. and Schäfer, K.: Influence of mixing layer height on air pollutant concentrations in an urban street canyon, Urban Clim., 22, 64–79, https://doi.org/10.1016/j.uclim.2015.11.001, https://doi.org/10.1016/j.uclim.2015.11.001, 2017.

Wang, G., Zhang, R., Gomez, M. E., Yang, L., Zamora, M. L., Hu, M., Lin, Y., Peng, J., Guo, S., Meng, J., Li, J., Cheng, C., Hu, T., Ren, Y., Wang, Y., Gao, J., Cao, J., An, Z., Zhou, W., Li, G., Wang, J., Tian, P., Marrero-Ortiz, W., Secrest, J., Du, Z., Zheng, J., Shang, D., Zeng, L., Shao, M., Wang, W., Huang, Y., Wang, Y., Zhu, Y., Li, Y., Hu, J., Pan, B., Cai, L., Cheng, Y., Ji, Y., Zhang, F., Rosenfeld, D., Liss, P. S., Duce, R. A., Kolb, C. E., and Molina, M. J.: Persistent sulfate formation from London Fog to Chinese haze, Proc. Natl. Acad. Sci. U. S. A., 113, 13 630–13 635, https://doi.org/10.1073/pnas.1616540113, 2016.

Wang, X., Dickinson, R. E., Su, L., Zhou, C., and Wang, K.: PM 2.5 Pollution in China and How It Has Been Exacerbated by Terrain and Meteorological Conditions, Bull. Am. Meteorol. Soc., 99, 105–119, https://doi.org/10.1175/BAMS-D-16-0301.1, http://journals.ametsoc. org/doi/10.1175/BAMS-D-16-0301.1, 2018.

Yang, M., Chu, C., Bloom, M. S., Li, S., Chen, G., Heinrich, J., Markevych, I., Knibbs, L. D., Bowatte, G., Dharmage, S. C., Komppula, M., Leskinen, A., Hirvonen, M.-R., Roponen, M., Jalava, P., Wang, S.-Q., Lin, S., Zeng, X.-W., Hu, L.-W., Liu, K.-K., Yang, B.-Y., Chen, W., Guo, Y., and Dong, G.-H.: Is smaller worse? New insights about associations of PM1 and respiratory health in children and adoles-cents, Environ. Int., 120, 516–524, https://doi.org/10.1016/j.envint.2018.08.027, http://dx.doi.org/10.1016/S2542-5196(17)30095-5https: //linkinghub.elsevier.com/retrieve/pii/S2542519617300955https://linkinghub.elsevier.com/retrieve/pii/S0160412018303866, 2018.

Yang, Y. Q., Wang, J. Z., Gong, S. L., Zhang, X. Y., Wang, H., Wang, Y. Q., Wang, J., Li, D., and Guo, J. P.: PLAM - A meteoro-logical pollution index for air quality and its applications in fog-haze forecasts in North China, Atmos. Chem. Phys., 16, 1353–1364, https://doi.org/10.5194/acp-16-1353-2016, 2016.

33

Eidesstattliche Versicherung

Eidesstattliche Versicherung gemäß § 6 Abs. 1 Ziff. 4 der Promotionsordnung des Karlsruher Instituts für Technologie für die Fakultät für Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften.

1. Bei der eingereichten Dissertation zu dem Thema "Environmental Influences on Patterns of Atmospheric Particulate Matter: a Quantitative Study Using Ground- and Satellite-Based Observations" handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Ort und Datum   Unterschrift