

A quantitative, satellite-based analysis of fog and low stratus clouds in the context of land-atmosphere interactions

Zur Erlangung des akademischen Grades einer
DOKTORIN DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT Fakultät für
Bauingenieur-, Geo- und Umweltwissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Eva Pauli

aus Bad Urach

Tag der mündlichen Prüfung: 16.12.2022

Referent: Professor Dr. Jan Cermak

Korreferent: Professor Dr. Jörg Bendix

Karlsruhe 2022

Danksagung

Als passionierte Läuferin komme ich nicht umhin, die letzten drei Jahre mit einem schönen und anstrengenden Berglauf zu vergleichen: Auch wenn es Anstiege gibt, die fast nicht zu enden scheinen, gibt am Ende eines Aufstieges immer wieder Ausblicke, die einem neue Energie geben und Abstiege, die sich fast so anfühlen wie Fliegen. Wie es auch im Sport so oft ist, wäre mein Weg zur Doktorarbeit ohne die Unterstützung einiger Personen nicht möglich gewesen die mich in dieser Zeit begleitet haben.

An erster Stelle möchte ich Prof. Dr. Jan Cermak und Prof. Dr. Jörg Bendix für die Begutachtung der vorliegenden Doktorarbeit danken.

Mein besonderer Dank gilt meinem Betreuer Jan Cermak für die Unterstützung, sein Vertrauen, seine Geduld und den permanenten wissenschaftlichen Austausch. Ich danke ihm, dass er mich in meinen zahlreichen Ideen unterstützt hat und mir die Richtung gewiesen hat wenn ich mich verirrt hatte.

Die Promotion wäre in dieser Form nicht möglich gewesen ohne die finanzielle Unterstützung der Landesgraduiertenförderung und der Graduiertenschule GRACE, welche mir zwei Forschungsaufenthalte im Ausland ermöglicht hat.

Den Mitgliedern der Satellitenklimatologie-Arbeitsgruppe möchte ich herzlich für ihr stetes Interesse an meiner Arbeit und für die hilfreichen Diskussionen danken. Ohne die positive Feedbackkultur und eure Unterstützung hätte ich mich nicht so gut entwickeln und entfalten können. Ein besonderer Dank gilt all denjenigen, die diese Arbeit in ihrer Rohfassung geduldig gelesen haben und mir wertvolles Feedback gegeben haben. Auch meinen Kollegen am IPF möchte ich für die hilfreichen interdisziplinären Gespräche zu meiner Arbeit danken.

Neben Kollegen am KIT gilt mein Dank auch Ryan Teuling für die Betreuung während meiner Zeit an der Universität Wageningen und Philip Stier und Alyson Douglas für die Betreuung während meiner Zeit an der Universität Oxford.

Zuletzt gilt mein Dank den wichtigsten Personen in meinem Leben. Meinen Eltern möchte ich für ihre uneingeschränkte Unterstützung und ihr unerschütterliches Vertrauen in mich danken. Ohne euren Glauben in mich und meine Fähigkeiten wäre ich heute nicht dort wo ich bin. Von ganzem Herzen möchte ich meinem Partner

Florian für seine Geduld und für seine Gelassenheit danken. Du schenkst mir Halt wenn ich drohe davonzuschweben und gibst mir die Klarheit die wichtigen Dinge im Leben zu sehen. Danke, dass du mir stets den Rücken freihältst.

Eva Pauli
Steinach, November 2022

Abstract

Fog and low stratus clouds (FLS) are a key component of the climate system, as they directly modify Earth's radiation balance and provide nutrients and water to ecosystems. They further have economic impacts via affecting traffic on land, at sea and in the air and pose a challenge for solar power prediction. FLS are influenced by a myriad of atmospheric and land surface processes, with magnitude, direction and interactions of these processes varying across temporal and spatial scales. Understanding the effect of environmental conditions on spatial and temporal patterns of FLS is critical for their prediction and also for parameterizations of land-atmosphere processes in weather and climate models. This thesis aims to advance the scientific understanding of the influence of meteorological and land surface drivers on FLS occurrence and life cycle. FLS processes are investigated in the context of land-atmosphere interactions and the influence of land cover types on FLS occurrence is analyzed. With a spatial focus on central Europe, four research questions are targeted in the research presented in this thesis:

1. *What are the main drivers of large-scale spatial and temporal fog and low stratus patterns?*
2. *How does the land surface influence spatial variations in fog and low stratus occurrence?*
3. *What are the climatological patterns of fog and low stratus formation and dissipation time?*
4. *What are fog and low stratus formation and dissipation regimes on regional to sub-regional scales?*

These research questions are addressed in four studies using satellite and reanalysis data sets and a set of statistical methods. Using satellite and reanalysis data allows to investigate how environmental conditions influence FLS on spatial and temporal scales previously not considered. The main findings of the thesis are described in the following.

For a study area spanning continental Europe, a statistical model is set up to predict daily means of FLS occurrence using meteorological and land surface predictors. The sensitivity analysis reveals that atmospheric proxies, in particular mean sea level pressure, wind speed and FLS occurrence on the previous day, are more important determinants of the FLS presence than land surface characteristics. The importance of evapotranspiration and land surface temperature for FLS prediction increases in high pressure conditions. The analysis reveals not only the importance of atmospheric proxies for the prediction of FLS occurrence but also highlights the potential for the analysis of land surface effects on FLS occurrence by filtering for specific weather situations.

In a sub-regional approach, the mean climatological nighttime FLS occurrence over a large European forest area compared to its surrounding agricultural land is investigated. The two independent satellite-based FLS data sets used in the analysis show significantly higher nighttime FLS occurrence over the forest area, especially in summer and fall. Lower wind speeds and land surface temperatures, together with a temperature inversion and biogenic organic compounds serving as cloud condensation nuclei are identified as potential drivers for higher FLS occurrence over the forest area. The results suggest that the land surface significantly influences FLS occurrence on the climatological scale considered.

Besides FLS occurrence, the FLS life cycle is analyzed by extracting FLS formation and dissipation time via the application of logistic regression to an existing satellite-based FLS data set spanning the central European land mass. The climatological analysis reveals distinct geographic patterns of FLS formation and dissipation pertaining to topography, distance to the coast and the solar cycle. The seasonal cycle reveals similar patterns in spring and summer and in winter and fall, with higher FLS persistence found in the latter two seasons. The therein created novel FLS formation and dissipation time data set provides a basis for the analysis of drivers influencing the FLS life cycle, from daily to decadal time scales.

From this novel data set, regional FLS life cycle regimes are analyzed by applying a hierarchical clustering approach to correlations of FLS formation and dissipation time with environmental conditions. The clustering approach reveals four main FLS formation and dissipation regimes across the central European land mass: central, maritime, Mediterranean and Baltic-Scandinavian. Sub-regional regimes are analyzed in terms of their average monthly mean correlations of FLS formation and dissipation time with mean surface pressure, wind speed, land surface temperature and evapotranspiration. The correlations display the sensitivities of the FLS life cycle to changes in environmental conditions and indicate a strong dependency

on the climatological and geographic background of the regime. Furthermore, the prevailing FLS type of a regime might be a main determinant of the observed sensitivity patterns, thus filtering for specific FLS types is critical going forward.

The findings of this thesis advance the understanding of the effect of atmospheric and land surface drivers and their interactions on fog and low stratus cloud occurrence and life cycle. The specific influences of these drivers vary depending on geographic position and also with the temporal scales considered: Meteorological influences dominate on daily to seasonal time scales, whereas by considering multi-year means or filtering for specific weather situations, meteorological variations decrease, making it possible to distill land cover influences on FLS occurrence. A promising path forward is the set-up of regionally specific models to investigate drivers of FLS persistence, which further holds promise to include previously not considered variables. Results of a preliminary study using such an approach were presented in the outlook of this thesis: In the Po valley region, meteorological conditions are found to be the main drivers of FLS persistence, but the aerosol loading of the previous day can prolong FLS duration up to 60 minutes during high pollution events. Regional specific evaluations of the effect of environmental conditions on the FLS life cycle thus have implications and learnings for regional FLS predictions, traffic safety and solar energy production.

Kurzfassung

Nebel und niedrige Stratuswolken (fog and low stratus, FLS) sind ein wichtiger Teil des Klimasystems, da sie die Strahlungsbilanz der Erde direkt beeinflussen und als Wasser- und Nährstoffquellen für Ökosysteme dienen. Darüber hinaus haben sie wirtschaftliche Auswirkungen, da sie den Verkehr beeinträchtigen und eine Herausforderung für die Erzeugung von Solarenergie darstellen. FLS werden durch eine Vielzahl von Prozessen in der Atmosphäre und auf der Landoberfläche beeinflusst, wobei die Stärke, Richtung und Wechselwirkungen dieser Prozesse auf zeitlichen und räumlichen Skalen variieren. Das Verständnis des Einflusses von Umweltfaktoren auf die räumlichen und zeitlichen Muster von Nebel und niedrigem Stratus ist für ihre Vorhersage sowie für die Parametrisierung der Wechselwirkungen zwischen Landoberfläche und Atmosphäre in Wetter- und Klimamodellen unerlässlich. Ziel der vorliegenden Arbeit ist es das wissenschaftliche Verständnis des Einflusses von meteorologischen und landoberflächenbasierten Faktoren auf das Auftreten und den Lebenszyklus von FLS zu verbessern. Hierbei werden FLS Prozesse im Kontext von Interaktionen zwischen Landoberfläche und Atmosphäre untersucht sowie der Einfluss der Landoberfläche auf das Vorkommen von Nebel und niedrigem Stratus analysiert. Die zentralen Orientierungspunkte der Teilstudien, mit räumlichen Fokus auf Mitteleuropa, sind vier Hauptforschungsfragen:

1. *Welche Faktoren beeinflussen die großskaligen räumlichen und zeitlichen Muster von Nebel und niedrigem Stratus?*
2. *Wie beeinflusst die Landoberfläche räumliche Variationen des Vorkommens von Nebel und niedrigem Stratus?*
3. *Was sind klimatologische Muster der Bildung und Auflösung von Nebel und niedrigem Stratus?*
4. *Was sind regionale und subregionale Regime der Bildung und Auflösung von Nebel und niedrigem Stratus?*

Diese Forschungsfragen werden in vier separaten Studien behandelt, in denen Satellitendatensätze und Reanalysedaten in Kombination mit verschiedenen statistischen Methoden verwendet werden. Durch die Verwendung von Satelliten- und Reanalysedaten ist es möglich den Einfluss von Umweltfaktoren auf FLS auf großen räumlichen und zeitlichen Skalen zu untersuchen, die in der bisherigen Forschung nicht berücksichtigt wurden. Die wichtigsten Ergebnisse werden im Folgenden erläutert.

In einer für Kontinentaleuropa durchgeführten Studie wird ein statistisches Modell aufgesetzt, welches Tagesmittel der FLS Bedeckung mithilfe von meteorologischen und landoberflächenbasierten Prädiktoren vorhersagt. Die anschließende Sensitivitätsstudie zeigt, dass atmosphärische Prädiktoren, insbesondere Luftdruck, Windgeschwindigkeit und die FLS Bedeckung des Vortages, für die Vorhersage der FLS Bedeckung wichtiger sind als Charakteristika der Landoberfläche. Verdunstung und Landoberflächentemperatur gewinnen an Bedeutung für das Modell, wenn ausschließlich Hochdruckwetterlagen berücksichtigt werden. Die Analyse zeigt nicht nur die Bedeutung von atmosphärischen Prädiktoren für das statistische Modell auf, sondern verdeutlicht auch das Potential der Auswahl von Hochdruckwetterlagen für die Analyse des Einflusses der Landoberflächen auf Nebel und niedrigen Stratus.

In einem subregionalen Ansatz wird die mittlere nächtliche FLS Bedeckung über einem großen Waldgebiet in Europa mit derjenigen über dem umliegenden landwirtschaftlichen Gebiet verglichen. Unter Verwendung zweier unabhängiger satellitengestützter FLS Datensätze findet sich über dem Waldgebiet eine signifikant höhere nächtliche FLS Bedeckung. Mögliche Gründe hierfür sind niedrige Windgeschwindigkeiten und Landoberflächentemperaturen sowie eine Temperaturinversion und eine höhere Konzentration von biogenen Kondensationskernen über dem Waldgebiet. Die Ergebnisse deuten darauf hin, dass die Landoberfläche auf den untersuchten klimatologischen Skalen einen erheblichen Einfluss auf das Auftreten von Nebel und niedrigem Stratus hat.

Zusätzlich zur Analyse der FLS Bedeckung folgt eine Analyse des FLS-Lebenszyklus hinsichtlich dessen Bildungs- und Auflösungszeitpunktes. Zu diesem Zweck wird eine logistische Regression auf einen bestehenden satellitengestützten FLS Datensatz über Mitteleuropa angewendet. Die klimatologische Analyse zeigt ausgeprägte geographische Muster der FLS-Bildung und Auflösung, die von der Topographie, der Nähe zur Küste und der Sonneneinstrahlung abhängen. Die saisonalen Muster ähneln sich in Frühling und Sommer und in Herbst und Winter, wobei die FLS-Persistenz in den letztgenannten Jahreszeiten höher ist. Der so entstandene neue Datensatz bildet die Grundlage für Analysen von Faktoren des Lebenszyklus von

Nebel und niedrigem Stratus von Fallstudien bis hin zu klimatologischen Analysen. Aus diesem Datensatz werden regionale Regime des FLS Lebenszyklus abgeleitet, indem ein hierarchischer Clustering Algorithmus auf Sensitivitäten der FLS Bildungs- und Auflösungszeit hinsichtlich Umweltfaktoren angewandt wird. In Mitteleuropa werden vier Hauptregime der FLS Bildung und Auflösung identifiziert: zentral, maritim, Mediterran und Baltisch-Skandinavisch. Subregionale Regime werden hinsichtlich ihrer mittleren monatlichen Sensitivitäten von Bildungs- und Auflösungszeitpunkt gegenüber Veränderungen in Luftdruck, Windgeschwindigkeit, Landoberflächentemperatur und Verdunstung analysiert. Die Sensitivitäten zeigen eine deutliche Abhängigkeit vom klimatischen und geographischen Hintergrund des Regimes. Darüber hinaus könnte der vorherrschende FLS Typ der bestimmende Faktor für die beobachteten Sensitivitätsmuster sein, sodass künftig eine Filterung nach spezifischen FLS Typen von Bedeutung sein wird.

Die Ergebnisse dieser Arbeit verbessern das Verständnis der Einflüsse atmosphärischer und landoberflächenbasierter Faktoren und ihrer Wechselwirkungen auf das Auftreten und den Lebenszyklus von Nebel und niedrigen Stratuswolken. Die spezifischen Einflüsse variieren standortabhängig und auf zeitlichen Skalen: Einflüsse meteorologischer Faktoren dominieren auf kürzeren Zeitskalen (mehrere Tage bis Monate), während die Verwendung von Jahresmittelwerten oder die Filterung spezifischer Wettersituationen die meteorologischen Schwankungen verringert und die Analyse von Landoberflächeneinflüssen auf die FLS Bedeckung ermöglicht. Ein vielversprechender Ansatz ist die Verwendung von regionalen Modellregionen und von zusätzlichen Variablen, um Faktoren der FLS-Persistenz zu untersuchen. Vorläufige Ergebnisse einer solchen Studie für die Po-Ebene zeigen, dass meteorologische Faktoren hauptverantwortlich für die Persistenz von Nebel und niedrigem Stratus sind, wobei die Aerosolbelastung des Vortages die Persistenz bei starker Luftverschmutzung um bis zu 60 Minuten erhöht. Mithilfe von regional spezifischen Studien über die Auswirkungen von Umweltfaktoren auf den Lebenszyklus von Nebel und niedrigen Stratus können somit Rückschlüsse und Lehren für deren regionale Vorhersage, die Verkehrssicherheit und die Erzeugung von Solarenergie gezogen werden.

Contents

Abstract	i
Kurzfassung	v
List of Figures	xiii
List of Tables	xvii
List of Acronyms	xix
1 Introduction and general motivation	1
1.1 General Motivation	1
1.2 Scientific basis	5
1.2.1 A microphysical perspective: Fog droplet formation	5
1.2.2 A geographic perspective: Fog types	7
1.2.3 A life cycle perspective: The role of the atmosphere and the land surface for the radiation fog life cycle	9
1.3 Observing fog and low stratus cloud occurrence	13
1.4 Overview of suitable data sets	16
1.5 Scientific aims and outline	18
2 Meteorological and land surface drivers of fog and low stratus occurrence	25
2.1 Specific Motivation and Aim	25
2.2 Data and methods	26
2.2.1 Study area	26
2.2.2 Data sets	26
2.2.3 Machine learning model technique	28
2.2.4 Data preprocessing	29
2.2.5 Study design	30
2.3 Results	33

2.3.1	Model performance	33
2.3.2	Feature importance	37
2.3.3	Spatial patterns of relevant features	38
2.3.4	One-variable partial dependence	40
2.3.5	Two-variable partial dependence	44
2.4	Discussion	46
2.4.1	Model performance	46
2.4.2	FLS drivers and their temporal and spatial patterns	49
2.5	Specific conclusions	52
3	Land surface driven variations in fog and low stratus occurrence	55
3.1	Specific Motivation and Aim	55
3.2	Data and methods	56
3.2.1	Data sets	56
3.2.2	Fog and low stratus detection	57
3.3	Results	58
3.3.1	Cross-validation of fog and low stratus products	58
3.3.2	Climatological means	59
3.3.3	Seasonal analysis	60
3.3.4	Vertical temperature profiles	62
3.4	Discussion	64
3.5	Specific conclusions	65
4	Climatological patterns of fog and low stratus formation and dissipation times	67
4.1	Specific Motivation and Aim	67
4.2	Data and methods	68
4.2.1	Fog and low stratus data set	68
4.2.2	Methods	70
4.2.3	Case illustration: 2011-02-07	74
4.3	Results	77
4.3.1	Number and duration of FLS events	77
4.3.2	Most frequent formation and dissipation times	78
4.3.3	Regional study: Upper Rhine valley, Germany	82
4.4	Discussion	84
4.5	Specific conclusions	86

5 Identifying and understanding fog and low stratus formation and dissipation regimes	89
5.1 Specific Motivation and Aim	89
5.2 Data and methods	90
5.2.1 Fog and low stratus formation and dissipation time data set	90
5.2.2 Clustering input and algorithm description	91
5.3 Results and Discussion	93
5.3.1 Regional cluster overview	93
5.3.2 Sub-regional cluster analysis	96
5.4 Specific conclusions	100
6 Concluding discussion and outlook	103
6.1 Concluding discussion	103
6.2 Outlook	112
A1 Scientific Appendix of Chapter 2	117
A2 Scientific Appendix of Chapter 3	119
A3 Scientific Appendix of Chapter 4	129
A4 Scientific Appendix of Chapter 5	133
Bibliography	135
Further appendices	161
Original publication: Pauli et al. (2020)	161
Original publication: Pauli et al. (2022a)	177
Original publication: Pauli et al. (2022b)	187

List of Figures

1.1	Processes of the coupled land-atmosphere system (Jia et al., 2019a). . . .	2
1.2	MODIS Aqua Scene from 2022-01-15 over southern Germany displaying a large FLS field.	3
1.3	Potential processes leading to an enhancement of convective cloud cover over temperate forests based on Teuling et al. (2017).	4
1.4	Phase diagram of water from Lamb and Verlinde (2011).	6
1.5	Köhler curve for a droplet containing an ammonium sulfate particle with a dry diameter of 200 nm (McFiggans et al., 2006).	7
1.6	Types of fog based on Bruijnzeel et al. (2005)	8
1.7	Development of a deep adiabatic radiation fog event starting from pre-fog conditions (Smith et al., 2018).	10
1.8	FLS detection scheme by Egli et al. (2017), based on the SOFOS scheme by Cermak (2006).	14
1.9	Schematic illustration of the four research chapters.	23
2.1	Average fog and low stratus hours per day from 2006-2015 in the central European study area based on the product by Egli et al. (2017).	27
2.2	Schematic of the model workflow.	31
2.3	R^2 of training and test set over all grid sizes and seasons using either all pressure situations or only high-pressure situations.	34
2.4	Map of deviation from the mean R^2 over all model units (domain average) in %, a topographic map (based on Hijmans et al. (2005)) of the study area and a correlation of these two.	36
2.5	Mean 10x10 feature importance over all model units for all features and all seasons.	38
2.6	Most important feature per 10x10 model unit in the full-year run over all model settings.	39
2.7	Most important feature per 10x10 model unit for different seasons. . . .	40

2.8	Partial dependence plot showing the mean response in modeled FLS occurrence to changes in all input features over all seasons, for the model run using data with seasonality.	42
2.9	Two-variable partial dependence on FLS (hours day ⁻¹) for six feature combinations for the 10x10 full-year run including the seasonality.	45
3.1	Example validation of SEVIRI-based fields with the corresponding CALIPSO profiles.	59
3.2	Spatial distribution of climatological mean FLS cover over the study region.	60
3.3	Climatological mean FLS hours by month based on true positive and true negative observations, and the corresponding ERA5 land wind speed and wind direction (u and v wind components) on the respective days (mean from 0–6 UTC).	61
3.4	Profiles of mean hourly temperature gradient in K 25 mb ⁻¹ for the true positive and true negative days in August and September from 825 to 1000 mb along 44.25° N and -0.25° W.	63
4.1	Mean fog and low stratus cover in central Europe from 2006–2015 (data set by Egli et al. (2017)) and mean height above sea level.	69
4.2	Workflow for the identification of FLS formation and dissipation times with logistic regression using an artificial example.	71
4.3	Logistic regression curves for three formation situations and three dissipation situations.	74
4.4	Mean fog and low stratus cover in the Upper Rhine valley from 2006–2015 (data set by Egli et al. (2017)) and mean height above sea level.	75
4.5	Binary 15-minute FLS masks (data set by Egli et al. (2017)) from 09:00 to 13:00 UTC on 2011-02-07.	76
4.6	Logistic regression curves for the dissipation events on 2011-02-07 at three locations.	77
4.7	Number of formation events identified by the algorithm and the median of the duration of all FLS events in which both formation and dissipation times could be determined.	78
4.8	Most frequent formation (a) and dissipation times (b) over the entire study period (2006–2015).	80
4.9	Most frequent formation and dissipation time for DJF, MAM, JJA and SON from 2006-2015.	81

4.10	Most frequent formation and dissipation time in the Upper Rhine valley and the corresponding % of values contained in the most frequent formation and dissipation class for the complete time period (All months) and in spring (MAM).	83
5.1	Correlation map for mean surface pressure (MSP) with formation (upper) and dissipation (lower) time for the month of February.	92
5.2	Dendrogram of the hierarchical clustering procedure, displaying the relationship of clusters and their joining distance.	94
5.3	Map of formation and dissipation clusters identified by the hierarchical clustering algorithm.	95
5.4	Sub-regional cluster covering the Po valley and monthly sensitivities of FLS formation and dissipation time to meteorological and land surface drivers.	97
5.5	Sub-regional cluster covering the Northwest of England and Wales and monthly sensitivities of FLS formation and dissipation time to meteorological and land surface drivers.	98
5.6	Sub-regional cluster covering the Jucar basin on the east coast of Spain and monthly sensitivities of FLS formation and dissipation time to meteorological and land surface drivers.	100
6.1	Mean FLS cover in hours per day during the general weather situation “High over Central Europe” (HM) and “Anticyclonic southwesterly” (SWA).	105
6.2	Mean seasonal nighttime (0–6 UTC) LST from 2006–2015 over the Landes area.	106
6.3	SHAP “beeswarm” plot, showing the impact of each feature on the predicted FLS duration.	113
6.4	SHAP dependence plots showing a) the main effect of aod_{prev} on the prediction of FLS duration and b) the interaction effects with $t_{amplitude}$	114
A1.1	MSE of training and test set over all grid sizes and seasons using either all pressure situations or only high-pressure situations.	117
A2.1	Digital elevation map of the study area.	121
A2.2	Climatological mean FLS hours by month and the corresponding ERA5 land wind speed and wind direction (u and v wind components) for January, February, March and April.	122

A2.3 Climatological mean FLS hours by month and the corresponding ERA5 land wind speed and wind direction (u and v wind components) for June, October, November, December.	122
A2.4 Profiles of mean hourly inversion strength in K 25 mb ⁻¹ for the true positive and true negative days in January and February, from 825 to 1000 mb along 44.25° N and -0.25° W.	123
A2.5 Profiles of mean hourly inversion strength in K 25 mb ⁻¹ for the true positive and true negative days in March and April, from 825 to 1000 mb along 44.25° N and -0.25° W.	124
A2.6 Profiles of mean hourly inversion strength in K 25 mb ⁻¹ for the true positive and true negative days in May and June, from 825 to 1000 mb along 44.25° N and -0.25° W.	125
A2.7 Profiles of mean hourly inversion strength in K 25 mb ⁻¹ for the true positive and true negative days in July and October, from 825 to 1000 mb along 44.25° N and -0.25° W.	126
A2.8 Profiles of mean hourly inversion strength in K 25 mb ⁻¹ for the true positive and true negative days in November and December, from 825 to 1000 mb along 44.25° N and -0.25° W.	127
A3.1 Mean diurnal course of FLS cover in winter over London, Paris and Milan and their respective rural surroundings from 2006–2015.	129
A3.2 Normalized mean FLS cover, normalized median duration of FLS event and normalized number of formation events over the complete study period (2006–2015).	129
A3.3 Seasonal number of formation events identified by the algorithm.	130
A3.4 Fraction of daytime formation and dissipation events out of all events, in the Upper Rhine valley, Germany.	130
A3.5 Most frequent formation (a) and dissipation times (b) over the entire study period (2006–2015) using daytime and nighttime dissipation.	131
A3.6 Most frequent formation and dissipation time for DJF, MAM, JJA and SON from 2006–2015 using daytime and nighttime dissipation.	131
A4.1 Map of formation and dissipation clusters identified by the hierarchical clustering algorithm using nearest neighbor as an interpolation technique.	133

List of Tables

1.1	Overview of data sets used in this thesis.	17
2.1	Parameter grid for the grid search and final hyperparameters used for model training.	32
4.1	Day- and nighttime classes with respective % ranges.	79
5.1	Example layout of the table used as an input for the hierarchical clustering algorithm.	93
A2.1	Cross-validation of the satellite products.	119
A2.2	P values of the two sample t-tests, testing the difference between FLS cover over forest vs. non-forest areas.	120
A2.3	P values of the two sample t-tests, testing the difference between wind speed over forest vs. non-forest areas.	120

List of Acronyms

ABL	Atmospheric Boundary Layer
ALB	Albedo
AOD	Aerosol optical depth
AVHRR	Advanced Very High Resolution Radiometer
BVOCs	Biogenic volatile organic compounds
CAD	Cloud and aerosol discrimination
CALIPSO	Cloud-Aerosol LiDAR and Infrared Pathfinder Satellite Observations
CCN	Cloud condensation nuclei
CM SAF	Satellite Application Facility on Climate Monitoring
DEM	Digital elevation model
DJF	December, January, February: winter
ECMWF	European Centre for Medium-Range Weather Forecast
ERA	ECMWF Re-Analysis
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
ET	Evapotranspiration
FLS	Fog and low stratus
FLS_{prev}	Fog and low stratus on the previous day

FVC	Fraction of vegetation cover
GBRT	Gradient boosted regression trees
GOES	Geostationary Operational Environmental Satellite
HILDA	HIstoric Land Dynamics Assessment
HM	High over Central Europe
JJA	June, July and August: summer
LEO	Low earth orbiting
LES	Large eddy simulations
LiDAR	Light detection and ranging
LSA-SAF	Land-Surface Analysis Satellite Applications Facility or EUMETSAT Satellite Application Facility on Land Surface Analysis
LST	Land surface temperature
LULCC	Land use and land cover change
LW	Longwave
LWP	Liquid water path
MAM	March, April, May: spring
METAR	Meteorological Aviation Routine Weather Report
MOR	Meteorological Optical Range
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	Mean squared error
MSP	Mean sea level pressure
MTG	Meteosat Third Generation
NAN	Not a Number
NDSI	Normalized Difference Snow Index

RH	Relative humidity
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SHAP	SHapley Additive exPlanations
SOA	Secondary organic aerosols
SOFOS	Satellite-based Operational Fog Observation Scheme
SOM	Self organizing maps
SON	September, October, November: fall
SST	Sea surface temperature
SWA	Anticyclonic southwesterly
TLCFs	Tropical lowland cloud forests
u	u wind vector
UTC	Universal Time Coordinated
v	v wind vector
XGB	EXtreme Gradient Boosting
WS	Wind speed

1 Introduction and general motivation

1.1 General Motivation

Land surface and climate interact in complex ways through fluxes of water and energy, greenhouse gases and aerosols with interactions occurring across different spatial and temporal scales (Pielke and Avissar, 1990; Pielke, 2001; Pitman, 2003; Pielke et al., 2011; Jia et al., 2019a). They can be described as biophysical interactions (exchanges of water and energy) and biogeochemical interactions (emissions of greenhouse gases and aerosols) (Jia et al., 2019a). The former strongly depend on reflectivity (albedo) of the earth's surface, the emissivity of longwave radiation by vegetation and soils, surface roughness and soil water (Jia et al., 2019a). These factors vary spatially and temporally, among land cover and land use types (Anderson et al., 2011; Jia et al., 2019a) and in magnitude and direction, leading to warming or cooling of the earth (Duveiller et al., 2018) (Fig. 1.1). Apart from the current state of the land surface, land use and land cover change (LULCC) strongly impact these biophysical and biogeochemical interactions, especially in the atmospheric boundary layer (Pielke and Avissar, 1990; Pielke, 2001; Pielke et al., 2011) but magnitude and direction depend on geographic location and season (Jia et al., 2019a).

The redistribution of water and energy through differences in landscape structure and changes in land cover affect the horizontal and vertical gradients of temperature, pressure and moisture. This consequently leads to changes in wind patterns, moisture transport, temperature advection and convection, resulting in different cloud and precipitation patterns (Avissar and Liu, 1996; Jia et al., 2019a). A cloud positioned directly at the earth's surface and heavily linked to surface-atmosphere exchanges (Bergot et al., 2005) is fog. Fog affects everyday human life, as the timing of its formation and dissipation is important for traffic on land, at sea and in the air (Leigh et al., 1998; Pagowski et al., 2004), and is critical for the quality of the prediction of solar power production (Köhler et al., 2017). Fog is also a crucial water source for various ecosystems around the world (Bruijnzeel et al., 2005; Gottlieb

et al., 2019) and plays an important role in climate processes (Vautard et al., 2009). From the satellite perspective and for conceptual purposes, fog can be viewed as a cloud touching the ground and is thus frequently allocated to one category with low stratus: Fog and low stratus (FLS) (Cermak and Bendix, 2011). In this thesis the terms fog and FLS are used interchangeably.

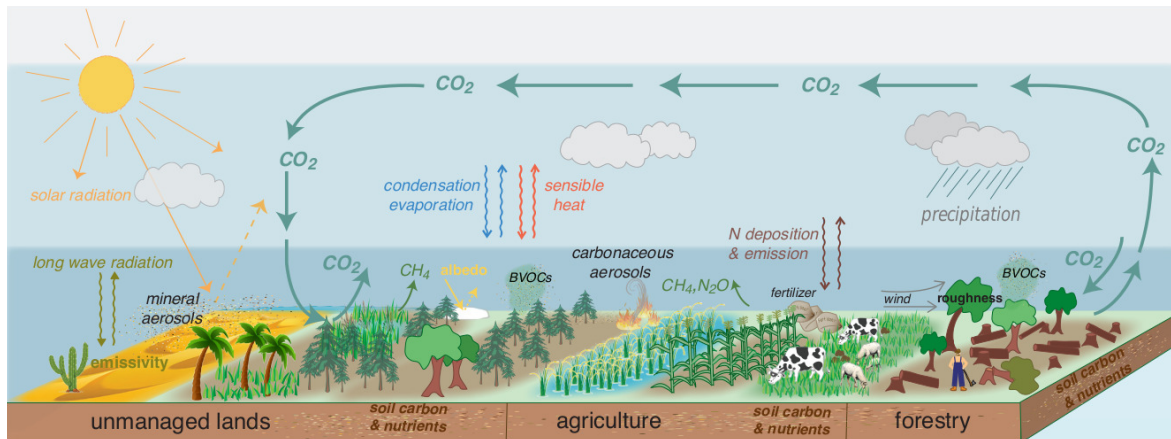


Figure 1.1: Processes of the coupled land-atmosphere system (Jia et al., 2019a).

In general, FLS is influenced by a myriad of atmospheric and land-surface processes: At large scales, the weather type (Bendix, 1994; Gultepe et al., 2007; Egli et al., 2019), particularly the pressure field (van Oldenborgh et al., 2010) influences FLS occurrence. On the local and small scale, low wind speeds, low temperatures and a stable boundary layer are beneficial for FLS formation (Cuxart and Jiménez, 2012; Pérez-Díaz et al., 2017; Price, 2019). Local to regional variations in FLS occurrence result from land surface effects on the FLS life cycle: FLS forms earlier in valleys whereas the mountain ranges stay FLS free (Fig. 1.2) (Bendix, 1994; Scherrer and Appenzeller, 2014). Higher temperatures over cities lead to a decrease in fog and low stratus occurrence over urban areas (Bendix, 1994; Sachweh and Koepke, 1995, 1997; Williams et al., 2015; Gautam and Singh, 2018; Izett et al., 2019; Fuchs et al., 2022). FLS can form earlier over fields compared to bare soil (Roach, 1995) and earlier over a homogeneous grass surface compared to a surface with trees (Mazoyer et al., 2017). A further important component of processes in the land-atmosphere system is the presence of aerosols (Fig. 1.1). Aerosols serve as cloud condensation nuclei (CCN) which act as the substrate on which water vapor condenses and fog droplets form (Ramanathan et al., 2001; Poku et al., 2019). Depending on the underlying land surface, different types of aerosols are present, for example mineral dust, black and organic carbon and biovolatile organic compounds (BVOCs). As the growth rate of

the fog droplet depends on the initial size of the aerosol and its solubility (Lamb and Verlinde, 2011; Poku et al., 2019), it is thus also indirectly dependent on the underlying land surface.

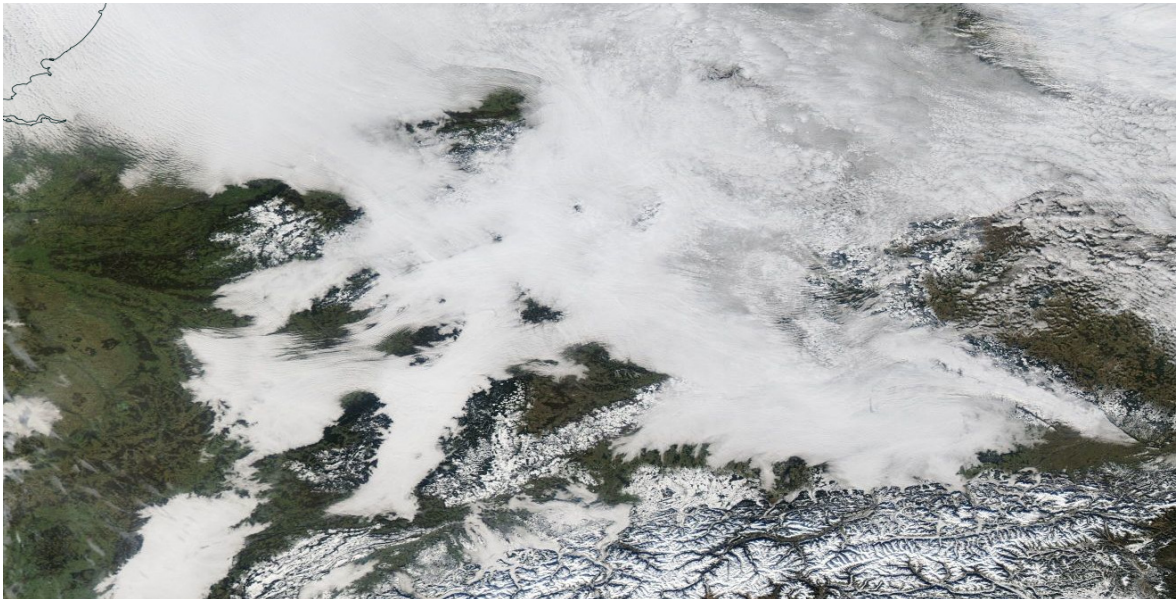


Figure 1.2: MODIS Aqua Scene from 2022-01-15 over southern Germany displaying a large FLS field (NASA Worldview, 2022).

While some insights on atmospheric and land surface effects on the FLS life cycle exist, these effects and their interactions are manifold, most likely vary across different temporal and spatial scales and are still loaded with considerable uncertainty. To identify potential important pathways between the land surface and FLS, patterns and processes of other boundary layer clouds, such as convective cumulus clouds, might give some insights. Convective cumulus clouds are, similar to FLS, in close proximity to the surface, thus pathways of land surface effects on convective cumulus clouds might be important for FLS as well.

In boreal and temperate regions, an enhancement of convective cumulus cloud cover over forests compared to the adjacent open land has been found (Teuling et al., 2017; Cerasoli et al., 2021; Petäjä et al., 2022; Xu et al., 2022). Possible pathways of the impact of those forests on clouds are the development of a mesoscale circulation (Avissar and Liu, 1996; Carleton et al., 2001; Bonan, 2008; Wang et al., 2009; Gambill and Mecikalski, 2011), a higher amount of cloud condensation nuclei (Spracklen et al., 2008; Petäjä et al., 2022), enhanced evapotranspiration (Petäjä et al., 2022) and the underlying moisture conditions (Rabin and Martin, 1996) (Fig. 1.3). Contrasting patterns have been detected over the tropics, in particular less convective cumulus cloud cover over forests and an enhancement over deforested areas (Durieux et al.,

2003; Wang et al., 2009; Xu et al., 2022). Here, the magnitude of the sensible heat flux has been found to be higher over open land than over forests, as in humid climates most available energy over forests goes into the latent heat flux (Xu et al., 2022). The stronger sensible heat flux over the non-forest area can lead to a development of a mesoscale circulation and an enhancement of cloud cover over deforested areas (Durieux et al., 2003; Bonan, 2008; Wang et al., 2009; Xu et al., 2022). Cumulus cloud cover enhancement has also been detected over cities (Theeuwes et al., 2019) and over agricultural areas (Ray et al., 2003). To sum up, the influence of the land-atmosphere interactions on convective cumulus clouds varies depending on region, season, land surface type and heterogeneity, water availability and the amount of BVOCs (Duveiller et al., 2021).

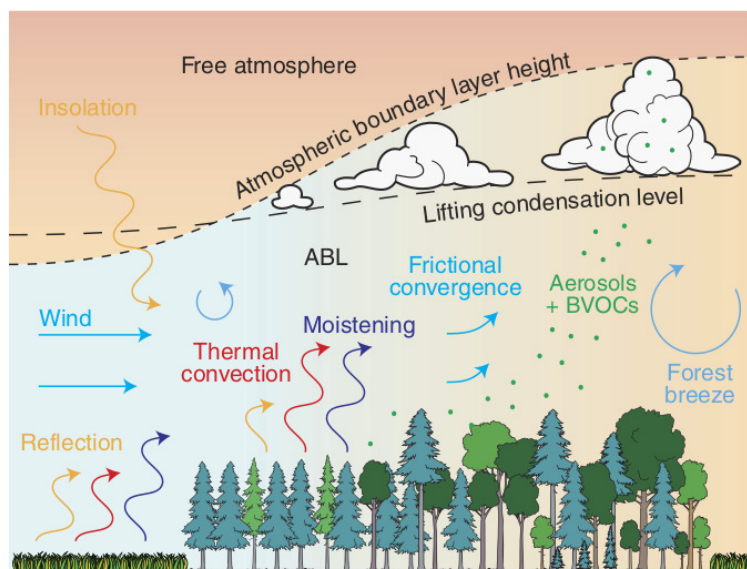


Figure 1.3: Potential processes leading to an enhancement of convective cloud cover over temperate forests based on Teuling et al. (2017). ABL is the atmospheric boundary layer.

The processes depicted for convective cumulus clouds are potentially important for FLS as well, but are most likely highly nonlinear and depend on the spatial and temporal scales considered. This thesis aims to quantitatively analyze FLS occurrence and life cycle in the context of land-atmosphere interactions across central Europe and over a large temporal extent (10 years). The following section 1.2 gives an overview on fog droplet formation, fog types and outlines the fog life cycle. The subsequent section (1.3) describes measurement techniques of fog and low stratus occurrence, with a focus on satellite observations, followed by a description of the applied data sets (section 1.4). Following the identification of open research questions derived from the existing literature, the scientific

aims are outlined in section 1.5. The chapters 2, 3, 4 and 5 contain the main research. Initially, the main atmospheric and land surface based drivers of FLS occurrence are identified and their individual and combined influences on FLS occurrence are analyzed (chapter 2). The focus of chapter 3 is the analysis of land-surface induced variations in FLS occurrence, particularly differences between forest and surrounding agricultural land. In chapter 4, a climatological analysis of FLS formation and dissipation time over central Europe provides the basis for investigations of the FLS life cycle in context of land-atmosphere interactions. The sensitivities of formation and dissipation times to environmental conditions and the regimes identified therein are analyzed in chapter 5. The findings of the research chapters are discussed and contrasted with the current literature in chapter 6 and an outlook on current and planned research is given.

1.2 Scientific basis

Fog is generally defined as a suspension of water droplets leading to visibilities < 1000 m (Glickman, 2000) and can be treated as a cloud touching the ground (compare Fig. 1.2). Thus, the microphysical processes of fog droplet formation are similar to those of cloud droplet formation, which are described in section 1.2.1. Differences in the evolution of different fog types and the radiation fog life cycle are outlined in section 1.2.2 and 1.2.3.

1.2.1 A microphysical perspective: Fog droplet formation

The basic requirements for fog formation, and all cloud formation in general, are the availability of moisture and aerosols and a process of cooling the air (Lamb and Verlinde, 2011). The latter is described by the *Clausius Clapeyron equation* (Manton, 1983) as the saturation vapor pressure of water e_s [hPa] strongly depends on temperature T [K]:

$$e_s(T) = e_0 * \exp\left(\frac{l_v}{R_v} * \left(\frac{1}{T_0} - \frac{1}{T}\right)\right) \quad [hPa] \quad (1.1)$$

(Manton, 1983; Lamb and Verlinde, 2011), where e_0 is the equilibrium vapor pressure (6.108 hPa), l_v [J kg⁻¹] is the latent heat of vaporization, R_v [J kg⁻¹ K⁻¹] is the gas constant for water vapor and T_0 is the freezing point temperature (273.15 K).

A phase diagram (Fig. 1.4) describes this relationship graphically, where the vapor pressure is plotted as a function of temperature. Both the saturation vapor pressure for ice (e_i) and for liquid (e_s) increase exponentially with temperature, as higher temperatures cause vibrations of water molecules, which break the bonds that hold water in its condensed phase (Lamb and Verlinde, 2011). Thus warmer air can take up more water vapor until it reaches saturation.

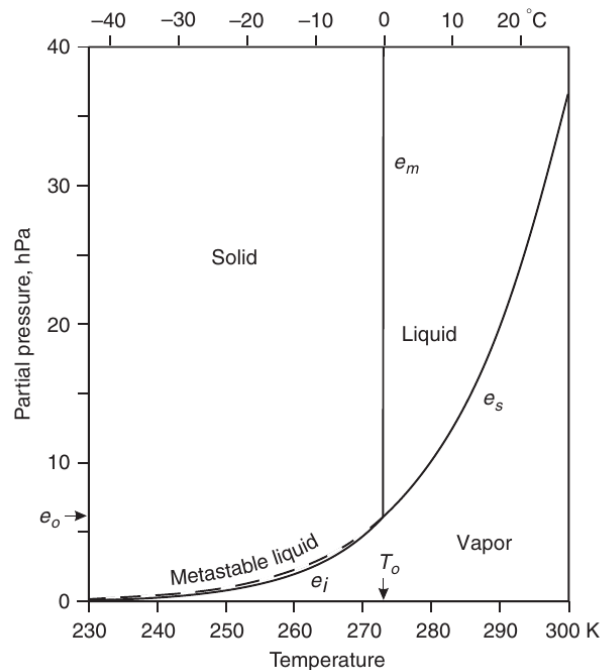


Figure 1.4: Phase diagram of water from Lamb and Verlinde (2011). e_s is the saturation vapor pressure of liquid water, e_i is the saturation vapor pressure of ice. T_0 depicts the freezing point temperature at 273.15 K, alas 0 °C with the respective saturation vapor pressure e_0 . e_m is the hypothetical vapor pressure of the melt water in equilibrium with ice.

As the Clausius-Clapeyron equation describes a plane water surface, it needs to be modified for cloud droplets by considering the effect of curvature (*Kelvin effect*) and the effects of solutes, i.e. aerosols (*Raoult effect*) (Manton, 1983). These effects are summarized in the *Köhler theory* and the combination of effects can be seen in a so called *Köhler curve* (Fig. 1.5). The *Kelvin effect* describes the increase in saturation vapor pressure due to the higher surface tension of droplets compared to a plane surface. The *Raoult effect* describes the decrease in saturation vapor pressure when soluble substances in an aerosol particle are present. This leads to a lower critical supersaturation and water vapor can condense even if the relative humidity does not exceed 100 % and a cloud droplet becomes activated (Lamb and Verlinde, 2011).

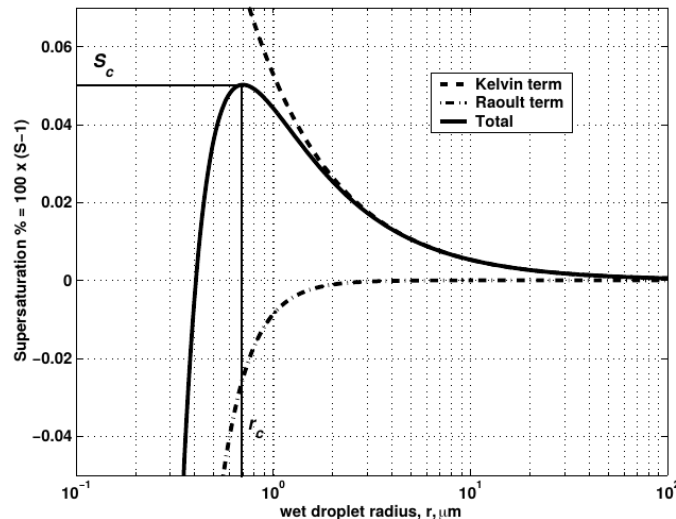


Figure 1.5: Köhler curve for a droplet containing an ammonium sulfate particle with a dry diameter of 200 nm (McFiggans et al., 2006). The contribution of both the Kelvin term (dashed line) and Raoult term (dot-dashed line) are shown. S_c is the critical supersaturation with the corresponding critical radius r_c . If one of these is exceeded, the cloud droplet becomes activated.

1.2.2 A geographic perspective: Fog types

The microphysical perspective on fog droplet formation has shown that an air parcel has to reach saturation for fog formation to occur. There are three main processes that lead to saturation, which dominate in the major fog types described below (Gultepe et al., 2007; Cotton et al., 2011):

- Cooling the air to its dewpoint (Radiation fog)
- Horizontal movement of air resulting in vertical mixing of moist air parcels with different temperatures (Advection fog)
- Addition of water vapor to the air (Frontal fog)

While the main process for the specific fog type might dominate, all processes described above (cooling, mixing, addition of water) occur in most fogs. The three fog types listed above, as well as specific sub-types which are most common in central Europe and relevant for the presented research are presented in the following.

Radiation fog is the most frequent fog type in central Europe (Fuzzi et al., 1992; Gultepe et al., 2007) and occurs mostly in the lower flatland regions of Europe (Egli et al., 2019). It forms through radiative cooling of the ground, a build-up of a temperature inversion and subsequent condensation of the excess water vapor

on aerosol particles present, and is sustained by radiative heat loss at the fog top (Roach, 1995; Bruijnzeel et al., 2005; Cotton et al., 2011; Pérez-Díaz et al., 2017). Clear skies, low wind speeds and anticyclonic conditions are beneficial for these processes (Gultepe et al., 2007; Pérez-Díaz et al., 2017). Geographically, radiation fog occurs more frequently in valleys (*valley fog*), compared to mountainous areas, as shown by Scherrer and Appenzeller (2014) for the Swiss Plateau and Fuzzi et al. (1992) and Bendix (1994) for the Po valley. Radiation fog occurs mostly in fall and winter, when the conditions described above are present (Bendix, 2002; Bruijnzeel et al., 2005; Cermak et al., 2009; Egli et al., 2017). *Orographic fog*, or *upslope fog*, can theoretically be considered as a sub-type of radiation fog (Cotton et al., 2011) since adiabatic cooling occurs as an air parcel rises along a mountain slope which results in cooling of the air to its dewpoint (Bruijnzeel et al., 2005).

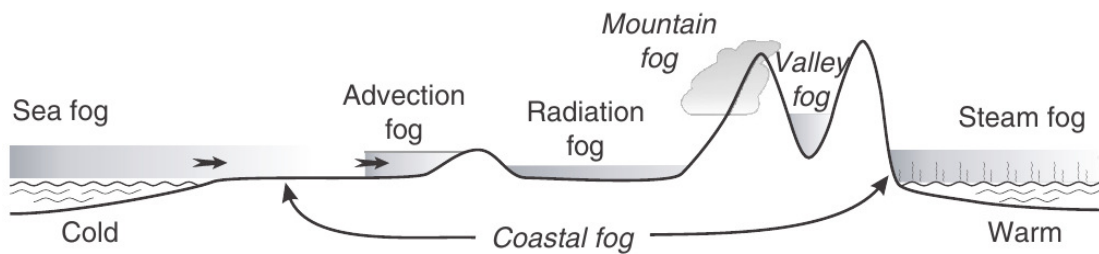


Figure 1.6: Types of fog based on Bruijnzeel et al. (2005).

Advection fog forms when a moist air mass is advected over a surface with a different temperature, resulting in vertical mixing and condensation (Gultepe et al., 2007; Cotton et al., 2011; Pérez-Díaz et al., 2017). In contrast to radiation fog formation, steady winds are necessary for the formation of advection fog (Bruijnzeel et al., 2005) and it is usually thicker than radiation fog and can persist for days. An example of an advection fog is *sea fog*, when humid air moves over a cold water surface and is thus cooled to its dew point (Bruijnzeel et al., 2005; Gultepe et al., 2007; Eugster, 2008). *Steam fog* can also be considered a type of advection fog, as it forms when cold air flows over warm water and condensation occurs since the cold air can not take up the evaporated water fast enough (Bruijnzeel et al., 2005; Eugster, 2008).

Mountain fog is a cloud which is being advected onto a mountain, with the land-surface positioned at or above the lifting condensation level and thus the cloud touches the ground (Eugster, 2008). This type of fog is of great importance for tropical montane cloud forests where it strongly affects the hydrologic, nutrient and pollution budget (Bruijnzeel, 2001; Bruijnzeel et al., 2005; Berry et al., 2015; Weathers et al., 2019).

Frontal fog forms, when water is added to a near saturated layer, for example by precipitation from an overlying layer (Cotton et al., 2011). Frontal fog occurs in warm and cold fronts and is strongly coupled to the large-scale weather situation (Egli et al., 2019).

While all fog types described above occur in central Europe, the focal point of this thesis is (lifted) radiation fog, as it is strongly coupled to the surface. To focus on radiation fog, specific filtering methods are applied in the research chapters. In the following, the life cycle of radiation fog is described.

1.2.3 A life cycle perspective: The role of the atmosphere and the land surface for the radiation fog life cycle

The development of a deep adiabatic radiation fog event can be subdivided into three stages, as shown in figure 1.7. The formation of radiation fog usually begins after sunset, when radiative cooling sets in and a temperature inversion develops near the surface (Roach, 1995; Smith et al., 2018) (Fig. 1.7a). Initially, the condensate thereby formed is deposited as dew due to turbulence (Roach, 1995; Cermak and Bendix, 2011) but as the inversion deepens, the static stability suppresses turbulence and fog droplets begin to form (Roach, 1995; Smith et al., 2018). This state of fog is referred to as shallow stable radiation fog (Smith et al., 2018; Toledo et al., 2021) (Fig. 1.7b) and contains a small amount of liquid water (Toledo et al., 2021). It can last up to 10 hours (Price, 2019) and usually dissipates after sunrise (Price, 2011).

When a substantial fraction of aerosols activate into fog droplets, the fog becomes optically thick and the primary location of radiative cooling is at the fog top (Roach, 1995; Price, 2011; Boutle et al., 2018; Smith et al., 2018; Poku et al., 2019) (Fig. 1.7c). The radiation inversion is detached from the ground (Roach et al., 1976) and the surface warms due to a positive heat flux from the soil (Price, 2011). Turbulence levels increase and the cooled air from the top of the fog layer sinks down to the bottom, which leads to the development of an adiabatic temperature profile (Roach et al., 1976; Price, 2011; Smith et al., 2018; Poku et al., 2019). This deepening takes approximately 2 hours (Smith et al., 2018) and the resulting deep adiabatic fog layer can persist throughout the day for up to 24 hours and more (Price, 2019). A development from shallow stable radiation fog to deep adiabatic radiation fog occurs in approximately 50 % of all cases (Price, 2011).

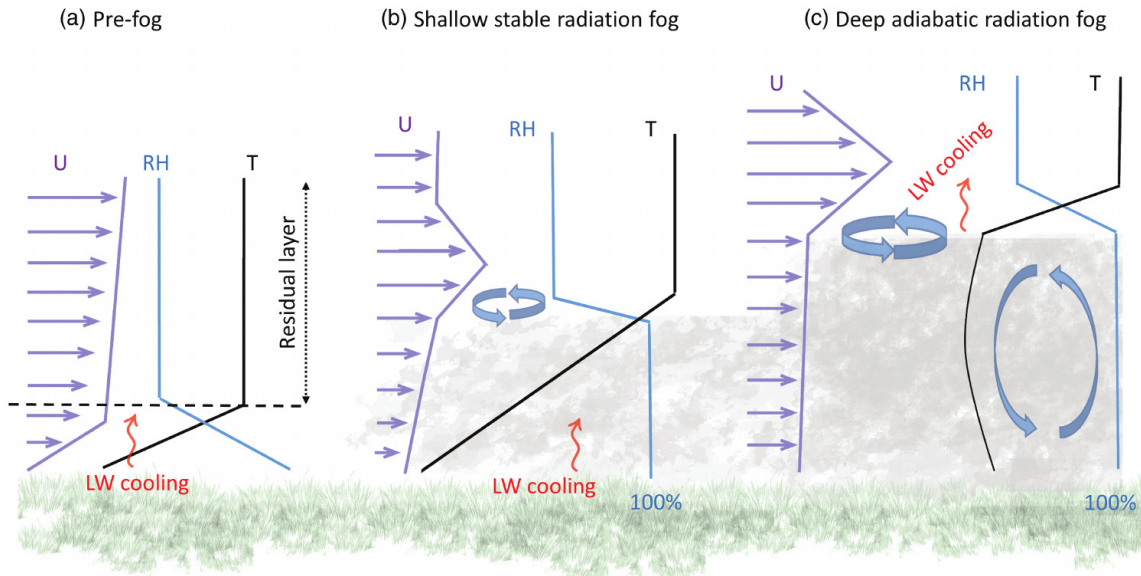


Figure 1.7: Development of a deep adiabatic radiation fog event starting from pre-fog conditions (Smith et al., 2018). The red arrows depict the region of largest longwave (LW) cooling, U is the u -wind component, T the temperature, RH the relative humidity. The blue arrows show turbulent mixing.

Dissipation of radiation fog can occur through various pathways: After sunrise, the increasing solar radiation leads to a “burn-off” of fog due to warming of air and mixing of the boundary layer (Roach, 1995; Haeffelin et al., 2010; Maalick et al., 2016). Shortwave radiative warming can also lead to a transition of fog into a low stratus cloud via fog dissipation at the ground and lifting of the stratus layer (Maalick et al., 2016). Turbulent heat fluxes can contribute strongly to the loss of liquid water content in the fog layer and thus to fog dissipation (Wærsted et al., 2019). Furthermore, higher wind speeds can lead to the erosion of fog at its top (Bergot, 2016), and high clouds moving over the fog can reduce radiative cooling of the cloud top, resulting in dissipation of the fog layer (Roach, 1995).

The role of the atmosphere

The radiation fog life cycle is the result of non-linear interactions of simultaneously occurring processes (Haeffelin et al., 2010), with interactions varying along the temporal course of the fog life cycle (Haeffelin et al., 2013). Along its life cycle, radiation fog can be influenced by multiple meteorological factors. In general, low temperatures in combination with sufficient moisture are beneficial for fog formation (Underwood et al., 2004; Ye, 2009; Cuxart and Jiménez, 2012; Pérez-Díaz et al., 2017; Boutle et al., 2018; Gray et al., 2019; Price, 2019; Mühlig et al., 2020).

A temperature decrease in the evening together with a dry layer above the boundary layer strengthens radiative cooling (Underwood et al., 2004; Haeffelin et al., 2010, 2013). The importance of temperature and moisture for fog occurrence can also be seen over urban areas, where an increase in temperature and a decrease in moisture leads to a reduction of fog and fog holes (Williams et al., 2015; Klemm and Lin, 2016; Gautam and Singh, 2018; Gray et al., 2019; Yan et al., 2020).

Low wind speeds and a stable boundary layer further improve the conditions for radiation fog formation (Cuxart and Jiménez, 2012; Pérez-Díaz et al., 2017; Price, 2019). The influence of wind direction greatly depends on the specific location and its topography (Wrzesinsky and Klemm, 2000; Błaś et al., 2002; Klemm and Wrzesinsky, 2007). In addition, the distribution of the atmospheric pressure fields as well as general weather situations associated with strong high pressure fields can be beneficial for fog formation (Ye, 2009; Egli et al., 2019).

The role of the land surface

Still, the short-term variability in meteorology can not explain the long-term trends in fog frequency (Gray et al., 2019), thus, other non-meteorological drivers have to be considered. A driver situated at the interface of the atmosphere and the land surface are aerosols, which can serve as CCN in fog (Roach, 1995; Haeffelin et al., 2010; Maalick et al., 2016; Boutle et al., 2018; Poku et al., 2019; Mühlig et al., 2020). When the relative humidity increases due to radiative cooling, aerosol particles grow hygroscopically into fog droplets (Haeffelin et al., 2010; Boutle et al., 2018). A higher amount of aerosols and thus available CCN has been shown to increase fog lifetime by delaying its dissipation (Maalick et al., 2016) and increase the likelihood of a deep adiabatic radiation fog (Poku et al., 2019). In addition, through aerosol-radiation interactions, aerosols can induce a more stable atmosphere, leading to favorable conditions for the accumulation of pollutants and fog formation (Gao et al., 2015).

On geographic scales, topography and landform strongly influence FLS patterns by modifying the influence of atmospheric dynamics on FLS occurrence, leading to a higher FLS cover in valleys compared to mountainous areas (Bendix, 1994; van Oldenborgh et al., 2010; Scherrer and Appenzeller, 2014). Gravity flows generated by mountain ranges influence the variability of wind and temperature at the local scale and thus the presence of fog (Cuxart et al., 2021). Furthermore, advection slightly above the fog layer, which is strongly linked to topography, controls the growing phase of radiation fog (Cuxart and Jiménez, 2012). Besides topography, surface induced turbulent mixing, soil temperature, conductivity and moisture

fluxes influence the fog life cycle (Maronga and Bosveld, 2017; Steeneveld and de Bode, 2018; Adhikari and Wang, 2020). Dew evaporation can lead to a longer persistence of fog (Pilié et al., 1975) and higher relative humidity values over irrigated cultivated areas can potentially induce fog formation (Montecinos et al., 2008). Similarly, an earlier onset of fog formation has been observed over fields compared to bare soil (Roach, 1995).

The formation of fog is also sensitive to the surface type as the location of larger roughness elements influences the local wind, temperature and humidity field (Roach, 1995). Large buildings have been found to delay fog formation (Bergot et al., 2015) and trees have been found to elevate fog formation to higher levels as the induced drag mixes the lower levels and fog water is deposited onto the trees (Mazoyer et al., 2017). In addition, an increase in forest area and water availability can lead to an increase in fog occurrence, but their influence is considerably lower than the influence of altitude or seasonality (Hůnová et al., 2021a, 2022).

Changing the perspective to investigate the influence of fog on the land surface shows that fog supplies water and nutrients to various ecosystems around the world, such as coastal deserts (Ebner et al., 2011; Roth-Nebelsick et al., 2012; Eckardt et al., 2013; Lehnert et al., 2018; Mitchell et al., 2020), tropical montane cloud forests (Berry et al., 2015), Californian redwood forests (Dawson, 1998) and temperate grasslands during dry spells (Li et al., 2021). Furthermore, fog can influence the productivity in cultivated crops by modifying radiation, e.g. leading to a higher water use efficiency in strawberries on foggy days (Baguskas et al., 2018).

While meteorology, topography and seasonality strongly influence the fog life cycle, additional factors, such as aerosols and the land surface, should not be left out of consideration. Interactions between these factors are complex and vary both in time and space. Studies on drivers of factors influencing the fog life cycle should ideally look at a combination of these drivers simultaneously to analyze their contributions in a quantitative manner.

1.3 Observing fog and low stratus cloud occurrence

From the ground, fog presence is traditionally measured by determining visibility, as per definition, fog is present if visibility drops below 1000 m (Glickman, 2000). The meteorological quantity to measure visibility is the transparency of the atmosphere or the Meteorological Optical Range (MOR), which is measured using transmissometers (WMO, 2018). Besides visibility measurements, the cloud base height and cloud top height, as well as vertical dynamics, can be determined using cloud radar and ceilometer measurements (Haeffelin et al., 2010). These ground-based measurements of fog and low stratus have a high accuracy and high temporal resolution, but they suffer from drawbacks such as sparse spatial sampling and high maintenance.

Compared to ground-based measurements, satellite data can provide information on fog over larger areas not covered by weather stations and has a large potential for automation. It has to be considered though, that from the satellite perspective, fog is frequently defined as a cloud touching the ground and a discrimination between (ground) fog and elevated fog (low stratus) is not possible with most sensors. Satellite-based fog studies thus consider fog and low stratus as one category, which, as previously indicated, summarizes fog and low stratus as FLS.

Early satellite-based fog detection algorithms have been created for low earth orbiting (LEO) satellite systems such as the Advanced Very High Resolution Radiometer (AVHRR) sensor using the brightness temperature difference between the IR3.7 μm and IR11 μm channel (Eyre et al., 1984; Bendix and Bachmann, 1991; Bendix, 2002). These channels are used since land or sea have the same brightness temperature in both the IR3.7 μm and IR11 μm channel, but fog has a lower emissivity in the IR3.7 μm channel and thus a lower brightness temperature (Turner et al., 1986). Similarly, fog can be detected during the day using a combination of solar and infrared bands from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the LEO-satellites Terra and Aqua (Bendix et al., 2006).

A drawback of using sensors on LEO-satellites, such as AVHRR and MODIS, is their low repeat rate, thus, the observation of the fog life cycle is not possible. Consequently, fog detection algorithms have been developed for geostationary satellites, which provide a high temporal resolution (15-minute repeat rate) and cover a large area. The first fog detection algorithm on a geostationary system has been implemented on the Geostationary Operational Environmental Satellite (GOES) system using the channel differences between the IR3.9 μm and IR11 μm channel at night (Ellrod, 1995) similar to the AVHRR fog detection algorithm. More

extensive algorithms have been developed for the geostationary Meteosat Spinning Enhanced Visible and Infrared Imager (SEVIRI), and have been successfully used for the detection of fog and low stratus in Europe (Cermak, 2006; Cermak et al., 2009; Cermak and Bendix, 2011; Egli et al., 2017) and the Namib desert (Andersen and Cermak, 2018).

The FLS product for Europe by Egli et al. (2017) is based on the Satellite-based Operational Fog Observation Scheme (SOFOS) algorithm by Cermak (2006). The SOFOS algorithm uses several threshold tests as depicted in figure 1.8 to detect pixels covered by fog or low stratus clouds. The SEVIRI data used as an input for the algorithm has a temporal resolution of 15 minutes and a spatial resolution of 3 km at nadir.

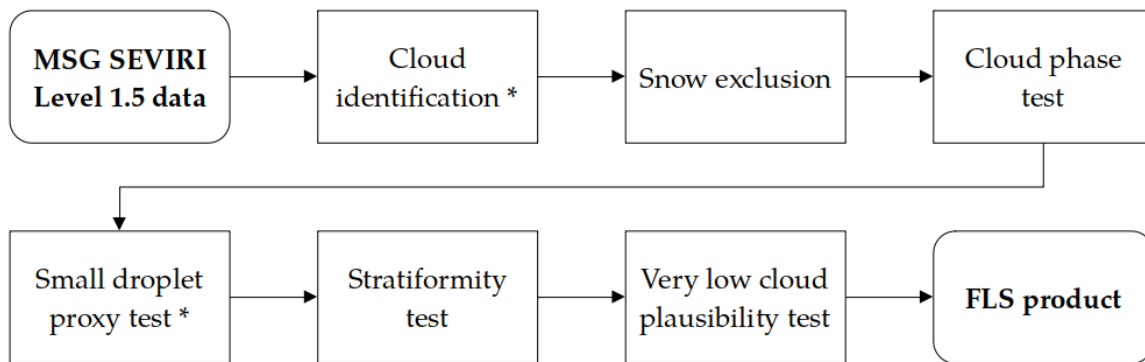


Figure 1.8: FLS detection scheme by Egli et al. (2017), based on the SOFOS scheme by Cermak (2006). Steps marked with a * have been adapted by Egli et al. (2017). The original illustration of the scheme can be found in Egli et al. (2017)

In a first step, cloudy pixels are separated from clear sky pixels by using the reflectivity properties in the $3.9\ \mu\text{m}$ and $10.8\ \mu\text{m}$ channel, as cloudy pixels have a higher reflectivity in the $3.9\ \mu\text{m}$ channel. Pixels covered by snow are excluded using the Normalized Difference Snow Index (NDSI) (Dozier, 1989). In the subsequent cloud phase test, non-liquid clouds are removed by making use of the difference in blackbody temperatures between the $8.7\ \mu\text{m}$ and $12.0\ \mu\text{m}$ channel. This difference is smaller for ice clouds, as ice absorbs much stronger in the $12.0\ \mu\text{m}$ channel. Then pixels with large sized cloud droplets are removed by using the emissive characteristics in the $3.9\ \mu\text{m}$ channel. The stratiformity of the cloud layer is eventually tested using the standard deviation of brightness temperature in the $10.8\ \mu\text{m}$ channel before a plausibility check is conducted which compares the brightness temperature of each cloud with the one of its surrounding cloud-free pixels.

The resulting FLS data set shows good agreement with Meteorological Aviation Routine Weather Reports (METAR), provides a new level of spatial detail and agrees well with previous studies. However, classification errors, small-scale FLS features and multiple cloud layers can lead to misclassifications (Cermak and Bendix, 2008; Cermak, 2018). These classification errors can be minimized when using active satellite data such as LiDAR (Light Detection And Ranging) data from CALIPSO (Cloud-Aerosol LiDAR and Infrared Pathfinder Satellite Observations) which also makes it possible to discriminate between ground fog and elevated fog, as information on the cloud base height is available (Vaughan et al., 2009; Cermak, 2018). Despite its low temporal sampling rate, CALIPSO data is still highly valuable for the study of fog and low cloud patterns, especially when combined with passive satellite data. While patterns of FLS found with passive and active satellite data are similar, about one quarter of FLS situations are not detected by passive satellite data due to multi-layer cloud situations (Cermak, 2018). Nevertheless, FLS data sets based on geostationary satellite data currently provide the best combination of temporal and spatial resolution, which makes them ideal for the study of FLS occurrence and life cycle on the continental scale.

1.4 Overview of suitable data sets

The choice of the data sets for this thesis is primarily motivated by the research questions but also inherently determined by the spatial and temporal coverage, as well as data accuracy. The study region of this thesis is central Europe, but the analysis is conducted at three different spatial scales: Regional (continental Europe, chapter 2), sub-regional (Western France, chapter 3) and continental (central Europe, chapter 4 and 5). At these scales, satellite and re-analysis data sets are the ideal choice due to their spatial coverage, as well as similar spatial and temporal resolution. It is further beneficial to use data sets derived from the same satellite sensor to ensure the same spatial and temporal resolution, since spatial and temporal resampling always introduces a certain level of uncertainty to the data. Table 1.1 provides an overview of the data sets used in this thesis.

The basis of this thesis is the FLS data set by Egli et al. (2017) which is currently available for 10 years (2006–2015). From this data set, a novel FLS formation and dissipation time data set is derived (Pauli et al., 2021). As the FLS data set by Egli et al. (2017) is based on passive satellite data it allocates fog and low stratus in one category. In the thesis at hand and in the analyses conducted therein, the results thus relate to both fog and low stratus processes. Nevertheless, filtering methods are applied to specifically concentrate on fog events. The results are discussed focusing on fog processes while considering also the impact of treating both fog and low stratus as one category.

For the quantitative analyses in the research chapters, the FLS data set is combined with other satellite data sets and with reanalysis data. Data on land surface temperature (LST), evapotranspiration (ET), fraction of vegetation cover (FVC) and albedo (ALB) is taken from the Land-Surface Analysis Satellite Applications Facility (LSA-SAF) and is based on Meteosat SEVIRI. The CALIPSO cloud layer product is based on the LiDAR aboard the polar-orbiting CALIPSO system. Meteorological data is taken from the ERA5-land and ERA5-pressure reanalysis data sets. As their spatial resolution differs from the FLS data, they are rescaled to SEVIRI resolution using nearest neighbor. This is also done for the HILDA (Historic Land Dynamics Assessment) land cover data.

Table 1.1: Overview of data sets used in this thesis.

Data set & Publication	Parameter	Chapter
FLS data set, Egli et al. (2017)	Fog and low stratus occurrence	2, 3, 4
FLS formation and dissipation time data set, Pauli et al. (2021)	Fog and low stratus formation and dissipation time	5, 6
CALIPSO level 2 1 km cloud layer product, NASA/LARC/SD/ASDC (2018)	Cloud top height, Cloud base height	3
ERA5-land, Muñoz Sabater (2019)	Surface pressure, Temperature, Specific humidity, u and v wind components	2, 3, 5, 6
ERA5-pressure, Hersbach et al. (2018)	Temperature, Specific humidity, u, v and w wind components	3, 6
LSA-SAF, Trigo et al. (2011)	Land surface temperature, Evapotranspiration, Fraction of vegetation cover, Albedo	2, 3*, 5
WorldClim, Hijmans et al. (2005)	Height above sea level (Digital Elevation Model: DEM)	2, 3, 4
HILDA, Winkler et al. (2021)	Land cover	4
CM SAF, Clerbaux et al. (2017)	Aerosol optical depth	6

*Fraction of vegetation cover map as background used in chapter 3

1.5 Scientific aims and outline

The fog life cycle is influenced by various meteorological and land surface drivers as shown in chapter 1.2.3. While local process studies of the influences of different drivers on the fog life cycle in Europe exist (e.g. Haeffelin et al., 2010; Cuxart and Jiménez, 2012; Steeneveld and de Bode, 2018; Price, 2019; Wærsted et al., 2019), their spatial and temporal scales considered are small. In addition, studies on how the land surface modulates meteorological conditions and thus fog occurrence are lacking. This is of high importance, as LULCC as well as climate change modulates the land surface, leading to land-atmosphere feedbacks whose magnitude and direction still have considerable uncertainty as feedbacks vary in time and space (Li et al., 2015; Jia et al., 2019a; Xu et al., 2022).

Therefore, the primary target of this thesis is to quantify the influence of meteorological and land surface drivers on FLS occurrence and life cycle over central Europe. A regional analysis of these drivers is complemented by a site specific analysis of the influence of specific land cover types on FLS occurrence. To explicitly analyze the fog life cycle, a novel FLS formation and dissipation data set is developed and FLS formation and dissipation time regimes are investigated. By combining different data sets, the application of different methods and the development of new data sets, different processes of fog and low stratus can be evaluated and specific patterns found in central Europe can be analyzed closely. The described objectives are targeted in the following research chapters:

1. Meteorological and land surface drivers of fog and low stratus occurrence
2. Land surface driven variations in fog and low stratus occurrence
3. Climatological patterns of fog and low stratus formation and dissipation times
4. Identifying and understanding fog and low stratus formation and dissipation regimes

The research questions and associated hypotheses are presented below and are summarized in figure 1.9.

(1) Meteorological and land surface drivers of fog and low stratus occurrence

Local measurement campaigns and modeling studies have identified meteorological drivers, such as wind speed and temperature, and land surface drivers, such as topography and soil moisture content, as important determinants of FLS occurrence (Bendix, 1994; Cuxart and Jiménez, 2012; Scherrer and Appenzeller, 2014; Pérez-Díaz et al., 2017; Steeneveld and de Bode, 2018; Price, 2019; Cuxart et al., 2021). Still, analyses over a large spatial and temporal scale and studies considering the interaction of these drivers are lacking. The analysis presented in chapter 2 provides an overview of the main drivers of FLS occurrence over continental central Europe, laying the foundation for further site-specific or process-specific analyses in the subsequent chapters. The main research question of this chapter is as follows:

What are the main drivers of large-scale spatial and temporal fog and low stratus patterns?

The guiding hypothesis is that meteorological drivers are the most important determinants of fog and low stratus occurrence. It is further hypothesized that land surface drivers are more important in spring and summer than in fall in winter, as they are strongly influenced by the growing season.

The following more specific research questions are derived from the main research question:

- *What are the magnitude and direction of the main drivers of FLS patterns?*
- *What are spatial and temporal patterns of these influences?*
- *What are the combined effects of multiple drivers on FLS occurrence?*

These research questions are targeted by constructing a machine-learning model to predict daily FLS occurrence over 10 years over continental central Europe. As the amount of data is large and interactions between FLS and its drivers are non-linear and complex, machine-learning is an ideal tool for such a multivariate analysis (Olden et al., 2008; Lary et al., 2016). Spatially explicit model units and seasonal models are created, which are quantitatively analyzed and compared considering their predictive power and sensitivities. In the study, daily means of meteorological and land surface drivers based on satellite and reanalysis data are used.

(2) Land surface driven variations in fog and low stratus occurrence

After the preceding analysis of the main drivers of FLS occurrence, this chapter presents a site-specific analysis of land surface effects on fog and low stratus occurrence. Land surface effects on cloud cover have been found for convective clouds, with variation both in time and space (Teuling et al., 2017; Theeuwes et al., 2019; Xu et al., 2022), but no study looking at the land surface effects on fog and low stratus clouds has been conducted to date. Thus, the following research question is the foci point of chapter 3:

How does the land surface influence spatial variations in fog and low stratus occurrence?

The guiding hypothesis is that when meteorological and topographic effects are minimized, the land surface significantly influences FLS occurrence. This should lead to a higher FLS occurrence over land cover types with lower wind speed, lower temperatures and a higher availability of CCN. The analysis is specifically conducted for a large forest area and the surrounding agricultural land.

On this basis, the following specific questions are addressed:

- *How does FLS occurrence vary spatially over forest compared to surrounding agricultural land?*
- *Are there seasonal variations?*
- *What are potential drivers of the patterns found?*

To answer the outlined questions, nighttime (0–6 UTC) FLS cover over a forested area in western France is compared with the surrounding agricultural land. To decrease the influence of misclassifications on the results, two independent FLS classifications, based on passive (Egli et al., 2017) and active satellite data (NASA/LARC/SD/ASDC, 2018; Cermak, 2018) are used. The influences of wind speed and temperature are analyzed and discussed.

(3) Climatological patterns of fog and low stratus formation and dissipation times

While the preceding research chapters focused on the analysis of mean FLS occurrence, the FLS life cycle, specifically its formation and dissipation time are the central point of chapter 4. Formation and dissipation time are usually analyzed using ground-based measurements and modeling studies (Duynderke, 1991; Haeffelin et al., 2010; Dupont et al., 2012; Steeneveld and de Bode, 2018; Wærsted et al., 2019; Karimi, 2020). A satellite-based analysis of FLS formation and dissipation time spanning a large spatial and temporal scale does not exist. By creating a novel FLS formation and dissipation time data set and analyzing the climatological patterns, the research presented in this chapter lays the foundation of the further analysis of drivers influencing the observed formation and dissipation time patterns. The following main research question is targeted:

What are the climatological patterns of fog and low stratus formation and dissipation time?

The guiding hypothesis is that the spatial patterns of fog and low stratus formation and dissipation time show a strong dependency on topography and the distance to the coast. Local to regional modulations of the meteorological situation further modify the observed timing of FLS formation and dissipation.

The following specific research questions are addressed:

- *What are the spatial and seasonal patterns of the timing of FLS formation and dissipation?*
- *How do these patterns relate to the geography of the study area?*

To obtain the FLS formation and dissipation time, logistic regression is applied to the existing FLS data set by Egli et al. (2017) spanning central Europe. The resulting FLS formation and dissipation times are converted into distinct day- and nighttime classes relative to sunrise and sunset and the analysis of climatological patterns of FLS formation and dissipation time is based on these classes. The spatial detail of the novel data set is shown in a regional case study over southern Germany.

(4) Identifying and understanding fog and low stratus formation and dissipation regimes

While FLS occurrence can be related to different meteorological and land surface based drivers (chapter 2), the influence of these drivers on FLS formation and dissipation processes varies in magnitude and direction across central Europe and different FLS types. Investigations of the influence of environmental conditions on the FLS life cycle are usually region- and FLS type specific. A further comparison of those site specific FLS life cycle sensitivities to other regions is lacking, especially on the continental scale. Fog and other cloud occurrence patterns have been grouped into specific regimes on the regional (Knerr et al., 2021), continental (Egli et al., 2019) and global (Douglas and Stier, 2021) scale using different clustering approaches on either FLS occurrence patterns (Egli et al., 2019; Knerr et al., 2021) or cloud controlling factors (Douglas and Stier, 2021). Transferring these approaches to group areas of similar sensitivities of FLS formation and dissipation to meteorological and land surface conditions allows to analyze the FLS life cycle in respect to FLS type, background climate and geographic location. In this research chapter, the following main research question is targeted:

What are fog and low stratus formation and dissipation regimes on regional to sub-regional scales?

The guiding hypothesis is that the sensitivities of the identified FLS regimes to changes in meteorological and land surface conditions depend on background climate and FLS type.

The following specific research questions are derived from the main research question:

- *What are the underlying regime-specific sensitivities of FLS formation and dissipation time to changes in meteorological and land surface conditions?*
- *How do these sensitivities vary depending on geographic position?*

The FLS formation and dissipation regimes are extracted by applying a hierarchical clustering algorithm to monthly correlations of FLS formation and dissipation time with meteorological and land surface variables. This is done for each land pixel over central Europe, making it possible to create distinct geographic clusters of similar sensitivities of FLS formation and dissipation time to variations in environmental conditions. The geographic position and relation to the background climate is

analyzed for the obtained FLS life cycle regimes on different hierarchy levels. Sub-regional FLS life cycle regimes are investigated according to their cluster-average monthly sensitivities of FLS formation and dissipation time to meteorological and land surface conditions. Possible FLS type specific formation and dissipation pathways are discussed.

The four main objectives constitute the four research chapters of this thesis, which aim to answer the research questions stated above. A schematic illustration of the research chapters is shown in figure 1.9. The results of the individual chapters are merged and discussed in chapter 6, followed by a conclusion and outlook for future work. The research presented in this thesis has already been published (chapter 2-4) or is ready for peer-reviewing (chapter 5) in scientific journals. The original publications are attached at the end of this thesis.

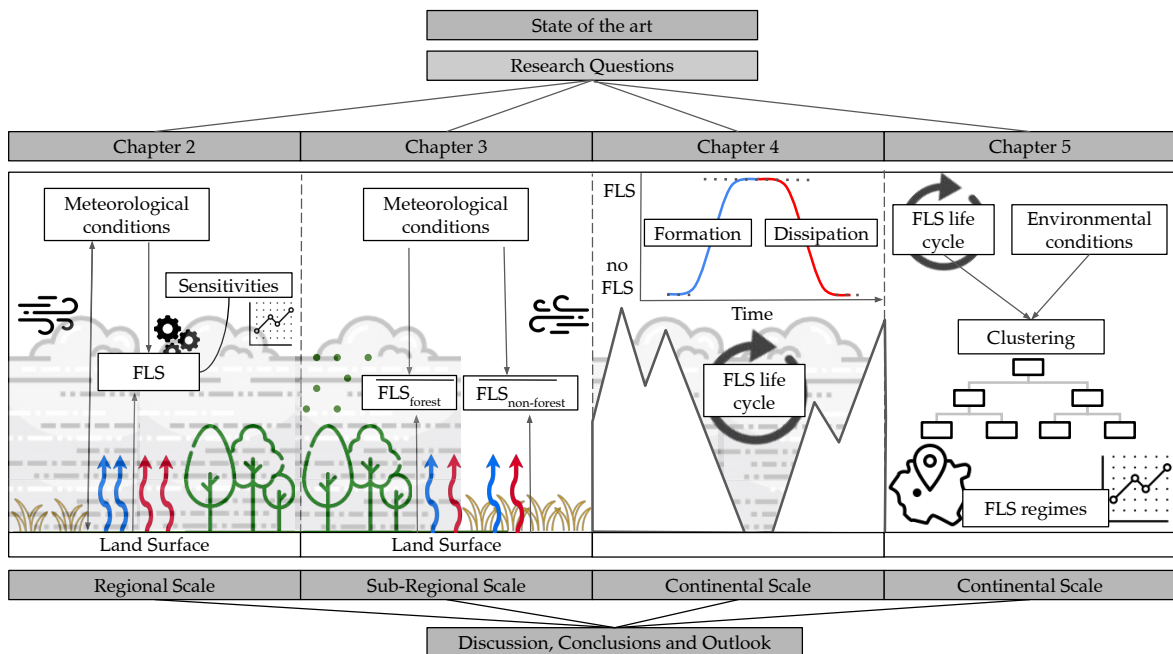


Figure 1.9: Schematic illustration of the four research chapters. The blue and red arrows depict moisture and energy fluxes from the land surface. The black arrows illustrate the wind and the green dots illustrate BVOCs. The length of the arrows is not to scale.

2 Meteorological and land surface drivers of fog and low stratus occurrence

2.1 Specific Motivation and Aim

FLS processes occur across a large range of different scales, from small scales of aerosol activation (10^{-7} m) to synoptic scales (10^6 m). These processes are influenced by various drivers relating to atmospheric and land surface processes (as described in chapter 1.2.3). Low temperatures, low wind speeds and a stable boundary layer are drivers beneficial for fog formation (Cuxart and Jiménez, 2012; Pérez-Díaz et al., 2017; Price, 2019), leading to a peak in fog and low stratus occurrence in winter and fall over continental Europe (Cermak and Bendix, 2007, 2008; Egli et al., 2017). The atmospheric drivers are further modulated by the land surface, specifically through influences of topography and landform (van Oldenborgh et al., 2010; Price, 2019; Ball and Tzanopoulos, 2020). Advected fog can accumulate in steep windward slopes (Ball and Tzanopoulos, 2020) and lower advected fog densities are found in leeward locations (Torregrosa et al., 2016; Ball and Tzanopoulos, 2020). The formation and persistence of radiation fog can also be influenced by topographically induced drainage flows and turbulence (Price, 2019). In addition, soil temperature and moisture (Maronga and Bosveld, 2017), as well as the availability of aerosols that can act as cloud condensation nuclei (Ramanathan et al., 2001; Stolaki et al., 2015), influence fog formation and dissipation.

While investigations about temporal and spatial FLS patterns over Europe exist (Cermak, 2006, 2018; Cermak and Bendix, 2007, 2008; Haeffelin et al., 2010; Egli et al., 2017), as well as many insights on specific local and non-local processes determining FLS formation and development (Roach, 1995; Haeffelin et al., 2010; Bergot, 2016), no explicit satellite-based investigation of the determinants of FLS development spanning multiple years and a large spatially coherent region exists to date.

The aim of this study is to quantitatively determine the relative importance of several land-cover and meteorological parameters for FLS occurrence in a large central European domain. This is done by using a machine learning technique, investigating its predictive performance concerning FLS occurrence, as well as analyzing model sensitivities. The high occurrence frequencies of FLS in central Europe during winter provide a good database for conducting this analysis. This study aims to unravel the question of what the main drivers of large-scale spatial and temporal FLS patterns in central Europe are and how these drivers influence FLS patterns. The guiding hypothesis is that meteorological drivers are the most important determinants of fog and low stratus occurrence in central Europe. It is further hypothesized that land surface drivers influence FLS occurrence more in spring and summer than in fall and winter due to their distinct seasonal cycle and close relation to the vegetation cycle.

2.2 Data and methods

2.2.1 Study area

The study is conducted in a large domain in continental Europe, from 48° N to 53° N and 5° E to 15° E including parts of the countries Austria, Belarus, Belgium, Czech Republic, France, Germany, Luxembourg, Netherlands, Poland, Slovakia and Ukraine (see Fig. 2.1). This continental study area was chosen to allow for a focus on continental situations unimpacted by the effects of local circulations such as land-sea and mountain breeze systems. Thus, large mountain ranges, such as the Alps and areas close to the coast, were excluded. Nonetheless, small local topographic differences and smaller topography-induced circulations still exist in the presented study area. The effects of such local-scale modulations on FLS occurrence and FLS-land surface interactions are reduced by subdividing the study area into a number of smaller units. In this study, two model unit sizes are tested (10x10 and 15x15 SEVIRI pixels), which are further described in chapter 2.2.5.

2.2.2 Data sets

The FLS product used in this study presented by Egli et al. (2017) is based on geostationary satellite data from the Meteosat Second Generation platforms Meteosat 8, 9 and 10. The SEVIRI system on board of the satellites scans the full hemisphere every 15 minutes. The sub-satellite resolution is 3 km in 11 spectral

bands with an additional high-resolution visible channel with 1 km resolution. The FLS product is a modification of the Satellite-based Operational Fog Observation Scheme (SOFOS) by Cermak (2006) and covers the years 2006 to 2015. To produce a data set of daily FLS occurrence, the original 15-minute product is locally averaged into FLS hours day⁻¹, which serves as the target quantity in this study. The mean FLS distribution for the chosen study area over the complete time period can be seen in figure 2.1.

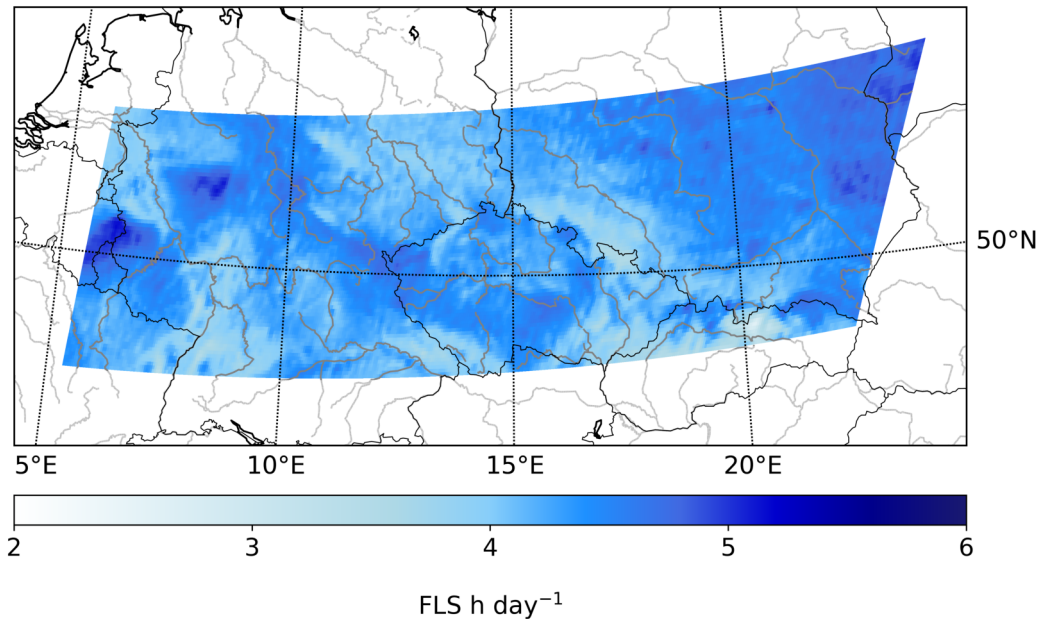


Figure 2.1: Average fog and low stratus hours per day from 2006-2015 in the central European study area based on the product by Egli et al. (2017).

To describe the physical state of the land surface and its possible influence on FLS distribution, several land surface features are included in this study, based on products from the Land-Surface Analysis Satellite Applications Facility (LSA-SAF) (Trigo et al., 2011). All of these products are created using data from Meteosat SEVIRI and thus have the same spatial resolution as the FLS product. To describe temperature and moisture availability of the surface, land surface temperature (LST) and evapotranspiration (ET) are included. Additional land surface characteristics relevant for energy and moisture fluxes are described by the fraction of vegetation cover (FVC) and the bi-hemispherical albedo for the total shortwave range (ALB). To describe atmospheric conditions, ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) is used. ERA5 is the follow-up of ERA-interim (Dee et al., 2011) and provides higher spatial resolution (0.25°) (Hersbach, 2016) which is especially useful for investigating land-atmosphere

interactions. In this study, mean sea level pressure (MSP) and the u and v component of wind at 10 meters are included. The u and v component of the wind are combined to calculate the near surface wind speed (WS). As a measure for the persistence of meteorological conditions that can lead to FLS, the FLS value of the previous day (FLS_{prev}) is included.

Additional information on the height above sea level, geographical position or land cover type were tested but are not included in the final model, as they did not lead to marked improvements in model performance and, equally to a location ID, could lead to overfitting (Meyer et al., 2018, 2019). Meyer et al. (2019) further assume that using predictor variables that describe the spatial location prevent the model from making reliable spatial predictions. As such, the final models are based on 7 predictors (ALB, ET, FLS_{prev} , FVC, LST, MSP, WS) to predict the duration of FLS cover.

2.2.3 Machine learning model technique

Interactions between the atmosphere and the land surface are often highly dimensional. With machine learning, non-linear and complex multivariate problems can be handled (Olden et al., 2008; Lary et al., 2016). Especially for problems where theoretical knowledge is limited but the amount of available observational data is large, machine learning can be an ideal tool for addressing these problems (Lary et al., 2016). In this study, a model was constructed using the machine learning technique Gradient Boosting Regression Trees (GBRT), aiming at predicting FLS hours per day on the basis of the meteorological and surface predictors described above. GBRTs use an ensemble of weak base learners (decision trees) which follow the negative gradient of the loss function (Friedman, 2001; Natekin and Knoll, 2013). They are flexible when using categorical and numerical data and have large predictive power, being able to represent nonlinear relationships between predictors and the predictand (Hastie et al., 2001). Furthermore, one of the key strengths of this machine learning technique is its high level of interpretability, since a number of interpretation methods exist for GBRTs (Elith et al., 2008). These properties have been exploited in recent studies to analyze complex atmospheric systems with GBRTs (Fuchs et al., 2018; Stirnberg et al., 2020). For this study, the GBRT implementation of the scikit-learn library in python was used (Pedregosa et al., 2011).

2.2.4 Data preprocessing

In a first step (Fig. 2.2 I)), the ERA5 data sets are rescaled to the SEVIRI pixel resolution of 3x3 km at nadir of the FLS data set using a nearest-neighbor approach. As described above, wind speed at 10 m height is calculated using the u and v components of the wind at 10 m height. Furthermore, daily averages are calculated for FLS, LST, ET, MSP and WS. FVC and ALB are already in the form of daily means. Daily averages are used, as the focus lies on the day-to-day variability of FLS occurrence.

Poor quality data and data with high error values, e.g. due to adjacent cloud-covered areas, is excluded using the following data quality control steps. For FVC, only data that is free of snow and water and with reliable input ranges for the FVC algorithm is included, as well as data with an absolute error smaller than 0.2. For ALB, data with an absolute error greater than 0.2 is excluded. For ET, data where input variables for the ET algorithm have insufficient quality or are missing, is excluded. For LST, cloud-filled pixels or pixels partly contaminated by clouds, snow or ice are excluded. Since some cloud-covered or cloud adjacent pixels with either implausible high or low surface temperatures are not caught by the quality flag, they are removed by setting the valid LST range depending on the month. For the full year, the valid data range is set to -60 °C to 60 °C. This range is then adjusted depending on the time of year: the maximum is adjusted to +40 °C from October to March and the minimum to -50 °C from March to September. With this, plausible LST data at the upper or lower ends of the temperature range can be kept in each season. The data excluded in the quality control as described above are marked as invalid in each data set. Any such flagging leads to the exclusion of the data point, as the model only considers data points with valid data in all data sets.

Since seasonal influence on FLS as well as on numerous other parameters is large, the models are run with two seasonality settings: 1) The seasonality is kept in the data and 2) the seasonal cycle is subtracted from the data. These two seasonality settings are used to investigate to what extent model performance and sensitivities are dependent on seasonal effects. When removing the seasonality, only the day-to-day variability is modeled, whereas keeping the seasonality in the data also models the seasonal interactions of FLS and its predictors. Here, the seasonality is defined as the mean over the investigated time period for every pixel for every day of the year and then smoothing this time series using a Savitzky-Golay filter (Savitzky and Golay, 1964). The seasonality is then subtracted from the data sets of each specific feature.

2.2.5 Study design

The study area is subdivided into spatially explicit model units, in which data is treated summarily at the SEVIRI resolution for training, testing and sensitivity analyses (compare Fig. 2.2 II)). Two spatial sizes of the model units are tested, with model units containing either 10x10 or 15x15 SEVIRI pixels. Due to the high latitudinal position and the rather shallow satellite viewing angle of 45° in the study area, one SEVIRI pixel corresponds to about a width of 4 km and a height of 7 km (EUMETSAT, 2013). For the 10x10 model unit, this leads to a mean model unit area of roughly 2800 km² and 6300 km² for the 15x15 unit, respectively. This model unit set-up is chosen to account for regional differences in the study area, while still choosing a unit size where some variability in land cover is present. Furthermore, this set-up allows for the evaluation of spatial differences in model performance and sensitivities. Thus, the 15x15 pixel units contain more data for training and testing, whereas the 10x10 units provide smaller and thus more homogeneous model units. To account for seasonal differences, seasonal model runs are conducted as well as full-year runs. These seasonal runs are winter (December, January and February: DJF), spring (March, April and May: MAM), summer (June, July and August: JJA) and fall (September, October and November: SON). In the considered time periods (full-year or seasons), separate models are trained using either all data or only the 20 % highest pressure situations in each model unit to focus on situations where FLS is favored. The pressure threshold is calculated from the distribution of pressure values for every model unit and model time period separately.

In each model unit, data is split into separate training and test data sets, containing 70 % and 30 % of the data, respectively. To ensure that training and test data contain comparable FLS distributions, the test-training split is coupled with a stratification of the FLS data set. In general, the number of available data points varies from model unit to model unit e.g. due to cloud contamination e.g. for LST and FVC data, but can be higher than 400000 (full-year, all pressure situations, 15x15 pixel model unit size), but also as low as 3800 (seasonal model, with pressure filter in a 10x10 pixel model unit).

Considering different seasonality and pressure filters, model unit sizes and modeling time periods (seasons or full-year), 30 different model settings are investigated in this study.

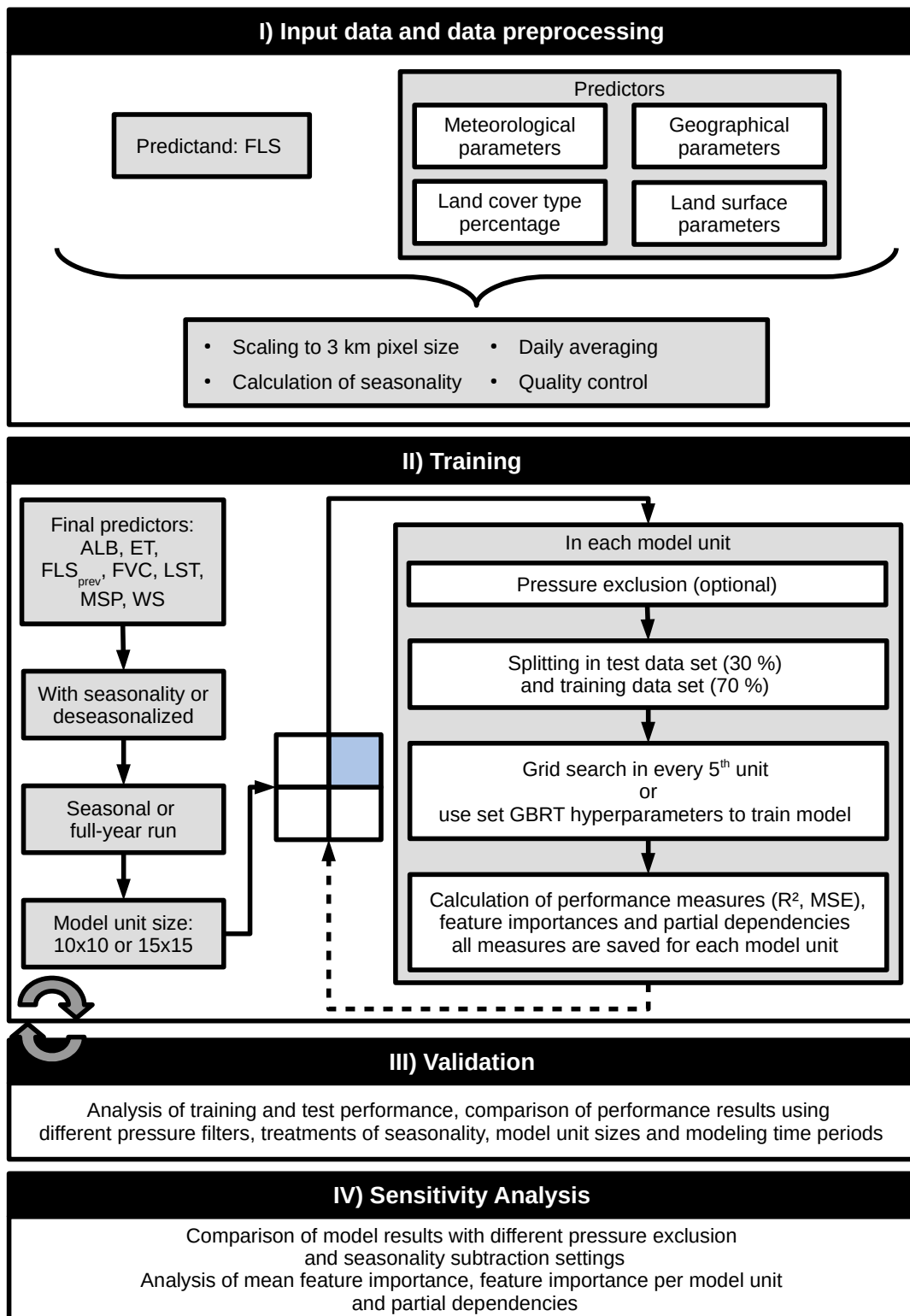


Figure 2.2: Schematic of the model workflow. Step II and III are conducted multiple times for different settings concerning pressure filters, treatments of seasonality, model unit sizes and modeling time periods.

Hyperparameter tuning

Hyperparameter tuning is conducted using grid search with a small range of possible values to avoid overfitting, e.g. by using only low values for the maximum depth (Tab. 2.1). Since grid search is computationally expensive, it is conducted in every 5th model unit over a range of feature combinations and model settings to investigate the influence on the hyperparameters. The hyperparameters proposed in the grid search are similar over model units, feature combinations and model settings, and the most common hyperparameters are used as the final settings.

Table 2.1: Parameter grid for the grid search and final hyperparameters used for model training.

	Number of estimators	Learning rate	Maximum depth	Minimum samples per leaf
parameter grid	500, 1000, 1500, 2000, 2500, 3000	0.1, 0.05, 0.03, 0.02, 0.01	2, 3	10, 14, 18
final settings	3000	0.1	3	10

Validation and sensitivity analysis

To test the influence of different model settings on model performance and sensitivities, the validation (Fig. 2.2 III)) and sensitivity analyses (Fig. 2.2 IV)) are conducted for all model settings described above. Model performance is evaluated using the coefficient of determination (R^2) and the mean squared error (MSE) between observed and predicted FLS in each model unit. The validation results are then evaluated concerning the differences between settings for each unit separately and over all units. Potential overfitting is analyzed by comparing R^2 and MSE for the test and training data sets. Correlations between variables are quantified using the Spearman's rho coefficient to account for effects of outliers and non-linearity in the investigated correlation.

To investigate the most important features for model performance, the permutation feature importance is calculated for all model units. The permutation feature importance measures the increase in prediction error after permuting (randomly shuffling) one feature (Breiman, 2001). This breaks the relationship between the feature and the target, thus indicating how much the model depends on that feature

(Strobl et al., 2007; Molnar, 2019). Compared to the often used impurity based feature importance (gini importance), the permutation importance is not as strongly biased towards variables with many categories and the continuous variable (Strobl et al., 2007). Thus, the term feature importance relates to permutation importance in this paper. The mean feature importance is calculated over all model units, but is also investigated for each model unit separately. To improve comparability, the permutation feature importance is scaled, so the sum of permutation importance over all features equals 1.

While the feature importance only displays the relative importance of a feature, the sign and nonlinearity of the predictand response relative to changes of each predictor variable can be analyzed using the partial dependence. In the context of this paper, the partial dependencies quantify the average change of the predicted FLS values relative to either one or two features (one-variable partial dependence and two-variable partial dependence) while accounting for the average effects of all other variables (Friedman, 2001; Elith et al., 2008; Fuchs et al., 2018). The partial dependence is calculated by gridding the investigated feature and calculating the corresponding average FLS prediction while the complement features are varied over their distribution (Molnar, 2019).

2.3 Results

2.3.1 Model performance

As a first overview of model performance, GBRT performance during training and validation with independent data is analyzed for all different model unit sizes as well as seasonality and pressure settings. The results of this validation are shown in figure 2.3. R^2 values are generally higher when filtering for high pressure situations, but are similar for runs containing seasonality versus runs that use deseasonalized data. The mean R^2 values of the model units in the 15x15 runs tend to be lower than in the 10x10 runs. The highest mean R^2 (averaged over all model units) is found in the pressure-filtered, deseasonalized 10x10 spring run (0.94), but is similarly high (~ 0.93) in other runs (e.g. in the pressure-filtered 15x15 winter run with seasonality). A clear seasonal pattern in model performance exists, with skill generally lowest in summer and highest in winter, albeit this seasonal pattern is weakened when only high pressure situations are considered. The difference in R^2 between the model runs with all pressure values (Fig. 2.3 A & C) versus those considering only high pressure situations (Fig. 2.3 B & D) is therefore lowest in

winter. The full-year models feature the lowest R^2 of runs without a pressure filter (mean R^2 0.6–0.7), but also reach values ~ 0.9 in the pressure-filtered runs. The regional variability of model performance, expressed as the spread in R^2 values, is highest in winter and for the full-year run, and lowest in spring and fall. The training R^2 is significantly ($p < 0.01$) higher than the test R^2 indicating some level of overfitting in the statistical models. Depending on the season, the training R^2 is on average slightly (4–8%) higher over all model units, and smaller than that in the 15x15 and full-year models. Still, the high test R^2 shows the ability of the models to generalize relationships between the predictors and FLS patterns learned during training. The difference between test and train R^2 is higher in the 10x10 model units and lowest in the full-year runs. The model predictions do not feature a noticeable bias (mean bias = 0.0, median bias ~ -0.3).

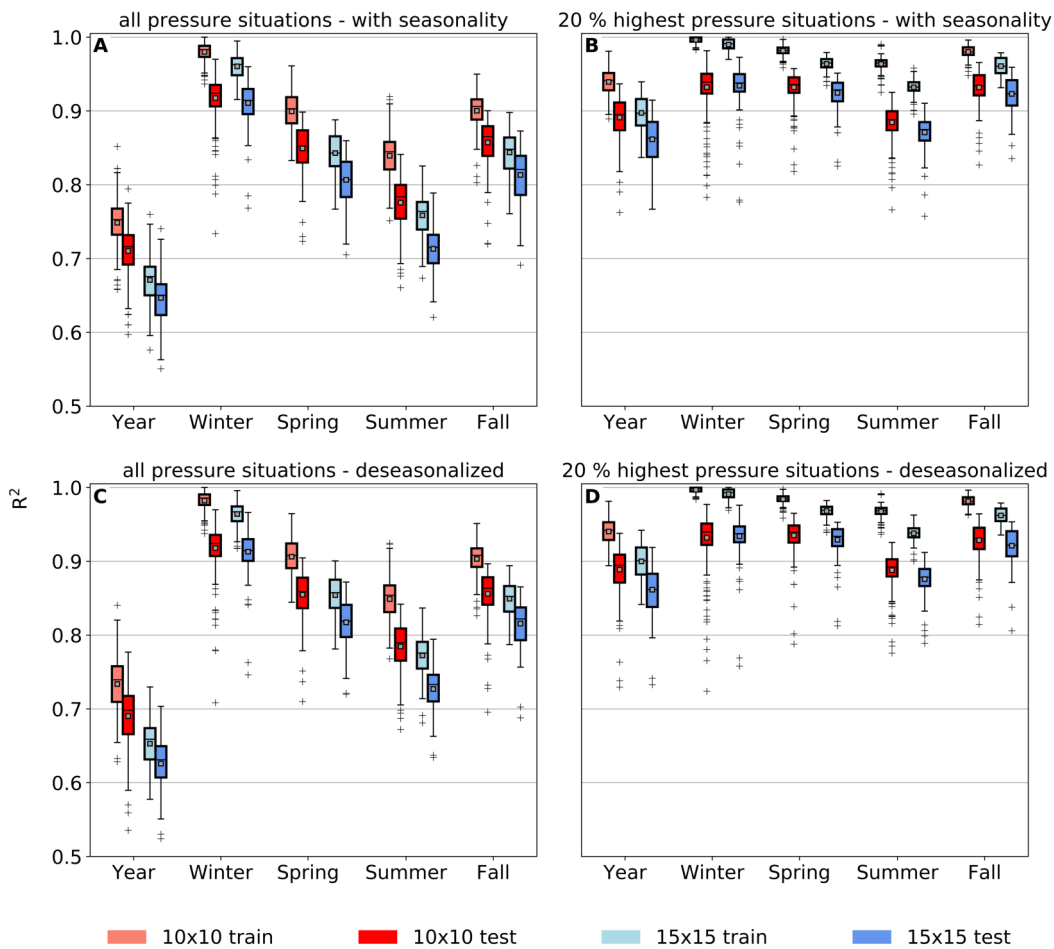


Figure 2.3: R^2 of training and test set over all grid sizes and seasons using either all pressure situations (left) or only high-pressure situations (right). The top row shows the results using the data with seasonality, while the bottom row displays the model results using deseasonalized data.

MSE patterns are similar to R^2 , with higher MSE values (thus worse performance) in the 15x15 runs and a larger difference of MSE between seasons than between model runs with different seasonality subtraction settings (Fig. A1.1). This means that model performance is more dependent on the modeled season than on seasonality subtraction settings. As the MSE is dependent on the absolute values, MSE patterns are influenced by the seasonal FLS distribution, thus leading to high MSE values in winter and fall, and lower MSE in spring and summer.

To investigate the reasons for overfitting, the difference between training and test R^2 and MSE is correlated with the number of data points available for training and testing. For the correlation with the absolute difference between training and test R^2 the Spearman's rho is approximately -0.5 ($p < 0.01$) especially in the full-year and winter run. For the MSE correlation, the Spearman's rho is -0.5 ($p < 0.01$) in the full-year, spring and fall runs. Thus, overfitting is partly controlled by the number of available data points, and models generalize better when more data points are available for training and testing.

Due to the good overall performance in the 10x10 model units and the higher spatial resolution, only the results of the 10x10 model units are presented in the following sections. Where not specifically pointed out, results of the 15x15 model units are similar, though.

Figure 2.4 shows spatial patterns of R^2 relative to the domain average R^2 for two different model settings (full-year models including seasonality with (Fig. 2.4 A) and without (Fig. 2.4 B) pressure filter). This highlights model units that either have distinctly better or worse performance than the overall mean. Additionally, in figure 2.4 C a topographic map is shown for the study area, together with two scatterplots of the R^2 deviation with the mean height per model unit (Fig. 2.4 D) for values west and east of 10° E separately. The corresponding Spearman's rho, p-value and R^2 for the complete correlations (negative and positive R^2 deviation values together) are added in the top right corner of the (Fig. 2.4 D) plot.

The domain average R^2 for the model run including all pressure situations is 0.71, for the model run using only high pressure situations it is 0.89. While the spatial patterns of skill are similar for both model settings, the models using only high-pressure situations feature a lower deviation from the domain average R^2 . Generally, in the north-eastern part of the model region, e.g. in Poland and the central Czech Republic and the south-eastern and north-eastern parts of Germany, the R^2 in the model units shows a positive deviation from the domain average.

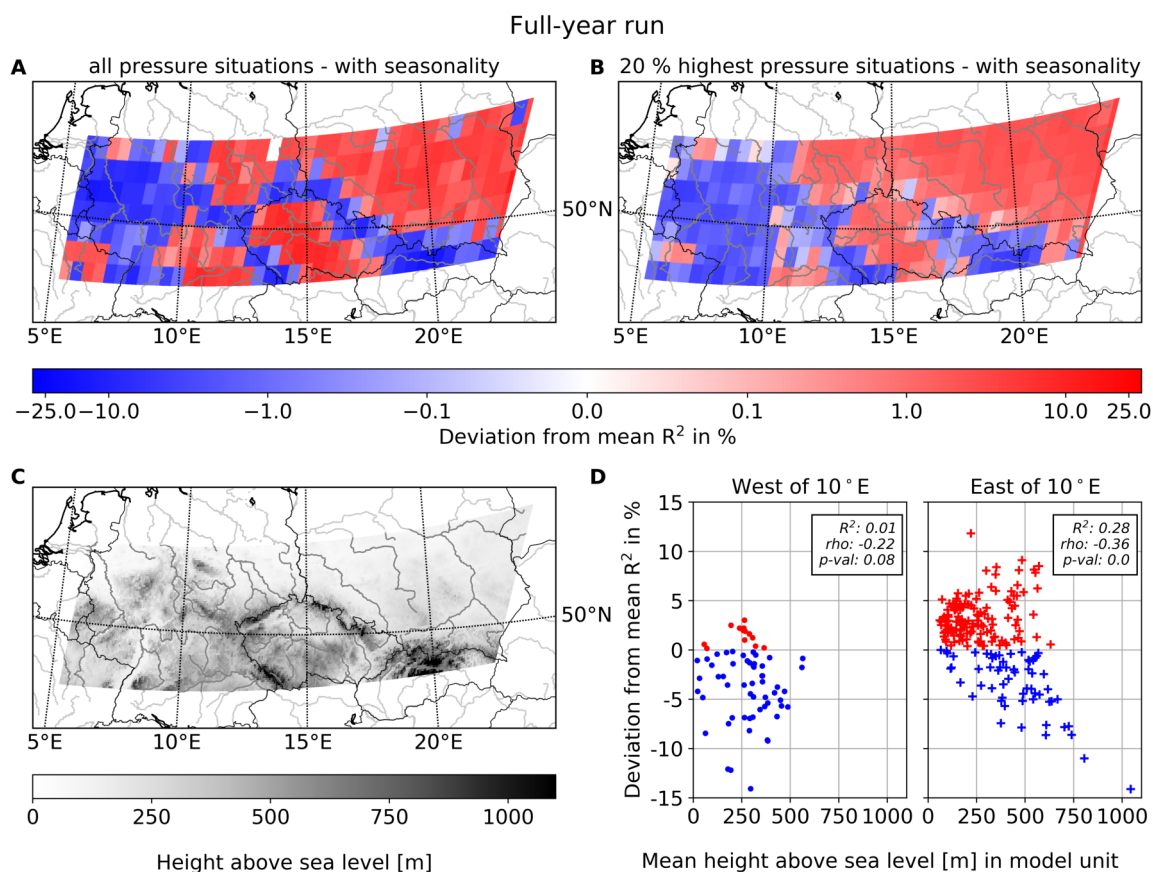


Figure 2.4: Map of deviation from the mean R^2 over all model units (domain average) in %, a topographic map (based on Hijmans et al. (2005)) of the study area and a correlation of these two. The left upper panel (A) shows the results from the model run using all pressure situations, the right upper panel (B) shows the results from the high pressure model run. Both include the seasonality. The domain average R^2 is computed for all model settings separately. For better visibility, the colormap is scaled logarithmic. In (C) the topographic map for the study area is shown. In (D) the results from (A) are correlated with the mean height per model unit for model units values west and east of 10° E separately. The corresponding Spearman's rho, p-value and R^2 of the correlation of the combined negative (blue) and positive (red) R^2 deviation values are plotted in each subplot.

In the western part of the model region (western Germany), as well as in parts of the south western part of the model region (northern parts of the Vosges, the upper Rhine valley, Black Forest) the R^2 is lower than the domain average. This is also the case for the mountain ranges surrounding the Czech Republic as well as in the south eastern part of the model region (High Tatras). The relationship between R^2 deviation and topography is visible in a) the visual relationship between the two (e.g. compare Fig. 2.4 A & C) and b) the correlation of the two (Fig. 2.4 D). With a Spearman's rho of -0.36 the latter is especially strong for values east of 10° E. A correlation of A and B with the distance to the coastline produced a Spearman's

rho close to 0 with a p-value > 0.05 . Small differences between pressure settings are visible in the north eastern and south western part of the model region, as well as the border triangle between Germany, Poland and the Czech Republic at 15° E. The apparent spatial patterns of skill are similar during all seasons, and independent of skill measure (R^2 or MSE) and training or test data set.

2.3.2 Feature importance

In figure 2.5 the domain-average feature importance is shown for all seasons and specific model settings as stacked barplots. The left-hand panel (Fig. 2.5 A) shows results for the model run that contains seasonality, while the right-hand panel shows the results of the deseasonalized model (Fig. 2.5 B). In these subplots, the darker-colored bars on the left display the results of the model runs containing all pressure situations, while the lighter-colored bars on the right show the results of the model using only high-pressure situations. Since results in the 15×15 model unit runs are similar, only 10×10 model unit results are shown and described here.

In general, clear similarities but also some distinct differences in mean feature importance are apparent: MSP, FLS_{prev} , WS and ET are among the most important features in all model runs. During winter and summer, MSP and WS are the most important features, while in the spring, fall and full-year run, ET and FLS_{prev} gain importance. When filtering for high pressure situations, ET becomes more important while the importance of MSP decreases. This is especially apparent in the spring, fall and full-year runs. LST reaches its highest feature importance in summer, but overall has little importance for model performance. ALB and FVC have little importance in all model runs however, this is slightly increased when filtering for high pressure situations in summer.

Differences between the model runs containing seasonality and the deseasonalized model runs are present in the exact values of feature importance but general patterns are similar. The largest differences in feature importance are apparent between the different seasons and the pressure exclusion runs. Model unit size and subtraction of seasonality only produce small differences in the mean permutation feature importance over all model units.

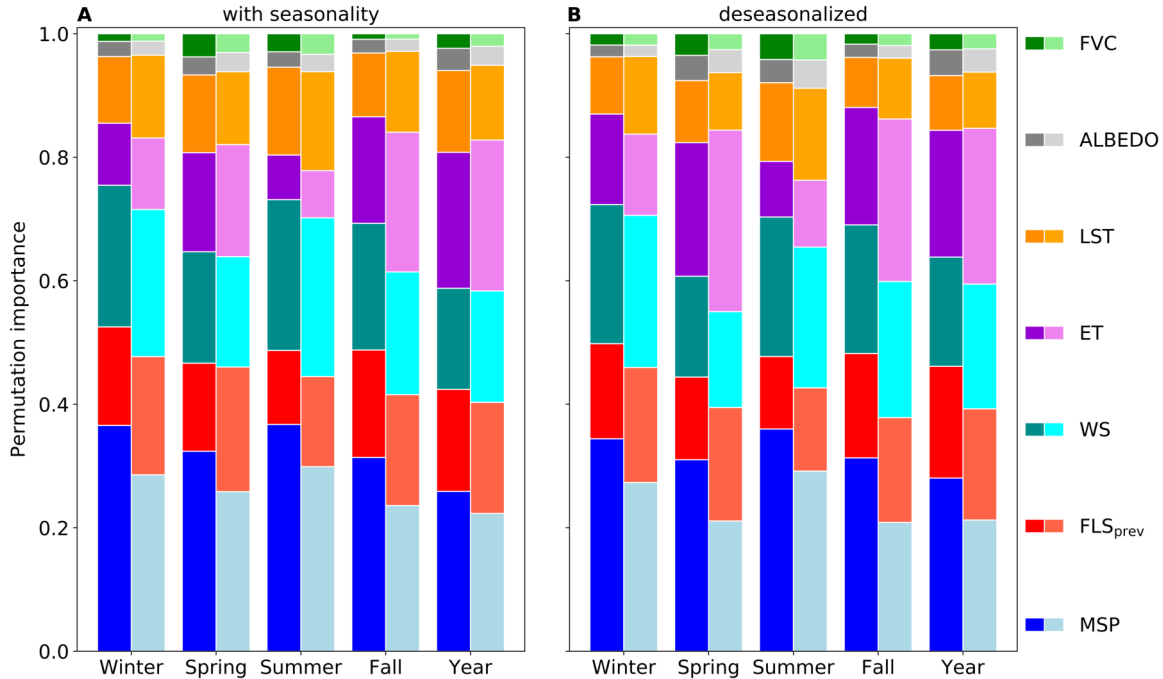


Figure 2.5: Mean 10x10 feature importance over all model units for all features and all seasons. The darker, left bars display the feature importance of the model run with all pressure situations, the right, brighter bars display the feature importance of the model run with high pressure situations. The left subplot (A) shows the feature importance from the model run including the seasonality, the right subplot (B) the one of the deseasonalized model run.

2.3.3 Spatial patterns of relevant features

In figure 2.6 the most important feature per model unit of the full-year run is shown for four different model settings. In general, the spatial patterns of the most important feature show clear differences between the model runs using all pressure situations (Fig. 2.6 A & C) versus those using the 20% highest pressure values (Fig. 2.6 B & D). When all pressure situations are used, MSP is the most important feature in most model units. This pattern is especially prevalent in the deseasonalized model run. Here, MSP is the most important feature in all model units except for some units in the eastern part of the study area, where ET is the most important feature. In the model run that contains seasonality (Fig. 2.6 A), ET is the most important feature in numerous model units east of 10° E, in parts of eastern Germany, most of the Czech Republic and parts of Poland. In three model units in the Czech Republic, FLS_{prev} is the most important feature.

As expected, in the high-pressure filtered model runs MSP is much less important. Instead, ET is the most important feature in large parts of the eastern study area. Other than ET, MSP, FLS_{prev} and WS are the most important features in numerous

model units across both seasonality settings (Fig. 2.6 B & D). MSP is the dominant feature in most parts of western Germany and eastern France and in some units in the Czech Republic and Poland. FLS_{prev} is dominant in the north western part of the Czech Republic, western Germany as well as parts of France and Poland. WS is the most important feature in some scattered model units of the model run including seasonality. In the deseasonalized model run, WS is dominant in more model units mostly in the western and the north western part of the study area.

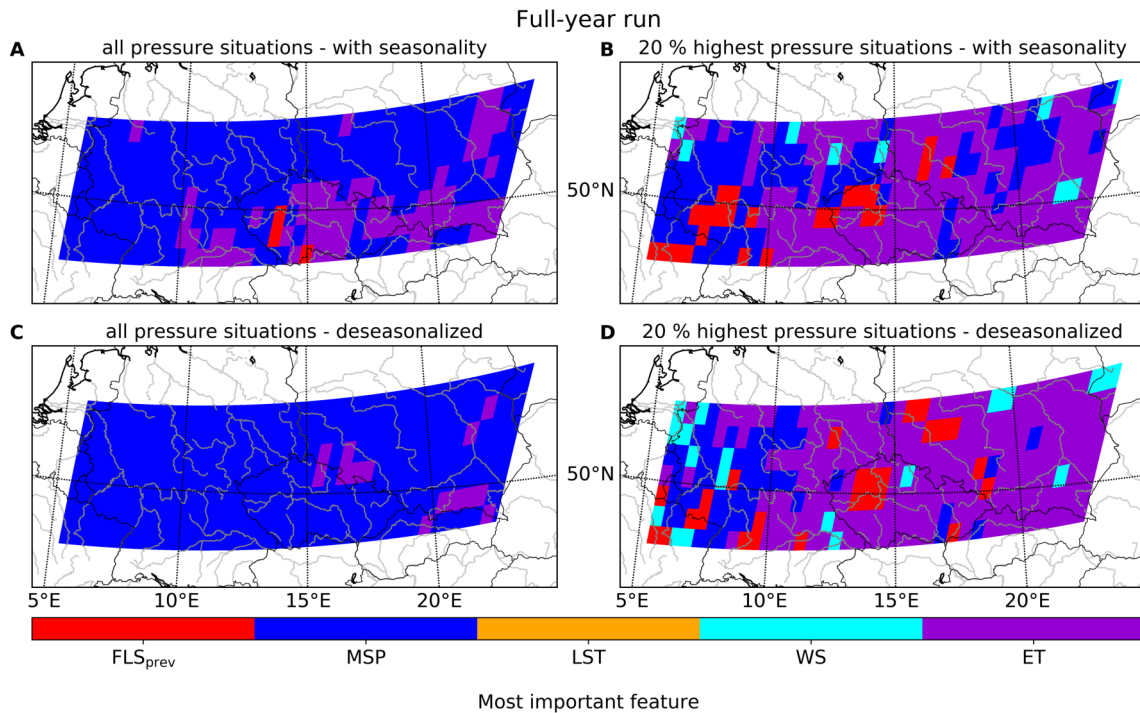


Figure 2.6: Most important feature per 10x10 model unit in the full-year run over all model settings. The left column shows the model runs including all pressure situations (A & C) while the right column shows the model runs using only the 20 % highest pressure situations (B & D). In the model runs of the top row, seasonality is included, in the bottom row it is subtracted.

The seasonal patterns of the most important feature (Fig. 2.7) are similar to the patterns observed in the full-year run. In the model runs using all pressure values (Fig. 2.7 A), MSP is dominant over all seasons, with only some exceptions in winter and fall. In high-pressure situations, though, seasonal differences are more pronounced (Fig. 2.7 B): in winter, FLS_{prev} is dominant in the western part of the study area (Germany and parts of France) and WS is dominant in the eastern part of the study area. MSP is the most important feature in between WS and FLS_{prev} in central and north western parts of the study area. Other than these three, ET and LST are the most important features in some scattered model units as well. In

spring, the most important features are FLS_{prev} in south western and ET in north eastern parts of the study area. Other than that, MSP is dominant in most of the study region, with some scattered model units where LST and WS are the most dominant features. In summer, MSP is the most important feature in most model units, with WS being dominant on the border of Germany and France, in Poland, in the Czech Republic and in some scattered model units in central Germany. Apart from that, only a few model units are dominated by FLS_{prev} and LST. In fall, MSP is dominant in the western part of the study area and ET is dominant in the eastern part of the study area (east of 10° E). WS is dominant in parts of eastern Germany and Poland and FLS_{prev} is the most important feature in some scattered model units in Germany, Poland and the Czech Republic.

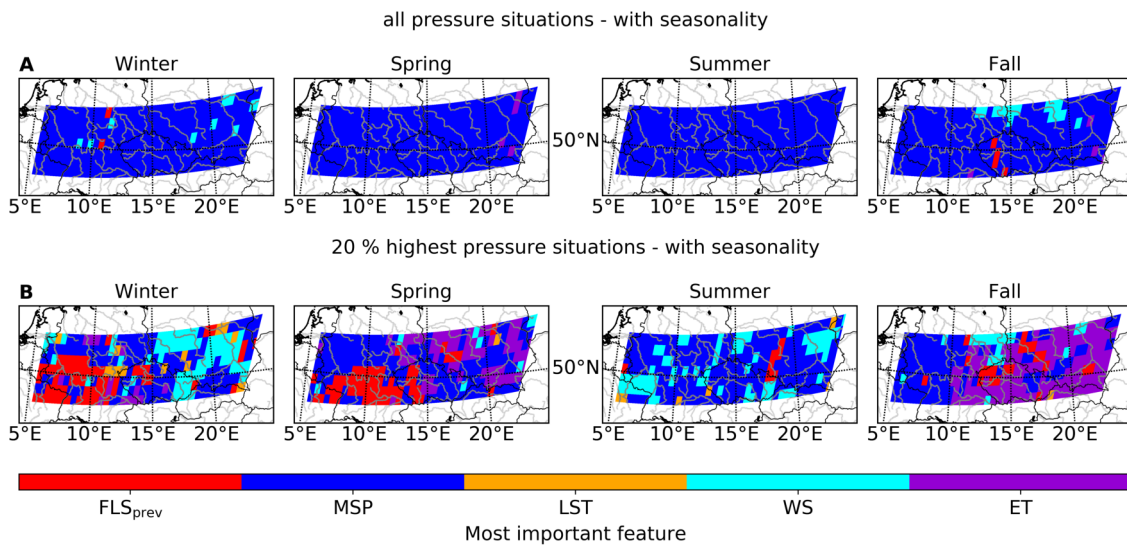


Figure 2.7: Most important feature per 10x10 model unit for different seasons. The top row shows the model run using all pressure situations while the bottom row shows high-pressure situations. Both model runs use the data that includes the seasonality.

2.3.4 One-variable partial dependence

In figure 2.8 the predictand responses relative to changes in all predictors are displayed as partial dependencies (see Sec. 2.2.5) for all considered time periods. Since the partial dependencies are calculated for each model unit separately, the mean and the interquartile ranges of the partial dependencies of all model units are presented here. Additionally, the distribution of values of each predictor is plotted below the corresponding partial dependence as a qualitative assessment of the representativeness of the obtained partial dependencies. The partial dependencies

presented here are from the model run where seasonality and all pressure values are included, since the partial dependencies are similar over seasonality and pressure-exclusion settings. The effect of the high pressure filter on the partial dependence of MSP is shown in the bottom right panel (all other partial dependencies of this run are similar to the ones shown).

The partial dependence can be understood as the deviation from the mean of the predictand (here FLS occurrence) that can be attributed to a specific predictor. Thus, a positive partial dependence at a given feature value indicates a positive influence on modeled FLS occurrence. Similarly, a negative partial dependence at a given feature value indicates a negative influence on modeled FLS occurrence. The magnitude of the partial dependence determines the magnitude of the influence. A partial dependence close to zero indicates very little or no influence on modeled FLS occurrence at the given feature value.

As a sensitivity measure, the range of the mean partial dependence over all model units for the full-year run is given in the text, together with the corresponding slope direction of the partial dependence curve. This provides a measure of influence on modeled FLS occurrence for each feature over the corresponding value distribution with a greater sensitivity measure indicating a large positive or negative influence on modeled FLS occurrence. Seasonal differences of sensitivity are shortly described.

The partial dependence of MSP (Fig. 2.8 A) is nonlinear, following a clear pattern: in low pressure situations (980 hPa–1005 hPa) the partial dependence is low, however, not many cases fall into this value range. In situations with $1005 \text{ hPa} < \text{MSP} < 1030 \text{ hPa}$, where most of the cases occur, the partial dependence increases substantially, reaching its maximum at $\sim 1030 \text{ hPa}$ or even higher pressure situations during winter, before it decreases again for pressure situations around 1040 hPa. However, only a few cases (most of them in winter) exist with such high pressure of $\approx 1040 \text{ hPa}$. The full-year sensitivity of MSP is +3.64 and the seasonal sensitivity is highest in winter and fall (+3.63) and lowest in summer (+1.73).

FLS on the previous day (Fig. 2.8 B) shows the most pronounced partial dependence pattern over all seasons and the full-year run with increasing values over the complete value distribution. The minimum partial dependence is thus reached for small FLS_{prev} values and the maximum partial dependence for high FLS_{prev} values. The sensitivity of FLS_{prev} also shows the strong influence of FLS_{prev} on modeled FLS occurrence and lies at +7.04 for the full-year run. The FLS_{prev} sensitivity reaches its seasonal maximum in fall (+7.94) and its minimum in spring (+4.95).

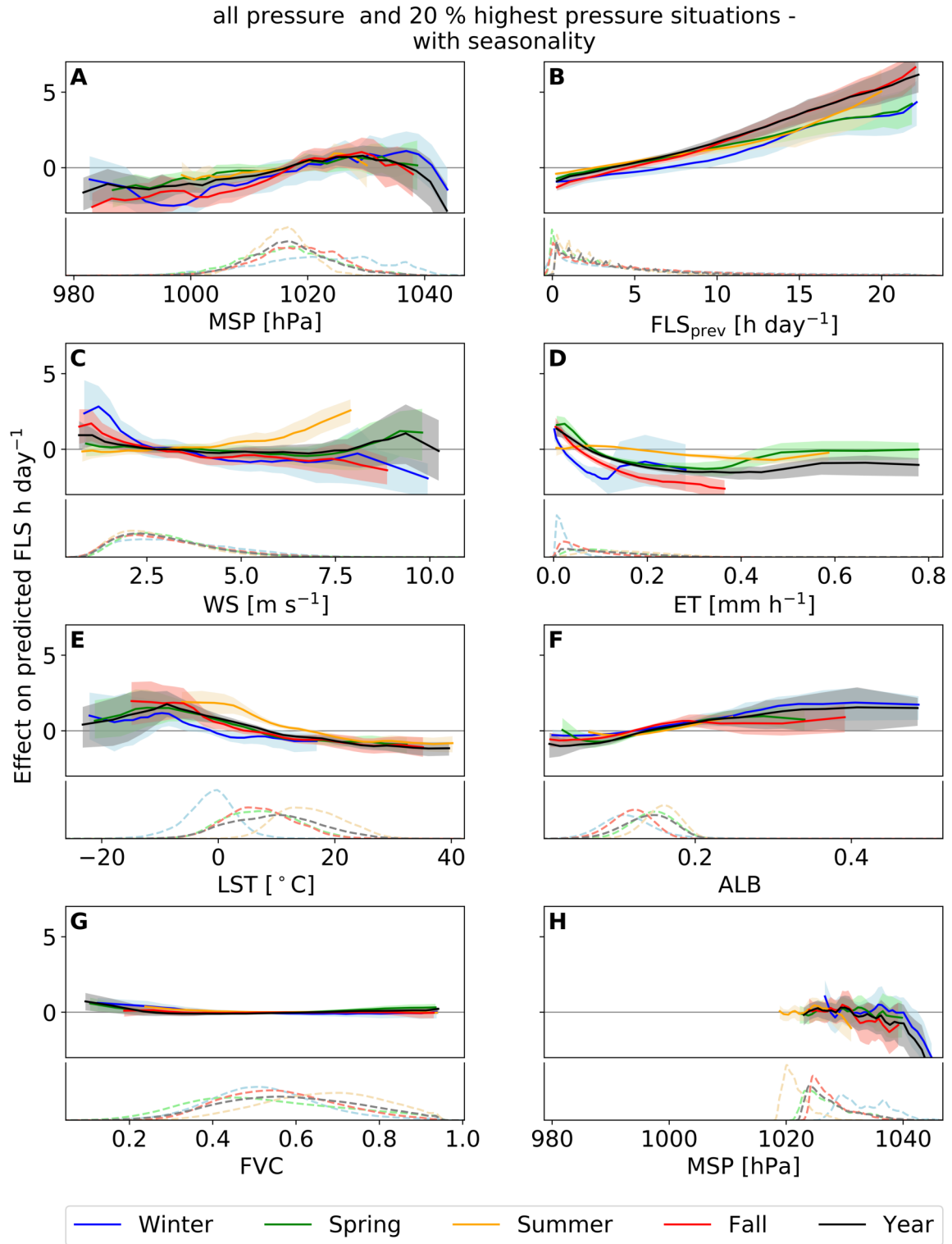


Figure 2.8: Partial dependence plot showing the mean response in modeled FLS occurrence to changes in all input features over all seasons, for the model run using data with seasonality. The top plot displays the partial dependence, the bottom plot shows a kernel density estimation of the data distribution. The shown features are: MSP in the model run using all pressure values (A), FLS_{prev} (B), WS (C), ET (D), LST (E), ALB (F) and FVC (G). In the additional (H) plot, the mean model response to changes in MSP in high pressure situations is shown.

The pattern of partial dependence of FLS on WS (Fig. 2.8 C) is strongly dependent on season and the distribution of values. At low wind speeds (0 to 2 m s⁻¹) the partial dependence decreases in winter, fall and the full-year run. In spring and summer it remains more or less constant at such low WS. In summer, it increases slowly, in spring and the full-year run it increases around 8 m s⁻¹ and for the winter and fall run it remains more or less constant. Still it has to be considered that any WS values greater than 6 m s⁻¹ are rare and the corresponding partial dependence pattern may not be as reliable and due to the low occurrence frequency not that relevant. The full-year sensitivity of WS is, compared to MSP and FLS_{prev} relatively low at -1.33 but reaches higher values in winter (-4.75). The seasonal minimum of WS sensitivity is reached in spring (+1.65).

The partial dependencies of the ET and LST are also both strongly dependent on the investigated season. In summer, the partial dependence of FLS on ET (Fig. 2.8 D) is more or less constant at a partial dependence of 0 predicted FLS h day⁻¹ over the complete value distribution. In all other seasons and the full-year run the partial dependence decreases at ET values from 0.0 to 0.1 mm h⁻¹. In spring and the full-year run it then slightly increases again at 0.4 mm h⁻¹. Due to the low distribution of ET values greater than 0.3 mm h⁻¹ the corresponding partial dependence values might be less reliable. The full-year sensitivity for ET is at -2.90, the seasonal maximum of sensitivity is found in fall (-4.09) and the minimum is found in summer (-0.94). The partial dependence of FLS on LST (Fig. 2.8 E) is highest when LSTs are below the freezing point, reaching a maximum at around -10 °C. At higher LSTs, the partial dependence of FLS on LST decreases. In summer, the LST value range and also these partial dependence patterns are shifted to higher temperatures. The full-year sensitivity of LST is similar to the ET sensitivity at -2.93. The seasonal maximum of LST sensitivity is found in fall (-3.06), the seasonal minimum is found in winter (-1.85).

Similar to their low feature importance, ALB (Fig. 2.8 F) and FVC (Fig. 2.8 G) both have small partial dependence values. The partial dependence of ALB slightly increases over the complete value distribution of all seasons and the full-year run. The partial dependence of FVC is low and constant over all seasons. Interestingly, the full-year sensitivity of ALB reaches +2.56, a similar high value to the sensitivity of ET and LST. The seasonal maximum is found in winter (+2.19) and the minimum is found in summer (+1.24). The full-year sensitivity of FVC is at +0.87 and thus the lowest of all features. The seasonal FVC sensitivity is highest in winter (+0.73) and lowest in fall (-0.19).

In the pressure-filtered model run, all partial dependencies show a similar pattern, except for the partial dependencies of MSP (Fig. 2.8 H). This is due to the reduced value range of the MSP values, thus only the decrease in the partial dependencies starting from around 1035 hPa is visible. The MSP sensitivity changes direction in the pressure filtered model and lies at -4.33 for the full-year run and is strongest in winter (-4.03) and lowest in spring (-0.74).

All in all, the response in modeled FLS occurrence is distinctive for the features FLS_{prev} , ET, LST, WS and MSP, with the latter showing a clear dependency on the application of the pressure filter. For changes in FVC and ALB, the response in modeled FLS occurrence is small. These partial dependence patterns confirm the feature importance patterns discussed above, with FLS_{prev} , MSP and ET being important features for model performance.

2.3.5 Two-variable partial dependence

While the isolated impacts of individual features can be investigated with the one-variable partial dependencies, two-variable partial dependence plots can show the combined effects of two variables on the response of modeled FLS occurrence, allowing for the analysis of possible co-dependencies. Based on the one-variable partial dependencies described above, six combinations of the most important features are presented in figure 2.9. These are MSP-WS (A), FLS_{prev} -MSP (B), FLS_{prev} -ET (C), ET-MSP (D), FLS_{prev} -WS (E) and ET-LST (F). In figure 2.9, the two-variable partial dependencies of the full-year run using all pressure situations including the seasonality are shown. Similar to the one-variable partial dependencies, the exact data preprocessing pertaining to seasonality treatment or pressure filtering does not substantially influence the two-variable partial dependencies. To outline characteristics of the distribution of the considered features and as an assessment for the representativeness of the results, two distribution borders are marked in the plots below: inside the outer dashed line 99 % of the data can be found, inside the inner dashed line 75 % of the data is present. Additionally, on the equivalent axes of the feature combinations, the 1st, 50th and 99th percentiles are marked as short lines.

The two-variable partial dependence of MSP and WS is displayed in figure 2.9 A. The values of the partial dependence are low over the complete value range of both features. A diagonal pattern, where low wind speeds and high pressure values lead to higher predicted FLS values is indicated. The combined effects of FLS_{prev} and MSP in figure 2.9 B are stronger, with FLS_{prev} being the more dominant variable. Still, a weak combined effect of higher MSP values and higher FLS_{prev} values leading to higher predicted FLS is visible.

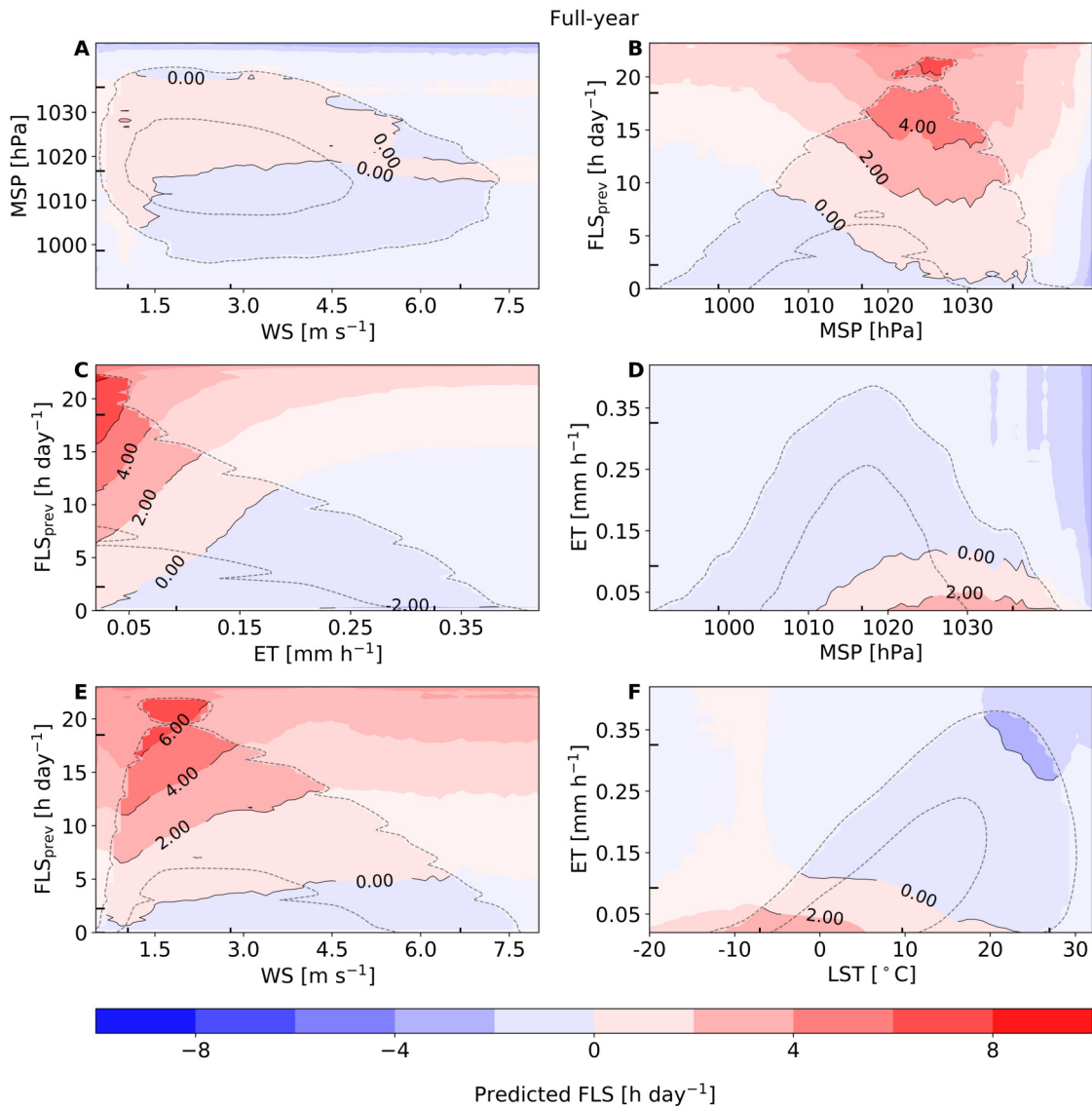


Figure 2.9: Two-variable partial dependence on FLS (hours day⁻¹) for six feature combinations for the 10x10 full-year run including the seasonality. The feature combinations are MSP and WS (A), FLS_{prev} and MSP (B), FLS_{prev} and ET (C), ET and MSP (D), FLS_{prev} and WS (E) and ET and LST (F). The outer dashed circle contains 99% of the data, the inner dashed circle contains 75% of the data. On the x and y axes of the feature combinations, the 1st, 50th and 99th percentile are marked as short lines.

In figure 2.9 C the combined effects of FLS_{prev} and ET are visible. While low FLS_{prev} values in combination with high ET values have a negative effect on predicted FLS, a positive effect on predicted FLS is visible for high FLS_{prev} and low ET values. The two-variable partial dependence of ET and MSP (Fig. 2.9 D) also shows a diagonal pattern of the combined effect of high MSP values with low ET values leading to higher predicted FLS. The combined effects of FLS_{prev} and WS (Fig. 2.9 E) are similar to FLS_{prev} and MSP: FLS_{prev} dominates the combined effects, with stronger influence of WS when FLS_{prev} values are higher, indicated by the more diagonal patterns of the two-variable partial dependence at FLS_{prev} values larger than 5 hours day⁻¹. For the two-variable partial dependencies of ET and LST (Fig. 2.9 F) low ET values with low LST values lead to high predicted FLS values.

In general, the seasonal differences of the two-variable partial dependence are governed by the seasonal differences in FLS occurrence as well as the range of values of the features. Since FLS occurrence is higher in winter and fall, the magnitude of the two-variable partial dependence is higher in winter and fall. Furthermore, the combined effects of MSP and WS are stronger in fall and winter thus confirm the combination of high MSP - low WS leading to higher predicted FLS indicated in the full-year plot. In summer and spring, two-variable partial dependencies are low.

Considering the patterns seen in the one-variable and two-variable partial dependence plots FLS_{prev} shows the most pronounced and clear effect on modeled FLS occurrence and dominates the combined effects in the two-variable partial dependence plots. This is especially interesting since MSP is the most important feature in most model runs (compare Fig. 2.5).

2.4 Discussion

2.4.1 Model performance

In general, the statistical models are capable of accurately predicting FLS using meteorological and land-cover predictors. While subtracting seasonality does not influence model performance substantially, filtering for high pressure situations boosts model performance significantly. Filtering for high pressure situations also leads to a higher relative frequency of radiation fog situations, which helps the model learn the relevant patterns, and thus improves its performance. Performance is best in winter and worst in summer. These patterns might both be due to more frequent high pressure situations in winter (e.g. Bartoszek, 2017), which can also be seen by significantly higher MSP values in winter compared to the other seasons (winter

mean MSP being 5 hPa higher) in the ERA5 data used in this study. Additionally, filtering for high pressure values makes it easier for the model to generalize and predict FLS. In summer and when using all pressure values, FLS occurrence may be due to local processes more frequently, such as orographic lifting (Egli et al., 2019) and not specifically tied to large-scale weather systems. This is in line with the results shown in van Oldenborgh et al. (2010) who found that atmospheric dynamics contribute up to 40 % to the variability of fog occurrences in winter, whereas in summer, the contribution of circulations on fog occurrence is more dependent on local factors.

The spatial differences of R^2 show lower model performance in units with high topographic variability such as parts of the mountain ranges surrounding the Czech Republic and the High Tatras. Furthermore, the deviation from the mean R^2 (Fig. 2.4) and MSE per model unit is greater in areas with greater height above sea level and with greater standard deviation of height in that model unit. The correlation between the mean height above sea level and deviation from the mean R^2 per model unit is especially prevalent east of 10° E, with a stronger negative deviation from the mean R^2 at larger heights (compare Fig. 2.4 D). This correlation is strongest in spring and fall. The decreasing performance in these either high altitude or topographically highly variable model units could be due to more pronounced small-scale, local processes that the model is not capable of reproducing. This could be also due to the missing terrain information in the selected predictors, e.g. according to Cuxart and Jiménez (2012) the advection flows generated by the topography are crucial for the growing phase of radiation fog. As stated in the Methods, using static geolocation variables such as height above sea level or geographic position can lead to overfitting and might be useful in training but not in making spatial predictions, as stated in Meyer et al. (2018) and Meyer et al. (2019). Terrain information (as height above sea level) is left out for this reason.

Besides the missing terrain information, the low resolution of the ERA5 data might lead to worse performance in topographically highly variable model units. In these units, the ERA5 data might be less accurate in general due to grid averaging. Here, the quality of the ERA5 may be improved by using a different interpolation technique like kriging or bicubic interpolation during data preprocessing. However, such a change is not expected to markedly change the results. Generally, the model performance in the eastern parts of the study area is high, likely due to rather low variability in terrain and a large influence of large-scale circulation patterns on fog occurrence leading to mostly radiation or advection fog. These processes are well captured by the chosen predictors and therefore well represented in the models.

Interestingly, the high-pressure filter can also lead to lower performance in some model units in winter, such as near the Harz and in parts of Poland. This is not the case at any other time of year, though. Two things might be responsible for this pattern: First, in the high pressure situations in winter, available data for training and testing has been reduced strongly. Thus, false predictions have a stronger influence on the R^2 of the model unit when less predictions are made in general. In model units with more available data points, false predictions have less influence on the model unit R^2 . Secondly, the lower performance in these units may indicate that in these regions, fog occurrence during winter may be caused by processes that are not inherently captured by the set of predictors. For example, terrain-induced processes may play a role as stated above, such as on the Mt. Brocken at 1142 m asl which is immersed in clouds about 50 % of the time (Acker et al., 2002).

A thorough investigation of the influence of the FLS distribution on model performance shows that in model units with higher FLS occurrence, model performance is better. The relationship of model performance and mean FLS occurrence for each model unit over all seasons is strong (Spearman's $\rho = 0.8$) and significant ($p < 0.01$) especially in the model runs using all pressure situations. This confirms the previously stated relationship of better performance in winter and worse performance in summer. These findings underline the influence of data availability and distribution on model performance.

Overfitting is apparent in all models, with larger differences between test and training performance in the 10x10 units than in the 15x15 units. This is likely due to the higher number of available data points in the 15x15 units for training and testing, shown with the decreasing difference between training and test performance with increasing number of data available in the model unit. More data thus makes it possible to learn more FLS situations and the subsequent behavior of the predictor variables. Furthermore, in a larger domain, a model needs to be able to represent rather general FLS patterns instead of specific local FLS patterns. A highly localized model might be also prone to overfitting. This could be investigated in future studies, by varying the model unit size while keeping the number of available training points constant. The different model unit sizes show the conflict between sufficient data for training and testing and small enough regional models to model smaller scale, regional FLS patterns. This conflict between data availability and spatial resolution is frequently encountered in observational studies. Conducting hyperparameter tuning in all model units for all model unit sizes in combination with regularization would lead to higher computation times, but could reduce the observed overfitting.

2.4.2 FLS drivers and their temporal and spatial patterns

Generally, fog formation and presence depend on multiple factors, such as the synoptic situation, smaller-scale atmospheric circulations and local orographic factors (Pérez-Díaz et al., 2017). The important role of the synoptic situation is underscored by the results of this study, with the high feature importance of MSP and WS, which are used as proxies for atmospheric dynamics. In high-pressure situations, often related to anticyclonic conditions, an inversion can develop during nighttime which is a prerequisite for the formation of radiation fog. The results obtained from the sensitivity analysis indicate that lower wind speed together with high pressure leads to a stronger stability and thus persisting FLS. This interplay of MSP and WS is also visible in figure 2.9. The influence of low wind speeds on turbulence generation during FLS development is not visible here as such phenomena occur below the spatial resolution of the ERA5 data. Wind direction has not been included in this study, since creating a general FLS - land surface model with a low number of predictors and minimum amount of rescaling has been a priority in this study, to first create a general model framework which can be further developed based on its first results in the future. For example, more detailed treatments of wind speed and direction in different atmospheric layers will be investigated in further studies. In the investigated study area, especially humid air masses from the west are important for fog occurrence (Wrzesinsky and Klemm, 2000; Błaś et al., 2002; Klemm and Wrzesinsky, 2007). The importance of atmospheric dynamics and air masses is also visible by the higher frequency of model units with WS and MSP as the most important feature in the western parts of the study region, which possibly indicates a decreasing relevance of westerly (moist) air masses with increasing continentality. As noted above, wind direction should thus be added as a predictor in further studies.

The positive relationship of FLS cover with FLS_{prev} is most likely due to persisting FLS over several days due to persisting atmospheric conditions. Additionally, feedback processes between FLS and the surface exist: in the presence of FLS, surface heating due to solar radiation is reduced during daytime which leads to lower near-surface temperatures in the subsequent night and a reduction of the difference between air and dew point temperature. This relationship might be stronger in basins and valleys where FLS tends to persist longer due to lower wind speeds inside the FLS layer and inversions or cold pool situations that can last for several days (Cuxart and Jiménez, 2012; Scherrer and Appenzeller, 2014). To investigate processes and sensitivities in larger river valleys, model units could be delineated specifically considering

topographic information in further studies. In these larger model units, SHAP (SHapley Additive exPlanations) values, introduced by Lundberg and Lee (2017) could be investigated, which allow for the investigation of individual predictions leading to a global understanding of the model (Lundberg et al., 2020), providing a tool to distinguish and individually analyze different FLS regimes.

Two land-surface based predictors that are closely related to each other due to the dependence of the saturation vapor pressure on temperature, described by the Clausius-Clapeyron equation (see chapter 1.2.1), are LST and ET. This means that higher land surface temperatures lead to higher evapotranspiration, with the latter being constrained by moisture availability (especially in summer), but also by solar radiation as well as the vapor pressure deficit. Concerning FLS processes, one would expect LST to have a negative relationship with FLS occurrence and ET to have a positive influence on FLS occurrence due to the moisture input into the boundary layer. However, in this study, both LST and ET show a negative relationship with modeled FLS, the latter one likely confounded by temperature, but also winds. FLS occurrence might be lower in high ET situations due to the larger vapor pressure deficit and thus drier air. Generally, ET is strongly influenced by the variability of atmospheric conditions (Teuling et al., 2010; Seneviratne et al., 2012) e.g. with anticyclonic circulation patterns leading to high ET values in Poland (Bogawski and Bednorz, 2016). At the scales considered in this study, the influence of atmospheric conditions on ET-FLS patterns is likely larger than the effects of moisture input via ET on FLS formation.

In general, temperature and humidity during FLS formation are connected via the process of condensation: Cooling leads to a rise in relative humidity, a reduction of the atmospheric saturation vapor pressure, and, once saturation is reached, condensation of water vapor to available cloud condensation nuclei and fog formation (Steenefeld and de Bode, 2018; Bergot and Lestringant, 2019). This effect is especially prevalent in winter, where fog formation is favored by a higher frequency of low temperatures together with high relative humidity conditions (Wrzesinsky and Klemm, 2000). Hůnová et al. (2018) list relative humidity and air pollutants as the most important features when modeling fog in the Czech Republic. As the presented statistical model presents a general working ground, a proxy for CCN has not been included yet but will be in further studies. The negative influence of higher LST on FLS cover in urban areas (Williams et al., 2015; Izett et al., 2019) can be suspected here, but is not visible in the results due to the size of the model units. Still, compared to LST, ET plays a stronger role in this study, especially in spring and fall, where the combined effects of low temperatures favorable for FLS and a

higher soil moisture leading to moderate evapotranspiration are prevalent. This is mainly visible in the eastern part of the study area, where ET is the most important feature in most model units in spring and fall. In these units, the moisture input by maritime westerly air masses might be limited, which could lead to a higher importance of local moisture sources. According to van Oldenborgh et al. (2010) fog during summer in the Netherlands only forms when enough moisture is present, with higher summer fog occurrences when moisture input in the preceding early summer or spring months has been larger. Here, in high pressure situations, day to day variations in ET are more important for predicting FLS occurrence than day to day variations in MSP, which is visible in the feature importance of ET and MSP in the deseasonalized model run.

In this study, the influence of radiative properties of the surface on FLS occurrence is only indicated by the small positive influence of increasing ALB values on FLS occurrence. Still, the feature importance and partial dependence of ALB is rather low and does not feature a clear pattern. The small effect of ALB on FLS occurrence might also be due to the multiple directions in which ALB might influence FLS occurrence. While surfaces with high ALB absorb less radiation during daytime and thus might lead to a lower air temperature and higher relative humidity above these surfaces, during nighttime, when radiation fog usually forms, darker surfaces might provide more moisture initially (such as water bodies or forests). Furthermore, FLS reduces the incoming solar radiation thus masking some expected radiative effects of ALB on FLS occurrence. In further studies, the effects of radiation on FLS occurrence could be further investigated by adding radiation fluxes as predictors to the model.

Similar to ALB, the influence of FVC on FLS occurrence is also not clearly visible in the results obtained from the study. Duynkerke (1991) lists vegetation as a driver for FLS occurrence, but this might not be visible here due to the spatial resolution and the stronger influence of meteorological drivers that mask such influences of the land surface and cover. Still, with the exclusion of low pressure situations and subsequent decreasing the influence of atmospheric dynamics can help the investigation of the influence of land surface parameters on FLS distribution. Filtering for specific weather situations, as in Egli et al. (2019) could decrease the variability and influence of the considered atmospheric predictors and may help isolate the influence of land surface parameters on FLS distribution under specific conditions.

As already stated, certainly not all factors influencing FLS distribution are considered in this study. In part, this study is limited by the spatial resolution of the data sets used, as local and regional morphology are important drivers for FLS occurrence (e.g. Błaś et al., 2002). Furthermore, the landform, turbulent boundary layer mixing,

above-cloud humidity and winds, soil conductivity, availability of CCN and the distance to the coast are listed as influential FLS drivers (e.g. Błaś et al., 2002, 2010; Hůnová et al., 2018; Steeneveld and de Bode, 2018). In further studies, the presented baseline model can be expanded by integrating these predictors to investigate their influence on model performance and sensitivities.

2.5 Specific conclusions

In this study, a machine learning technique, gradient boosting regression trees, was used to predict observed FLS occurrence over continental Europe, using meteorological and land surface parameters as predictors. To analyze spatial patterns of model skill and sensitivities, spatially explicit 10x10 and 15x15 SEVIRI pixel model units were created covering the entire study area. Additionally, models were applied for different model set-ups and data preprocessing procedures. The models were then applied to the entire, full-year data, as well as in different seasons. In general, the statistical models were able to accurately predict FLS occurrence in all regions of the study area, with R^2 values between 0.6 and 0.94 during validation with independent data. Model skill was observed to be highest in winter and lowest in the full-year run. Model performance increases when only high pressure situations are considered and when smaller (10x10 SEVIRI pixels) model units are used. Some overfitting is apparent in all models and depends on model unit size, with larger modeling domains featuring less overfitting. Using deseasonalized data only has a small effect on model performance.

Analyses of feature importance reveal that features pertaining to atmospheric dynamics are more relevant to predict FLS than surface characteristics. The most important features for FLS prediction are MSP, WS, FLS on the previous day, ET and to some extent LST. Albedo and FVC are less important in the statistical models. When only considering high pressure situations, MSP becomes less important, while ET and WS gain importance, the latter especially in spring and fall. Spatial patterns of the most important feature show the dominance of MSP in most of the model units when using all pressure situations. When only high pressure situations are considered, ET is dominant in central and eastern parts of the study area while WS, FLS_{prev} and MSP are dominant in western parts of the study area.

A sensitivity analysis was conducted with the statistical models, using the partial dependence technique. While there are some differences in partial dependencies when excluding low pressure situations, seasonality and model unit size settings

do not seem to have a marked influence. A positive influence of FLS_{prev} and MSP and a negative influence of ET, LST and WS on modeled FLS were found. These patterns are also confirmed in analyses of two-variable partial dependencies, which were used to study the combined effects of MSP, WS, ET, LST and FLS_{prev} . This analysis showed that especially the combination of high FLS_{prev} , high MSP and low WS values leads to high predicted FLS values. This is also the case for low LST and low ET values.

Considering the modeling framework, there are several limitations that should be noted. First of all, using spatially explicit model units reduces the available number of data points for training and testing. Thus, the available data puts a constraint on the spatial resolution at which such an analysis can be conducted. Still, there is potential for future improvements: Creating larger model units could provide more data that could also be used for validation and reduce the risk of overfitting. Grid search or a different methodology for hyperparameter tuning could be applied in all units separately and with fewer model units, less computing time would be consumed. While the generalization of hyperparameters is useful in this model set-up, conducting a thorough grid search in all model units would lead to a more complete picture of ideal hyperparameters for the model but still may have little influence on model skill and overfitting. It has to be noted that FLS is influenced by a large number of parameters, some of which are not represented in the model, such as topography, humidity, wind direction, soil moisture and aerosol loading. Additionally, FLS formation is also influenced by the state of higher atmospheric layers for example through moisture advection and mixing. While the model set-up presented here provides a general working ground, these features may be integrated in more complex statistical frameworks in the future. Furthermore, FLS processes take place from small scales of aerosol activation (10^{-7} m) to synoptic scales (10^6 m). Accurately addressing the issues concerning scales will thus be a critical aspect going forward, specifically as processes that are not important on one scale might be crucial on a different scale.

To gain further insights into the relationship between FLS and the land surface, FLS properties such as cloud top height and liquid water path can also be analyzed with the presented modeling framework. While the model units can be varied in size, they can also be grouped into areas of similar topography or land cover which could help in further analyzing FLS - land cover dependencies.

3 Land surface driven variations in fog and low stratus occurrence

3.1 Specific Motivation and Aim

The analysis conducted in chapter 2 showed the high importance of meteorological drivers for the prediction of FLS occurrence. Compared to the effect of meteorological variations on FLS occurrence, the effect of variations in the land surface, specifically albedo and the fraction of vegetation cover, was found to be small. Yet, studies looking at interactions between the land surface and clouds have found land surface effects on other cloud types, such as convective cumulus clouds. For instance, an enhancement of convective cumulus cloud cover has been reported over forests (Teuling et al., 2017), mega-cities (Theeuwes et al., 2019) and natural bushland (Ray et al., 2003). Xu et al. (2022) found enhanced cloud cover over most temperate and boreal forests, but decreased cloud cover over forests in Amazonia, Central Africa and the Southeast US.

Several potential mechanisms have been identified to be involved in interactions between forests and clouds. These include a higher evaporation and sensible heat flux over forests (Gambill and Mecikalski, 2011; Gentine et al., 2013; Bosman et al., 2019), as well as a higher aerodynamic roughness leading to the development of a forest breeze (Mahrt and Ek, 1993). Further potential reasons for differences in cloud cover over forest and non-forest areas are differences in the surface energy balance (Teuling et al., 2017) and the development of a mesoscale circulation (Xu et al., 2022). Another important factor identified to contribute to cloud formation over forests is the emission of biovolatile organic compounds (BVOCs) by trees. BVOCs can form secondary organic aerosols (SOA), which can act as cloud condensation nuclei (CCNs) and thereby favor cloud formation (Spracklen et al., 2008; Pöschl et al., 2010; Shrivastava et al., 2017). BVOC emissions further influence warm cloud microphysics and cloud-aerosol interactions (Petäjä et al., 2022). Considering these processes described for convective boundary clouds could give further insights when

studying FLS in the context of land-atmosphere interactions. However, temporally and spatially extensive studies looking specifically at the influence of land cover on FLS are rare.

To isolate potential local effects of forests on FLS occurrence, the strong influence of other confounding factors, like topography and meteorology, should be minimized. The Landes forest in southern France is a suitable area for such an approach as it has low topographic variability (cf. Fig. A2.1) and large spatial extent. In addition, BVOC emissions have been detected over the Landes forest (Kammer et al., 2018, 2020), providing the potential to link FLS occurrence with BVOC emissions over the region. In the following analysis, nighttime (0–6 UTC) FLS cover over the Landes forest over ten years is analyzed. FLS data based on passive and active satellite data is used, minimizing the influence of potential misclassifications on the results. By comparing climatological means over forest and non-forested areas, the influence of the day-to-day variability in meteorological conditions is reduced. The seasonal differences of FLS cover over the area are investigated and the influence of wind and temperature on the observed patterns is analyzed. This long-term analysis contributes to the understanding of potential interactions of FLS and forests, independent of satellite sensor and short-term fluctuations in FLS cover. The main hypothesis investigated here is that FLS occurrence is significantly influenced by the underlying land surface through land surface effects on wind speed, temperature and CCN availability.

3.2 Data and methods

3.2.1 Data sets

The Landes forest covers an area of about 12,000 km² in southern France directly at the Atlantic coast. The forest is mainly composed of maritime pine (*PINUS PINASTER*) (Kammer et al., 2018) and shows a distinct contrast to its surrounding land cover types (compare Fig. 3.2d). The study site has previously been used for the investigation of daytime cumulus clouds using satellite data, showing its potential for the investigation of land-atmosphere interactions (Teuling et al., 2017).

The primary fog and low cloud data set used in this study was created by Egli et al. (2017). It uses passive satellite data from Meteosat SEVIRI following the Satellite-based Operational Fog Observation Scheme (SOFOS) by Cermak (2006) using the approach presented in Cermak and Bendix (2007) for nighttime data. The resulting data set provides a binary FLS mask for every 15-minute time step covering

the entire central European land mass and the years 2006–2015. The data set has been validated against ground observations, showing that 80–90 % of FLS and no-FLS situations are classified correctly (Egli et al., 2017). The binary FLS mask is available for daytime and nighttime hours but not during, as well as shortly before and after twilight, due to sensor and algorithm constraints. For the purpose of this study, nighttime observations (0–6 UTC) of FLS are used. This is in line with BVOC emissions observed during night over the Landes forest (Kammer et al., 2018), which can potentially serve as CCN. The time frame of 0–6 UTC was chosen to minimize the amount of missing observations during twilight and to select a time frame with nighttime observations across all seasons. As a plausibility check for the FLS data set by the passive satellite sensor, an additional FLS classification based on activate satellite data from CALIPSO was created. For this, the CALIPSO level 2 1 km cloud-layer product (Version 4.20) (NASA/LARC/SD/ASDC, 2018) was used. The cloud and aerosol discrimination (CAD) algorithm classifies 90 % of layers correctly (Liu et al., 2009). Over the ten years, 186 nighttime overpasses over the Landes region were available. Due to missing data in both data sets, 179 of those overpasses could be used for comparison.

For the surface (10 m) winds, ERA5-land reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) was used (Muñoz Sabater, 2019). For temperature and wind data at different pressure levels ERA5 was used as well (Hersbach et al., 2018). The land cover data plotted in the results is taken from HILDA (Winkler et al., 2021). The Fraction of Vegetation Cover data serving as the background for the binary FLS maps are based on data from the EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA-SAF) (Trigo et al., 2011).

3.2.2 Fog and low stratus detection

The CALIPSO overpasses usually took place between 02:15 and 02:30 in the night. The CALIPSO FLS mask was derived similarly to the approach presented in Cermak (2018). First, the cloud layer altitude was calculated by subtracting the terrain altitude from the observed feature altitude. Then all cloud layers with a cloud top height equal to or smaller than 2.5 km and a cloud base height equal to or smaller than 2 km were defined as FLS. The thresholds differ slightly to those used in Cermak (2018) to include not only fog but also low stratus clouds.

To compare the SEVIRI based data set with the FLS mask derived from CALIPSO, the FLS observations of both data sets at the location of the CALIPSO overpass were contrasted by creating a confusion matrix. The time step of the SEVIRI-based data

set used for comparison was 02:15 UTC. To take into account the larger pixel size of SEVIRI and the small mismatch of observation time of both data sets, the comparison was done for the complete CALIPSO swath area across the forest area and not pixel-by-pixel. For example, an observation was marked as true positive, when any pixel along the CALIPSO swath for both data sets showed FLS. If there was no FLS pixel in both data sets, the observation was marked as true negative.

This comparison was followed by creating nighttime averages of FLS using the data from 0–6 UTC of the SEVIRI based data set. The respective days belonging to one of the categories of the confusion matrix were flagged accordingly.

3.3 Results

3.3.1 Cross-validation of fog and low stratus products

Out of the 179 used CALIPSO observations, 50 were identified as true positive and 83 as true negative. On 12 observations, no FLS was identified by CALIPSO but FLS was identified by the SEVIRI based data set (false positive). On 34 observations, FLS was identified by CALIPSO, but no FLS was present in the SEVIRI data set (false negative). Possible reasons for this could be multi-layer cloud situations or classification errors in the FLS data set (cf. Cermak and Bendix, 2008; Cermak, 2018). Further reasons are also described in the discussion. The confusion matrix can be found in table A2.1.

Two true positive cases are displayed in figure 3.1. On 2008-07-14 (Fig. 3.1a) most of the forest is covered by a large FLS patch which is also visible in the CALIPSO profile, where the cloud top is situated at approximately 2.5 km and cloud base at or below 2 km. A slight mismatch of the two products is visible at the CALIPSO swath at 44° N, where no FLS is present in the SEVIRI-based product, but FLS is present in the CALIPSO based product. On 2015-09-20 (Fig. 3.1b) both FLS patches over the forest and south of the forest are present in both data sets.

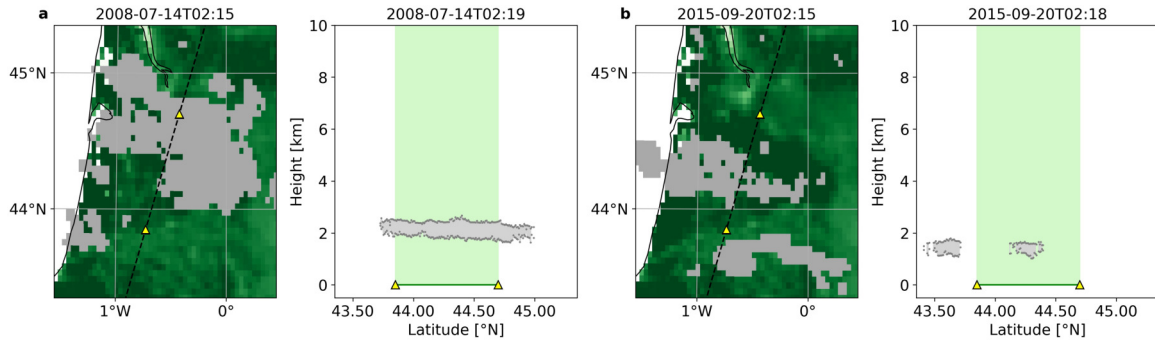


Figure 3.1: Example validation of SEVIRI-based fields with the corresponding CALIPSO profiles. Shown are two days, 2008-07-14 (a) and 2015-09-20 (b), with the SEVIRI based FLS maps on the left and the corresponding CALIPSO profiles on the right. The gray pixels in the maps of the SEVIRI-based data set display FLS, the background is a map of mean fraction of vegetation cover of the study area. For orientation, the CALIPSO swath is plotted as a dashed line in the maps. In the CALIPSO profiles, the cloud layers are plotted as a gray area at their corresponding height and with their latitudinal extent. The green line, as well as the light green background in the profiles mark the forest area. The yellow triangles in both the maps and the profiles mark the beginning of the forest area.

3.3.2 Climatological means

To decrease and visualize the influence of potential misclassifications on the climatological mean of nighttime FLS cover, three types of climatological means were constructed using the FLS data set:

1. Climatological mean over all days of the SEVIRI based data set by Egli et al. (2017) (3652 days)
2. Climatological mean over days with CALIPSO overpasses (179 days)
3. Climatological mean over days where the FLS observations of the SEVIRI and CALIPSO based data set match (true positives and true negatives) (133 days).

The corresponding maps, together with a land cover map of the Landes region, are shown in Fig. 3.2. In all three climatology maps, mean FLS cover from 0–6 UTC is higher over the forest, with differences in mean FLS cover between forest and non-forest most distinct in the northern forest area. Nighttime FLS cover is up to 1 hour longer over the forest than over the surrounding areas and local patches of enhanced FLS cover over the forest are visible (e.g. south of 44° N). The difference in nighttime FLS cover over forest vs. other land cover types is significant (two sample t-test, $P < 0.05$, Tab. A2.2) for all three climatologies, and strongest for the true positive and

negative days (Fig. 3.2c). This significance, independent of calculated climatology, shows the robustness of the results.

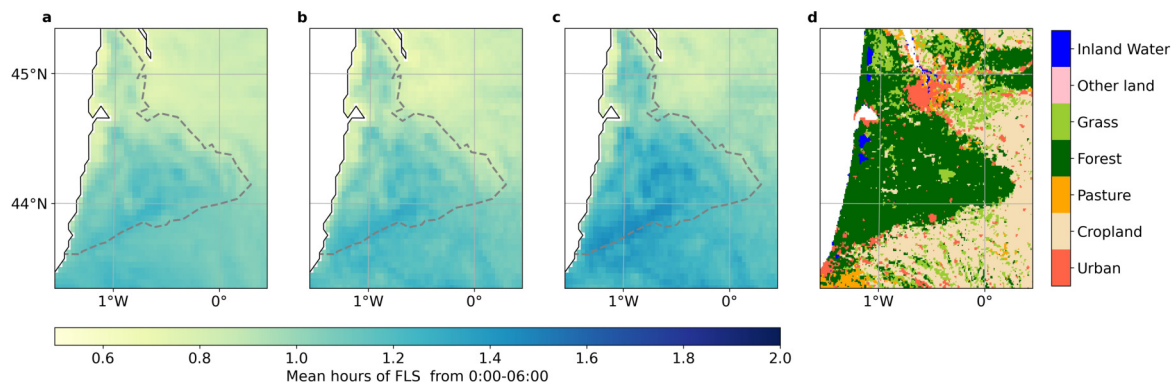


Figure 3.2: Spatial distribution of climatological mean FLS cover over the study region. In a) all observations from 2006-2015 are used to calculate mean FLS cover (3652 days), in b) only days with CALIPSO overpasses (179 days), in c) only true positive and true negative observations (133 days), and d) shows HILDA land cover for the year 2006. The gray dashed line approximately marks the forest border.

3.3.3 Seasonal analysis

A seasonal analysis of FLS cover (only true positive and true negative observational days) shows that nighttime FLS enhancement over the forest is significant ($P < 0.05$) for all months except January, February, June and October (see Tab. A2.3 for all t-test results and Fig. A2.2 & A2.3 for all monthly maps). The difference in mean FLS cover between forest and non-forest areas is shown here for the months of May, July, August and September (Fig. 3.3a), when FLS is most likely the result of more localized processes, as opposed to the winter months. Nonetheless, differences in FLS cover between forest and non-forest areas can still be significant in winter but are most likely due to higher FLS cover over the study area but lower FLS cover in the Pyrenees (south of 43.5° N) (cf. Egli et al., 2017). In May, especially the central Landes forest shows enhanced nighttime FLS compared to its surroundings. In July, enhanced FLS cover extends to the south of the forest, towards the Pyrenees. In August, the shape of the forest is quite well replicated by the pattern of enhanced FLS cover over the forest. In September, nighttime FLS cover is enhanced mostly in the central and western parts of the forest. Similar to the patterns depicted in the full-year means in figure 3.2, local patches of enhanced FLS cover are apparent inside the forest area in all of those monthly plots.

To investigate the potential reasons for high FLS cover over the Landes region, maps of mean wind speed are created. They show lower wind speed over the forest compared to the surrounding area (around 0.5 m s^{-1}) (Fig. 3.3b) and wind speed is highest directly at the coastline. The differences in mean wind speed above forest and non-forest pixels are significant (two sample t-test, $P < 0.05$, Tab. A2.3) in March, April, June, July and August. The main wind direction is west, with variations over the different months and over the study area. Patterns of wind direction also seem to be split into two subpatterns, with westerly winds prevailing near the Pyrenees (south of 44° N) and changing wind directions north of 44° N , e.g. with northerly winds in July and August.

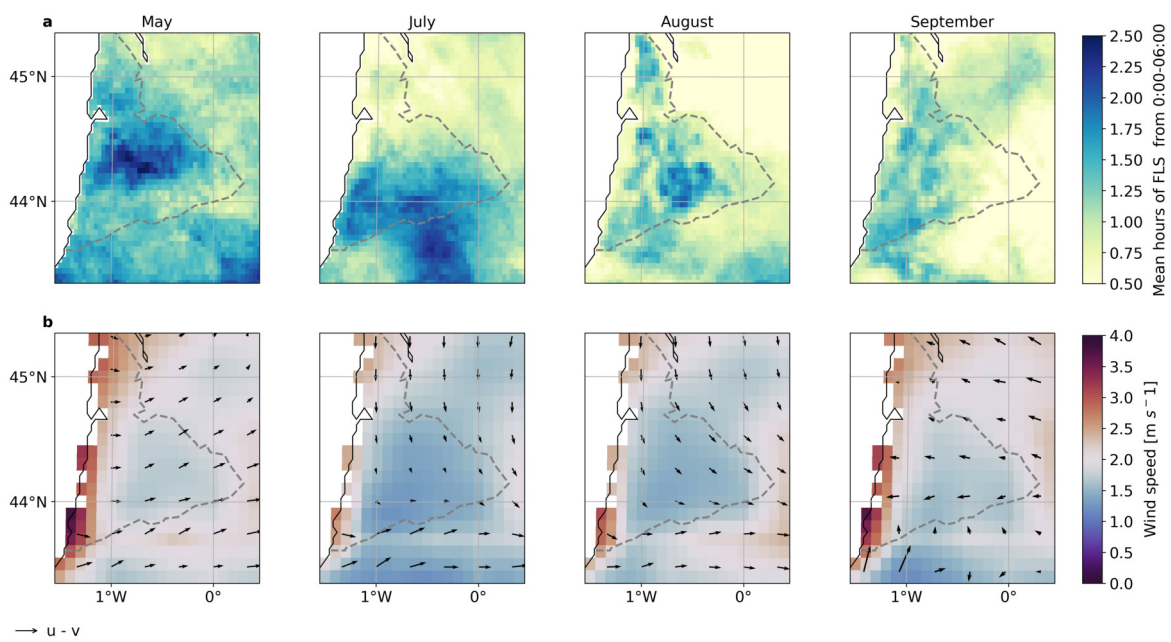


Figure 3.3: Climatological mean FLS hours by month based on true positive and true negative observations (a), and the corresponding ERA5 land wind speed and wind direction (u and v wind components) on the respective days (mean from 0–6 UTC) (b). The number of true negative and true positive days available to calculate the climatologies is as follows: 9 (May), 11 (July), 12 (August), 13 (September). The gray dashed line approximately marks the forest border.

3.3.4 Vertical temperature profiles

To better understand boundary layer effects on enhanced FLS cover over the Landes region, vertical profiles of temperature gradient in combination with wind vectors on the different pressure levels are analyzed. For this study, the temperature gradient is calculated by subtracting the temperature of a pressure level from the temperature of its overlying layer, i.e. positive values indicate a temperature inversion. In figure 3.4, hourly means over the respective true positive and true negative days in August (Fig. 3.4a and b) and September (Fig. 3.4c and d) are presented. Both longitudinal (Fig. 3.4a) and latitudinal (Fig. 3.4b) vertical profiles show a temperature inversion (positive temperature gradient) over the forest compared to the surrounding areas, especially at 05:00 and 06:00 UTC. In August, the temperature gradient reaches values up to $+0.67 \text{ K } 25 \text{ mb}^{-1}$ at -0.25° W in the longitudinal profile and $+0.57 \text{ K } 25 \text{ mb}^{-1}$ at 44.25° N in the latitudinal profile. In September, the temperature gradient reaches values up to $+1 \text{ K } 25 \text{ mb}^{-1}$ at 44.25° N and -0.5° W (both in the latitudinal and longitudinal profile). The u-w component vectors plotted in the longitudinal plots are very weak, whereas the v-w component vectors (latitudinal plot) show wind from the north in the pressure levels up to 850 mb during the night. Especially at 00:00 and 01:00 UTC air rises at around 44° N , just before the Pyrenees. Over the night the v-w wind component weakens.

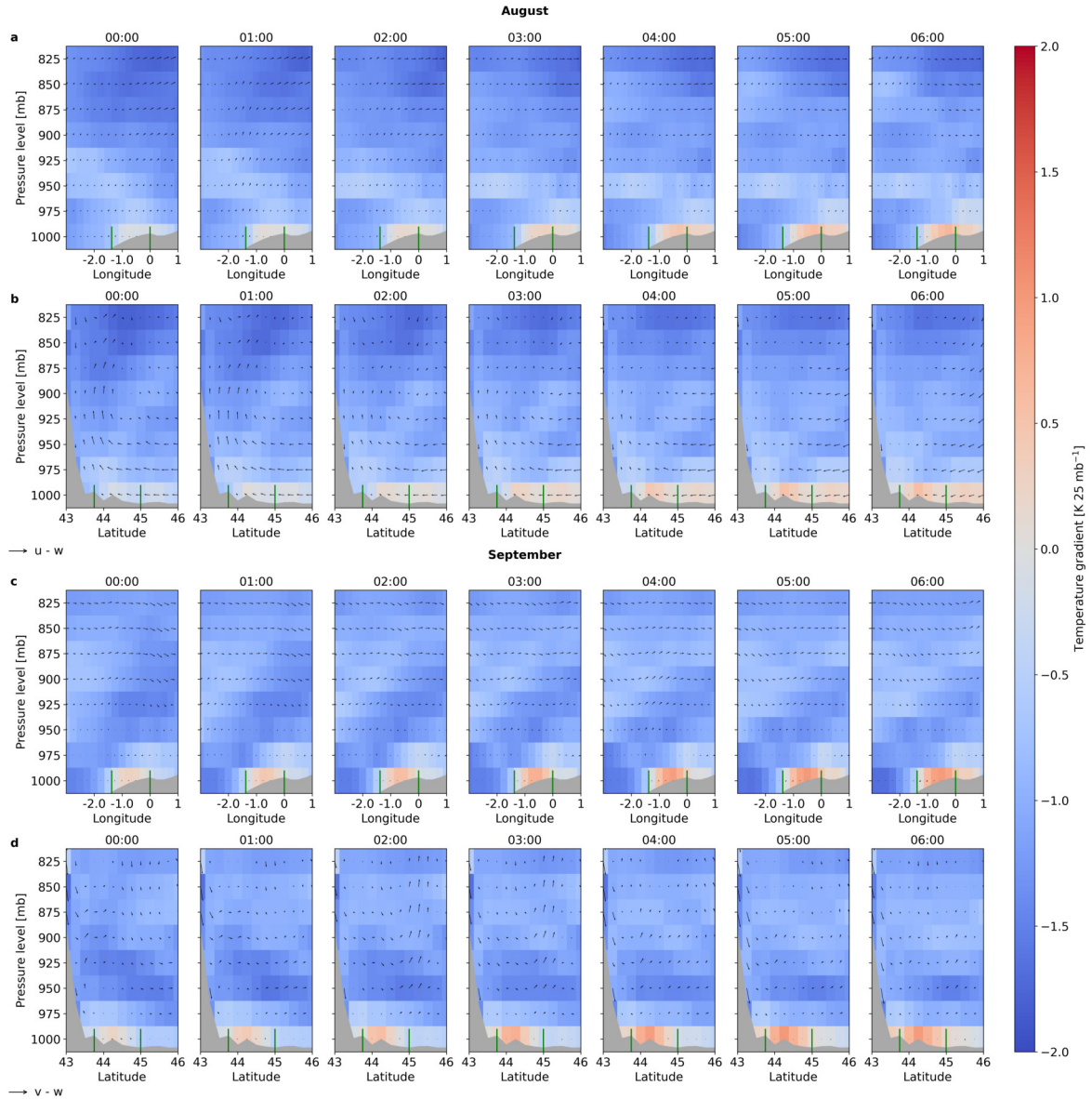


Figure 3.4: Profiles of mean hourly temperature gradient in $\text{K } 25 \text{ mb}^{-1}$ for the true positive and true negative days in August (a, b) and September (c, d), from 825 to 1000 mb along 44.25° N (a, c) and -0.25° W (b, d). In the longitudinal profile (a, c) $u-w$ wind vectors are plotted, in the latitudinal profile (b, d) $v-w$ wind vectors are plotted. For visibility reasons, the w vector is enhanced by a factor of 20. The location of the Landes forest is marked in both plots as green vertical lines. The topography is plotted in gray.

3.4 Discussion

The passive and active satellite data used in this study reveal enhanced FLS cover over the Landes forest compared to the surroundings. Both data sets agree well in most cases (74%), disagreement is potentially due to misclassifications in the FLS data set (cf. Cermak, 2018) or due to difficult FLS detection in the transitional zone (“twilight zone”) between aerosols and clouds (Koren et al., 2007). Local patterns of higher FLS cover over the forest area are visible, especially in the southern part of the Landes forest. This is similar to the findings by Teuling et al. (2017), who found enhanced daytime cumulus cover over the Landes forest with local maxima in the southern part of the forest. FLS enhancement over the forest compared to the surrounding areas is primarily visible in summer and fall, when local processes are potentially more important than in winter.

Various drivers are important for the development of FLS (Pauli et al., 2020). Here, the roles of wind speed and temperature are investigated. Based on the results of the t-test, lower wind speeds over the forest are identified as a potential driver of higher FLS occurrence, especially in the summer months. While some turbulence is required for the formation of a stable fog layer (Haeffelin et al., 2010), lower wind speeds are generally beneficial for FLS development (Roach, 1995; Gradstein et al., 2011; Bergot, 2016; Bergot and Lestringant, 2019; Pauli et al., 2020). Similar to the processes described in Gradstein et al. (2011) for lowland cloud forests, low wind speeds combined with nighttime cooling and saturation of air are a potential pathway leading to the enhanced FLS cover described in this study.

The vertical profiles of temperature gradient show lower temperatures and a positive temperature gradient in the forest area at night, which is unusual for temperate forests since they usually have higher nighttime temperatures than their surroundings due to turbulence and the storage of heat (Li et al., 2015; Schultz et al., 2017). Still, in both of these studies, the Landes region seems to be an exception, showing lower temperatures over the forest compared to unforested areas. A potential explanation is a strong nighttime cooling through evapotranspiration, similar to nighttime cooling over forests in tropical regions (Li et al., 2015). In combination with the observed temperature inversion, nighttime cooling increases the relative humidity over the area, supporting the development of fog and low stratus.

Further likely reasons for enhanced FLS cover over the Landes forest are the interplay between BVOC emissions and high evapotranspiration over the forest area. High loadings of natural aerosols from late spring to early fall have been found for boreal

forests (Tunved et al., 2006). It has been shown that secondary organic aerosol emissions together with evapotranspiration over forests can lead to an increase in liquid water path and cloud droplet number concentration in low-level liquid clouds (Petäjä et al., 2022). This could also be a potential pathway in the Landes forest, where BVOC emissions have been measured in summer (Kammer et al., 2018). A higher number of CCN and therefore potentially a higher number of small cloud droplets could lead to more FLS identified by both FLS detection algorithms. Measurements of BVOC emissions and fog and low cloud occurrences in the Landes forest could test the interactions between BVOCs and FLS in the future.

The patterns and drivers of higher FLS occurrence over the forest area are further modified by the general synoptic situation and geographic position of the Landes forest, with the Atlantic ocean to the west and the Pyrenees to the south. While the former is a source of moisture, the latter might enhance stationarity of air masses in the region, preventing high wind speeds and supporting the build-up of atmospheric moisture.

3.5 Specific conclusions

In this study, nighttime fog and low stratus cloud cover over the Landes forest in southwestern France using activate and passive remote sensing products was analyzed. FLS cover is significantly higher over the forest compared to non-forest areas and lower wind speed and a temperature inversion over the forest are potential drivers for this enhancement. As these parameters only partially explain the enhanced FLS cover over the forest, further atmospheric and biophysical drivers should be included into the analysis in the future, such as soil moisture, evapotranspiration and BVOC emissions. In addition, a systematic approach combining modeling and sensitivity studies could be used for future work to further quantify the role of forests for fog and low stratus cloud formation over varying geographic and synoptic backgrounds.

4 Climatological patterns of fog and low stratus formation and dissipation times

4.1 Specific Motivation and Aim

While the analyses presented in chapter 2 and chapter 3 are focused on the drivers and land surface specific patterns of FLS occurrence, an analysis of the FLS life cycle, namely its formation time, dissipation time and persistence is lacking. In this chapter, a novel FLS formation and dissipation time data set is presented which will further provide the foundation to study the drivers influencing the fog life cycle and its persistence on an unprecedented spatial and temporal scale.

Previous studies looking at the formation and dissipation time of fog rely strongly on ground-based observational data and localized process studies with numerical models, such as large eddy simulations (LES). These have been conducted for example in France (e.g. Haeffelin et al., 2010; Dupont et al., 2012; Wærsted et al., 2019; Karimi, 2020) or in the Netherlands (Duynderke, 1991; Steeneveld and de Bode, 2018) on time scales ranging from 6 days (Dupont et al., 2012) to up to 7 years (Wærsted et al., 2019). According to these studies, radiation fog usually forms during the night through nocturnal cooling (Roach, 1995) and dissipates a few hours after sunrise (Haeffelin et al., 2010; Bergot, 2016; Steeneveld and de Bode, 2018). While local studies using ground-based measurements provide insights into the small-scale processes of fog formation and dissipation, large-scale processes play a major role as well. Thus, data at large spatial scales are necessary to obtain knowledge on fog formation and dissipation processes across different landscapes. In this study, both formation and dissipation times are extracted from an existing, well validated satellite-based binary FLS data set by Egli et al. (2017) with a logistic regression. The goal is to analyze and discuss the spatial patterns of continental FLS formation and dissipation times over central Europe. Seasonal differences of

the extracted formation and dissipation times are analyzed across Europe and the regional applicability and detail is shown in a regional study in southern Germany. The guiding hypothesis is that the spatial patterns of FLS formation and dissipation time strongly depend on topography and the distance to the coast. The resulting novel data set gives information on FLS formation and dissipation times over continental Europe and over 10 years, a previously not investigated spatial and temporal scale. The applicability of the data set to study drivers of FLS formation, dissipation and persistence is presented in chapter 5 and in the outlook.

4.2 Data and methods

4.2.1 Fog and low stratus data set

The basis of this study is an FLS data set by Egli et al. (2017) which covers central Europe and the years 2006–2015. The FLS data set is created using data from the Meteosat Spinning Enhanced Visible and Infrared Imager (SEVIRI) and a number of threshold tests that are based on the Satellite-based Operational Fog Observation Scheme (SOFOS) by Cermak (2006). FLS in the Egli et al. (2017) data set is defined as a low stratiform cloud with liquid droplets not exceeding a size of 20 μm (Egli et al., 2017) but radiation fog and other stratiform clouds are not discriminated.

The FLS data set contains binary information that is 1 if FLS is present, and 0 if FLS is not present. The data set is available for every 15 minute time slot of SEVIRI at its native resolution (3x3 km at nadir). At twilight, no FLS detection is possible, leading to “FLS not present” (0) values in the original data set during about 4-5 15-minute time steps during each twilight episode. After exploring different treatments of those twilight values (keeping them at 0, setting them to NAN (not a number), temporal interpolation), setting them to NAN has been chosen as the most suitable method in this study, as this avoids the introduction of false information. However, on the basis of the Egli et al. (2017) data set, no extraction of formation and dissipation time at twilight is possible.

To provide an overview of the data used for the identification of FLS formation and dissipation time, the mean FLS cover is shown in figure 4.1a. The mean FLS cover over the study domain of central Europe generally shows higher FLS frequencies over the northern latitudes and in large-scale river valleys (Fig. 4.1a). In mountain ranges as well as in the Mediterranean, FLS frequencies are low. The dependency of FLS cover on topography can also be seen by looking at both mean FLS cover (Fig. 4.1a) and mean height above sea level over the study area (Fig. 4.1b). The data set can be used in various research applications, such as the investigation of large-scale drivers of FLS (Pauli et al., 2020).

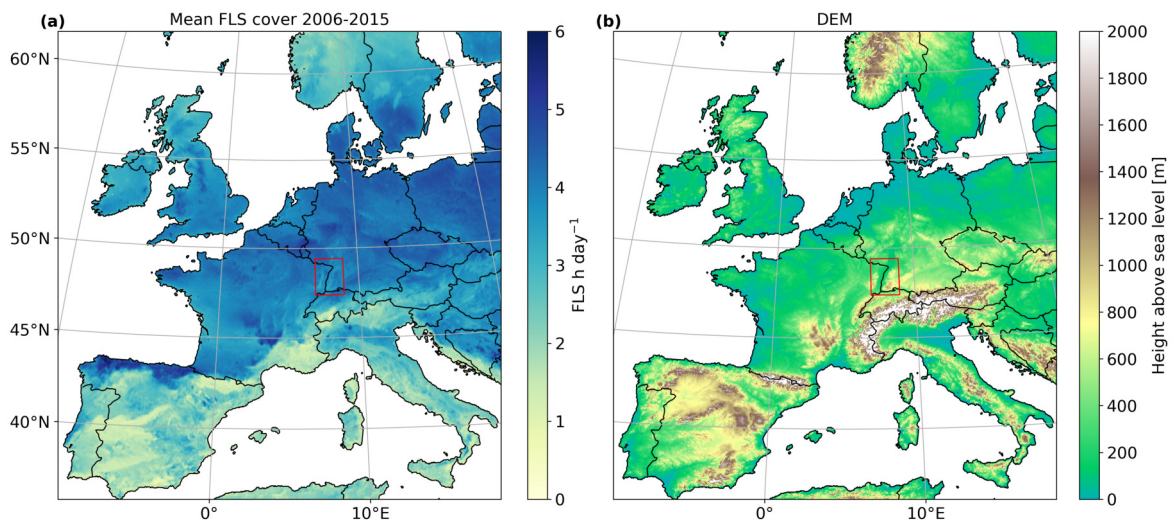


Figure 4.1: Mean fog and low stratus cover in central Europe from 2006–2015 (data set by Egli et al. (2017)) (a) and mean height above sea level (b). The Upper Rhine valley, used as a regional study area for the demonstration of the algorithm and to show the spatial detail of the results, is depicted with the red rectangle.

Summary of FLS data set validation

In the work presented here, the existing and thoroughly validated FLS data set by Egli et al. (2017) is used to extract FLS formation and dissipation times by applying logistic regression. While a novel FLS formation and dissipation times data set is created, the technique used is not an FLS detection technique but a statistical interpretation of an already existing, thoroughly validated, satellite-based FLS data set. This section provides an overview of the validation results of studies using the SOFOS algorithm for the detection of fog and low clouds (Cermak, 2006; Cermak and Bendix, 2007, 2008; Egli et al., 2017). Details on the validation procedures can be found in the relevant publications.

For the validation of the SOFOS algorithm, Meteorological Aviation Routine Weather Reports (METAR) were used, as these are available with a high temporal frequency (at least every hour) and well distributed over continental central Europe. In general, the SOFOS algorithm detects FLS very accurately (Cermak, 2006; Cermak and Bendix, 2007, 2008; Egli et al., 2017). Cermak (2006) and Cermak and Bendix (2008) found that 70 to 85 % of FLS situations are detected by the scheme. Situations where FLS is not detected can to a large degree be explained by overlying clouds, which are present above FLS about 25 % of the time in the study domain (Cermak, 2018). False alarms are rare, i.e. a pixel classified as FLS is almost never either clear or covered by a different cloud type (Cermak and Bendix, 2008). Similar to the other studies using the SOFOS algorithm, about 80-90 % of FLS and no-FLS situations are classified correctly in the Egli et al. (2017) data set used for the extraction of FLS formation and dissipation time. The geographic patterns of validation scores show higher validation scores in continental areas with radiative FLS events as opposed to coastal areas, where advective FLS events prevail (Egli et al., 2017). Thus, the constraints of a satellite-based FLS climatology compared to ground-observations are acknowledged in this study. Nevertheless, the good validation results show that this FLS data set can be used to derive an FLS formation and dissipation climatology as shown in the study at hand.

4.2.2 Methods

In this study, logistic regression is applied to statistically interpret the binary FLS data set by Egli et al. (2017) to determine FLS formation and dissipation times. After the algorithm is completed, a novel FLS formation and dissipation data set exists, which is derived from the original binary FLS data set by Egli et al. (2017). The analysis of this new data set of formation and dissipation time is shown in the results section. In the following, a general overview on logistic regression is given and then the algorithm is described, which applies logistic regression to the binary FLS time series described above.

Logistic regression is used for binary or categorical data and in this study predicts the probability of a data point belonging to one of the binary classes (Bisong, 2019; Lever et al., 2016). Mathematically, the probability can be expressed as:

$$p(t) = \frac{1}{1 + \exp(-b_0 - b_t * t)} \quad (4.1)$$

where p is the probability, t the time, b_0 the intercept and b_t the slope. While multiple predictor variables can be used in logistic regression, in this study, time is the only predictor. The transition from one state to another and its associated predictor value is called “decision boundary” (Bisong, 2019). Here, this decision boundary is used to determine the time of FLS formation and dissipation.

In the atmospheric sciences, logistic regression has previously been used for the detection of hail (López and Sánchez, 2009) or to forecast the probability of extreme precipitation events (Applequist et al., 2002; Herman and Schumacher, 2018). It has also been applied to predict the occurrence of orographic cloud cover (Wu and Zhang, 2013) or to forecast the probabilities of low visibility conditions at an airport site (Kneringer et al., 2019).

Algorithm

The algorithm which identifies FLS formation and dissipation time based on the binary cloud mask consists of 6 steps. These steps are similar for formation and dissipation time, but are conducted separately. The algorithm is applied to each valid FLS event (defined below) and to each pixel over the years 2006–2015. The steps described below are marked with the corresponding numbers in Fig. 4.2.

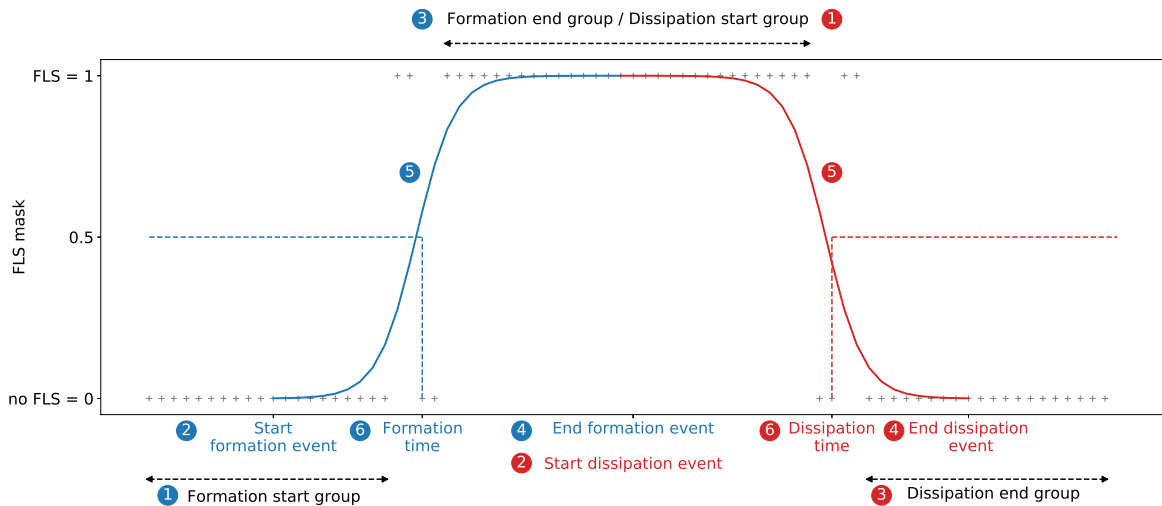


Figure 4.2: Workflow for the identification of FLS formation and dissipation times with logistic regression using an artificial example. Steps for the identification of formation time are plotted in blue, the corresponding steps for the identification of dissipation time are plotted in red. Black crosses mark the binary FLS values.

To apply the logistic regression to the data set, at first, an FLS formation/dissipation event has to be identified (step 1). In the case of FLS formation, this means that over a period of at least 10 consecutive 15-minute time steps (2.5 hours) no FLS has to be present (cloud mask = 0). In the case of FLS dissipation, the requirement is similar (FLS present (cloud mask = 1) for 10 consecutive 15-minute time steps). By choosing a minimum value of 2.5 hours, random misclassifications in the original FLS data set as well as advective FLS events with fast changes between FLS and no-FLS are likely excluded. One should note that the results can thus only be interpreted with respect to such longer lasting FLS events, and may not be representative of FLS events with a shorter life cycle. After identification of such a start group, its temporal center point is then used as the start of the FLS formation or dissipation event (step 2). Looking at the artificial example in figure 4.2, the start group of the formation event (blue curve) has 20 consecutive 15-minute time steps of no FLS. The starting point is then defined as the center, in this case at the 10th time step of no FLS.

The end group of a given FLS formation or dissipation event is defined similarly to the start group: In the case of FLS formation this means that for 10 consecutive 15-minute time steps FLS has to be present (cloud mask = 1), and in case of dissipation FLS has to be absent for 10 consecutive 15-minute time steps (step 3). The end group has to be present within 30 hours from the start of the formation or dissipation event. By choosing a time period of 30 hours, it is possible to capture FLS events that persist for longer than a day. If no end group can be identified within 30 hours of the start of the formation/dissipation event (step 2), the formation/dissipation event is omitted, and no time is determined by the algorithm. Similar to the determination of the start of the formation/dissipation event, the end of such an event is then set to be the center of the corresponding end group (step 4). In the artificial example in figure 4.2 the end group of the formation event contains 30 15-minute time steps with FLS, with the end point of formation defined in the center of the group at time step 15. During twilight, the groups of consecutive FLS or no-FLS are interrupted by NAN values as outlined in section 4.2.1. If the remaining groups then do not fulfill the requirement of 10 consecutive FLS or no-FLS values, this leads to a potential loss of start and end groups.

Following the identification of the start and end points of the FLS formation/dissipation events, logistic regression is applied to the binary time series between the start and end points to model the probability of FLS presence (step 5). Then, the decision boundary is used to depict the formation and dissipation times (step 6). For formation, the first 15-minute time period in which the probability modeled by the logistic regression exceeds 0.5 is defined as the formation time. For dissipation,

the first 15-minute time period, where the probability falls below 0.5, is defined as the dissipation time.

After the calculation of the formation/dissipation time, these 6 steps are repeated for each identified start group for each pixel separately. While the formation and dissipation situation displayed in figure 4.2 can be considered as an ideal example, with a dissipation event following a formation event right away, there is also the possibility that a dissipation event is not identified right after a formation event. This occurs when an end of the dissipation event can not be identified within 30 hours, due to alternating FLS and no-FLS values. If the requirements discussed above are not fulfilled by either a potential formation or dissipation event it is possible that multiple FLS formation or dissipation events are identified after one another. This leads to slight differences in the number of formation and dissipation events identified by the algorithm, which is discussed in section 4.3.1.

Discussion of algorithm uncertainties

It is clear that the quality of the derived FLS formation and dissipation time products is dependent on the quality of the underlying FLS data set. Therefore, systematic and random errors in the FLS data set will propagate to the classification carried out in this study. By focusing on temporally persistent FLS events, however, the effects of random misclassifications in the data set are reduced. Another uncertainty of the derived FLS formation and dissipation time products is introduced by differences in the characteristics of the binary variability of each specific FLS event (Fig. 4.3). An abrupt, temporally coherent change between the binary classes (little alternation between FLS and no-FLS values), leads to a higher absolute value of the slope and a steep curve fit. This leads to either the classic sigmoid shape (Fig. 4.3a and d) or a very steep transition (Fig. 4.3b and e), both of which are easy to interpret with respect to FLS formation and dissipation time. In case of frequent changes between FLS and no-FLS values of an FLS event considered, however, the fitted logistic curve is flat and the slope value approaches zero (Fig. 4.3c and f). These flat curves are difficult to interpret with respect to an FLS formation/dissipation time, and are thus excluded from the data set. These valid situations are defined to feature a slope of $> |0.1|$. This threshold was defined by conducting a thorough visual analysis of different events and their corresponding curve shapes. The exclusion of these high-uncertainty events leads to an average reduction of 9 % of FLS events per pixel.

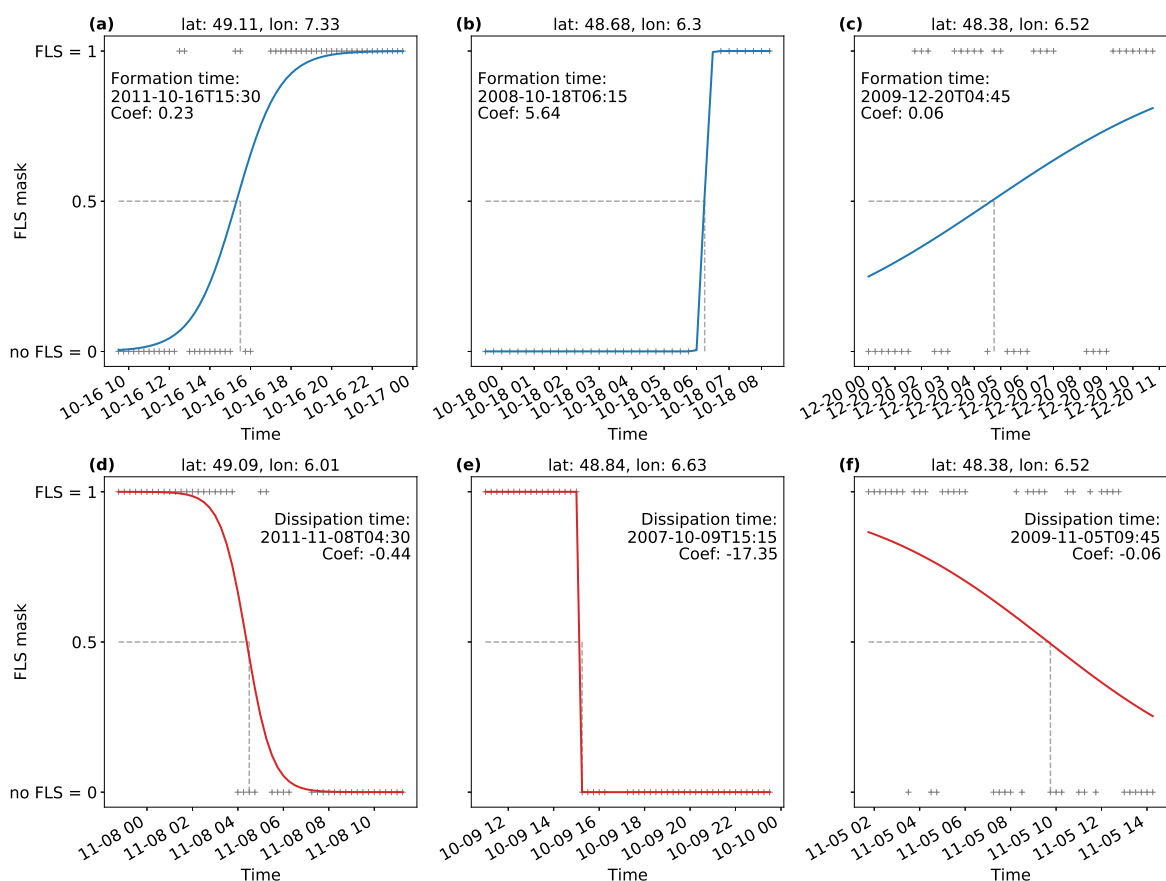


Figure 4.3: Logistic regression curves for three formation situations (a–c, blue curves) and three dissipation situations (d–f, red curves). The situations in the left-hand column (a and d) have a classic sigmoid shape and the situations in the center (b and e) have a sharp transition. The curves in the right-hand column (c and f) do not fulfill the slope requirements (coefficient $> |0.1|$). The “coef” parameter displays the slope of the logistic regression curve.

4.2.3 Case illustration: 2011-02-07

To illustrate how formation and dissipation time are derived from the binary FLS masks, a dissipation case from 2011-02-07 over the Upper Rhine valley in Germany is presented in the following. The mean FLS cover and topography of the region are shown in figure 4.4.

Radiation fog is a frequent phenomenon in the Upper Rhine valley, especially in the colder months (Kalthoff et al., 1998; Bendix, 2002; Bendix et al., 2006; Egli et al., 2017). In this rift valley between the Vosges mountains in the west and the Black Forest mountains in the east (visible in Fig. 4.4b and roughly depicted in Fig. 4.4 with the 500 m contour), FLS occurs more frequently in the valley than on the mountain ridges (Fig. 4.4a).

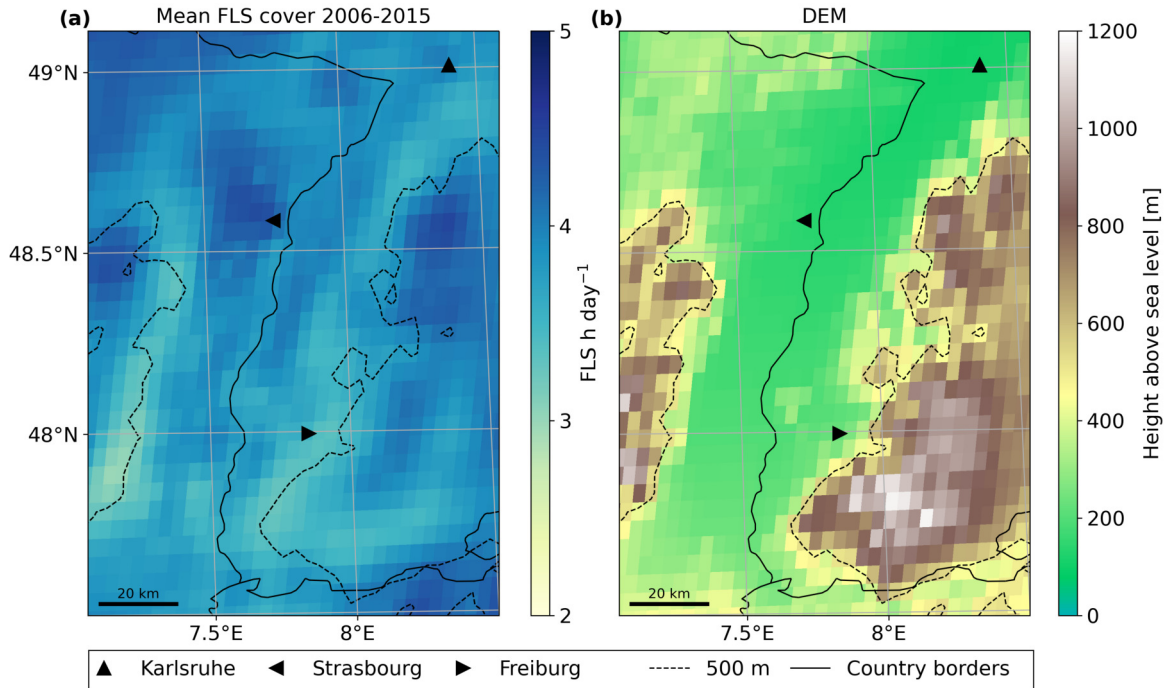


Figure 4.4: Mean fog and low stratus cover in the Upper Rhine valley from 2006–2015 (data set by Egli et al. (2017)) (a) and mean height above sea level (b). The country border between France and Germany also marks the course of the river Rhine. The black dashed line depicts the 500 m above sea level height.

The case illustration of 2011-02-07 focuses on three pixels in the Upper Rhine valley (at locations A, B, C), all of them located in France. As can be seen in the binary FLS maps (Fig. 4.5), FLS is present in mainly the northern part of the valley on the morning of 2011-02-07 and then dissipates during the day. To illustrate the derivation of the dissipation time, the binary time series at the three locations is extracted (Fig. 4.6). A logistic regression is calculated for all time series and the time stamp where the probability of the binary value being equal to 1 falls below 0.5 is assigned to be the dissipation time. In line with the binary FLS masks in figure 4.5, FLS at location B dissipates first (09:30 UTC) followed by location A (dissipation at 11:45 UTC). FLS at location C is most stable and dissipates in the afternoon (14:30 UTC).

The approach presented here for the identification of dissipation time at the three locations is then applied to each pixel of the study area, for each identified FLS formation and dissipation event.

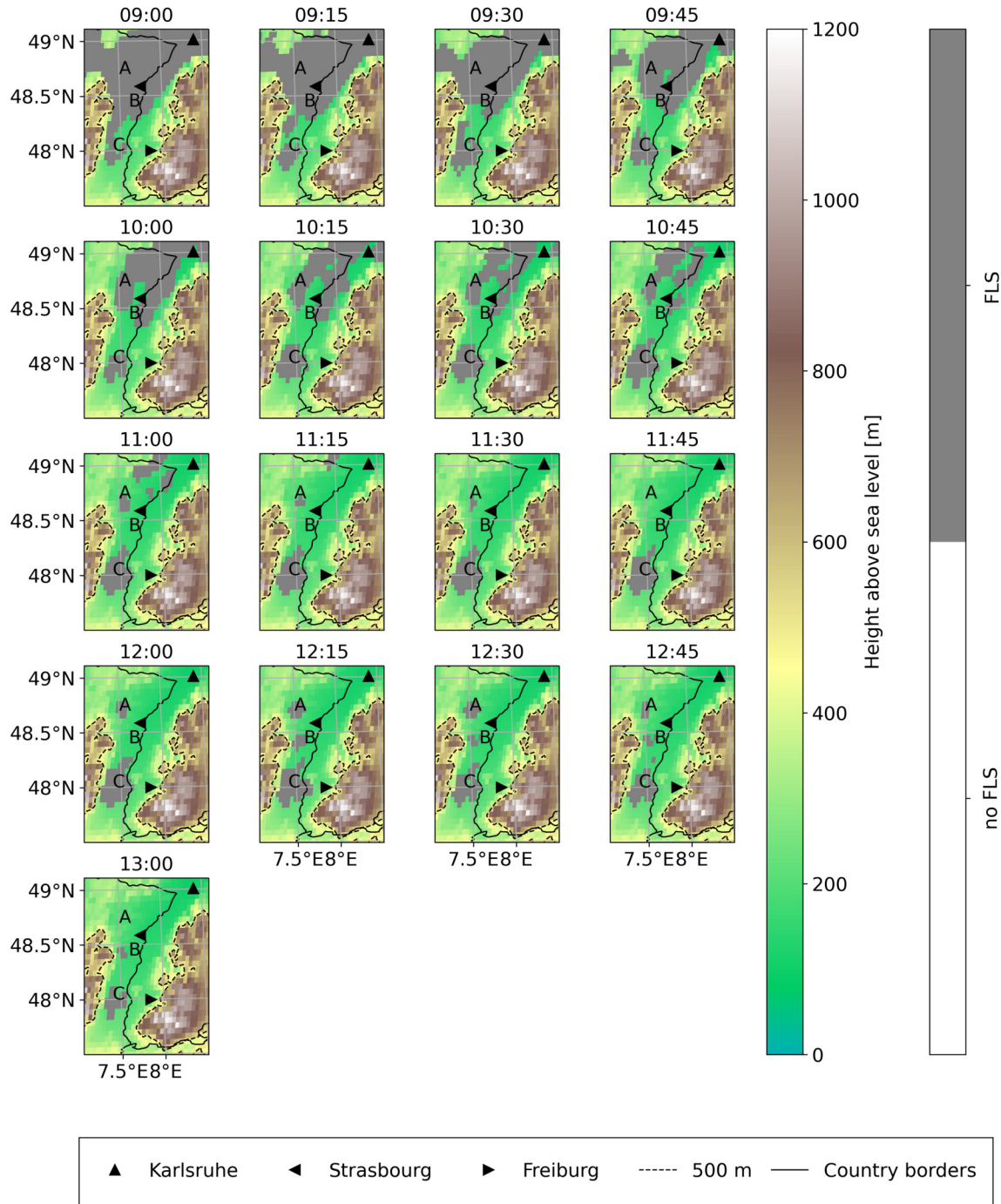


Figure 4.5: Binary 15-minute FLS masks (data set by Egli et al. (2017)) from 09:00 to 13:00 UTC on 2011-02-07. Pixels with FLS cover are gray, pixels without FLS are transparent. The background is a topographic map of the region. The country border between France and Germany also marks the course of the river Rhine. The black dashed line depicts the 500 m above sea level height.

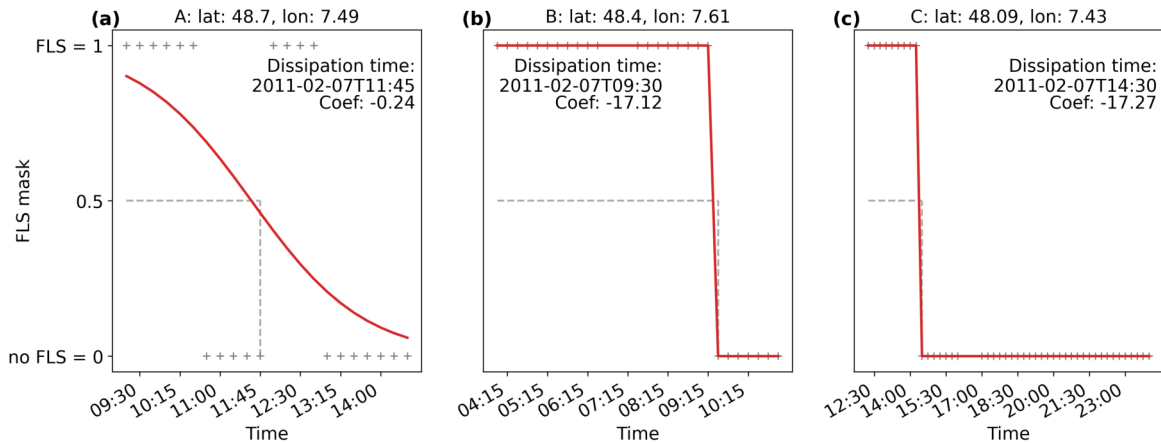


Figure 4.6: Logistic regression curves for the dissipation events on 2011-02-07 at the three locations A, B, C marked in figure 4.5. The “coef” parameter displays the slope of the logistic regression curve.

4.3 Results

4.3.1 Number and duration of FLS events

As a first view of the data set, the number of formation events, for which a formation time was calculated is shown (Fig. 4.7a). The number of formation events is representative for the number of dissipation events, which is very similar with a difference around ± 10 events per pixel. As discussed above, this slight difference is present since not every formation event is followed by a dissipation event which fulfills the requirements discussed above and vice versa. The geographic patterns of the number of detected formation events are similar to the geographic patterns of mean FLS cover (Fig. 4.1). A high number of formation events are identified in the north east of the study area (1000–1200 events over 10 years), as well as in the Po valley and at the northern coast of Spain. Moreover, a higher number of formation events are detected over cities, which are especially visible over Milan and Paris. This is potentially a systematic error of the FLS data set used, as higher FLS cover over cities can also be seen in figure 4.1a. This overestimation of FLS cover over cities is likely due to their high reflectance in the $3.9 \mu\text{m}$ channel due to solar contamination during daytime, which might then lead to a false identification of clouds during this time (EUMETSAT CM SAF, 2019). Higher daytime FLS cover over urban pixels compared to the rural surroundings can be observed in the mean diurnal course of FLS cover over London, Paris and Milan (Fig. A3.1).

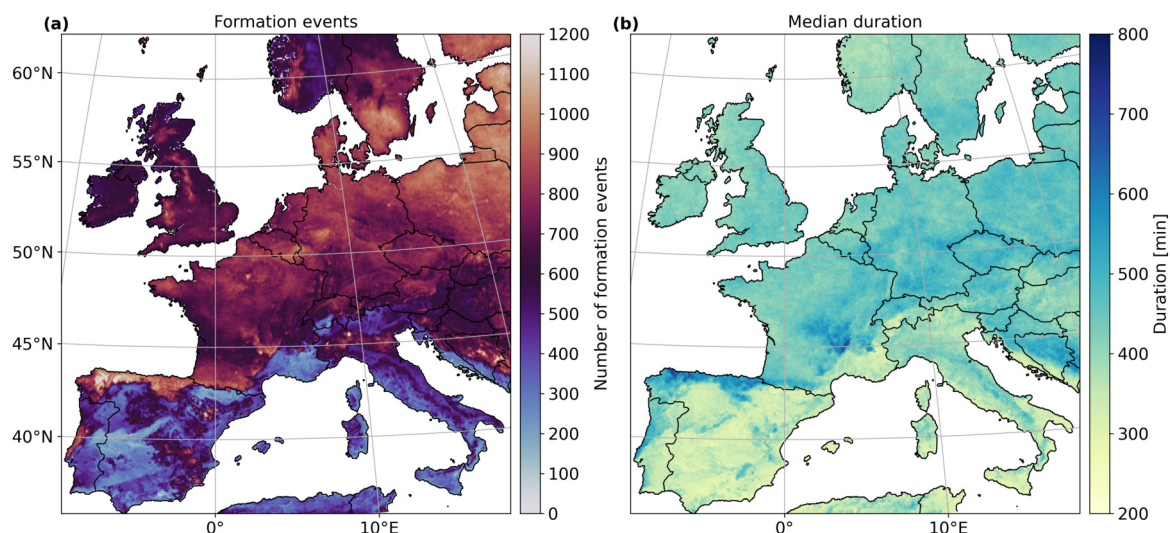


Figure 4.7: Number of formation events identified by the algorithm (a) and the median of the duration of all FLS events in which both formation and dissipation times could be determined (b). Both are calculated over the entire period (2006–2015).

A low number of formation events (300–400 events over 10 years) is detected over the Mediterranean and over mountain ranges (Alps, Pyrenees). The same geographic patterns described here for the number of formation events apply to the number of dissipation events.

The FLS duration is calculated as the difference between dissipation time and formation time for one FLS event for each pixel. To decrease the influence of outliers, the median instead of the mean is shown here. The geographic patterns of duration are similar to the FLS mean map and the number of dissipation events. FLS events tend to be longer in north-eastern part of the study area (400–600 minutes) and shorter in the Mediterranean (200–300 minutes). The geographic patterns of median duration are also similar to those of the mean FLS cover and the number of formation events. This is also visible when normalizing the three quantities (Fig. A3.2) to enhance comparability. Still, the spread of values is lower for the median duration compared to the mean FLS cover and the number of formation events.

4.3.2 Most frequent formation and dissipation times

For further analysis, the timestamps of formation and dissipation times are converted into % values as a function of day length (sunrise–sunset) for formation or dissipation during the day or as a function of night length (sunset–sunrise) for nighttime formation and dissipation. This is done to make formation and dissipation time comparable across seasons and latitudes.

To facilitate interpretation, these % values are then assigned to different classes of day- or nighttime (Tab. 4.1). To produce climatological maps of the most frequent formation and dissipation time, the mode of these classes is used for each pixel over the entire time period (Fig. 4.8) and depending on the season (Fig. 4.9). As dissipation most frequently occurs during the day (see Fig. A3.4 for the Upper Rhine valley region), the most frequent dissipation time is shown here for daytime dissipation values only. The corresponding plots of most frequent dissipation time using daytime and nighttime dissipation over the entire study period and for each season can be found in the appendix chapter A3 in Fig. A3.5 and Fig. A3.6.

Table 4.1: Day- and nighttime classes with respective % ranges.

Daytime			Nighttime		
class number	class name	% range	class number	class name	% range
1	Sunrise	0–12.4	1	Sunrise	87.5–100
2	Morning	12.5–37.4	8	Night	62.5–87.4
3	Midday	37.5–62.4	7	Midnight	37.5–62.4
4	Afternoon	62.5–87.4	6	Evening	12.5–37.4
5	Sunset	87.5–100	5	Sunset	0–12.4

When considering the entire study period (Fig. 4.8), FLS forms most frequently around midnight in large parts of the study area such as parts of France, Germany and Italy, where secondary mountain ranges as well as rather flat areas, such as the Danube valley are present. In the high-altitude mountain ranges (Alps, Pyrenees) and on the north coast of Spain, formation occurs most frequently around sunset. In the inner plateau of central Spain, western France and in the Po valley, FLS forms most frequently around sunrise and in the morning.

The geographic distribution of most frequent dissipation time shows similar spatial patterns of equal dissipation time as was seen for formation time. At the north coast of Spain, and in high-altitude mountains such as the Pyrenees and the Alps, FLS dissipates most frequently around sunrise or in the morning. In secondary mountain ranges such as the Massif Central in southern France and the pre-alpine areas of southern Germany, dissipation occurs mainly in the morning or around midday. In the low-altitude areas of central Europe, such as northern France, northern Germany and Poland but also in the inner plateau of Central Spain, dissipation occurs most frequently in the afternoon.

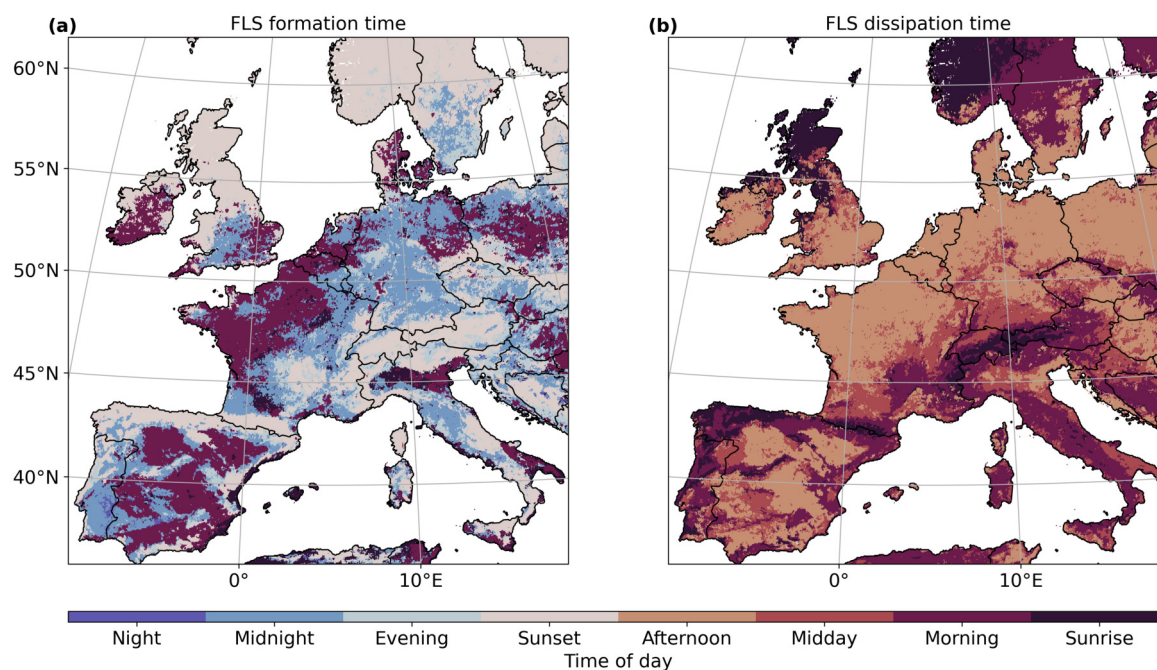


Figure 4.8: Most frequent formation (a) and dissipation times (b) over the entire study period (2006–2015).

The seasonal patterns of formation and dissipation times tend to be similar in winter (DJF) and fall (SON) and in spring (MAM) and summer (JJA) (Fig. 4.9). In winter and fall, FLS formation occurs most frequently in the evening or during the night and dissipation around midday or in the afternoon. In spring and summer, FLS typically forms later, most frequently around sunrise, but does not last as long as during winter and dissipates in the morning. The detailed geographic patterns for each season are described below.

In winter, FLS forms most frequently in the evening or around midnight, except for large areas in western France and the inner-plateau of central Spain, where formation most frequently takes place in the morning. During this time, FLS dissipates most frequently in the afternoon in most regions, and earlier (in the morning or around midday) in parts of the Alps, in the secondary mountain ranges of southern Germany or southern Italy. In spring, the most frequent formation time shifts to sunrise or to the morning hours in large parts of France, central Spain and in the Po valley. On the north coast of Spain, southern France, in the Pyrenees and in the Alps FLS forms most frequently at sunset. Dissipation most often takes place in the morning or around midday in most parts of the study area. In the Po valley and in large parts of Poland, FLS is more persistent and most frequently dissipates in the afternoon.

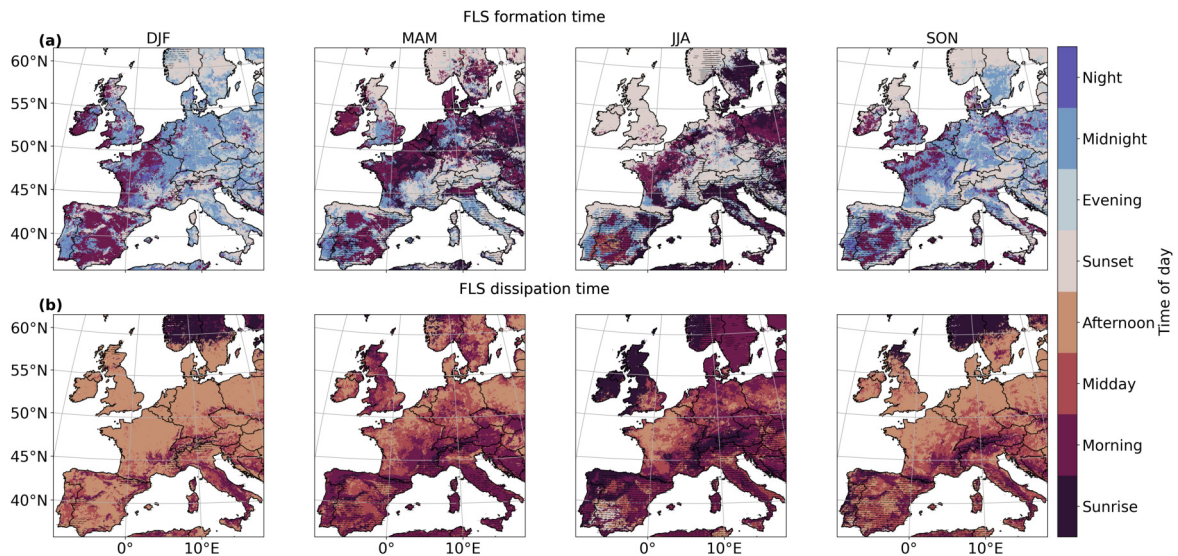


Figure 4.9: Most frequent formation (a) and dissipation time (b) for DJF, MAM, JJA and SON from 2006-2015. Pixels with a mean seasonal FLS occurrence of less than 2 hours day⁻¹ are marked with a horizontal line.

In summer, FLS formation patterns are spatially diverse, with formation at sunset (central Germany), around sunrise (Po valley) or in the morning (France). This could be influenced by the lower number of formation events in summer compared to other seasons in most parts of the study area (compare Fig. A3.3). A clear pattern of formation at sunrise is visible at the coast of the Mediterranean sea. Dissipation in summer is most frequent in the morning in large parts of the central study area. FLS dissipates earlier (around sunrise) at the northern coast of Spain and in the Alps, and later (in the afternoon) in western France. Still, these patterns should be interpreted with care, as FLS occurrence is low in most of the Mediterranean in spring and summer (compare hatched areas of Fig. 4.9). In fall, FLS formation shifts again towards nighttime hours in most of the study area, except for parts of France and Spain, where FLS forms most frequently in the morning. FLS dissipation also shifts to afternoon hours in most parts of the study area. In the Alps and the Pyrenees FLS most frequently dissipates at sunrise, in southern Germany around midday.

4.3.3 Regional study: Upper Rhine valley, Germany

While the formation and dissipation times data set can be used to produce climatological maps for Europe, it can also be used for a more thorough investigation of formation and dissipation patterns in regional studies. Such an example for a regional study is presented here by looking closely at formation and dissipation patterns in the Upper Rhine valley (depicted by the red rectangle in Fig. 4.1 and introduced in section 4.2.3) in southern Germany. The FLS patterns present in that region are likely to be influenced by local to regional modulations of the synoptic-scale weather patterns. Thus, the Upper Rhine valley is an ideal region to showcase the level of spatial detail provided by the novel data set.

The most frequent formation and dissipation time over the complete period (Fig. 4.10a) shows two distinct patches of formation and dissipation anomalies (at $\sim 7.5^\circ$ E and 48° N, and at around 48.7° N). At these locations, formation is observed to be in the morning and dissipation in the afternoon (Fig. 4.10a and e). Thus, FLS formation and dissipation at these patches are much later than in other places of the Upper Rhine Valley, where FLS usually forms around sunset or around midnight and dissipates in the morning or around midday. While the anomalies are clearly visible in the annual averages, they are especially pronounced in MAM, with formation in the morning and dissipation in the afternoon extending over the eastern slopes of the Vosges (Fig. 4.10b and f).

The percentage of values that fall into the most frequent formation and dissipation time classes relative to all formation and dissipation situations provides a measure for the representativeness of the mode as a proxy for the typical formation and dissipation time for each pixel (Fig. 4.10c, d, g, h). The percentage of values in the most frequent class lies around 15-20% for the formation over the complete year and rises to values around 20-25% in MAM and is highest in the anomalous patches described above. Considering the dissipation, the geographic distribution of % values is similar but on average 10% higher compared to formation, showing that the temporal variability in dissipation time is lower than the formation time.

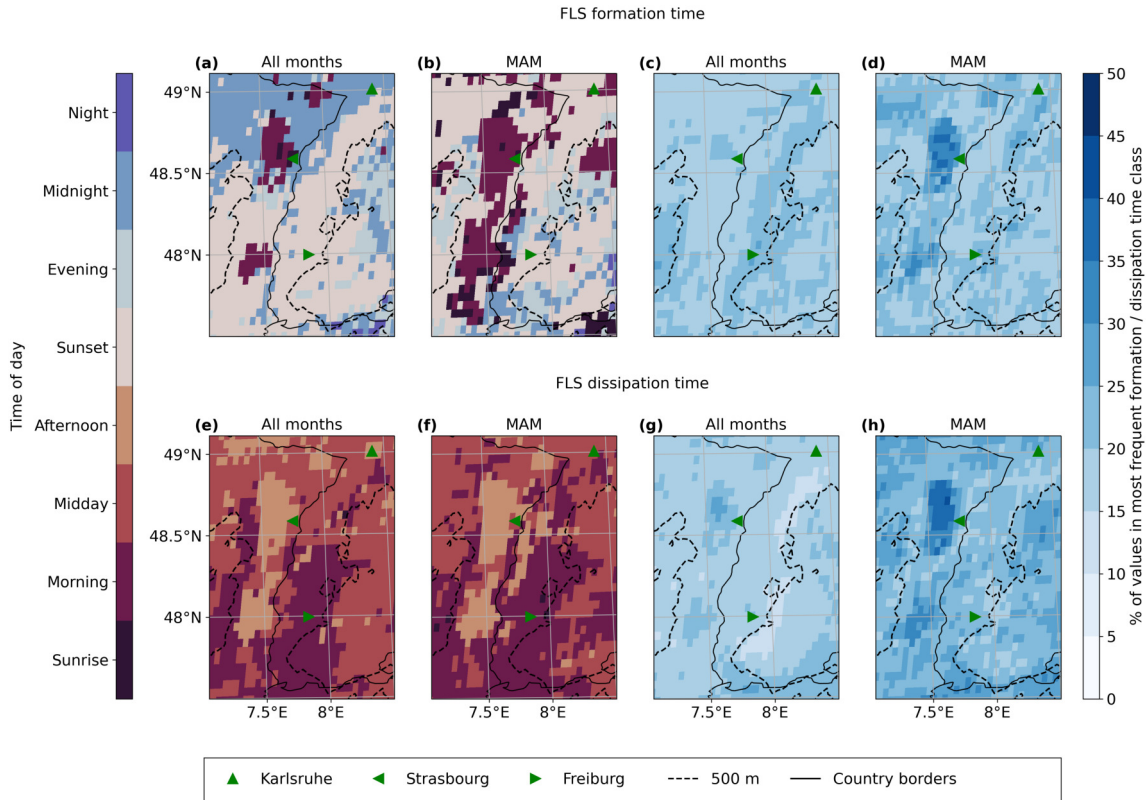


Figure 4.10: Most frequent formation (a and b) and dissipation time (e and f) in the Upper Rhine valley and the corresponding % of values contained in the most frequent formation (c and d) and dissipation class (g and h) for the complete time period (All months) and in spring (MAM). The country border between France and Germany also marks the course of the river Rhine. The black dashed line depicts the 500 m above sea level height.

This case study provides two important insights into the novel formation and dissipation data set. First of all, the relatively high spatial and temporal resolution (native SEVIRI resolution, see Sec. 4.2.1) makes it a useful product to study regional formation and dissipation patterns. Secondly, the dissipation time features a lower temporal variability than the formation time, as evidenced by the systematically higher fraction of events in the most frequent class. Over most of the study area, more than two thirds of dissipation events occur during the day, whereas formation time is equally distributed between daytime and nighttime in large parts of the study area (Fig. A3.4). This is likely due to a higher number of possible formation pathways and thus formation times whereas dissipation is strongly influenced by solar radiation. This is further evaluated in the discussion below.

4.4 Discussion

The geographic patterns of formation and dissipation time clearly underline the role of topography for the occurrence and development of FLS events. In large mountain ranges such as the Alps or the Pyrenees FLS forms earlier (around sunset) compared to lower terrain but also dissipates earlier (around sunrise). These geographic patterns may be interpreted as the signature of regionally characteristic processes influencing FLS formation and dissipation. It is likely that formation at sunset in those mountain ranges is due to advected FLS layers, especially on the windward slopes of those ranges, e.g. on the northern slopes of the Pyrenees. After sunrise, these FLS layers are likely to be “burned-off” as the sun reaches the mountain tops first. In the Po valley in Italy or in the inner plateau of Central Spain, FLS formation is likely to be due to nocturnal cooling and a subsequent transition into a low stratus cloud (Roach, 1995). Other potential involved processes could be topography induced drainage flows and turbulence (Price, 2019).

In smaller mountain ranges (e.g. the Black Forest), the dependence of formation and dissipation time on topography is not always clear. A potential explanation for this could be that the local modulation of the meteorological parameters that determine the FLS life cycle (e.g. relative humidity, wind) scales with the topographic features. In areas with moderate topography, the topography-induced local modulation of the meteorological setting would then have a weaker effect on the resulting geographic patterns of formation and dissipation times than in mountainous areas. In general, topographic features are only modulating the meteorological drivers responsible for FLS formation and dissipation. As has been shown in previous studies (chapter 2, Pauli et al., 2020) meteorological factors are the main drivers of FLS occurrence over central Europe. In addition, the presented formation and dissipation times should be interpreted with care over areas with a low sample size or high topographic variability, as the pixel size of the product (3–5 km, depending on the exact position) is too large to be able to depict small-scale variations in FLS.

Besides topography, the proximity to the sea is important for the timing of FLS formation and dissipation as well. In general, coastal fog is strongly influenced by the meteorological conditions and ocean-atmosphere interactions (Gultepe et al., 2021). The results presented here show a clear pattern of FLS formation at sunrise at the coast of the Mediterranean sea in summer (Fig. 4.9). According to Azorin-Molina et al. (2009) humid winds from the Mediterranean sea in combination with mountain ranges close to the coastline can lead to FLS occurrence at the coast. Therefore, at the Mediterranean coast, the diurnal cycle of the coastal circulation is likely a main driver

of the observed patterns in FLS formation and dissipation times. At the Atlantic coast of northern Spain, blocking of FLS by the Cantabrian mountains (Egli et al., 2017) and upwelling (Alvarez et al., 2010) might be important for FLS formation, as the latter plays an important role in other FLS prone regions such as at the south western African coast (Olivier and Stockton, 1989; Cermak, 2012; Andersen et al., 2020). Close to the french Atlantic coast over the Landes forest south of 45° N FLS forms earlier over the forest (around midnight) compared to its surroundings (FLS formation in the morning), particularly in spring and summer. This is in line with enhanced nighttime FLS occurrence over this forest compared to its surroundings shown in Pauli et al. (2022b) (chapter 3), potentially due to enhanced emissions of biogenic volatile organic compounds over the forest, which can serve as cloud condensation nuclei (Spracklen et al., 2008; Kammer et al., 2018).

Another driving factor for the observed patterns is the solar radiation. As stated above, this is especially true for the dissipation time, where processes related to downwelling solar radiation (absorption inside the FLS layer, sensible heat flux) more strongly influence the dissipation of FLS compared to other, more subtle drivers (Wærsted et al., 2019). Formation can occur through various pathways during the day or during the night, with formation patterns showing a much higher variability. To focus on specific FLS events and to unravel different formation and dissipation pathways, one could filter for stationary FLS events using image detection techniques or filter for meteorological conditions.

The higher number of formation and dissipation events over cities does not transfer to a difference of formation and dissipation time between cities and surrounding land. In contrast to the literature (cf. Yan et al., 2020) earlier dissipation over cities is not visible in the shown climatological means. On the one hand, this may be due to the assignment of the dissipation and formation times in % to different daytime and nighttime classes (Tab. 4.1) and the subsequent calculation of the mode, which could be investigated in more detail by looking at the raw formation and dissipation times over cities and surrounding areas. On the other hand, as discussed above, the FLS data set is likely to be flawed over cities during daytime, and thus misclassifications are likely to superimpose the actual patterns in the first place. Applying the proposed logistic regression algorithm to a robust high resolution cloud mask over cities (Fuchs et al., 2022) could add to our knowledge of the FLS life cycle over urban areas.

As these discussions on the possible processes underlying the geographic formation and dissipation patterns remain speculative, explicit regional analyses on the drivers of FLS formation and dissipation time are necessary. In addition, the potential

influence of multi-layer cloud situations and misclassifications on the presented FLS formation and dissipation time has to be considered for the interpretation of the results. In a regional study, the FLS data set by Egli et al. (2017) used as a basis for this study has been shown to agree well with active sensor data (chapter 3, Pauli et al., 2022b). The formation and dissipation time patterns shown with this novel FLS formation and dissipation data set are also in line with LES and modeling studies over Europe, showing FLS formation in the night and dissipation after sunrise (cf. Roach, 1995; Haeffelin et al., 2010; Bergot, 2016; Steeneveld and de Bode, 2018). While modeling, LES studies and local measurements display the FLS life cycle with high temporal resolution at a specific site and can also include the vertical component of an FLS event, the data set presented here provides a geographic perspective on formation and dissipation time over complete central Europe.

4.5 Specific conclusions

The central aim of this study was to investigate spatial patterns of FLS formation and dissipation times over central Europe. For this purpose, an algorithm was designed, which applies logistic regression to a binary satellite-based FLS cloud mask. With the novel data set, FLS formation and dissipation times were investigated, largely confirming known patterns of formation during the night and dissipation in the morning or in the afternoon. In general, FLS occurrence, formation and dissipation are dependent on various drivers (cf. Roach, 1995; Gultepe et al., 2007; Price, 2019; Pauli et al., 2020). The results presented here underline the importance of topography-induced modulations of meteorology for FLS formation and dissipation: In mountain ranges, FLS forms most frequently at sunset and dissipates in the morning. At lower altitudes, such as in large-scale river valleys, FLS most commonly forms around sunrise and dissipates in the afternoon. Furthermore, a higher variability in formation times compared to dissipation times is found, with the latter being much more dependent on solar radiation. The data set adds a geographic component to our knowledge of FLS formation and dissipation and provides a basis for future studies.

In the future, a preprocessing step could be implemented by adding a filter to more specifically study stationary FLS situations that may be indicative of radiation fog, e.g. using image detection techniques or focusing on specific meteorological conditions. The new algorithm can also be applied to other existing satellite-based FLS data sets with a high temporal resolution over regions where FLS are

an important component of the climate and environment and station measurements are lacking such as in the Namib Desert (Andersen and Cermak, 2018), and can be compared to diurnal patterns identified there so far (Andersen et al., 2019). Furthermore, this data set holds promise to help better understand the drivers of FLS formation and dissipation at continental scales.

5 Identifying and understanding fog and low stratus formation and dissipation regimes

5.1 Specific Motivation and Aim

Temporal and spatial patterns of fog and low stratus formation and dissipation vary considerably across continental Europe and were shown to be influenced by topography, the distance to the coast and the solar cycle in chapter 4. The meteorological and land surface drivers of FLS occurrence identified in chapter 2 most likely also play an important role for the FLS life cycle. As such, the research of the chapter at hand merges the knowledge gained in chapter 2 and 4 to identify geographically distinct FLS regimes and quantify sensitivities of FLS formation and dissipation time to changes in meteorological and land surface parameters.

Previous studies have identified FLS regimes based on their daily mean occurrence (Egli et al., 2019) or their daily frequency cycle (Knerr et al., 2021) using clustering algorithms such as self-organizing maps (SOMs) (Kohonen, 1982) and hierarchical clustering (Müllner, 2011, 2013). The resulting clusters were shown to be determined by the influence of topography and atmospheric conditions (Egli et al., 2019; Knerr et al., 2021). Besides using occurrence means as an input, sensitivities presenting relationships of the investigated variable with environmental conditions can be used as a basis for clustering, to group regions with similar sensitivity patterns (Douglas and Stier, 2021; Zuidema et al., 2022). Douglas and Stier (2021) created global regimes of cloud controlling factors by using the sensitivities obtained from a machine learning set-up in a k-means clustering approach. Similar approaches have also been applied in research fields outside the atmospheric sciences, for example, to analyze tree climate response using monthly correlations of tree growth with environmental variables as an input for SOMs (Zuidema et al., 2022). By delineating regimes of FLS formation and dissipation based on their sensitivities to

environmental conditions will shed new light onto the geographic variability of FLS processes.

The goal of the study is to identify and delineate FLS formation and dissipation regimes over central Europe and to quantify the sensitivity of the FLS life cycle to changes in meteorological and land surface based drivers identified in chapter 2 (Pauli et al., 2020). This is done by using pixel-based correlations of FLS formation and dissipation time with meteorological and land surface drivers as an input into a hierarchical clustering algorithm. Region-specific correlations are analyzed for each cluster and on different hierarchy levels. The guiding hypothesis is that the sensitivities of the identified FLS regimes to changes in environmental conditions depend on background climate and FLS type. This analysis contributes to the understanding of the spatial variations of FLS regimes, outlines the differences in the influence of meteorological and land surface drivers on the FLS life cycle and provides a basis for further FLS region-specific sensitivity studies.

5.2 Data and methods

5.2.1 Fog and low stratus formation and dissipation time data set

The basis of the study is the satellite-based FLS formation and dissipation time data set by Pauli et al. (2021), in particular data on daytime dissipation and nighttime formation to specifically investigate the radiation fog life cycle. This formation and dissipation time data set has been created by applying logistic regression to an existing FLS data set based on Meteosat SEVIRI (Spinning Enhanced Visible and Infrared Imager) data (Egli et al., 2017). Details on the logistic regression algorithm applied can be found in Pauli et al. (2022a). The spatial resolution of the Pauli et al. (2021) data set is 3 km at nadir (around 5 km in central Europe) and FLS formation and dissipation times are given as percentage values as a function of day or night length.

To describe the prevailing meteorological condition, daily means of mean surface pressure (MSP) and wind speed (WS) from ERA5 land are used (Muñoz Sabater, 2019). The state of the land surface is described using daily means of land surface temperature (LST) and evapotranspiration (ET) from the Land-Surface Analysis Satellite Applications Facility (LSA-SAF) (Trigo et al., 2011). Rescaling to SEVIRI resolution and quality control is carried out as described in Pauli et al. (2020). All data sets span the years 2006–2015 and the area of central Europe.

5.2.2 Clustering input and algorithm description

Pixel-based correlation calculation

As a first step, the correlation of daytime dissipation and nighttime formation with daily means of MSP, WS, ET and LST is quantified by calculating the corresponding Spearman's rho correlations and p-values for each pixel over central Europe. To obtain information on the seasonal cycle of the correlations, these are performed for each month separately. In addition, a moving window approach based on a 3x3 pixel region is implemented to enhance the number of observations used for the correlation analyses.

The significance of the obtained Spearman's rho correlation values is considered by interpolating over the non-significant ($p > 0.05$) pixels using linear interpolation, as applying a strict p-value threshold would lead to the exclusion of pixels in the subsequent clustering analysis. All correlation values of one pixel (with each variable, for each month) have to be valid to assign the pixel to a cluster. Different interpolation techniques, specifically nearest neighbor, were found to provide less spatial detail in the correlation maps than linear interpolation. To show the influence of the interpolation method on the resulting clusters, the appendix chapter A4 shows the results of the clustering procedure using nearest neighbor.

The effect of the interpolation on the input data is shown with exemplary maps over Germany displaying the correlation of FLS formation and dissipation time with mean surface pressure for the month of February in figure 5.1. While the variability of the spatial patterns is high, the correlations indicate a shift of sensitivities from maritime to continental regions which is further discussed in section 5.3. While the linear interpolation slightly smooths the spatial patterns of the correlation values, the difference to the raw correlation values (left column) is small.

Hierarchical clustering

In order to group areas in central Europe with a similar formation and dissipation regime and to analyze their sensitivities to changes of the four drivers, a hierarchical, agglomerative clustering algorithm is applied to the correlations calculated for each pixel. Agglomerative clustering is an unsupervised machine learning technique which separates a data set into singleton nodes and merges the closest nodes until only one node is left (Everitt, 2011; Müllner, 2011). Merging can be performed using various linkage strategies, which define the measure of distance (or similarity). In the context of this study, the *ward* linkage is applied, which aims to minimize the

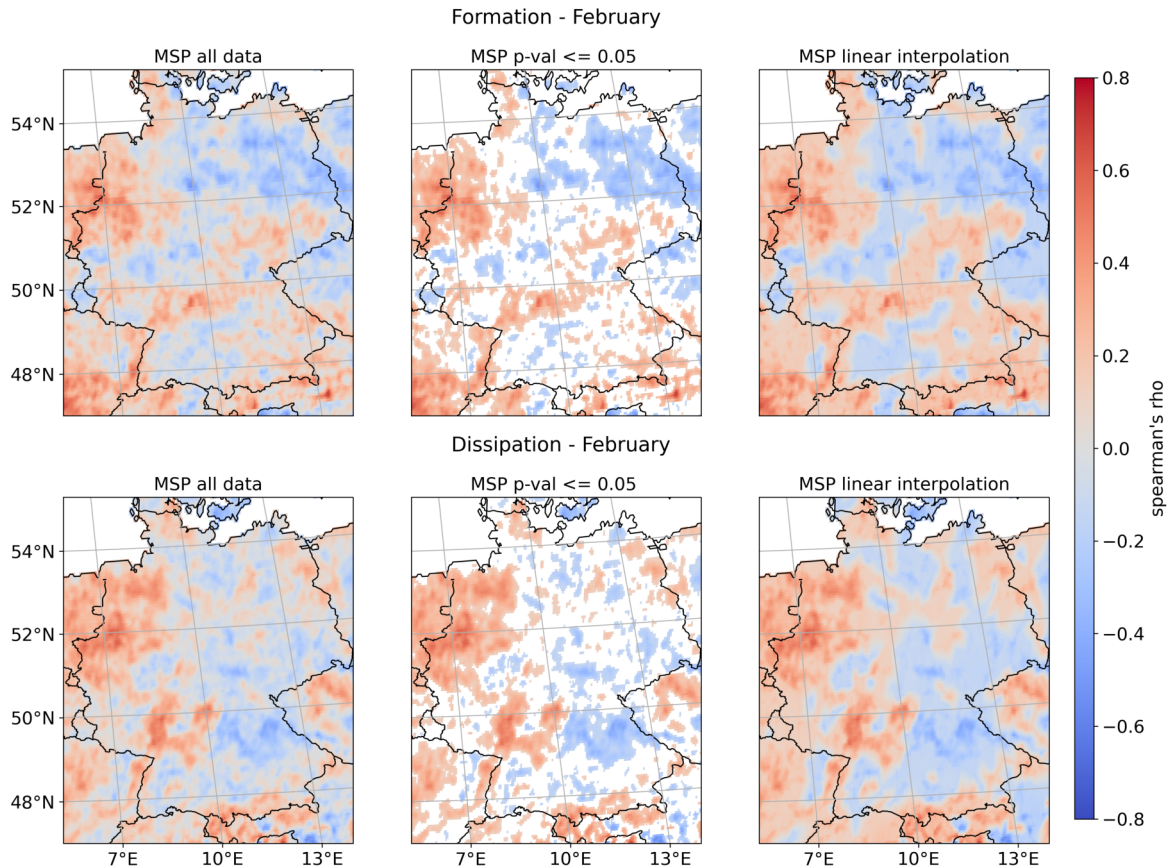


Figure 5.1: Correlation map for mean surface pressure (MSP) with formation (upper) and dissipation (lower) time for the month of February. The left column displays the raw Spearman's rho values, the center column those with a p -value ≤ 0.05 and the right column the result of the linear interpolation values.

increase in the error sum of squares within a cluster summed over all variables (Everitt, 2011).

One of the major advantages of using a hierarchical clustering approach is that it is possible to decide on the level of hierarchy where the clusters are chosen from and to view how clusters on the same or different hierarchy levels are related. In the analysis conducted here, this is especially useful as it makes it possible to analyze how clusters of similar FLS life cycle regimes relate geographically. Nevertheless, the main disadvantages of agglomerative hierarchical clustering is that if two observations have been joined they can not be separated in a later step (Everitt, 2011).

In the context of this study, the fastcluster package in python developed by Müllner (2011, 2013) is used, as it clusters a large input data set efficiently. The fastcluster package has been used successfully to cluster general weather situations into groups of similar fog patterns in Egli et al. (2019). The input data for the fastcluster

algorithm consists of one row for each (land-) pixel and one column for each monthly correlation value of each variable for both formation and dissipation (Tab.5.1).

Table 5.1: Example layout of the table used as an input for the hierarchical clustering algorithm. The correlation values are for visualization purposes only. The input table used in the clustering algorithm contains the correlation values with all variables (MSP, WS, ET, LST).

n_{Pixel}	corr_{MSP} form Jan	...	corr_{MSP} form Dec	...	corr_{MSP} diss Jan	...	corr_{MSP} diss Dec	...	corr_{LST} diss Dec
1	0.50	...	0.40	...	0.31	...	-0.15	...	-0.32
2	0.01	...	0.72	...	0.83	...	0.24	...	-0.13
3	0.45	...	-0.37	...	-0.23	...	0.02	...	0.8
...
206095	0.73	...	0.42	...	-0.7	...	0.21	...	0.27

The output of the hierarchical clustering procedure is a dendrogram, which is displayed as a rooted tree, with leaves as the initial nodes and internal nodes depicting where two clusters are joined (Müllner, 2011, 2013). The length of the stems (vertical lines) represents the distances at which the clusters are joined (Everitt, 2011), which helps to identify cluster partitions with a large increase in distance measure.

5.3 Results and Discussion

5.3.1 Regional cluster overview

The dendrogram of the hierarchical clustering procedure shown in figure 5.2 clearly delineates four main clusters, as shown by the increase in distance measure after the data is merged into these four clusters. A spatial map of the resulting four FLS formation and dissipation regimes is shown in figure 5.3a, whereas smaller, more regional clusters are visible when using a hierarchy level of 15 clusters (Fig. 5.3b). The spatial map using nearest neighbor as an interpolation technique is shown in figure A4.1. The identified clusters when using different interpolation techniques are similar and the spatial details described in the following for the most part also apply to the clustering results when using nearest neighbor for the input data interpolation.

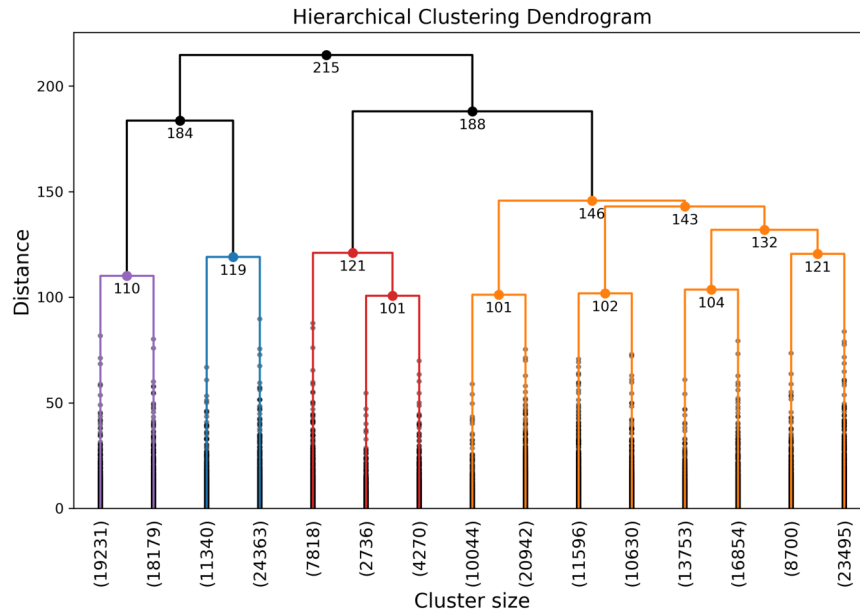


Figure 5.2: Dendrogram of the hierarchical clustering procedure, displaying the relationship of clusters and their joining distance. The y-axis shows the distance of data points to other members of the cluster (here the intra-cluster variance) with numbers at the nodes highlighting the distance at which the clusters are joined. The clusters are colored to mark the four main clusters. The points on the vertical lines show the position of the joining nodes. Cluster sizes (number of pixels contained) on the x-axis are given for the 15 clusters shown.

The FLS life cycle regimes obtained by the hierarchical clustering approach generally match well to known regional and subregional FLS types and climate zones. Different background climates are most likely reflected in the sensitivities of formation and dissipation time to changes in meteorological and land surface drivers, which can also lead to the assignment of geographically close regions to different clusters. Potential FLS types and background climatology of the four main clusters and their relation to the 15 subclusters are discussed in the following.

The largest cluster identified by the hierarchical clustering algorithm (orange, further referred to as “central cluster” as it covers most of central Europe) contains most of the data points and covers large parts of Italy and southeastern Europe, as well as parts of Germany, France, Portugal and Spain. The central cluster is derived from multiple subclusters (Fig. 5.3b) showing regions of known FLS regimes, such as the north-western coast of the Iberian peninsula (Egli et al., 2017), the Swiss Plateau and the Po valley (Bendix, 1994; Cermak et al., 2009). In these regions, FLS of radiative and advective origin are present.

The cluster in the north western part of the study area contains the UK, as well as coastal areas of France, Belgium, the Netherlands and Germany (blue, further referred to as “maritime cluster”). It is formed by two subclusters (Fig. 5.3b), which

also separate the east coast of the UK from the west coast, with the latter being much more prone to the incoming westerlies and higher rainfall (Mayes and Wheeler, 2013). According to Mayes (2013), fog on the west coast of England and Wales is more likely of advective than of radiative origin.

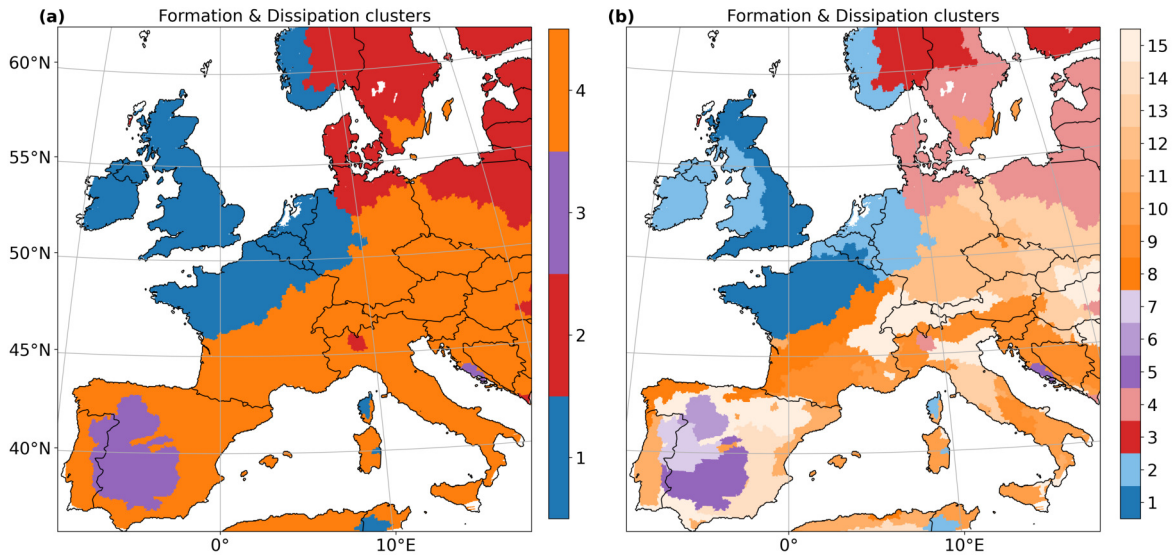


Figure 5.3: Map of formation and dissipation clusters identified by the hierarchical clustering algorithm. Two different hierarchy levels are displayed, 4 clusters (a) and 15 clusters (b). The white areas in some coastal regions are due to missing input values, as these are situated outside the convex hull used in the interpolation procedure.

The cluster situated in the north eastern part of the study area (red, further referred to as “Baltic-Scandinavian cluster”) covers most Baltic and Scandinavian states. FLS occurrence has been shown to be especially large in this region during prevailing anticyclonic conditions (Egli et al., 2017), leading to the advection of warm air over cold and potentially snow-covered land and the formation of advection fog (Avotniece et al., 2015). Besides advection fog, radiation fog can form as well in this region, especially in anticyclonic conditions in fall (Avotniece et al., 2015).

The fourth major cluster covers parts of Spain, Portugal and Croatia (purple, Mediterranean cluster). In summer and early fall, FLS on the Iberian Peninsula forms mainly due to advection of clouds onto the mountain ranges (Estrela et al., 2008), whereas in winter, FLS patterns are due to frontal systems (Valiente et al., 2011). Nevertheless, major parts of the Iberian Peninsula are assigned to the central cluster. While the climate of the north coast is oceanic and strongly influenced by the Atlantic, the climate in the central Plateau of the Iberian Peninsula has continental characteristics (Royé et al., 2018), leading to different sensitivities of formation and dissipation time to the investigated drivers.

5.3.2 Sub-regional cluster analysis

With the identification of distinct FLS regimes, their specific characteristics can be analyzed with greater detail regarding meteorological and land surface drivers of the FLS life cycle. In the following, the sensitivity of FLS formation and dissipation times to changes in meteorological and land surface drivers is analyzed for three sub-regional clusters identified by the hierarchical clustering algorithm. The sensitivities are given as the mean monthly Spearman's rho value for all pixels with $p \leq 0.05$. A positive correlation indicates a shift to later (earlier) formation and dissipation times with increasing (decreasing) feature value, whereas a negative correlation indicates a shift to earlier (later) formation and dissipation times with increasing (decreasing) feature values. All clusters analyzed below have been extracted from a hierarchy level of 80 clusters, as this level provides distinct subregional clusters with a considerable spatial extent.

Po valley, Italy

The first sub-regional cluster is part of the large central cluster described above (Fig. 5.3) and covers large parts of the Po valley in northern Italy (Fig. 5.4). This region is known for high FLS occurrence (Fuzzi et al., 1992; Bendix, 1994; Cermak et al., 2009; Egli et al., 2017) with a typical radiation FLS regime of formation before sunrise and dissipation in the afternoon (chapter 4, Pauli et al., 2022a). As shown here, the hierarchical clustering algorithm is thus able to extract the Po valley region by solely using the calculated correlations. Sensitivities of formation and dissipation time to variations in LST and ET are low during winter and fall and increase in spring and summer with a maximum of 0.5 for ET in June. This indicates that higher temperatures together with high evapotranspiration lead to a delayed FLS formation during the night and delayed FLS dissipation during the day. A potential explanation for this shift to later FLS formation times in spring and summer is, that due to generally higher daytime temperatures, it takes longer for the air to cool down to its dewpoint during the night. Increasing LST values in December lead to an earlier dissipation, which is in accordance with studies looking at the influence of temperature or solar radiation on the FLS life cycle (Haeffelin et al., 2010; Wærsted et al., 2019; Pauli et al., 2020). Generally, moist, Mediterranean air masses are important for FLS formation in the Po valley, as cold air at the ground mixes with the overlying moist air (Bendix, 1994). Looking at the variability of temperature and

moisture between different pressure levels in the future could further give insight into these processes.

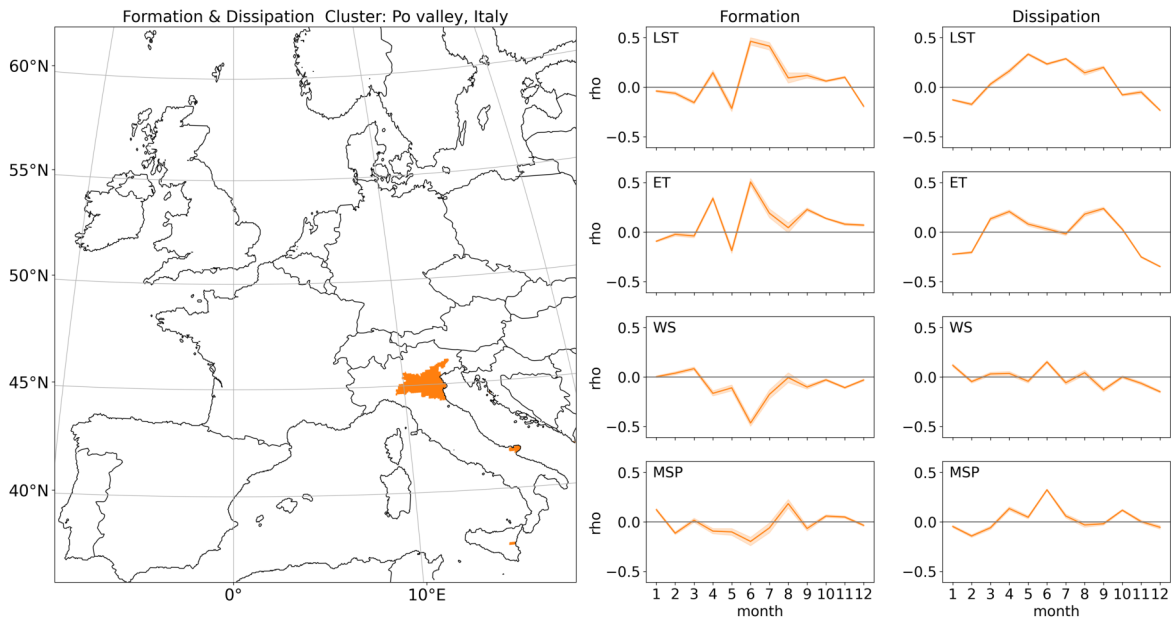


Figure 5.4: Sub-regional cluster covering the Po valley and monthly sensitivities of FLS formation and dissipation time to meteorological and land surface drivers. The sensitivities are given as the mean monthly Spearman's rho value for all pixels with $p \leq 0.05$. The 95th percentile is displayed as the shaded background.

The correlations of FLS formation and dissipation time with wind speed and mean surface pressure are generally low in magnitude except for summer: In the summer months, low WS and low MSP decrease the duration of FLS events by shifting FLS formation time to later times in the night and FLS dissipation time to earlier times in the morning. The sensitivities of FLS formation and dissipation to changes in MSP are in line with other studies, which have shown that high mean surface pressure (as found in anticyclonic conditions) leads to more persistent FLS and thus longer FLS duration (Ye, 2009; Egli et al., 2019). Concerning wind speed, the results are contrasting to other studies which identified low wind speeds to increase the duration of FLS (Cuxart and Jiménez, 2012; Pérez-Díaz et al., 2017; Price, 2019). Nevertheless, the influence of wind speed on FLS formation and dissipation time is most likely region-specific, dependent on wind direction, FLS type and the differentiation of low stratus and ground fog. Additionally, studying the interactive effects of MSP and WS (similarly as in chapter 2, Fig. 2.9) on the FLS life cycle would give further insights into the involved processes.

Northwest of England & Wales

The second sub-regional cluster is part of the maritime cluster and covers the coastal areas of the northwest of England and Wales (Fig. 5.5). Both FLS formation and dissipation show similar sensitivities to variations in LST and ET with the highest correlations present in summer, which lead to a shift of the FLS life cycle to both later formation and dissipation. In winter, correlations mostly have lower magnitudes but display a shift to earlier dissipation for high LST values.

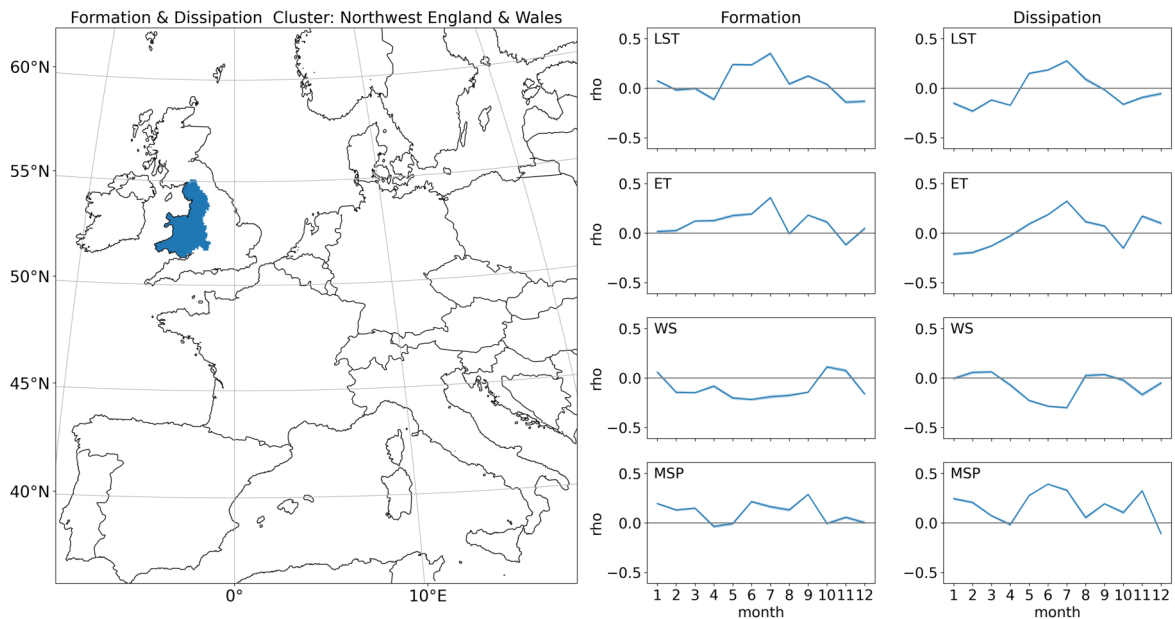


Figure 5.5: Sub-regional cluster covering the Northwest of England and Wales and monthly sensitivities of FLS formation and dissipation time to meteorological and land surface drivers. The sensitivities are given as the mean monthly Spearman's rho value for all pixels with $p \leq 0.05$. The 95th percentile is displayed as the shaded background.

The influence of MSP and WS on FLS formation and dissipation time varies strongly between formation and dissipation and the two variables. In high wind speed conditions, FLS forms earlier but also dissipates earlier in the summer months. In fall, the pattern is different: FLS forms later in high wind speed conditions and dissipates earlier. This could be due to high wind speeds associated with storm systems, leading to fast changes in low stratus cover and fog in that region. Unfortunately, no information on wind direction is available here, which makes it not possible to say if the different sensitivities are due to a change in circulation patterns. According to Kenworthy (2014), wind direction plays a major role for FLS persistence in the northern part of the cluster region, where low stratus clouds persist after the passage of a cold front when the following wind direction is northwesterly.

The correlation of FLS formation and dissipation time with daily means of MSP is positive throughout the year and highest in summer. A shift of FLS dissipation to later times in high pressure conditions has been shown in previous studies (Ye, 2009; Egli et al., 2019). With the highly dynamic weather in this region, together with a large frequency of storms during fall and winter (Mayes, 2013; Kenworthy, 2014), other drivers besides those investigated here are likely to influence the detected FLS regimes and sensitivities as well. One of these drivers is likely the sea surface temperature (SST), which has been shown to influence FLS formation at the south and east coast of England, especially in spring and early summer when SSTs are still low (Perry, 2014; Fallmann et al., 2019).

Jucar basin, Spain

The third sub-regional cluster is located near Valencia in the Jucar basin, at the Mediterranean east coast of Spain (Fig. 5.6). The correlation of LST and ET with FLS formation time varies between 0.2 and -0.2 with highest magnitudes in March and November for LST and in May for ET. These mostly negative correlations indicate that FLS forms earlier with higher LST and ET feature values. The correlation of FLS dissipation time with LST and ET is mostly negative, suggesting that FLS dissipates earlier when LST and ET are higher. As discussed above, the positive influence of higher temperatures and solar radiation on FLS dissipation has been shown in chapter 2 and in other studies as well (Haeffelin et al., 2010; Wærsted et al., 2019; Pauli et al., 2020).

FLS formation and dissipation sensitivity to changes in wind speed and mean surface pressure varies considerably over the course of the year. The correlations of formation time with wind speed are of low magnitude and are positive in winter and negative in summer. Earlier FLS formation in high wind speed situations (negative correlation) could relate to FLS advection from the sea onto the mountain ranges near the coast (Estrela et al., 2008), especially during easterly circulation types in summer and early fall (Valiente et al., 2011; Royé et al., 2018). FLS dissipation is shown to be later in high wind speed situations over all seasons, which also suggests a connection to specific circulation types, wind direction and topography of the region. FLS duration is prolonged in high pressure situations as it forms earlier and dissipates later. Still, correlations vary substantially in magnitude over the course of the year, which makes it difficult to extract a clear dependency of the FLS life cycle on both wind speed and surface pressure.

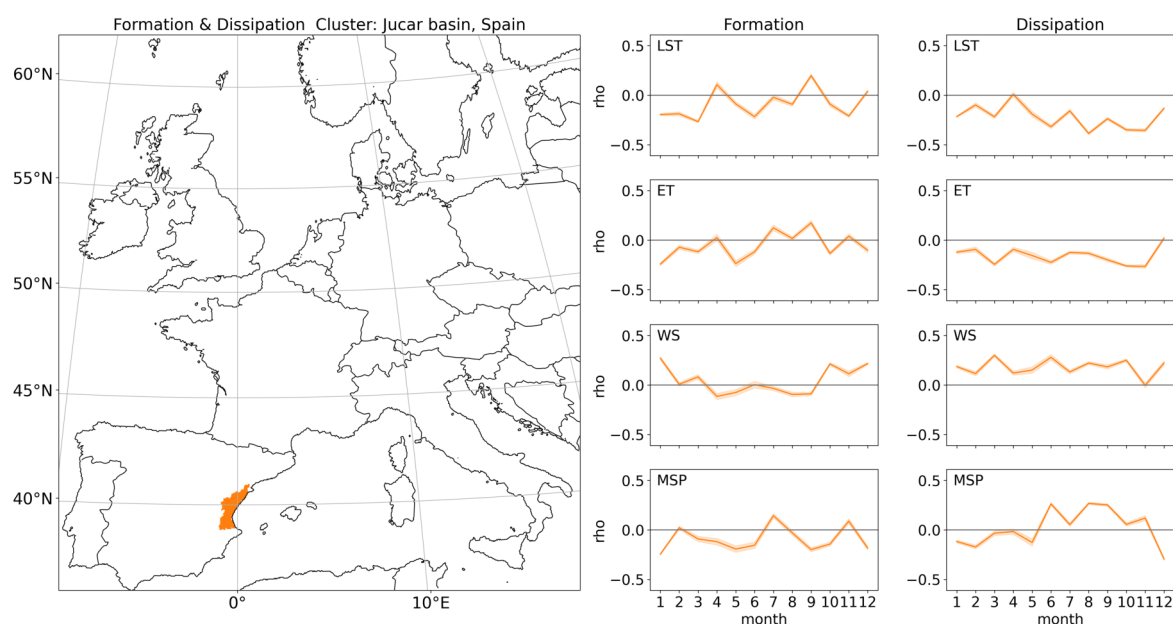


Figure 5.6: Sub-regional cluster covering the Jucar basin on the east coast of Spain and monthly sensitivities of FLS formation and dissipation time to meteorological and land surface drivers. The sensitivities are given as the mean monthly Spearman's rho value for all pixels with $p \leq 0.05$. The 95th percentile is displayed as the shaded background.

5.4 Specific conclusions

The aim of this study was to identify FLS life cycle regimes over central Europe by applying hierarchical clustering to correlations of FLS formation and dissipation time with meteorological and land surface drivers. The differences in the relationship of the FLS life cycle to environmental conditions across different subregional FLS life cycle regimes were outlined and knowledge on the main drivers for FLS occurrence gained in chapter 2 (Pauli et al., 2020) was applied. Distinct geographic regimes were found, which are most likely dependent on the relation of FLS formation and dissipation time with the background climate, as well as the proximity to the coast and topography given by the sensitivities used as the input data. The sensitivities of three exemplary FLS regimes show the potential of this approach to understand region-specific FLS life cycles as a function of land surface characteristics and meteorological conditions: An increase in land surface temperatures leads to a shift to later dissipation times in summer in the Po valley, but to earlier dissipation at the eastern coast of the Iberian Peninsula. Similarly, the influence of mean surface pressure and wind speed on the FLS life cycle varies across clusters: Increasing mean surface pressure values leads to a shift to later formation times over the Northwest of England and Wales but to earlier formation in the Po valley and the eastern coast

of the Iberian Peninsula. In addition, while LST seems to be highly important for the FLS life cycle in the Po valley, FLS formation and dissipation processes on the British Isles seem to depend more on the meteorological variability displayed by changes in wind speed and mean surface pressure.

Apart from the variation in sensitivities pertaining to cluster location, the type of fog (e.g. radiation or advection fog) and the distinction between (ground) fog and lifted fog (stratus) most likely also influences the observed patterns. Applying weather type filters or using active satellite data for the selection of (ground) fog days could help in distinguishing fog type specific sensitivities, which is critical going forward. The clustering procedure can be further improved by using daytime and nighttime means for the correlations. This will further clarify formation sensitivities, as these show a larger variability than dissipation sensitivities, similar to what has been shown for the climatological patterns in chapter 4 (Pauli et al., 2022a). Including previously not considered variables, such as wind direction as well as temperature and moisture on different pressure levels, will potentially further help in evaluating region-specific processes of the FLS life cycle.

The FLS regimes identified in this study provide an ideal working ground to further analyze region-specific FLS life cycle drivers. A possible approach would be to apply machine learning to predict FLS formation, dissipation or duration over multiple regions and compare performance and sensitivities. Combining this approach with a filter for FLS types will further help to quantify the influence of meteorological and land surface drivers on region-specific fog and low stratus life cycle patterns.

6 Concluding discussion and outlook

6.1 Concluding discussion

Fog and low stratus processes occur on a large range of scales, from aerosol processes on microphysical scales, to processes related to topography on the landscape scale. The life cycle of FLS is influenced across these scales by a range of atmospheric and land surface processes and their magnitude and direction depend on location, season and fog type. In this thesis, the central aim was to quantify the influence of meteorological and land surface drivers on FLS occurrence and FLS life cycle over central Europe. This was done by applying several statistical methods to satellite-based data sets and reanalysis data. The understanding of fog and low stratus processes, the role of the land surface and knowledge of fog and low stratus formation and dissipation processes was improved under the guidance of the following four research questions identified in chapter 1.5:

1. *What are the main drivers of large-scale spatial and temporal fog and low stratus patterns?*
2. *How does the land surface influence spatial variations in fog and low stratus occurrence?*
3. *What are the climatological patterns of fog and low stratus formation and dissipation time?*
4. *What are fog and low stratus formation and dissipation regimes on regional to sub-regional scales?*

The main findings and results of this thesis concerning these questions are discussed in the following.

(1) Meteorological and land surface drivers of fog and low stratus occurrence

In a regional study over continental central Europe, FLS occurrence was modeled using meteorological and land surface parameters. In agreement with the guiding hypothesis proposed in chapter 1.5, atmospheric proxies were identified as the main determinants of FLS presence, in particular mean sea level pressure, near-surface wind speed and FLS occurrence on the previous day. Of the land surface characteristics considered, evapotranspiration was shown to play a crucial role in the prediction of FLS occurrence, especially in high pressure conditions and in spring and fall (Fig. 2.7). Local moisture input was found to be especially important for locations further inland, highlighting the role of horizontal moisture advection for FLS occurrence and the interactions of atmospheric and land surface drivers.

The findings of this study support results of previous more localized process studies while providing insights on the spatial variation of drivers of FLS occurrence. The effect of variations in LST on FLS occurrence shows that low temperatures positively influence FLS occurrence as has been hypothesized previously (Underwood et al., 2004; Ye, 2009; Cuxart and Jiménez, 2012; Pérez-Díaz et al., 2017; Boutle et al., 2018; Gray et al., 2019; Price, 2019; Mühlig et al., 2020). The positive effect of low wind speeds, high pressure and specific weather patterns providing such conditions on FLS occurrence and persistence (Ye, 2009; Price, 2011; Cuxart and Jiménez, 2012; Scherrer and Appenzeller, 2014; Pérez-Díaz et al., 2017; Roskopf and Scherrer, 2017; Egli et al., 2019; Knerr et al., 2021) has also been detected in this study. To further analyze the role of large-scale atmospheric conditions and the related pressure fields, the daily average FLS occurrence is assigned to general weather situations after Hess and Brezowsky (1977). FLS occurrence maps by general weather situation reveal regions of high radiation fog occurrence such as river valleys (Po, Danube) and large basins (Pannonian basin) (Fig. 6.1). Applying this filtering approach in future studies could make it possible to create FLS type specific models and to subsequently analyze FLS type specific drivers.

The influence of land surface characteristics (namely albedo and fraction of vegetation cover) was found to be low. This is potentially due to the low day-to-day variability of these drivers compared to atmospheric drivers or LST and ET. Extending this concept further suggests that the day-to-day variability of FLS occurrence is mostly determined by the day-to-day variability of the atmospheric setting. Hence, the influence of the land surface on FLS occurrence might be detectable when climatological means of FLS occurrence are analyzed (cf. Teuling

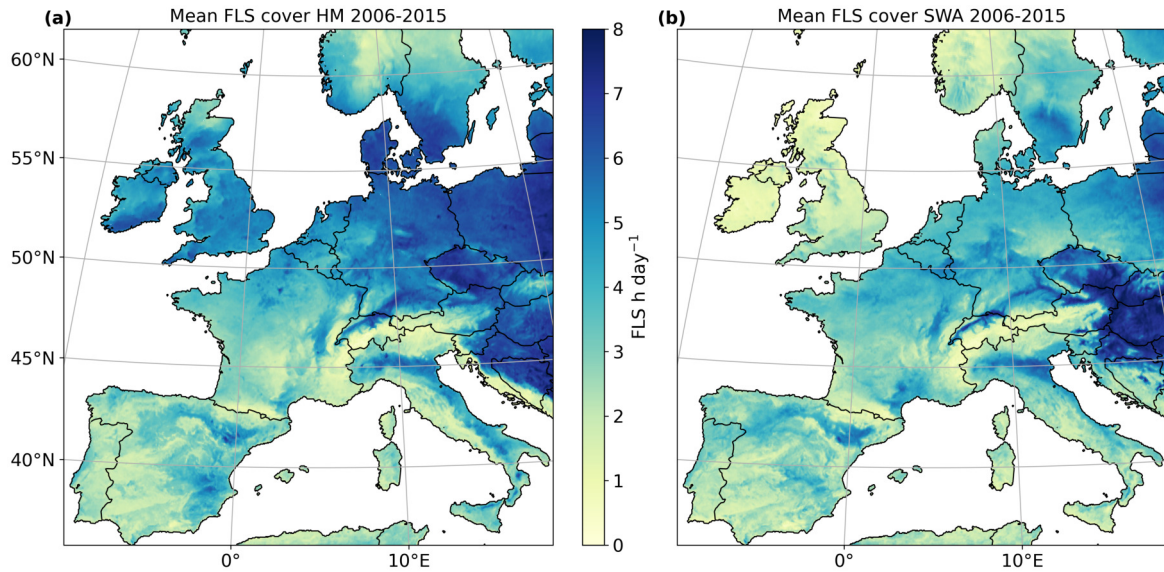


Figure 6.1: Mean FLS cover in hours per day during the general weather situation “High over Central Europe” (HM) (a) and “Anticyclonic southwesterly” (SWA) (b).

et al. (2017) and chapter 3). Furthermore, the low spatial resolution of the FLS data set used (3 km pixel size at nadir) might superimpose land cover specific influences on the FLS occurrence. Using high resolution satellite FLS products (Fuchs et al., 2022) could enhance the possibilities of the analysis of land surface effects on fog and low stratus clouds.

At the temporal and spatial scales of the study conducted in chapter 2, meteorological drivers are the most influential when predicting FLS occurrence. The influence of the land surface was found to be secondary, but is potentially dependent on the temporal averages considered. Furthermore, this study underlines the applicability of tree-based machine learning algorithms to atmospheric problems and their high level of interpretability.

(2) Land surface driven variations in fog and low stratus occurrence

In chapter 2, meteorology was identified as the main driver in determining daily means of FLS occurrence, whereas the influence of the land surface was found to be marginal. In this chapter, multi-year means of nighttime FLS occurrence, using two independent satellite products, were analyzed over a western European forest and the surrounding agricultural land to detect land cover influences on FLS occurrence. Over the forest area considered in the analysis, significantly higher nighttime FLS

occurrence compared to the surrounding agricultural land was detected. Potential reasons were found to be lower wind speeds, a temperature inversion and a higher availability of BVOCs serving as CCN over the forest.

The approach used in this chapter is similar to a space-for-time approach, which is often used to compare cloud or temperature patterns over neighboring forest and non-forest pixels to derive potential effects of land cover change on the variable of interest (Li et al., 2015; Teuling et al., 2017; Xu et al., 2022). Thus, the results of chapter 3 would suggest, that deforestation could decrease nighttime FLS occurrence, similar to what has been detected for convective cloud cover in boreal regions (Xu et al., 2022). However, projections of the findings of the study on the effects of the change in land cover on FLS occurrence are difficult, as potential drivers most likely vary depending on forest location and climatic zone.

An important driver for FLS occurrence identified in chapter 2 is the land surface temperature. Further analysis of LST over the study region shows lower nighttime values over the Landes forest compared to the surrounding agricultural land (Fig. 6.2), especially in summer and fall. Lower nighttime temperatures in forests compared to nearby non-forest areas are usually observed in the tropics, where strong cooling over the forest due to evapotranspiration decreases nighttime temperatures (Li et al., 2015; Schultz et al., 2017). Over parts of the Landes forest, transpiration rates have been found to be relatively low (Moreaux et al., 2013), thus, other drivers such as the surface turbulent heat fluxes and the ground heat flux might lead to lower nighttime temperatures over the forest area (Chen and Dirmeyer, 2020).

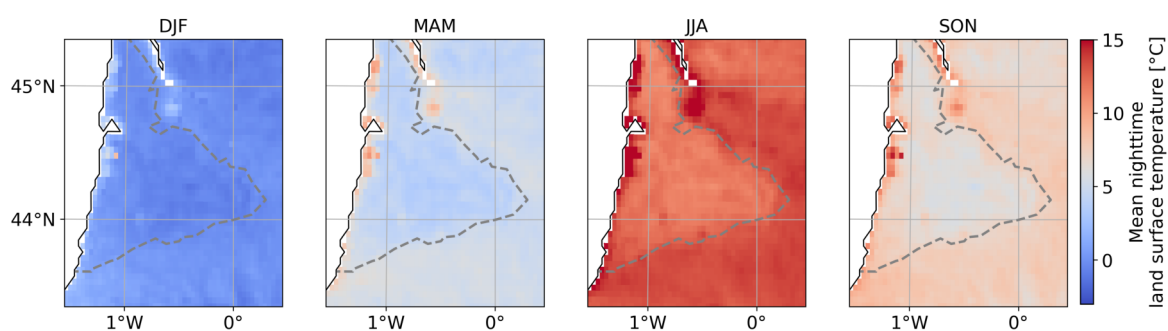


Figure 6.2: Mean seasonal nighttime (0–6 UTC) LST from 2006–2015 over the Landes area. The forest boundary is depicted by a dashed gray line.

The positive influence of low temperatures on FLS formation has also been found in studies looking at FLS formation in tropical lowland cloud forests (TLCFs) (Gradstein et al., 2011). In addition to the effect of temperature, the relief has been shown to influence nocturnal FLS patterns over the TLCFs of South America as well (Pohl et al., 2021). Extending the analysis of chapter 3 to other forest areas and analyzing the factors leading to enhanced FLS cover could also support conservation efforts of TLCF environments.

Similar to what has been shown for daytime convective cloud cover in Xu et al. (2022), the magnitude of the surface heat fluxes and the partitioning of the turbulent heat fluxes into their latent and sensible part also plays a significant role in the fog life cycle: Whereas the latent surface heat flux positively influences the liquid water path (LWP) budget in fog, the sensible surface heat flux is the most significant negative term and thus strongly contributes to fog dissipation (Wærsted et al., 2019). Wærsted et al. (2019) further suggest that the Bowen ratio plays a major role in determining the influence of the surface latent and sensible heat flux on the fog LWP. The Bowen ratio is influenced by the availability of liquid water on the surface and thus the type of underlying vegetation, as the latter strongly impacts water distribution by wind, throughfall and soil infiltration (Wærsted et al., 2019). Thus, comparing magnitude and direction of the turbulent heat fluxes over forests and other nearby land surface types would further contribute to the understanding of the role of land cover on FLS processes.

The findings of the study present observational evidence of land cover influence on FLS on a spatial and temporal scale previously not shown and support the hypothesis presented in chapter 1.5. The influence of the land surface is found to occur via biophysical interactions (i.e. exchanges of energy) and biogeochemical interactions (i.e. emissions of aerosols). The results further emphasize the need for the quantification of the effect of the responsible drivers, such as the analysis of BVOC and land surface temperature effects on FLS occurrence over the forest area, which both likely belong to the main determinants of the observed land cover specific FLS patterns. Moreover, the relationship of the size of the area with uniform land cover and the spatial scale of the data used has to be considered in future studies: When relatively large pixel sizes are used (3–5 km), patterns might only be detectable when the area with uniform land cover is relatively large (in the case of the Landes forest: 12,000 km²). Future analyses have to test these assumptions by first applying the methods used here to other European forest areas, secondly detecting the drivers of the observed patterns and finally using data with different spatial and temporal characteristics.

(3) Climatological patterns of fog and low stratus formation and dissipation times

In chapter 2 and 3 the occurrence of fog and low stratus and its meteorological and land surface drivers were investigated. To further analyze FLS in the context of land-atmosphere interactions, knowledge of the life cycle of FLS, in particular its timing of formation and dissipation, is necessary. Therefore, a novel FLS formation and dissipation time data set was developed in chapter 4 by applying logistic regression to a binary satellite-based FLS data set. An analysis of the temporal and spatial patterns of FLS formation and dissipation time revealed a clear dependency of these on topography. In river valleys, a clear diurnal cycle was found, with FLS formation in the evening and dissipation in the morning. Seasonal patterns of FLS formation and dissipation further reveal the importance of the distance to the coast and the solar cycle on FLS formation and dissipation time.

The findings of the study presented in chapter 4 agree well with LES and modeling studies over Europe (Roach, 1995; Haeffelin et al., 2010; Bergot, 2016; Steeneveld and de Bode, 2018) which show fog formation in the evening and dissipation in the morning, with a strong dependency of dissipation on the solar cycle. The large advantage of this study is its large spatial and temporal extent: Unlike local and mostly temporal constrained measurement and modeling studies, the novel data set created in chapter 4 of this thesis gives information on FLS formation and dissipation over the complete land mass of central Europe. This provides a data basis for large-scale studies on drivers of the FLS life cycle.

On the basis of the geographic patterns of formation and dissipation time, three potential drivers of these patterns have been identified: Topography, distance to the coast and the solar cycle. Pertaining to processes related to topography, non-local drainage flows as well as cold air flow and pooling have been previously identified to influence the FLS life cycle (Müller et al., 2010; Roskopf and Scherrer, 2017; Ducongé et al., 2020). Besides altitude, the slope, aspect and the position of the site in relation to local and regional morphology are important (Błaś et al., 2002; Hůnová et al., 2021b; Pohl et al., 2021). Topography further modulates meteorological drivers such as wind and temperature (Cuxart et al., 2021), and these influences might scale with the size of the topographic features. Furthermore, the influence of topography on FLS occurrence most likely depends on the season (Hůnová et al., 2021b) which is also visible in figure 4.9.

The identified patterns related to topography might also be a feature of local to regional wind patterns, leading to advected FLS layers at upwind locations and

less FLS at downwind locations. Such topography induced FLS patterns have been identified over Corsica, where FLS formation times shift depending on the site position in relation to the complex topography (Knerr et al., 2021). Currently, it is not possible to determine if the FLS life cycle patterns identified in chapter 4 relate to a specific FLS type. While the algorithm excludes fast changes in between FLS and no FLS situations (see 4.2.2) a distinction between advection and radiation fog is currently not possible with this data set.

The processes leading to the observed FLS formation and dissipation patterns at the Mediterranean coast are potentially similar to processes which influence the FLS life cycle at other coastal locations over the globe, for example in coastal California. FLS at the Californian coast has been shown to be influenced by the pressure field distribution over the ocean and land, ocean currents, coastal upwelling, as well as topography near the coastline (Leipper, 1994; Iacobellis and Cayan, 2013; Rastogi et al., 2016; Samelson et al., 2021). Coastal upwelling has also been found at the Cantabrian coast of northern Spain (Alvarez et al., 2010) and thus potentially plays a role in FLS formation processes at this location (compare Fig. 4.9). It has further been hypothesized that FLS layers form over the Atlantic in humid and relatively cool conditions and are then advected over the Bay of Biscay onto the Cantabrian mountains (Egli et al., 2017; Royé et al., 2018). The relief of the surrounding areas might influence coastal FLS formation as well: In arid regions, a basin which can heat up during the day can increase the sea breeze, enhancing the advection of FLS towards topographic features close to the coast (Schemenauer and Cereceda, 1994). Further processes at the coast which could influence the FLS life cycle are mixing of air masses with different temperatures (Bardoel et al., 2021) and the advection of radiation fog formed inland (Bari et al., 2015). Still, it has to be taken into account that the influence of sea or freshwater on the FLS life cycle could be superimposed by other mechanisms related to topography or seasonality (Hůnová et al., 2022). Therefore, the exact processes of FLS formation and dissipation in the coastal region of the Mediterranean are most likely dependent on a combination of different drivers, topography and climate and thus are subject of future investigations.

The third major driver identified to influence the climatological patterns of FLS formation and dissipation time is the solar cycle, which has been also frequently addressed in measurement and modeling studies (Roach, 1995; Stolaki et al., 2009; Haeffelin et al., 2010; Price, 2011; Stolaki et al., 2015; Bergot, 2016; Wærsted et al., 2019). FLS frequently forms around sunset or during the night when the air reaches its dewpoint due to radiative cooling (Roach, 1995; Stolaki et al., 2009; Price, 2011; Stolaki et al., 2015). After sunrise, FLS dissipates through the absorption of solar

radiation by the surface along with turbulent mixing caused by the increase in temperature. Wærsted et al. (2019) list the increase in the surface sensible heat flux after sunrise as the most significant negative term contributing to the loss of LWP, whereas the direct absorption of solar radiation is only secondary for LWP loss.

As suggested in the hypothesis in chapter 1.5, topography, the distance to the coast and the solar cycle have been identified as drivers of the climatological patterns of FLS formation and dissipation over central Europe. The effect of meteorological or land surface based drivers is potentially superimposed by the background climate and modified by the underlying topography. The land surface is presumably involved in a myriad of dissipation processes, such as turbulence and moisture (Wærsted et al., 2019). It is further possible that a specific land surface type has characteristics leading to counteracting influences on FLS formation and dissipation processes. Extracting the individual and combined effects of meteorological and land surface drivers on FLS formation and dissipation time is thus critical going forward.

The novel FLS formation and dissipation time data set provides a data basis to apply both analyses of FLS occurrence conducted in chapter 2 and 3 to the FLS life cycle. The spatial and temporal extent of the novel FLS formation and dissipation time data set is unprecedented, making it possible to analyze the FLS life cycle without ground observations over the complete European land mass. Furthermore, the logistic regression algorithm can be applied to any other cloud masks with high temporal resolution such as to novel high spatially resolved FLS masks (Fuchs et al., 2022) or to data sets over other FLS prone regions such as the Namib desert (Andersen and Cermak, 2018).

(4) Identifying and understanding fog and low stratus formation and dissipation regimes

The knowledge of drivers influencing FLS occurrence gained in chapter 2 and the novel FLS formation and dissipation time data set created in chapter 4 provide the foundation for the investigation of drivers influencing the FLS life cycle. Hence, in chapter 5, FLS formation and dissipation time regimes were identified and investigated with respect to their sensitivities to changes in environmental conditions. This was done using pixel-based monthly correlations of FLS formation and dissipation time with meteorological and land surface parameters in a hierarchical clustering approach over central Europe. FLS life cycle sensitivities to variations in meteorological and land surface conditions for the identified regimes

were found to differ seasonally, geographically and depending on background climate. Moreover, similar meteorological conditions were found to lead to different sensitivities in different regimes. This is potentially a pattern of the prevailing FLS types, but further analysis is necessary to quantify the effect of FLS type on the sensitivities accordingly. Nevertheless, the clustering analysis provides a basis for the analysis and discussion of FLS formation and dissipation sensitivities to meteorological and land surface conditions.

In general, the study builds on findings of chapter 2 and confirms the effect of lower wind speed, higher surface pressure and higher temperatures in delaying FLS dissipation in some identified FLS regimes. A deviation from the sensitivity patterns identified in chapter 2 is most likely due to differences in background climate and FLS type, confirming the hypothesis in chapter 1.5. The variation of sensitivities across regimes and seasons is more strongly pronounced for FLS formation, with a high variation of sensitivities of FLS formation time to changes in wind speed. Besides the influence of wind direction discussed in chapter 5, the variation in the influence of wind speed on the FLS life cycle might further depend on the depth of the fog layer and height of the increased wind intensity (Bergot, 2016). Low wind speeds have been shown to promote FLS formation (Cuxart and Jiménez, 2012; Bergot, 2016; Pérez-Díaz et al., 2017; Price, 2019), but low wind speeds near the fog top could lead to a decrease in LWP of the fog layer and a decrease in fog duration (Bergot, 2016).

The FLS formation and dissipation regimes identified provide ideal regional and sub-regional study areas for the application of site specific machine learning models. In addition, the agglomerative clustering approach used here can also be applied to other FLS studies, such as the extraction of temporal FLS type specific patterns and their link to the atmospheric background condition (similar to Egli et al., 2019). Furthermore, the partial dependencies calculated in chapter 2 could be used as input data for the clustering algorithm in future studies, to identify FLS regimes pertaining to model sensitivities. Therefore, the combination of the data sets, methods and learnings from the research chapters of this thesis, provide a myriad of possibilities for future analysis of fog and low stratus clouds.

6.2 Outlook

The results of this thesis show that meteorological and land surface processes, along with their interactions, influence FLS occurrence and its life cycle. This influence varies across regions and seasons and is further dependent on the temporal and spatial scales considered. The research presented here enhances process and system understanding pertaining to climatological occurrences of FLS in addition to its formation and dissipation processes. However, it has not been possible to quantify the effect of a change in land use and land cover on FLS in this thesis, though pathways identified for convective cumulus clouds in the literature (Xu et al., 2022) might be important for FLS as well. Furthermore, the role of scale should be kept in mind: FLS processes and land-atmosphere interactions occur across a large range of scales, from aerosol processes and leaf-atmosphere exchanges to synoptic processes and ecosystem exchanges (Sun et al., 2015; Koracin and Dorman, 2017). Thus, the results of this thesis inherently scale with the variation of the temporal and spatial scales of the analyses, leading to a shift in the importance of drivers from one scale to another.

Further important elements of the FLS life cycle, which have not been considered in this thesis, are variations in the type and amount of aerosols. It has been found that aerosols promote FLS via increasing LWC, increasing droplet concentration and decreasing droplet effective radius (Yan et al., 2020). Increasing the CCN concentration results in a higher number of small droplets in FLS and a decrease in visibility (Poku et al., 2019; Yan et al., 2021). This effect is called the Twomey effect and describes that aerosols can act as CCN and increase the number of small cloud droplets, resulting in an increase in cloud optical thickness (Twomey, 1977). The increase in CCN can further lead to stronger radiative cooling at the cloud top, which speeds up the vertical growth and delays dissipation of the resulting higher and denser FLS layer (Maalick et al., 2016; Jia et al., 2019b).

To further understand the effect of aerosols and their interactions with other drivers on FLS and its life cycle, a promising path forward could be to use the formation and dissipation time data set created in chapter 4 in a machine learning set-up. Such an analysis would provide insights if the FLS life cycle is influenced by aerosols, in which weather condition this influence is strongest and how this influence compares to other meteorological or land surface based drivers.

In a preliminary study, an extreme gradient boosting model was set up which predicts the duration of an FLS event using a set of meteorological drivers from ERA5 (Hersbach et al., 2018; Muñoz Sabater, 2019) and daily means of aerosol optical

depth (AOD) from the Satellite Application Facility on Climate Monitoring (CM SAF) (Clerbaux et al., 2017) on the day before FLS formation. Extreme gradient boosting (XGB) is similar to the GBRT model used in chapter 2 with the advantage of shorter run times and built-in regularization techniques (Chen and Guestrin, 2016). The model is run for high pressure conditions (>101200 Pa) in winter and fall of 2006–2012 for an area in the Po valley region, which was a preliminary clustering result of the procedure described in chapter 5. The spatial extent is similar to the area shown in figure 5.4, thus the climatological patterns described there apply here as well.

The XGB model is able to skillfully predict FLS duration with an R^2 of 0.86 using the training set and an R^2 of 0.73 when using the test set. Model sensitivities are calculated using SHapley Additive exPlanations (SHAP) values, which quantify the contribution of each feature to each individual prediction, enabling a global understanding of the model (Lundberg and Lee, 2017; Lundberg et al., 2020). In figure 6.3 the final feature set and the SHAP value for each feature depending on the feature value is shown, with features sorted by their mean absolute SHAP value in descending order.

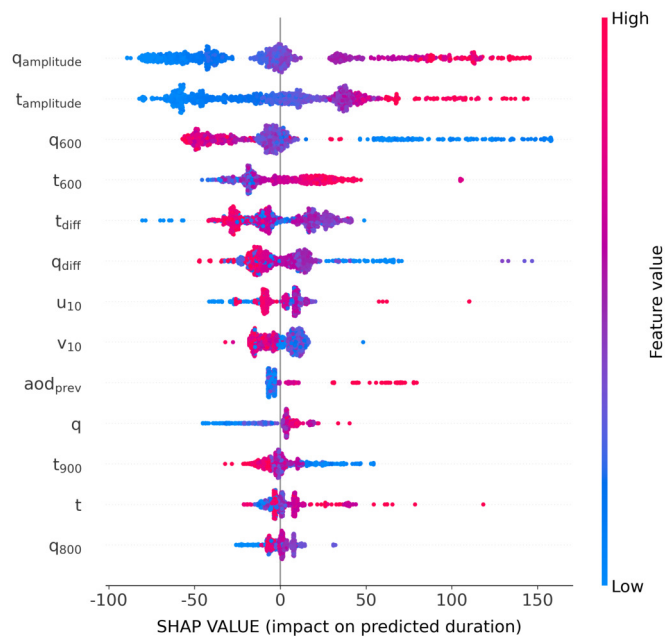


Figure 6.3: SHAP “beeswarm” plot, showing the impact of each feature on the predicted FLS duration. Each dot represents an observation and is colored dependent on its feature value. The position of the dot along the x-axis depicts its influence on the model prediction. The features are sorted by their mean absolute SHAP value in descending order. Multiple observations for one SHAP value are stacked horizontally. t is the temperature, q the specific humidity, u and v are the horizontal wind vectors and aod_{prev} is the aerosol load on the day before formation. The subscripts are described in the text.

The features of the final feature set are used either at 1 hour after formation at a specific height (10 m for u and v or on pressure levels for t and q), describe the displacement of a variable (e.g. $q_{\text{amplitude}}$) during the FLS event or describe the difference of feature value at dissipation time in relation to its value at formation time (e.g. t_{diff}). The most important features for the FLS duration prediction are $q_{\text{amplitude}}$ and $t_{\text{amplitude}}$. For both features the impact on predicted duration is higher for higher feature values. For $t_{\text{amplitude}}$ this can be explained as the FLS layer shields the ground from radiative heat loss (when FLS occurs during the night), leading to a positive offset in the detrended temperature curve. Furthermore, a dry layer above the FLS layer, i.e. low q values at 600 hPa (q_{600}), seems to be beneficial for FLS duration, as indicated by the positive SHAP values for q_{600} . This dry layer potentially enhances radiative cooling at the FLS top, resulting in a more persistent FLS layer (Wærsted et al., 2017; Andersen et al., 2020).

While the mean absolute SHAP value of aod_{prev} is comparably low, some high aod_{prev} situations have a high impact on the predicted FLS duration (Fig. 6.3). The impact on FLS duration solely attributable to aod_{prev} (main effect), is shown in figure 6.4a. High pollution events on the previous day seem to prolong FLS duration up to 60 minutes. This is in-line with findings from Yan et al. (2021) who found that aerosol-cloud interactions in fog enhance its duration by about 1 hour. The impact of the interaction of aod_{prev} with $t_{\text{amplitude}}$ increases the predicted FLS duration by an additional 30 minutes (Fig. 6.4b) for high $t_{\text{amplitude}}$ values. This implies an increasing impact of high aod_{prev} values in FLS situations which are potentially dense (high $t_{\text{amplitude}}$) and of large vertical extent.

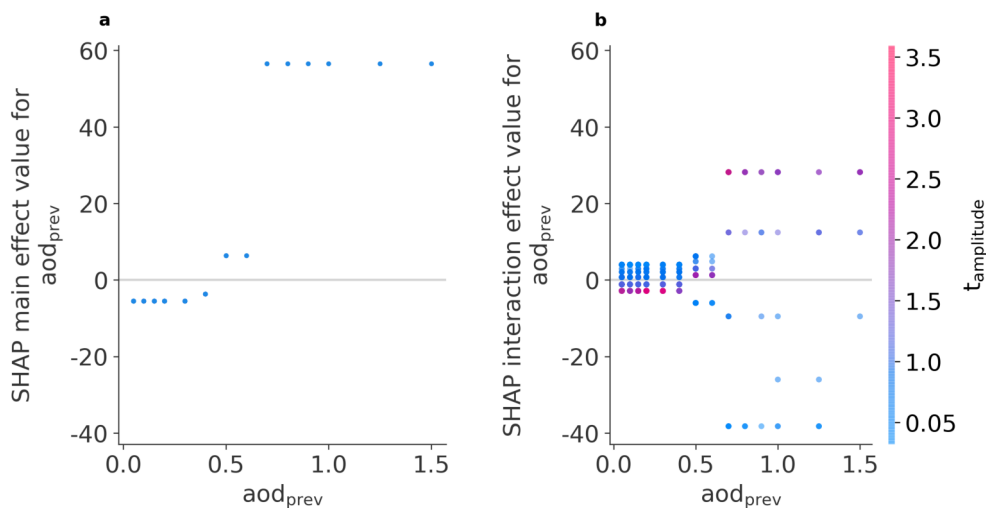


Figure 6.4: SHAP dependence plots showing a) the main effect of aod_{prev} on the prediction of FLS duration and b) the interaction effects with $t_{\text{amplitude}}$.

The SHAP values investigated here provide insights into the mechanisms that govern the FLS duration in the Po valley with the option to further analyze specific observations. The sensitivities shown improve our understanding of the important meteorological processes and particularly the role of aerosols. The model framework applied to the Po valley can be applied to other areas as well, providing a tool to analyze and compare the role of meteorology and aerosols on FLS duration over varying geographic backgrounds.

This thesis showed that atmospheric drivers determine the day-to-day variability in FLS occurrence, whereas climatological FLS patterns on larger, multiyear time scales are measurably influenced by the land surface. Future research efforts can use novel high-resolution satellite sensors such as the soon-launching Meteosat Third Generation (MTG) satellites and novel FLS data sets. These provide the data basis to further extract the influence of land cover and land cover change on FLS occurrence and life cycle on higher spatial and temporal scales. Using a combination of satellite observations, surface measurements and machine learning over varying temporal and spatial scales will provide a basis for future analyses relating to fog and low stratus processes in the climate system. Such an analysis can be extended to investigate FLS processes in a changing climate, as the associated changes in the distribution of energy and moisture in the atmosphere will potentially impact FLS life cycle and occurrence. In addition, knowledge of the FLS life cycle is further highly valuable for the prediction of solar power production, which becomes more and more important as the need for renewable energies increases.

A1 Scientific Appendix of Chapter 2

Mean squared error (MSE) of all models

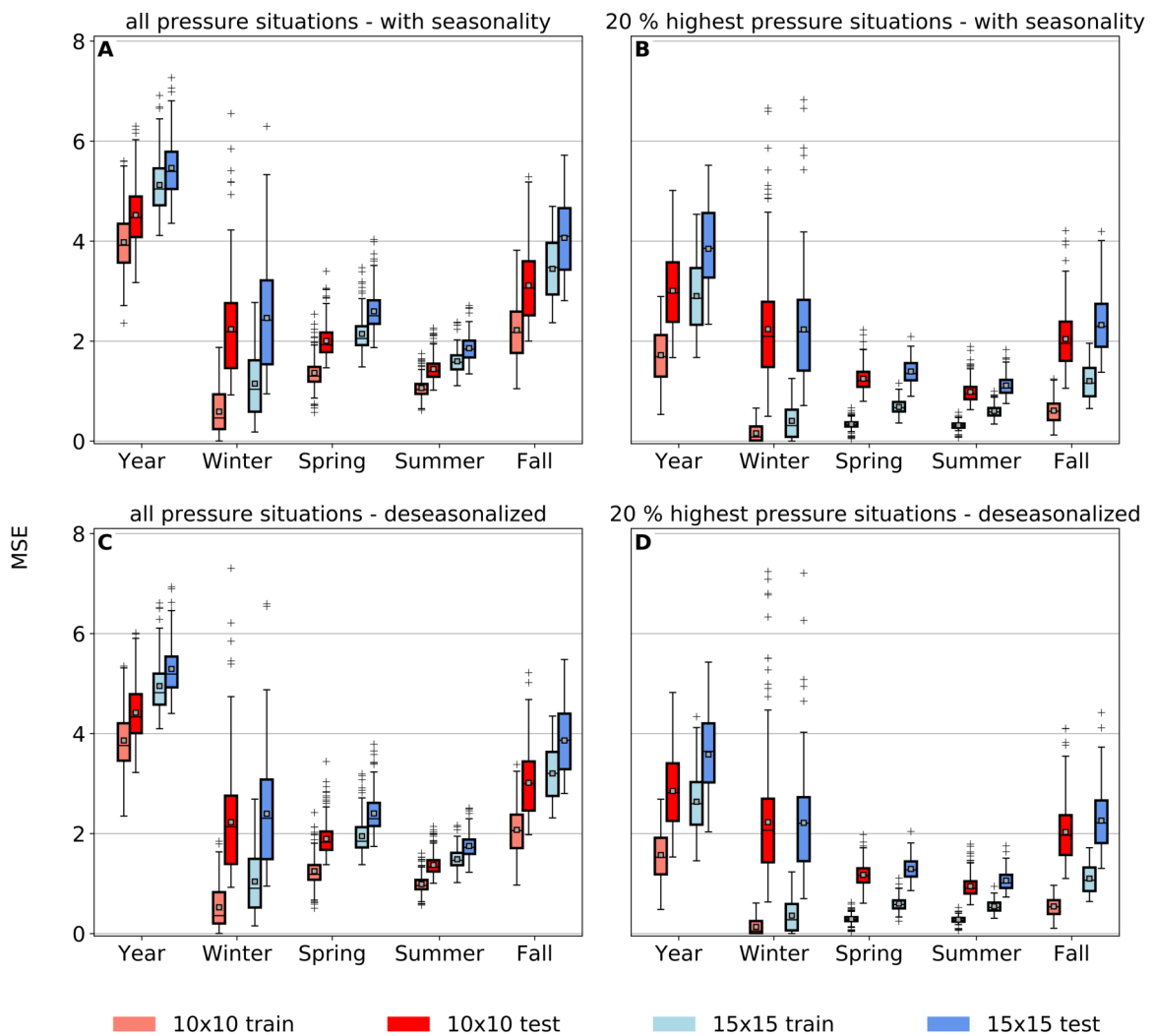


Figure A1.1: MSE of training and test set over all grid sizes and seasons using either all pressure situations (left) or only high-pressure situations (right). The top row shows the results using the data with seasonality, while the bottom row displays the model results using deseasonalized data.

A2 Scientific Appendix of Chapter 3

Introduction

This scientific appendix of chapter 3 provides the confusion matrix of the cross-validation of the two FLS products (Tab. A2.1), P values on all conducted t-tests in the analysis (Tab. A2.2 & A2.3), a topographic map of the area (Fig. A2.1) as well as figures for monthly mean FLS cover and wind speed (Fig. A2.2 & A2.3) and monthly mean profiles of mean hourly inversion strength (Fig. A2.4, A2.5, A2.6, A2.7 & A2.8). The figures A2.2-A2.8 show the plots for those months which have not been shown in the paper.

Cross-validation of the satellite products

Table A2.1: Cross-validation of the satellite products. Values indicate the confusion matrix for the two data sets. The number of true positives (50) and true negatives (83) is written in **bold**.

	FLS (SEVIRI)	no FLS (SEVIRI)
FLS (CALIPSO)	50	34
no FLS (CALIPSO)	12	83

Two sample t-tests results

Table A2.2: *P* values of the two sample t-tests, testing the difference between FLS cover over forest vs. non-forest areas for the three different SEVIRI based FLS climatologies (a, b and c) in figure 3.2. For the t-test 500 random pixels from each group are used. *P* values smaller than 0.05 are in **bold**.

	P value
a	1.05e⁻¹³
b	3.16e⁻⁹
c	2.00e⁻¹⁸

Table A2.3: *P* values of the two sample t-tests, testing the difference between FLS cover and wind speed (*ws*) over forest vs. non-forest areas for all months. For the FLS cover t-test 500 random pixels from each group are used, for the wind speed t-test 150 random pixels, due to the larger pixel size of the wind speed data. *P* values smaller than 0.05 are in **bold**.

Month	P value FLS	P value ws
1	0.30	0.79
2	4.12e⁻⁵	0.02
3	0.07	0.0007
4	0.10	0.002
5	2.83e⁻¹⁸	0.5
6	0.01	0.002
7	1.45e⁻¹⁵	0.004
8	2.75e⁻²⁶	2.61e⁻¹²
9	3.15e⁻³¹	0.2
10	1.86e⁻⁸	0.07
11	1.97e⁻¹⁵	0.17
12	1.54e⁻²⁶	0.67

Digital elevation map of the study area

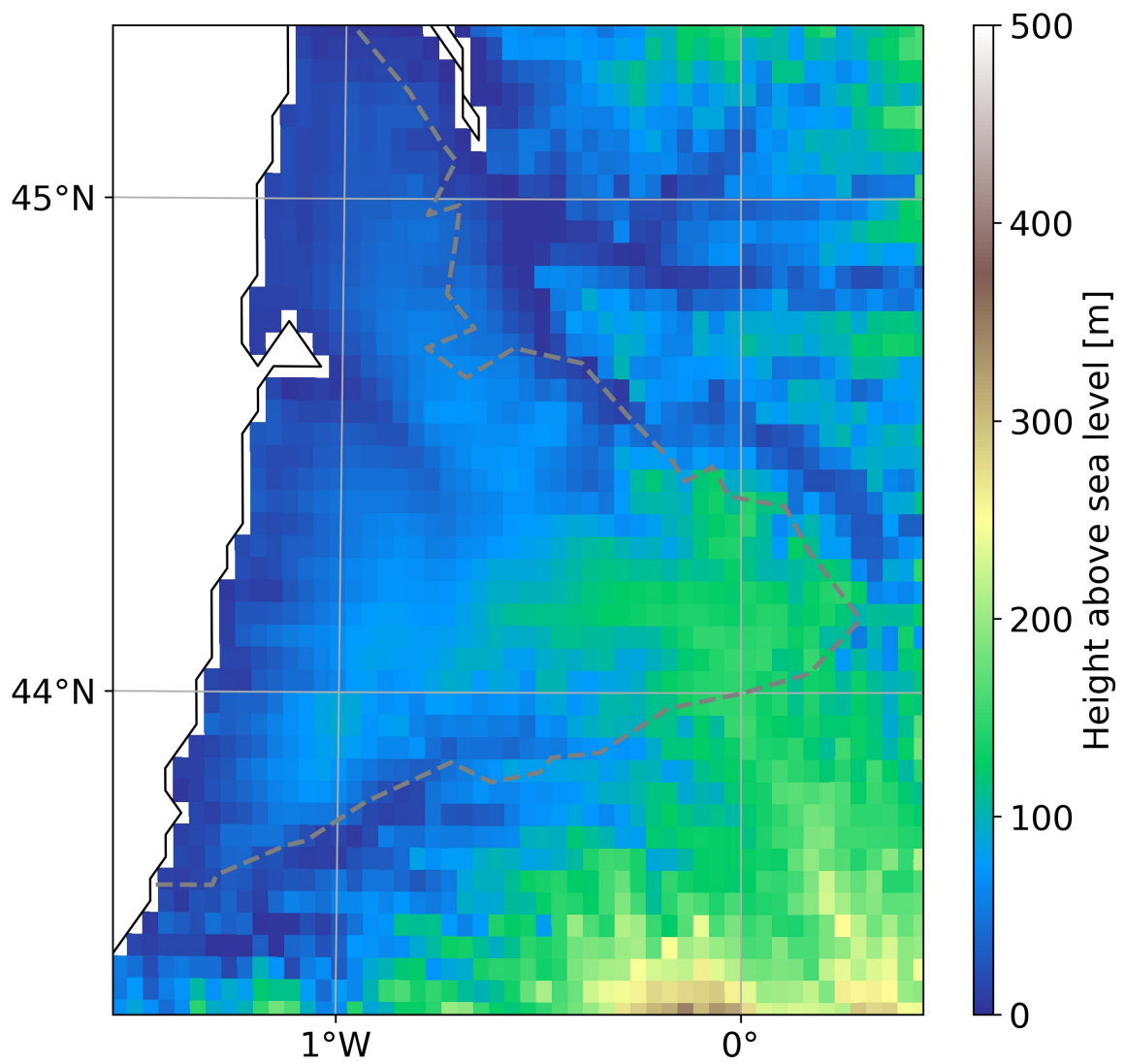


Figure A2.1: Digital elevation map of the study area. The gray dashed line approximately marks the forest border.

Monthly mean FLS cover and wind speed

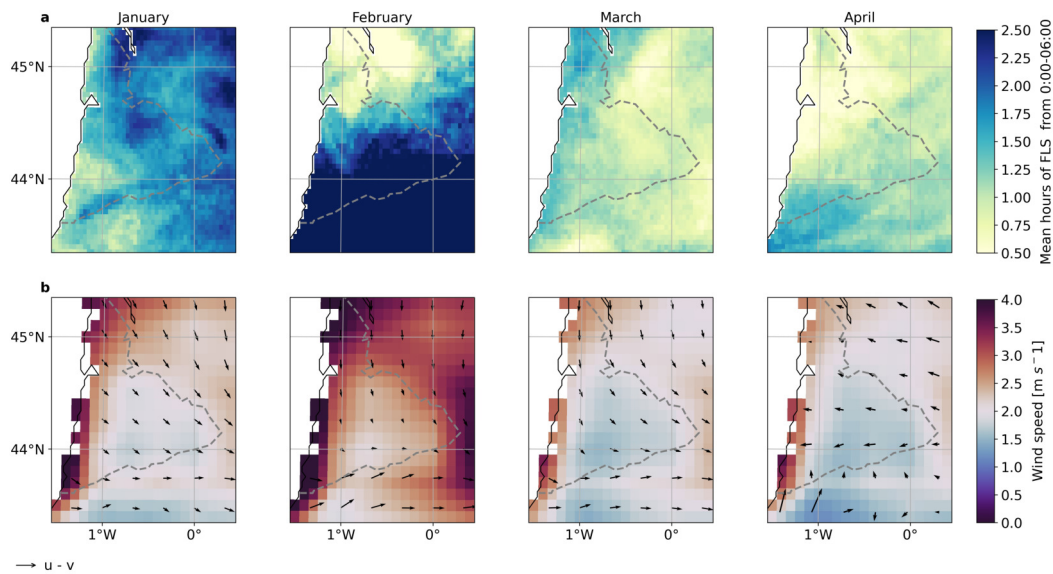


Figure A2.2: Climatological mean FLS hours by month based on true positive and true false observations (a), and the corresponding ERA5 land wind speed and wind direction (u and v wind components) on the respective days (mean from 0–6 UTC) (b) for January, February, March and April. The gray dashed line approximately marks the forest border.

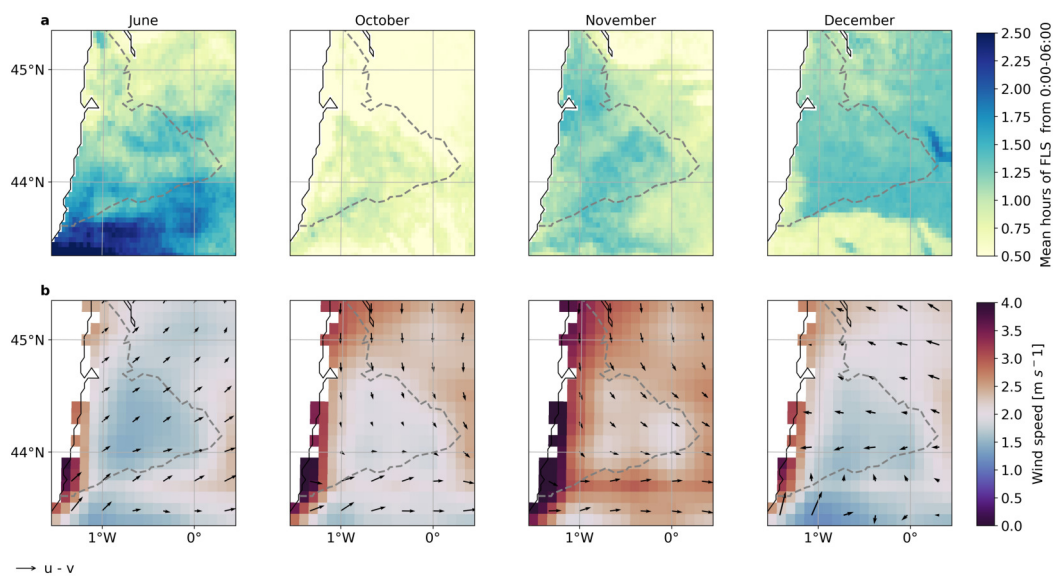


Figure A2.3: Climatological mean FLS hours by month based on true positive and true false observations (a), and the corresponding ERA5 land wind speed and wind direction (u and v wind components) on the respective days (mean from 0–6 UTC) (b) for June, October, November, December. The gray dashed line approximately marks the forest border.

Monthly mean of mean hourly inversion strength

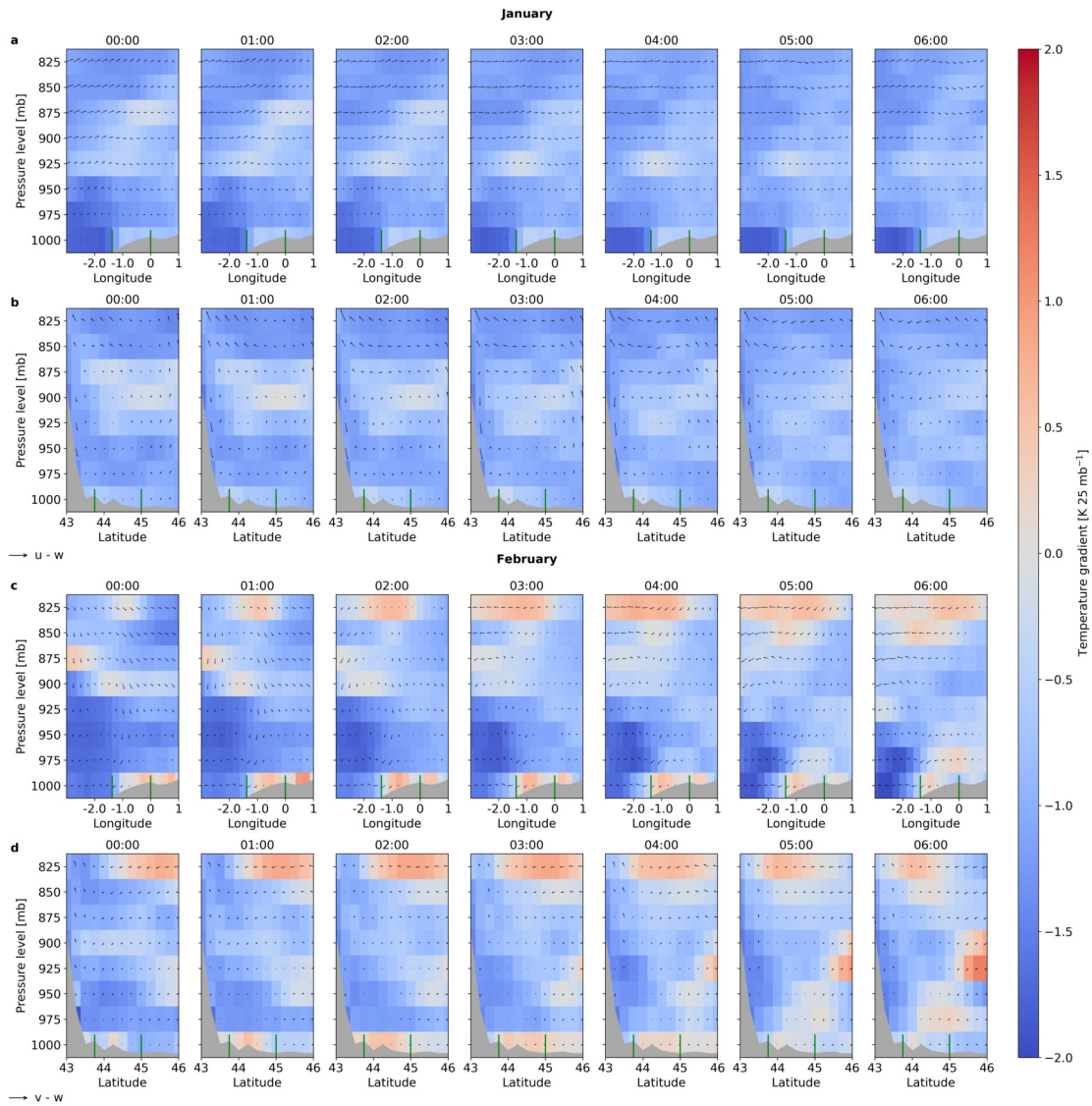


Figure A2.4: Profiles of mean hourly inversion strength in $K 25 \text{ mb}^{-1}$ for the true positive and true negative days in January (a, b) and February (c, d), from 825 to 1000 mb along 44.25° N (a, c) and -0.25° W (b, c). In the longitudinal profile (a, c) $u-w$ wind vectors are plotted, in the latitudinal profile (b, d) $v-w$ wind vectors are plotted. For visibility reasons, the w vector is enhanced by a factor of 20. The location of the Landes forest is marked in both plots as green vertical lines. The topography is plotted in gray.

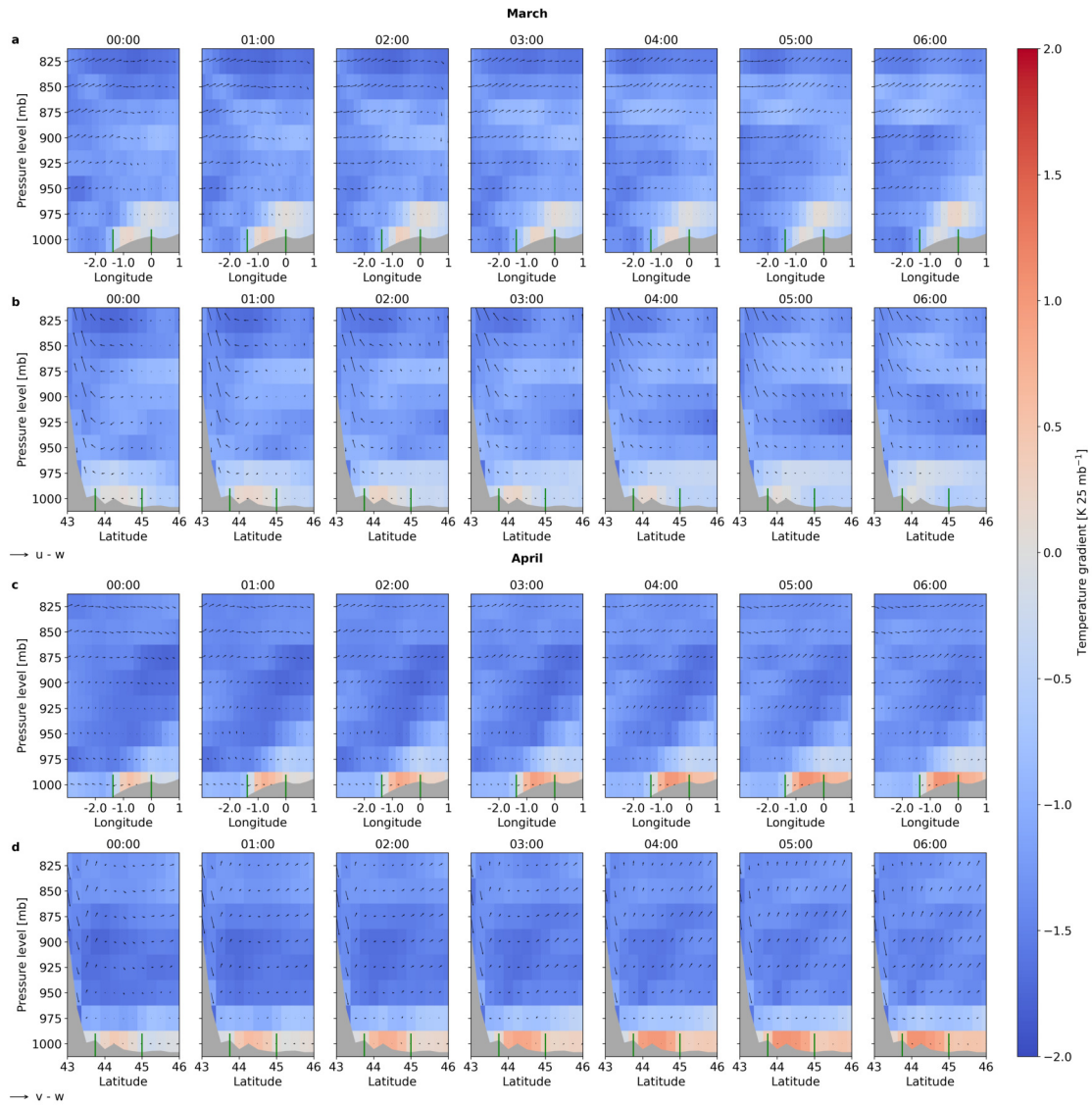


Figure A2.5: Profiles of mean hourly inversion strength in $K\ 25\ mb^{-1}$ for the true positive and true negative days in March (a, b) and April (c, d), from 825 to 1000 mb along $44.25^\circ N$ (a, c) and $-0.25^\circ W$ (b, c). In the longitudinal profile (a, c) $u-w$ wind vectors are plotted, in the latitudinal profile (b, d) $v-w$ wind vectors are plotted. For visibility reasons, the w vector is enhanced by a factor of 20. The location of the Landes forest is marked in both plots as green vertical lines. The topography is plotted in gray.

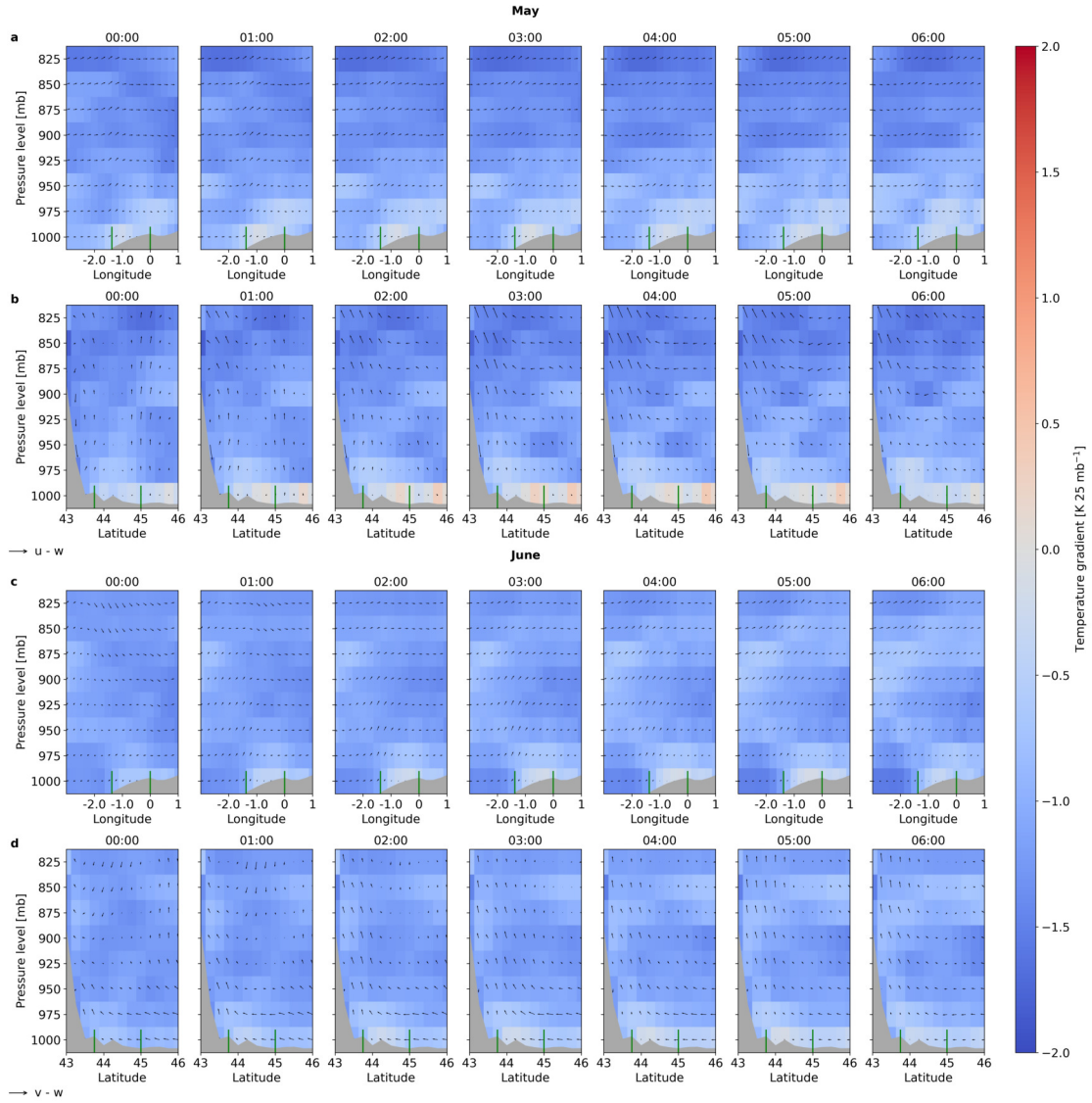


Figure A2.6: Profiles of mean hourly inversion strength in $\text{K } 25 \text{ mb}^{-1}$ for the true positive and true negative days in May (a, b) and June (c, d), from 825 to 1000 mb along 44.25° N (a, c) and -0.25° W (b, c). In the longitudinal profile (a, c) $u-w$ wind vectors are plotted, in the latitudinal profile (b, d) $v-w$ wind vectors are plotted. For visibility reasons, the w vector is enhanced by a factor of 20. The location of the Landes forest is marked in both plots as green vertical lines. The topography is plotted in gray.

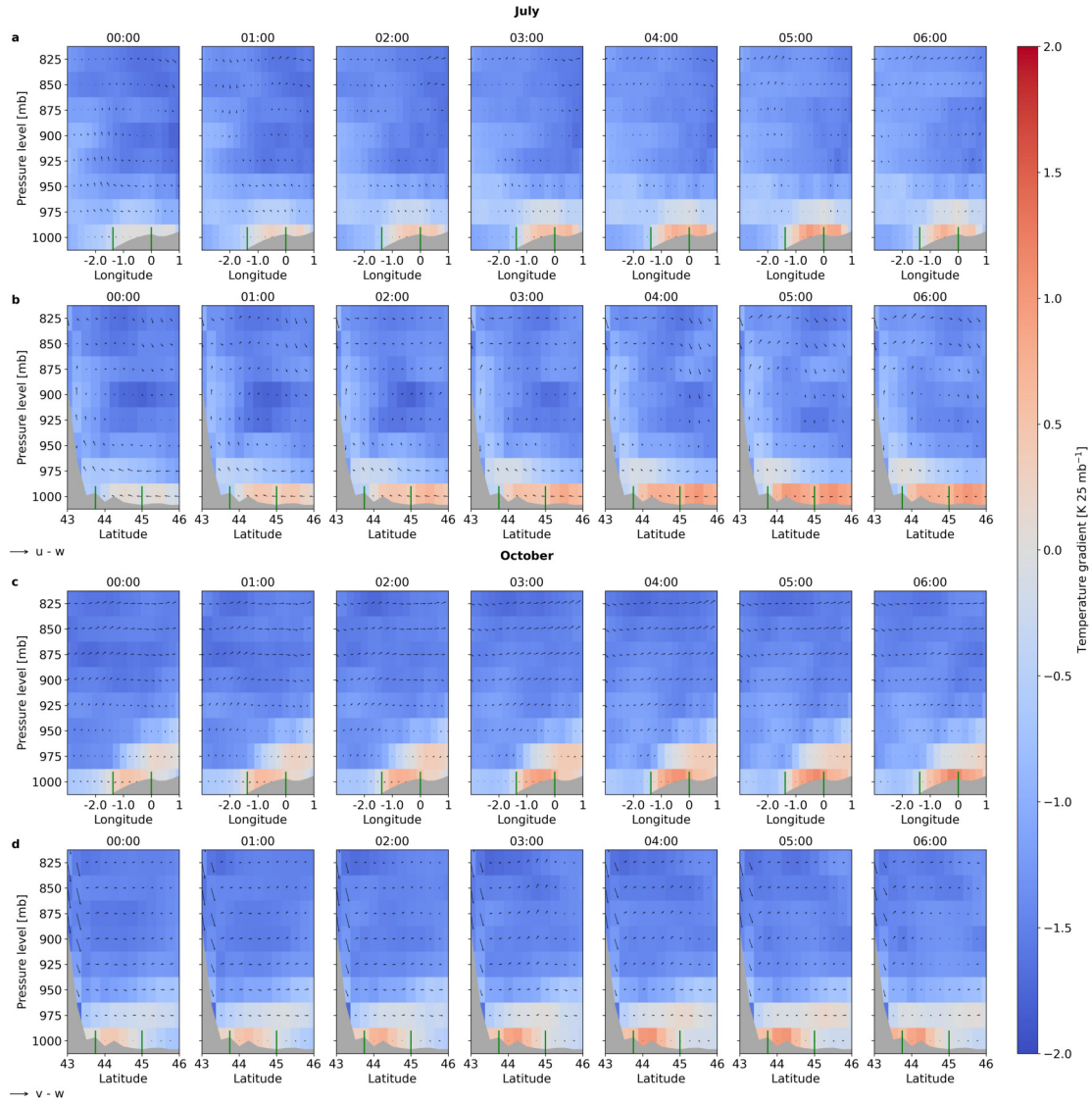


Figure A2.7: Profiles of mean hourly inversion strength in $K 25 \text{ mb}^{-1}$ for the true positive and true negative days in July (a, b) and October (c, d), from 825 to 1000 mb along 44.25° N (a, c) and -0.25° W (b, c). In the longitudinal profile (a, c) $u-w$ wind vectors are plotted, in the latitudinal profile (b, d) $v-w$ wind vectors are plotted. For visibility reasons, the w vector is enhanced by a factor of 20. The location of the Landes forest is marked in both plots as green vertical lines. The topography is plotted in gray.

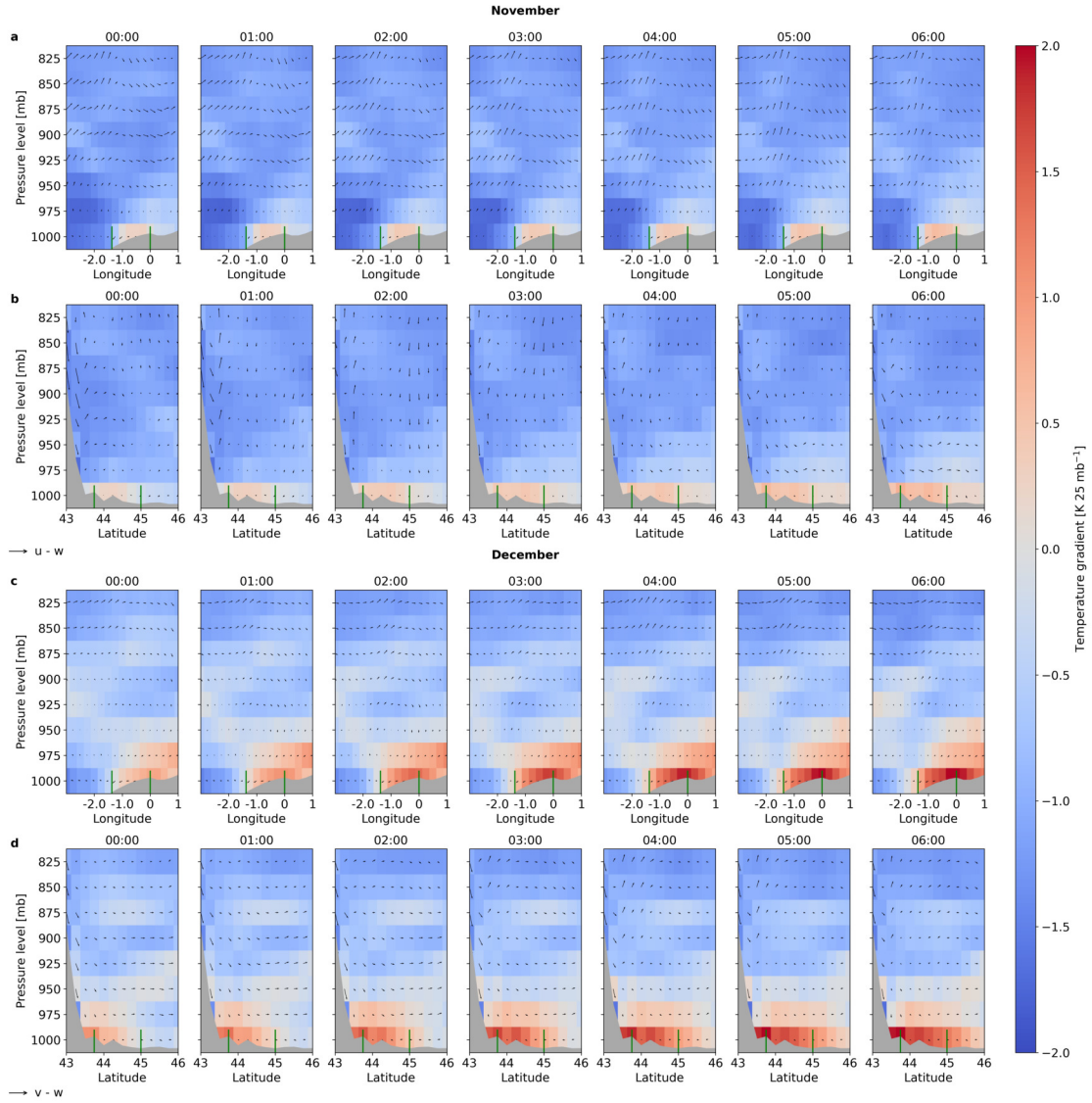


Figure A2.8: Profiles of mean hourly inversion strength in $K 25 \text{ mb}^{-1}$ for the true positive and true negative days in November (a, b) and December (c, d), from 825 to 1000 mb along 44.25° N (a, c) and -0.25° W (b, c). In the longitudinal profile (a, c) $u-w$ wind vectors are plotted, in the latitudinal profile (b, d) $v-w$ wind vectors are plotted. For visibility reasons, the w vector is enhanced by a factor of 20. The location of the Landes forest is marked in both plots as green vertical lines. The topography is plotted in gray.

A3 Scientific Appendix of Chapter 4

Diurnal course of FLS cover over urban and rural sites

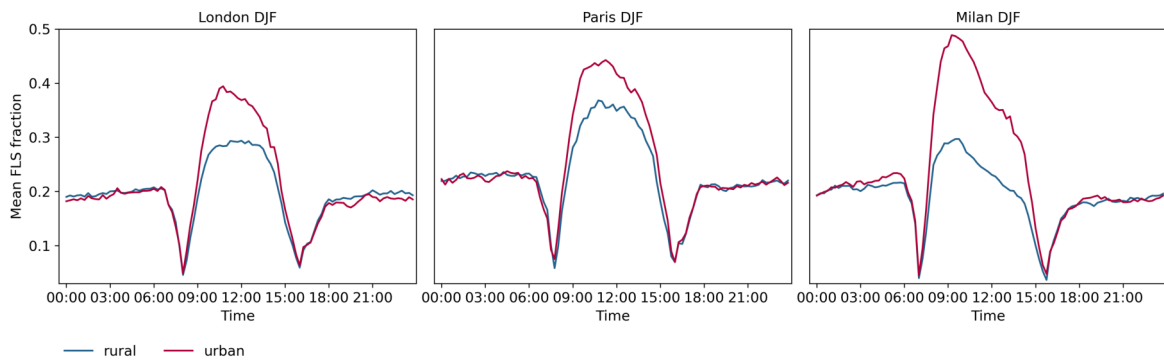


Figure A3.1: Mean diurnal course of FLS cover in winter over London, Paris and Milan and their respective rural surroundings from 2006–2015. For each city, the mean diurnal course is calculated over 10 different pixels for urban and rural pixels each. The sudden decrease in mean FLS fraction around 8:00 and 16:00 is due to the no-FLS values at twilight, serving as an reference for sunrise and sunset.

Normalized FLS cover, duration and number of formation events

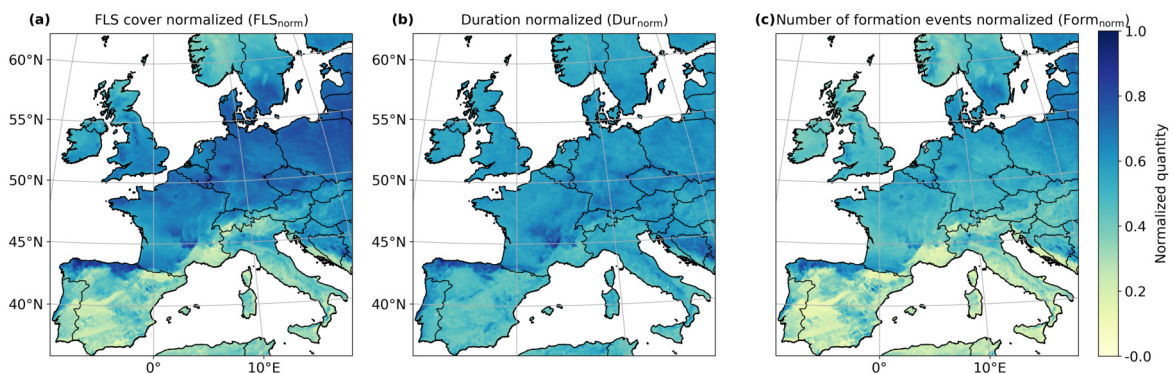


Figure A3.2: Normalized mean FLS cover (a), normalized median duration of FLS event (b) and normalized number of formation events (c) over the complete study period (2006–2015).

Seasonal number of formation events

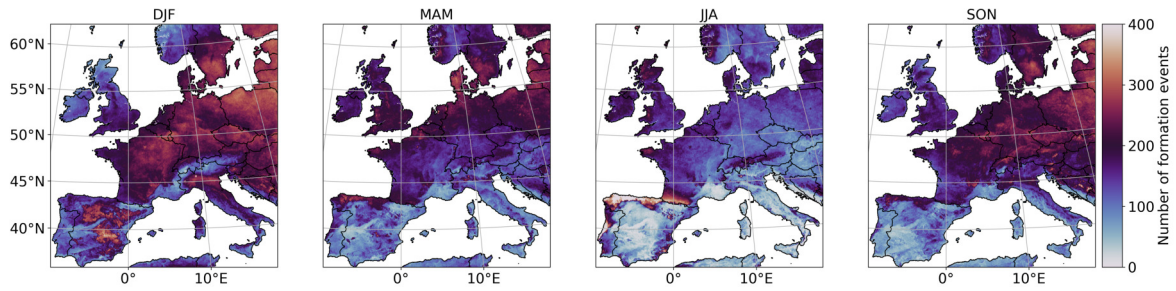


Figure A3.3: Seasonal number of formation events identified by the algorithm.

Fraction of daytime formation and dissipation events

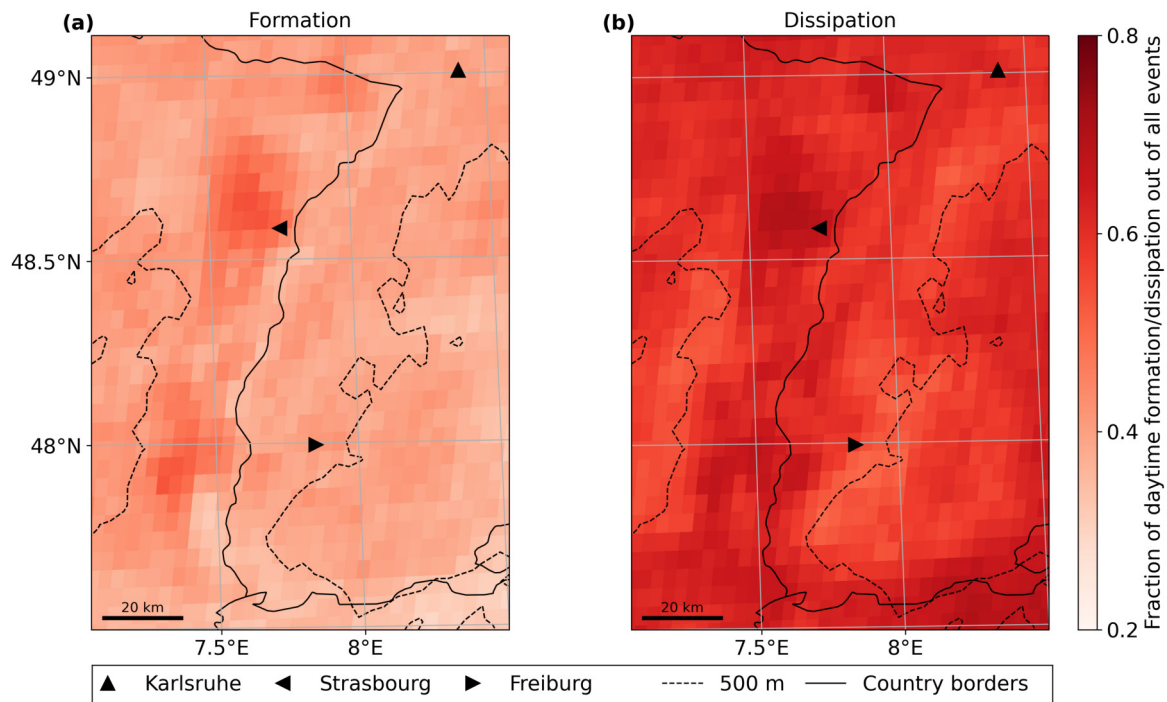


Figure A3.4: Fraction of daytime formation (a) and dissipation (b) events out of all events, in the Upper Rhine valley, Germany. The country border between France and Germany also marks the course of the river Rhine. The black dashed line depicts the 500 m above sea level height.

Most frequent formation and dissipation time - All day

The following plots show the most frequent formation and dissipation time plots as shown in chapter 4, but with the most frequent dissipation time including nighttime dissipation.

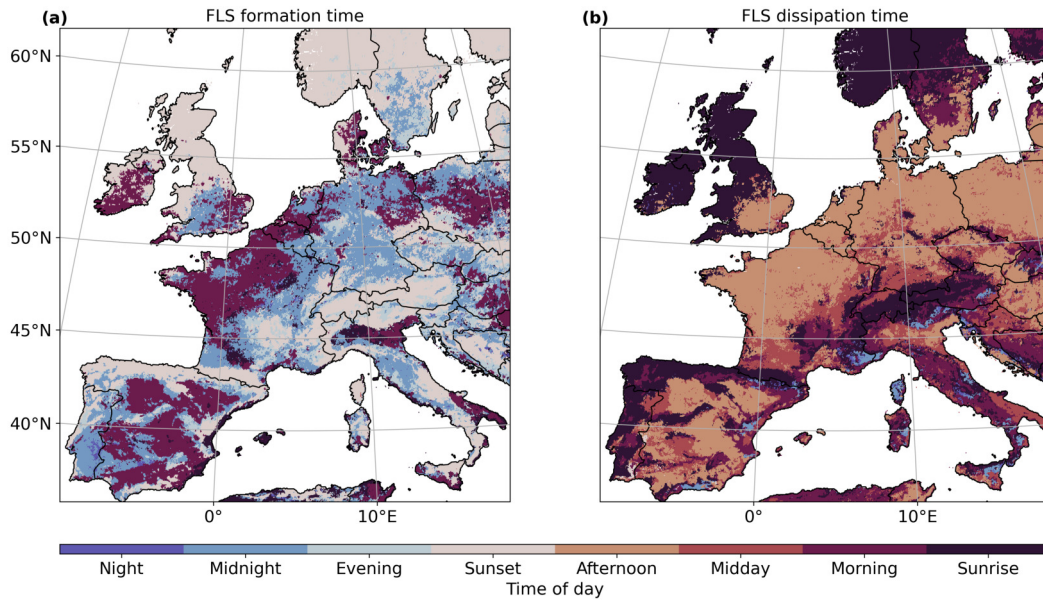


Figure A3.5: Most frequent formation (a) and dissipation times (b) over the entire study period (2006–2015) using daytime and nighttime dissipation.

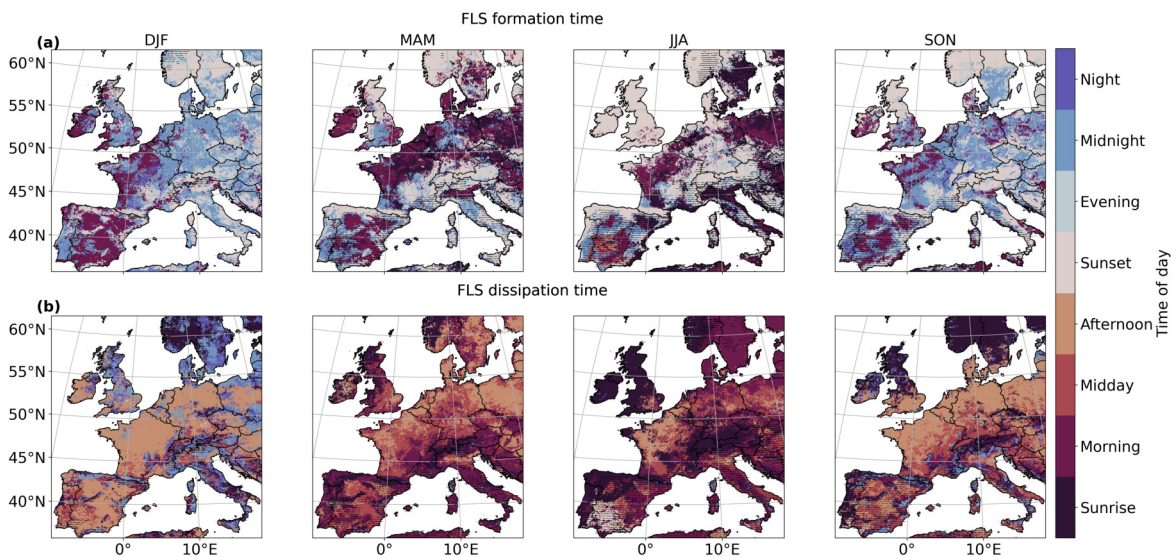


Figure A3.6: Most frequent formation (a) and dissipation time (b) for DJF, MAM, JJA and SON from 2006–2015 using daytime and nighttime dissipation. Pixels with a mean seasonal FLS occurrence of less than 2 hours day^{-1} are marked with a horizontal line.

A4 Scientific Appendix of Chapter 5

Introduction

Figure A4.1 displays the hierarchical clustering results when interpolating over non-significant ($p > 0.05$) Spearman's rho values using nearest neighbor before running the clustering algorithm. The regional clusters identified are similar compared to figure 5.3, with some differences in cluster location in the Mediterranean region. In general, the interpolation technique used has little influence on the obtained regional to sub-regional clusters.

Regional cluster overview

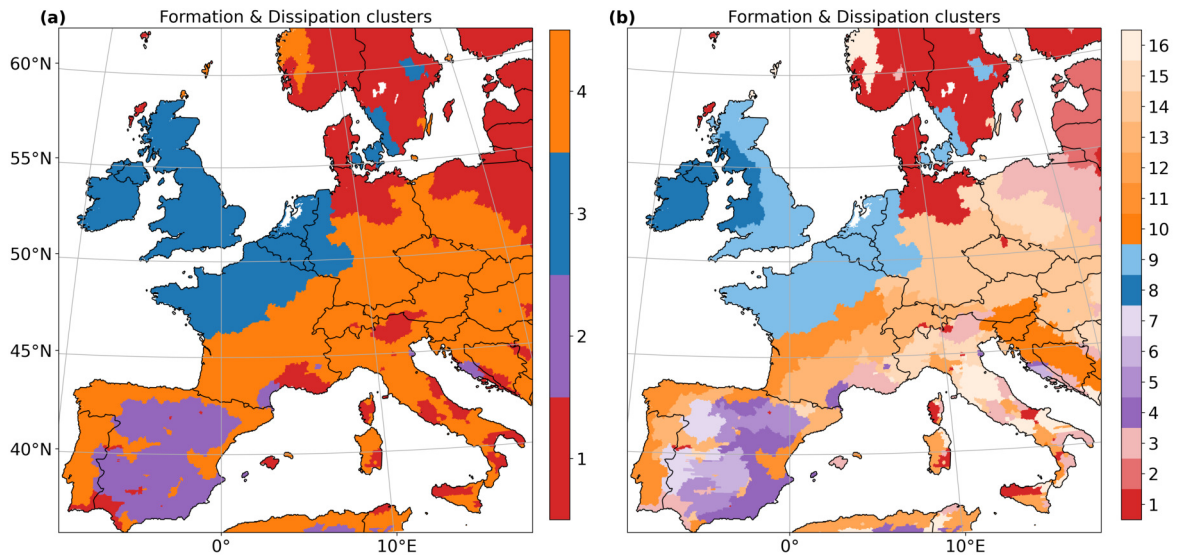


Figure A4.1: Map of formation and dissipation clusters identified by the hierarchical clustering algorithm using nearest neighbor as an interpolation technique. Two different hierarchy levels are displayed, 4 clusters (a) and 16 clusters (b).

Bibliography

- Acker, K., Mertes, S., Möller, D., Wieprecht, W., Auel, R., Kalaß, D., 2002. Case study of cloud physical and chemical processes in low clouds at Mt. Brocken. *Atmospheric Research* 64, 41 – 51. doi:10.1016/S0169-8095(02)00078-9.
- Adhikari, B., Wang, L., 2020. The potential contribution of soil moisture to fog formation in the Namib Desert. *Journal of Hydrology* 591, 125326. doi:10.1016/j.jhydrol.2020.125326.
- Alvarez, I., Gomez-Gesteira, M., deCastro, M., Gomez-Gesteira, J., Dias, J., 2010. Summer upwelling frequency along the western Cantabrian coast from 1967 to 2007. *Journal of Marine Systems* 79, 218–226. doi:10.1016/j.jmarsys.2009.09.004.
- Andersen, H., Cermak, J., 2018. First fully diurnal fog and low cloud satellite detection reveals life cycle in the Namib. *Atmospheric Measurement Techniques* 11, 5461–5470. doi:10.5194/amt-11-5461-2018.
- Andersen, H., Cermak, J., Fuchs, J., Knippertz, P., Gaetani, M., Quinting, J., Sippel, S., Vogt, R., 2020. Synoptic-scale controls of fog and low-cloud variability in the Namib Desert. *Atmospheric Chemistry and Physics* 20, 3415–3438. doi:10.5194/acp-20-3415-2020.
- Andersen, H., Cermak, J., Solodovnik, I., Lelli, L., Vogt, R., 2019. Spatiotemporal dynamics of fog and low clouds in the Namib unveiled with ground- and space-based observations. *Atmospheric Chemistry and Physics* 19, 4383–4392. doi:10.5194/acp-19-4383-2019.
- Anderson, R.G., Canadell, J.G., Randerson, J.T., Jackson, R.B., Hungate, B.A., Baldocchi, D.D., Ban-Weiss, G.A., Bonan, G.B., Caldeira, K., Cao, L., Diffenbaugh, N.S., Gurney, K.R., Kueppers, L.M., Law, B.E., Luysaert, S., O'Halloran, T.L., 2011. Biophysical considerations in forestry for climate protection. *Frontiers in Ecology and the Environment* 9, 174–182. doi:10.1890/090179.

- Applequist, S., Gahrs, G.E., Pfeffer, R.L., Niu, X.F., 2002. Comparison of Methodologies for Probabilistic Quantitative Precipitation Forecasting. *Weather and Forecasting* 17, 783–799. doi:10.1175/1520-0434(2002)017<0783:COMFPQ>2.0.CO;2.
- Avissar, R., Liu, Y., 1996. Three-dimensional numerical study of shallow convective clouds and precipitation induced by land surface forcing. *Journal of Geophysical Research: Atmospheres* 101, 7499–7518. doi:10.1029/95JD03031.
- Avotniece, Z., Klavins, M., Lizuma, L., 2015. Fog climatology in Latvia. *Theoretical and Applied Climatology* 122, 97–109. doi:10.1007/s00704-014-1270-4.
- Azorin-Molina, C., Sanchez-Lorenzo, A., Calbo, J., 2009. A climatological study of sea breeze clouds in the Southeast of the Iberian Peninsula (Alicante, Spain). *Atmosfera* 22, 33–49.
- Baguskas, S.A., Clemesha, R.E., Loik, M.E., 2018. Coastal low cloudiness and fog enhance crop water use efficiency in a California agricultural system. *Agricultural and Forest Meteorology* 252, 109 – 120. doi:10.1016/j.agrformet.2018.01.015.
- Ball, L., Tzanopoulos, J., 2020. Interplay between topography, fog and vegetation in the central South Arabian mountains revealed using a novel Landsat fog detection technique. *Remote Sensing in Ecology and Conservation* 6, 498–513. doi:10.1002/rse2.151.
- Bardoel, S.L., Horna Muñoz, D.V., Grachev, A.A., Krishnamurthy, R., Chamorro, L.P., Fernando, H.J.S., 2021. Fog Formation Related to Gravity Currents Interacting with Coastal Topography. *Boundary-Layer Meteorology* 181, 499–521. doi:10.1007/s10546-021-00638-w.
- Bari, D., Bergot, T., El Khlifi, M., 2015. Numerical study of a coastal fog event over Casablanca, Morocco. *Quarterly Journal of the Royal Meteorological Society* 141, 1894–1905. doi:10.1002/qj.2494.
- Bartoszek, K., 2017. The main characteristics of atmospheric circulation over East-Central Europe from 1871 to 2010. *Meteorology and Atmospheric Physics* 129, 113–129. doi:10.1007/s00703-016-0455-z.
- Bendix, J., 1994. Fog climatology of the Po Valley. *Rivista di meteorologia aeronautica* 54, 25–36.

- Bendix, J., 2002. A satellite-based climatology of fog and low-level stratus in Germany and adjacent areas. *Atmospheric Research* 64, 3–18. doi:10.1016/S0169-8095(02)00075-3.
- Bendix, J., Bachmann, M., 1991. Ein operationell einsetzbares Verfahren zur Nebelerkennung auf der Basis von AVHRR-Daten der NOAA-Satelliten. *Meteorologische Rundschau* 43, 169–178.
- Bendix, J., Thies, B., Nauß, T., Cermak, J., 2006. A feasibility study of daytime fog and low stratus detection with TERRA/AQUA-MODIS over land. *Meteorological Applications* 13, 111–125. doi:10.1017/S1350482706002180.
- Bergot, T., 2016. Large-eddy simulation study of the dissipation of radiation fog. *Quarterly Journal of the Royal Meteorological Society* 142, 1029–1040. doi:10.1002/qj.2706.
- Bergot, T., Carrer, D., Noilhan, J., Bougeault, P., 2005. Improved Site-Specific Numerical Prediction of Fog and Low Clouds: A Feasibility Study. *Weather and Forecasting* 20, 627–646. doi:10.1175/WAF873.1.
- Bergot, T., Escobar, J., Masson, V., 2015. Effect of small-scale surface heterogeneities and buildings on radiation fog: Large-eddy simulation study at Paris–Charles de Gaulle airport. *Quarterly Journal of the Royal Meteorological Society* 141, 285–298. doi:10.1002/qj.2358.
- Bergot, T., Lestringant, R., 2019. On the Predictability of Radiation Fog Formation in a Mesoscale Model: A Case Study in Heterogeneous Terrain. *Atmosphere* 10. doi:10.3390/atmos10040165.
- Berry, Z.C., Johnson, D.M., Reinhardt, K., 2015. Vegetation-zonation patterns across a temperate mountain cloud forest ecotone are not explained by variation in hydraulic functioning or water relations. *Tree Physiology* 35, 925–935. doi:10.1093/treephys/tpv062.
- Bisong, E., 2019. Logistic regression, in: *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Apress, Berkeley, CA, pp. 243–250. doi:10.1007/978-1-4842-4470-8_20.
- Bogawski, P., Bednorz, E., 2016. Atmospheric conditions controlling extreme summertime evapotranspiration in Poland (central Europe). *Natural Hazards* 81, 55–69. doi:10.1007/s11069-015-2066-2.

- Bonan, G.B., 2008. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science* 320, 1444–1449. doi:10.1126/science.1155121.
- Bosman, P.J.M., van Heerwaarden, C.C., Teuling, A.J., 2019. Sensible heating as a potential mechanism for enhanced cloud formation over temperate forest. *Quarterly Journal of the Royal Meteorological Society* 145, 450–468. doi:10.1002/qj.3441.
- Boutle, I., Price, J., Kudzotsa, I., Kokkola, H., Romakkaniemi, S., 2018. Aerosol–fog interaction and the transition to well-mixed radiation fog. *Atmospheric Chemistry and Physics* 18, 7827–7840. doi:10.5194/acp-18-7827-2018.
- Breiman, L., 2001. Random Forests. *Machine learning* 45, 5–32. doi:10.1023/A:1010933404324.
- Bruijnzeel, L., 2001. Hydrology of tropical montane cloud forests: A reassessment. *Land Use and Water Resources Research* 1, 1–18. doi:10.22004/ag.econ.47849.
- Bruijnzeel, L.A., Eugster, W., Burkard, R., 2005. Fog as a Hydrologic Input, in: Anderson, M., McDonnell, J. (Eds.), *Encyclopedia of Hydrological Sciences*. doi:10.1002/0470848944.hsa041.
- Błaś, M., Żaneta Polkowska, Sobik, M., Klimaszewska, K., Nowiński, K., Namieśnik, J., 2010. Fog water chemical composition in different geographic regions of Poland. *Atmospheric Research* 95, 455 – 469. doi:10.1016/j.atmosres.2009.11.008.
- Błaś, M., Sobik, M., Quiel, F., Netzel, P., 2002. Temporal and spatial variations of fog in the Western Sudety Mts., Poland. *Atmospheric Research* 64, 19 – 28. doi:10.1016/S0169-8095(02)00076-5.
- Carleton, A.M., Adegoke, J., Allard, J., Arnold, D.L., Travis, D.J., 2001. Summer season land cover — Convective cloud associations for the midwest U.S. “Corn Belt”. *Geophysical Research Letters* 28, 1679–1682. doi:10.1029/2000GL012635.
- Cerasoli, S., Yin, J., Porporato, A., 2021. Cloud cooling effects of afforestation and reforestation at midlatitudes. *Proceedings of the National Academy of Sciences* 118, e2026241118. doi:10.1073/pnas.2026241118.
- Cermak, J., 2006. SOFOS - A new Satellite-based Operational Fog Observation Scheme. Ph.D. thesis. Philipps-Universität Marburg. doi:10.17192/z2006.0149.

- Cermak, J., 2012. Low clouds and fog along the South-Western African coast — Satellite-based retrieval and spatial patterns. *Atmospheric Research* 116, 15–21. doi:10.1016/j.atmosres.2011.02.012.
- Cermak, J., 2018. Fog and Low Cloud Frequency and Properties from Active-Sensor Satellite Data. *Remote Sensing* 10, 1209. doi:10.3390/rs10081209.
- Cermak, J., Bendix, J., 2007. Dynamical Nighttime Fog/Low Stratus Detection Based on Meteosat SEVIRI Data: A Feasibility Study. *Pure and Applied Geophysics* 164, 1179–1192. doi:10.1007/s00024-007-0213-8.
- Cermak, J., Bendix, J., 2008. A novel approach to fog/low stratus detection using Meteosat 8 data. *Atmospheric Research* 87, 279 – 292. doi:10.1016/j.atmosres.2007.11.009.
- Cermak, J., Bendix, J., 2011. Detecting ground fog from space – a microphysics-based approach. *International Journal of Remote Sensing* 32, 3345–3371. doi:10.1080/01431161003747505.
- Cermak, J., Eastman, R.M., Bendix, J., Warren, S.G., 2009. European climatology of fog and low stratus based on geostationary satellite observations. *Quarterly Journal of the Royal Meteorological Society* 135, 2125–2130. doi:10.1002/qj.503.
- Chen, L., Dirmeyer, P.A., 2020. Distinct Impacts of Land Use and Land Management on Summer Temperatures. *Frontiers in Earth Science* 8. doi:10.3389/feart.2020.00245.
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, New York, NY, USA. p. 785–794. doi:10.1145/2939672.2939785.
- 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. URL: https://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=MSG_{_}AOD_{_}V001, doi:10.5676/EUM_SAF_CM/MSG_AOD/V001.
- Cotton, W.R., Bryan, G., van den Heever, S.C., 2011. Fogs and Stratocumulus Clouds, in: Cotton, W., Bryan, G., van den Heever, S.C. (Eds.), *Storm and Cloud*

- Dynamics. Academic Press. volume 99 of *International Geophysics*, pp. 179–242. doi:10.1016/S0074-6142(10)09912-2.
- Cuxart, J., Jiménez, M., 2012. Deep Radiation Fog in a Wide Closed Valley: Study by Numerical Modeling and Remote Sensing. *Pure and Applied Geophysics* 169, 911–926. doi:10.1007/s00024-011-0365-4.
- Cuxart, J., Telisman Prtenjak, M., Matjacic, B., 2021. Pannonian Basin Nocturnal Boundary Layer and Fog Formation: Role of Topography. *Atmosphere* 12. doi:10.3390/atmos12060712.
- Dawson, T.E., 1998. Fog in the California redwood forest: ecosystem inputs and use by plants. *Oecologia* 117, 476–485. doi:10.1007/s004420050683.
- 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.M., 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. *Quarterly Journal of the Royal Meteorological Society* 137, 553–597. doi:10.1002/qj.828.
- Douglas, A., Stier, P., 2021. Using the learnings of machine learning to distill cloud controlling environmental regimes from satellite observations, in: EGU General Assembly 2021, pp. EGU21–16443. doi:10.5194/egusphere-egu21-16443.
- Dozier, J., 1989. Spectral Signature of Alpine Snow Cover from the Landsat Thematic Mapper. *Remote Sensing of Environment* 28, 9–22. doi:10.1016/0034-4257(89)90101-6.
- Ducongé, L., Lac, C., Vié, B., Bergot, T., Price, J.D., 2020. Fog in heterogeneous environments: the relative importance of local and non-local processes on radiative-advective fog formation. *Quarterly Journal of the Royal Meteorological Society* 146, 2522–2546. doi:10.1002/qj.3783.
- Dupont, J.C., Haeffelin, M., Protat, A., Bouniol, D., Boyouk, N., Morille, Y., 2012. Stratus-Fog Formation and Dissipation: A 6-Day Case Study. *Boundary-layer meteorology* 143, 207–225. doi:10.1007/s10546-012-9699-4.

- Durieux, L., Machado, L.A.T., Laurent, H., 2003. The impact of deforestation on cloud cover over the Amazon arc of deforestation. *Remote Sensing of Environment* 86, 132–140. doi:10.1016/S0034-4257(03)00095-6.
- Duveiller, G., Filipponi, F., Ceglar, A., Bojanowski, J., Alkama, R., Cescatti, A., 2021. Revealing the widespread potential of forests to increase low level cloud cover. *Nature Communications* 12, 4337. doi:10.1038/s41467-021-24551-5.
- Duveiller, G., Hooker, J., Cescatti, A., 2018. The mark of vegetation change on Earth's surface energy balance. *Nature Communications* 9, 679. doi:10.1038/s41467-017-02810-8.
- Duynkerke, P.G., 1991. Radiation Fog: A Comparison of Model Simulation with Detailed Observations. *Monthly Weather Review* 119, 324–341. doi:10.1175/1520-0493(1991)119<0324:RFACOM>2.0.CO;2.
- Ebner, M., Miranda, T., Roth-Nebelsick, A., 2011. Efficient fog harvesting by *Stipagrostis sabulicola* (Namib dune bushman grass). *Journal of Arid Environments* 75, 524 – 531. doi:10.1016/j.jaridenv.2011.01.004.
- Eckardt, F., Soderberg, K., Coop, L., Muller, A., Vickery, K., Grandin, R., Jack, C., Kapalanga, T., Henschel, J., 2013. The nature of moisture at Gobabeb, in the central Namib Desert. *Journal of Arid Environments* 93, 7 – 19. doi:10.1016/j.jaridenv.2012.01.011.
- Egli, S., Thies, B., Bendix, J., 2019. A spatially explicit and temporally highly resolved analysis of variations in fog occurrence over Europe. *Quarterly Journal of the Royal Meteorological Society* 145, 1721–1740. doi:10.1002/qj.3522.
- Egli, S., Thies, B., Drönner, J., Cermak, J., Bendix, J., 2017. A 10 year fog and low stratus climatology for Europe based on Meteosat Second Generation data. *Quarterly Journal of the Royal Meteorological Society* 143, 530–541. doi:10.1002/qj.2941.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 802–813. doi:10.1111/j.1365-2656.2008.01390.x.
- Ellrod, G.P., 1995. Advances in the Detection and Analysis of Fog at Night Using GOES Multispectral Infrared Imagery. *Weather and Forecasting* 10, 606 – 619. doi:10.1175/1520-0434(1995)010<0606:AITDAA>2.0.CO;2.

- Estrela, M.J., Valiente, J.A., Corell, D., Millán, M.M., 2008. Fog collection in the western Mediterranean basin (Valencia region, Spain). *Atmospheric Research* 87, 324–337. doi:10.1016/j.atmosres.2007.11.013.
- Eugster, W., 2008. Fog research. *Die Erde - Journal of the Geographical Society of Berlin* 139, 1–10.
- EUMETSAT, 2013. MSG level 1.5 image data format description. Technical Report. European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). Darmstadt, Germany.
- EUMETSAT CM SAF, 2019. Algorithm Theoretical Basis Document (ATBD) Meteosat Solar Surface Radiation and effective Cloud Albedo Climate Data Records - Heliosat (SARAH-2). Technical Report. European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Satellite Application Facility on Climate Monitoring (CM-SAF). Darmstadt, Germany. doi:10.5676/EUM_SAF_{CM}/SARAH/V002₀1.
- Everitt, B. (Ed.), 2011. Cluster analysis. Wiley Series in Probability and Statistics. 5th ed., Wiley, Chichester.
- Eyre, J., Brownscombe, J., Allam, R., 1984. Detection of fog at night using Advanced Very High Resolution Radiometer (AVHRR) imagery. *Meteorological Magazine* 113, 266–271.
- Fallmann, J., Lewis, H., Sanchez, J.C., Lock, A., 2019. Impact of high-resolution ocean–atmosphere coupling on fog formation over the North Sea. *Quarterly Journal of the Royal Meteorological Society* 145, 1180–1201. doi:10.1002/qj.3488.
- Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29, 1189–1232.
- Fuchs, J., Andersen, H., Cermak, J., Pauli, E., Roebeling, R., 2022. High-resolution satellite-based cloud detection for the analysis of land surface effects on boundary layer clouds. *Atmospheric Measurement Techniques* 15, 4257–4270. doi:10.5194/amt-15-4257-2022.
- 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, 16537–16552. doi:10.5194/acp-18-16537-2018.

- Fuzzi, S., Facchini, M.C., Orsi, G., Lind, J.A., Wobrock, W., Kessel, M., Maser, R., Jaeschke, W., Enderle, K.H., Arends, B.G., Berner, A., Solly, I., Krusiz, C., Reischl, G., Pahl, S., Kaminski, U., Winkler, P., Ogren, J.A., Noone, K.J., Hallberg, A., Fierlinger-Oberlininger, H., Puxbaum, H., Marzorati, A., Hansson, H.C., Wiedensohler, A., Svenningsson, I.B., Martinsson, B.G., Schell, D., Georgii, H.W., 1992. The Po Valley Fog Experiment 1989. *Tellus B* 44, 448–468. doi:10.1034/j.1600-0889.1992.t01-4-00002.x.
- Gambill, L.D., Mecikalski, J.R., 2011. A Satellite-Based Summer Convective Cloud Frequency Analysis over the Southeastern United States. *Journal of Applied Meteorology and Climatology* 50, 1756 – 1769. doi:10.1175/2010JAMC2559.1.
- Gao, Y., Zhang, M., Liu, Z., Wang, L., Wang, P., Xia, X., Tao, M., Zhu, L., 2015. Modeling the feedback between aerosol and meteorological variables in the atmospheric boundary layer during a severe fog-haze event over the North China Plain. *Atmospheric Chemistry and Physics* 15, 4279–4295. doi:10.5194/acp-15-4279-2015.
- Gautam, R., Singh, M.K., 2018. Urban heat island over Delhi punches holes in widespread fog in the Indo-Gangetic Plains. *Geophysical Research Letters* 45, 1114–1121. doi:10.1002/2017GL076794.
- Gentine, P., Ferguson, C.R., Holtslag, A.A.M., 2013. Diagnosing evaporative fraction over land from boundary-layer clouds. *Journal of Geophysical Research: Atmospheres* 118, 8185–8196. doi:10.1002/jgrd.50416.
- Glickman, T., 2000. *Glossary of Meteorology*. 2nd ed., American Meteorological Society: Boston.
- Gottlieb, T.R., Eckardt, F.D., Venter, Z.S., Cramer, M.D., 2019. The contribution of fog to water and nutrient supply to *Arthroerua leubnitziae* in the central Namib Desert, Namibia. *Journal of Arid Environments* 161, 35 – 46. doi:10.1016/j.jaridenv.2018.11.002.
- Gradstein, S., Obregon, A., Gehrig, C., Bendix, J., 2011. Tropical lowland cloud forest: a neglected forest type, in: *Tropical montane cloud forests: Science for conservation and management*, Cambridge University Press. pp. 130–133.
- Gray, E., Gilardoni, S., Baldocchi, D., McDonald, B.C., Facchini, M.C., Goldstein, A.H., 2019. Impact of Air Pollution Controls on Radiation Fog Frequency in the

- Central Valley of California. *Journal of Geophysical Research: Atmospheres* 124, 5889–5905. doi:10.1029/2018JD029419.
- Gultepe, I., Heymsfield, A.J., Fernando, H.J.S., Pardyjak, E., Dorman, C.E., Wang, Q., Creegan, E., Hoch, S.W., Flagg, D.D., Yamaguchi, R., Krishnamurthy, R., Gaberšek, S., Perrie, W., Perelet, A., Singh, D.K., Chang, R., Nagare, B., Wagh, S., Wang, S., 2021. A Review of Coastal Fog Microphysics During C-FOG. *Boundary-Layer Meteorology* 181, 227–265. doi:10.1007/s10546-021-00659-5.
- Gultepe, I., Tardif, R., Michaelides, S., Cermak, J., Bott, A., Bendix, J., Müller, M.D., Pagowski, M., Hansen, B., Ellrod, G., et al., 2007. Fog Research: A Review of Past Achievements and Future Perspectives. *Pure and Applied Geophysics* 164, 1121–1159. doi:10.1007/s00024-007-0211-x.
- Haeffelin, M., Bergot, T., Elias, T., Tardif, R., Carrer, D., Chazette, P., Colomb, M., Drobinski, P., Dupont, E., Dupont, J.C., Gomes, L., Musson-Genon, L., Pietras, C., Plana-Fattori, A., Protat, A., Rangognio, J., Raut, J.C., Rémy, S., Richard, D., Sciare, J., Zhang, X., 2010. Parisfog: Shedding new Light on Fog Physical Processes. *Bulletin of the American Meteorological Society* 91, 767–783. doi:10.1175/2009BAMS2671.1.
- Haeffelin, M., Dupont, J.C., Boyouk, N., Baumgardner, D., Gomes, L., Roberts, G., Elias, T., 2013. A Comparative Study of Radiation Fog and Quasi-Fog Formation Processes During the ParisFog Field Experiment 2007. *Pure and Applied Geophysics* 170, 2283–2303. doi:10.1007/s00024-013-0672-z.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*. Springer Series in Statistics New York, NY, USA. doi:10.1007/978-0-387-84858-7.
- Herman, G.R., Schumacher, R.S., 2018. Dendrology in Numerical Weather Prediction: What Random Forests and Logistic Regression Tell Us about Forecasting Extreme Precipitation. *Monthly Weather Review* 146, 1785–1812. doi:10.1175/MWR-D-17-0307.1.
- Hersbach, H., 2016. The ERA5 Atmospheric Reanalysis, in: *AGU Fall Meeting Abstracts*, 12–16 December 2016, San Fransisco.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci,

- C., Dee, D., T.J.N., 2018. ERA5 hourly data on pressure levels from 1979 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi:10.24381/cds.bd0915c6.
- Hess, P., Brezowsky, H., 1977. Katalog der Grosswetterlagen Europas 1881-1976. 3. verbesserte und ergänzte Auflage. Berichte des Deutschen Wetterdienstes 113, 1–140.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978.
- Hůnová, I., Brabec, M., Geletič, J., Malý, M., Dumitrescu, A., 2021a. Statistical analysis of the effects of forests on fog. *Science of The Total Environment* 781, 146675. doi:10.1016/j.scitotenv.2021.146675.
- Hůnová, I., Brabec, M., Geletič, J., Malý, M., Dumitrescu, A., 2022. Local fresh- and sea-water effects on fog occurrence. *Science of The Total Environment* 807, 150799. doi:10.1016/j.scitotenv.2021.150799.
- Hůnová, I., Brabec, M., Malý, M., Dumitrescu, A., Geletič, J., 2021b. Terrain and its effects on fog occurrence. *Science of The Total Environment* 768, 144359. doi:10.1016/j.scitotenv.2020.144359.
- Hůnová, I., Brabec, M., Malý, M., Valeriánová, A., 2018. Revisiting fog as an important constituent of the atmosphere. *Science of The Total Environment* 636, 1490–1499. doi:10.1016/j.scitotenv.2018.04.322.
- Iacobellis, S.F., Cayan, D.R., 2013. The variability of California summertime marine stratus: Impacts on surface air temperatures. *Journal of Geophysical Research: Atmospheres* 118, 9105–9122. doi:10.1002/jgrd.50652.
- Izett, J.G., van de Wiel, B.J.H., Baas, P., van Hooft, J.A., Schulte, R.B., 2019. Dutch fog: On the observed spatio-temporal variability of fog in the Netherlands. *Quarterly Journal of the Royal Meteorological Society* 145, 2817–2834. doi:10.1002/qj.3597.
- Jia, G., Shevliakova, E., Artaxo, P., De-Docoudré, N., Houghton, R., House, J., Kitajima, K., Lennard, C., Popp, A., Sirin, A., et al., 2019a. Land–climate interactions, in: *Special Report on Climate Change and Land: An IPCC Special Report on climate change, desertification, land degradation, sustainable land*

- management, food security, and greenhouse gas fluxes in terrestrial ecosystems. *IPCC*, pp. 133–206.
- Jia, X., Quan, J., Zheng, Z., Liu, X., Liu, Q., He, H., Liu, Y., 2019b. Impacts of Anthropogenic Aerosols on Fog in North China Plain. *Journal of Geophysical Research: Atmospheres* 124, 252–265. doi:10.1029/2018JD029437.
- Kalthoff, N., Binder, H.J., Kossmann, M., Vöglin, R., Corsmeier, U., Fiedler, F., Schlager, H., 1998. Temporal evolution and spatial variation of the boundary layer over complex terrain. *Atmospheric Environment* 32, 1179–1194. doi:10.1016/S1352-2310(97)00193-3.
- Kammer, J., Flaud, P.M., Chazeaubeny, A., Ciuraru, R., Le Menach, K., Geneste, E., Budzinski, H., Bonnefond, J., Lamaud, E., Perraudin, E., Villenave, E., 2020. Biogenic volatile organic compounds (BVOCs) reactivity related to new particle formation (NPF) over the Landes forest. *Atmospheric Research* 237, 104869. doi:10.1016/j.atmosres.2020.104869.
- Kammer, J., Perraudin, E., Flaud, P.M., Lamaud, E., Bonnefond, J., Villenave, E., 2018. Observation of nighttime new particle formation over the French Landes forest. *Science of The Total Environment* 621, 1084–1092. doi:10.1016/j.scitotenv.2017.10.118.
- Karimi, M., 2020. Direct Numerical Simulation of Fog: The Sensitivity of a Dissipation Phase to Environmental Conditions. *Atmosphere* 11, 12. doi:10.3390/atmos11010012.
- Kenworthy, J.M., 2014. Regional weather and climates of the British Isles – Part 7: North West England and the Isle of Man. *Weather* 69, 87–93. doi:10.1002/wea.2256.
- Klemm, O., Lin, N.H., 2016. What Causes Observed Fog Trends: Air Quality or Climate Change? *Aerosols and Air Quality Research* 16, 1131–1142. doi:10.4209/aaqr.2015.05.0353.
- Klemm, O., Wrzesinsky, T., 2007. Fog deposition fluxes of water and ions to a mountainous site in Central Europe. *Tellus B: Chemical and Physical Meteorology* 59, 705–714. doi:10.1111/j.1600-0889.2007.00287.x.
- Kneringer, P., Dietz, S.J., Mayr, G.J., Zeileis, A., 2019. Probabilistic nowcasting of low-visibility procedure states at Vienna International Airport during cold season. *Pure and Applied Geophysics* 176, 2165–2177. doi:10.1007/s00024-018-1863-4.

- Knerr, I., Trachte, K., Egli, S., Barth, J.A., Bräuning, A., Garel, E., Häusser, M., Huneau, F., Juhlke, T.R., Santoni, S., Szymczak, S., van Geldern, R., Bendix, J., 2021. Fog - low stratus (FLS) regimes on Corsica with wind and PBLH as key drivers. *Atmospheric Research* 261, 105731. doi:10.1016/j.atmosres.2021.105731.
- Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43, 59–69. doi:10.1007/BF00337288.
- Koracin, D., Dorman, C.E., 2017. Introduction, in: *Marine Fog: Challenges and Advancements in Observations, Modeling, and Forecasting*. doi:10.1007/978-3-319-45229-6.
- 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. doi:10.1029/2007GL029253.
- Köhler, C., Steiner, A., Saint-Drenan, Y.M., Ernst, D., Bergmann-Dick, A., Zirkelbach, M., Bouallègue, Z.B., Metzinger, I., Ritter, B., 2017. Critical weather situations for renewable energies – Part B: Low stratus risk for solar power. *Renewable Energy* 101, 794–803. doi:10.1016/j.renene.2016.09.002.
- Lamb, D., Verlinde, J., 2011. *Physics and chemistry of clouds*. Cambridge University Press.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers* 7, 3–10. doi:10.1016/j.gsf.2015.07.003.
- Lehnert, L.W., Thies, B., Trachte, K., Achilles, S., Osses, P., Baumann, K., Bendix, J., Schmidt, J., Samolov, E., Jung, P., Leinweber, P., Karsten, U., Büdel, B., 2018. A Case Study on Fog/Low Stratus Occurrence at Las Lomitas, Atacama Desert (Chile) as a Water Source for Biological Soil Crusts. *Aerosol and Air Quality Research* 18, 254–269. doi:10.4209/aaqr.2017.01.0021.
- Leigh, R.J., Drake, L., Thampapillai, D.J., 1998. An Economic Analysis of Terminal Aerodrome Forecasts with Special Reference to Sydney Airport. *Journal of Transport Economics and Policy* 32, 377–392.
- Leipper, D.F., 1994. Fog on the U.S. West Coast: A Review. *Bulletin of the American Meteorological Society* 75, 229–240. doi:10.1175/1520-0477(1994)075<0229:FOTUWC>2.0.CO;2.

- Lever, J., Krzywinski, M., Altman, N., 2016. Logistic regression. *Nature Methods* 13, 541–542. doi:10.1038/nmeth.3904.
- Li, Y., Aemisegger, F., Riedl, A., Buchmann, N., Eugster, W., 2021. The role of dew and radiation fog inputs in the local water cycling of a temperate grassland during dry spells in central Europe. *Hydrology and Earth System Sciences* 25, 2617–2648. doi:10.5194/hess-25-2617-2021.
- Li, Y., Zhao, M., Motesharrei, S., Mu, Q., Kalnay, E., Li, S., 2015. Local cooling and warming effects of forests based on satellite observations. *Nature communications* 6, 1–8.
- Liu, Z., Vaughan, M., Winker, D., Kittaka, C., Getzewich, B., Kuehn, R., Omar, A., Powell, K., Trepte, C., Hostetler, C., 2009. The CALIPSO Lidar Cloud and Aerosol Discrimination: Version 2 Algorithm and Initial Assessment of Performance. *Journal of Atmospheric and Oceanic Technology* 26, 1198–1213. doi:10.1175/2009JTECHA1229.1.
- 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. *Nature machine intelligence* 2, 2522–5839. doi:10.1038/s42256-019-0138-9.
- Lundberg, S.M., Lee, S.I., 2017. A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems* 30, 4765–4774. doi:10.48550/arXiv.1705.07874.
- López, L., Sánchez, J., 2009. Discriminant methods for radar detection of hail. *Atmospheric Research* 93, 358–368. doi:10.1016/j.atmosres.2008.09.028.
- Maalick, Z., Kühn, T., Korhonen, H., Kokkola, H., Laaksonen, A., Romakkaniemi, S., 2016. Effect of aerosol concentration and absorbing aerosol on the radiation fog life cycle. *Atmospheric Environment* 133, 26–33. doi:10.1016/j.atmosenv.2016.03.018.
- Mahrt, L., Ek, M., 1993. Spatial variability of turbulent fluxes and roughness lengths in HAPEX-MOBILHY. *Boundary-layer meteorology* 65, 381–400.
- Manton, M.J., 1983. The physics of clouds in the atmosphere. *Reports on Progress in Physics* 46, 1393–1444. doi:10.1088/0034-4885/46/12/001.

- Maronga, B., Bosveld, F.C., 2017. Key parameters for the life cycle of nocturnal radiation fog: a comprehensive large-eddy simulation study. *Quarterly Journal of the Royal Meteorological Society* 143, 2463–2480. doi:10.1002/qj.3100.
- Mayes, J., 2013. Regional weather and climates of the British Isles - Part 5: Wales. *Weather* 68, 227–232. doi:10.1002/wea.2149.
- Mayes, J., Wheeler, D., 2013. Regional weather and climates of the British Isles - Part 1: Introduction. *Weather* 68, 3–8. doi:10.1002/wea.2041.
- Mazoyer, M., Lac, C., Thouron, O., Bergot, T., Masson, V., Musson-Genon, L., 2017. Large eddy simulation of radiation fog: impact of dynamics on the fog life cycle. *Atmospheric Chemistry and Physics* 17, 13017–13035. doi:10.5194/acp-17-13017-2017.
- McFiggans, G., Artaxo, P., Baltensperger, U., Coe, H., Facchini, M.C., Feingold, G., Fuzzi, S., Gysel, M., Laaksonen, A., Lohmann, U., Mentel, T.F., Murphy, D.M., O'Dowd, C.D., Snider, J.R., Weingartner, E., 2006. The effect of physical and chemical aerosol properties on warm cloud droplet activation. *Atmospheric Chemistry and Physics* 6, 2593–2649. doi:10.5194/acp-6-2593-2006.
- 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. 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. *Ecological Modelling* 411, 108815. doi:10.1016/j.ecolmodel.2019.108815.
- Mitchell, D., Henschel, J.R., Hetem, R.S., Wassenaar, T.D., Strauss, W.M., Hanrahan, S.A., Seely, M.K., 2020. Fog and fauna of the Namib Desert: past and future. *Ecosphere* 11, e02996. doi:10.1002/ecs2.2996.
- Molnar, C., 2019. Interpretable Machine Learning. <https://christophm.github.io/interpretable-ml-book/>.
- Montecinos, S., Favier, V., Astudillo, O., Tracol, Y., Börsch-Supan, W., Bischoff-Gauß, I., Kalthoff, N., 2008. The Impact of Agricultural Activities on Fog Formation in an

- Arid Zone of Chile. *Die Erde - Journal of the Geographical Society of Berlin* 139, 71–89.
- Moreaux, V., O’Grady, A.P., Nguyen-The, N., Loustau, D., 2013. Water use of young maritime pine and Eucalyptus stands in response to climatic drying in south-western France. *Plant Ecology & Diversity* 6, 57–71. doi:10.1080/17550874.2012.668228.
- Mühlig, A.C., Klemm, O., Gonçalves, F.L.T., 2020. Fog, Temperature and Air Quality Over the Metropolitan Area of São Paulo: a Trend Analysis from 1998 to 2018. *Water, Air, & Soil Pollution* 231, 535. doi:10.1007/s11270-020-04902-6.
- Muñoz Sabater, J., 2019. ERA5-Land hourly data from 1981 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi:10.24381/cds.e2161bac.
- Müller, M.D., Masbou, M., Bott, A., 2010. Three-dimensional fog forecasting in complex terrain. *Quarterly Journal of the Royal Meteorological Society* 136, 2189–2202. doi:10.1002/qj.705.
- Müllner, D., 2011. Modern hierarchical, agglomerative clustering algorithms. doi:10.48550/ARXIV.1109.2378.
- Müllner, D., 2013. fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python. *Journal of Statistical Software* 53, 1–18. doi:10.18637/jss.v053.i09.
- NASA Worldview, 2022. MODIS Aqua Scene from 2022-01-15 over southern Germany. URL: <https://go.nasa.gov/3DWZQ7i>.
- NASA/LARC/SD/ASDC, 2018. CALIPSO Lidar Level 2 1 km Cloud Layer, V4-20. doi:10.5067/CALIOP/CALIPSO/LID_L2₀1KMCLAY – STANDARD – V4 – 20.
- Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics* 7, 21. doi:10.3389/fnbot.2013.00021.
- Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine Learning Methods Without Tears: A Primer for Ecologists. *The Quarterly Review of Biology* 83, 171–193. doi:10.1086/587826.
- van Oldenborgh, G.J., Yiou, P., Vautard, R., 2010. On the roles of circulation and aerosols in the decline of mist and dense fog in Europe over the last 30 years. *Atmospheric Chemistry and Physics* 10, 4597–4609. doi:10.5194/acp-10-4597-2010.

- Olivier, J., Stockton, P.L., 1989. The influence of upwelling extent upon fog incidence at Lüderitz, southern Africa. *International Journal of Climatology* 9, 69–75. doi:10.1002/joc.3370090106.
- Pagowski, M., Gultepe, I., King, P., 2004. Analysis and Modeling of an Extremely Dense Fog Event in Southern Ontario. *Journal of Applied Meteorology* 43, 3–16. doi:10.1175/1520-0450(2004)043<0003:AAMOE>2.0.CO;2.
- Pauli, E., Andersen, H., Bendix, J., Cermak, J., Egli, S., 2020. Determinants of fog and low stratus occurrence in continental central Europe - a quantitative satellite-based evaluation. *Journal of Hydrology* 591, 125451. doi:10.1016/j.jhydrol.2020.125451.
- Pauli, E., Cermak, J., Andersen, H., 2021. Satellite-based fog and low stratus cloud formation and dissipation times data set. doi:10.5445/IR/1000141293.
- Pauli, E., Cermak, J., Andersen, H., 2022a. A satellite-based climatology of fog and low stratus formation and dissipation times in central Europe. *Quarterly Journal of the Royal Meteorological Society* 148, 1439–1454. doi:10.1002/qj.4272.
- Pauli, E., Cermak, J., Teuling, A.J., 2022b. Enhanced nighttime fog and low stratus occurrence over the Landes forest, France. *Geophysical Research Letters* 49, e2021GL097058. doi:10.1029/2021GL097058.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830. doi:10.48550/arXiv.1201.0490.
- Perry, A.H., 2014. Regional weather and climates of the British Isles – Part 8: South West England and the Channel Islands. *Weather* 69, 208–212. doi:10.1002/wea.2269.
- Petäjä, T., Tabakova, K., Manninen, A., Ezhova, E., O’Connor, E., Moisseev, D., Sinclair, V.A., Backman, J., Levula, J., Luoma, K., Virkkula, A., Paramonov, M., Rätty, M., Äijälä, M., Heikkinen, L., Ehn, M., Sipilä, M., Yli-Juuti, T., Virtanen, A., Ritsche, M., Hickmon, N., Pulik, G., Rosenfeld, D., Worsnop, D.R., Bäck, J., Kulmala, M., Kerminen, V.M., 2022. Influence of biogenic emissions from boreal forests on aerosol–cloud interactions. *Nature Geoscience* 15, 42–47. doi:10.1038/s41561-021-00876-0.

- Pielke, R.A., 2001. Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Reviews of Geophysics* 39, 151–177. doi:10.1029/1999RG000072.
- Pielke, R.A., Avissar, R., 1990. Influence of landscape structure on local and regional climate. *Landscape Ecology* 4, 133–155. doi:10.1007/BF00132857.
- Pielke, R.A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., Goldewijk, K.K., Nair, U., Betts, R., Fall, S., Reichstein, M., Kabat, P., de Noblet, N., 2011. Land use/land cover changes and climate: modeling analysis and observational evidence. *WIREs Climate Change* 2, 828–850. doi:10.1002/wcc.144.
- Pilié, R.J., Mack, E.J., Kocmond, W.C., Rogers, C.W., Eadie, W.J., 1975. The Life Cycle of Valley Fog. Part I: Micrometeorological Characteristics. *Journal of Applied Meteorology* 14, 347–363.
- Pitman, A.J., 2003. The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology* 23, 479–510. doi:10.1002/joc.893.
- Pohl, M.J., Lehnert, L., Bader, M.Y., Gradstein, S.R., Viehweger, J., Bendix, J., 2021. A new fog and low stratus retrieval for tropical South America reveals widespread fog in lowland forests. *Remote Sensing of Environment* 264, 112620. doi:10.1016/j.rse.2021.112620.
- Poku, C., Ross, A.N., Blyth, A.M., Hill, A.A., Price, J.D., 2019. How important are aerosol–fog interactions for the successful modelling of nocturnal radiation fog? *Weather* 74, 237–243. doi:10.1002/wea.3503.
- Pöschl, U., Martin, S.T., Sinha, B., Chen, Q., Gunthe, S.S., Huffman, J.A., Borrmann, S., Farmer, D.K., Garland, R.M., Helas, G., Jimenez, J.L., King, S.M., Manzi, A., Mikhailov, E., Pauliquevis, T., Petters, M.D., Prenni, A.J., Roldin, P., Rose, D., Schneider, J., Su, H., Zorn, S.R., Artaxo, P., Andreae, M.O., 2010. Rainforest Aerosols as Biogenic Nuclei of Clouds and Precipitation in the Amazon. *Science* 329, 1513–1516. doi:10.1126/science.1191056.
- Price, J., 2011. Radiation Fog. Part I: Observations of Stability and Drop Size Distributions. *Boundary-Layer Meteorology* 139, 167–191. doi:10.1007/s10546-010-9580-2.

- Price, J., 2019. On the Formation and Development of Radiation Fog: An Observational Study. *Boundary-Layer Meteorology* 172, 167–197. doi:10.1007/s10546-019-00444-5.
- Pérez-Díaz, J., Ivanov, O., Peshev, Z., Álvarez Valenzuela, M., Valiente-Blanco, I., Evgenieva, T., Dreischuh, T., Gueorguiev, O., Todorov, P., Vaseashta, A., 2017. Fogs: Physical Basis, Characteristic Properties, and Impacts on the Environment and Human Health. *Water* 9, 807. doi:10.3390/w9100807.
- Rabin, R.M., Martin, D.W., 1996. Satellite observations of shallow cumulus coverage over the central United States: An exploration of land use impact on cloud cover. *Journal of Geophysical Research: Atmospheres* 101, 7149–7155. doi:10.1029/95JD02891.
- Ramanathan, V., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D., 2001. Aerosols, Climate, and the Hydrological Cycle. *Science* 294, 2119–2124. doi:10.1126/science.1064034.
- Rastogi, B., Williams, A.P., Fischer, D.T., Iacobellis, S.F., McEachern, K., Carvalho, L., Jones, C., Baguskas, S.A., Still, C.J., 2016. Spatial and Temporal Patterns of Cloud Cover and Fog Inundation in Coastal California: Ecological Implications. *Earth Interactions* 20, 1–19. doi:10.1175/EI-D-15-0033.1.
- Ray, D.K., Nair, U.S., Welch, R.M., Han, Q., Zeng, J., Su, W., Kikuchi, T., Lyons, T.J., 2003. Effects of land use in Southwest Australia: 1. Observations of cumulus cloudiness and energy fluxes. *Journal of Geophysical Research: Atmospheres* 108. doi:10.1029/2002JD002654.
- Roach, W., 1995. Back to basics: Fog: Part 2 - The formation and dissipation of land fog. *Weather* 50, 7–11.
- Roach, W.T., Brown, R., Caughey, S.J., Garland, J.A., Readings, C.J., 1976. The physics of radiation fog: I – a field study. *Quarterly Journal of the Royal Meteorological Society* 102, 313–333. doi:10.1002/qj.49710243204.
- Rosskopf, Y., Scherrer, S., 2017. On the relationship between fog and low stratus (FLS) and weather types over the Swiss Plateau. Technical Report MeteoSwiss 266, 1–40.
- Roth-Nebelsick, A., Ebner, M., Miranda, T., Gottschalk, V., Voigt, D., Gorb, S., Stegmaier, T., Sarsour, J., Linke, M., Konrad, W., 2012. Leaf surface structures enable the endemic Namib desert grass *Stipagrostis sabulicola* to

- irrigate itself with fog water. *Journal of The Royal Society Interface* 9, 1965–1974. doi:10.1098/rsif.2011.0847.
- Royé, D., Lorenzo, N., Rasilla, D., Martí, A., 2018. Spatio-temporal variations of cloud fraction based on circulation types in the Iberian Peninsula. *International Journal of Climatology* 39, 1716–1732. doi:10.1002/joc.5914.
- Sachweh, M., Koepke, P., 1995. Radiation fog and urban climate. *Geophysical Research Letters* 22, 1073–1076. doi:10.1029/95GL00907.
- Sachweh, M., Koepke, P., 1997. Fog dynamics in an urbanized area. *Theoretical and Applied Climatology* 58, 87–93. doi:10.1007/BF00867435.
- Samelson, R.M., de Szoeke, S.P., Skillingstad, E.D., Barbour, P.L., Durski, S.M., 2021. Fog and Low-Level Stratus in Coupled Ocean–Atmosphere Simulations of the Northern California Current System Upwelling Season. *Monthly Weather Review* 149, 1593–1617. doi:10.1175/MWR-D-20-0169.1.
- Savitzky, A., Golay, M.J., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36, 1627–1639. doi:10.1021/ac60214a047.
- Schemenauer, R.S., Cereceda, P., 1994. Fog collection's role in water planning for developing countries. *Natural Resources Forum* 18, 91–100. doi:10.1111/j.1477-8947.1994.tb00879.x.
- Scherrer, S.C., Appenzeller, C., 2014. Fog and low stratus over the Swiss Plateau - a climatological study. *International Journal of Climatology* 34, 678–686. doi:10.1002/joc.3714.
- Schultz, N.M., Lawrence, P.J., Lee, X., 2017. Global satellite data highlights the diurnal asymmetry of the surface temperature response to deforestation. *Journal of Geophysical Research: Biogeosciences* 122, 903–917. doi:10.1002/2016JG003653.
- Seneviratne, S.I., Lehner, I., Gurtz, J., Teuling, A.J., Lang, H., Moser, U., Grebner, D., Menzel, L., Schrott, K., Vitvar, T., Zappa, M., 2012. Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event. *Water Resources Research* 48. doi:10.1029/2011WR011749.
- Shrivastava, M., Cappa, C.D., Fan, J., Goldstein, A.H., Guenther, A.B., Jimenez, J.L., Kuang, C., Laskin, A., Martin, S.T., Ng, N.L., Petäjä, T., Pierce, J.R., Rasch, P.J.,

- Roldin, P., Seinfeld, J.H., Shilling, J., Smith, J.N., Thornton, J.A., Volkamer, R., Wang, J., Worsnop, D.R., Zaveri, R.A., Zelenyuk, A., Zhang, Q., 2017. Recent advances in understanding secondary organic aerosol: Implications for global climate forcing. *Reviews of Geophysics* 55, 509–559. doi:10.1002/2016RG000540.
- Smith, D.K.E., Renfrew, I.A., Price, J.D., Dorling, S.R., 2018. Numerical modelling of the evolution of the boundary layer during a radiation fog event. *Weather* 73, 310–316. doi:10.1002/wea.3305.
- Spracklen, D.V., Bonn, B., Carslaw, K.S., 2008. Boreal forests, aerosols and the impacts on clouds and climate. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366, 4613–4626. doi:10.1098/rsta.2008.0201.
- Steenefeld, G.J., de Bode, M., 2018. Unravelling the relative roles of physical processes in modelling the life cycle of a warm radiation fog. *Quarterly Journal of the Royal Meteorological Society* 144, 1539–1554. doi:10.1002/qj.3300.
- Stirnberg, R., Cermak, J., Fuchs, J., Andersen, H., 2020. Mapping and Understanding Patterns of Air Quality Using Satellite Data and Machine Learning. *Journal of Geophysical Research: Atmospheres* 125. doi:10.1029/2019JD031380.
- Stolaki, S., Haeffelin, M., Lac, C., Dupont, J.C., Elias, T., Masson, V., 2015. Influence of aerosols on the life cycle of a radiation fog event. A numerical and observational study. *Atmospheric Research* 151, 146–161. doi:10.1016/j.atmosres.2014.04.013.
- Stolaki, S.N., Kazadzis, S.A., Foris, D.V., Karacostas, T.S., 2009. Fog characteristics at the airport of Thessaloniki, Greece. *Natural Hazards and Earth System Sciences* 9, 1541–1549. doi:10.5194/nhess-9-1541-2009.
- 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. doi:10.1186/1471-2105-8-25.
- Suni, T., Guenther, A., Hansson, H., Kulmala, M., Andreae, M., Arneth, A., Artaxo, P., Blyth, E., Brus, M., Ganzeveld, L., Kabat, P., de Noblet-Ducoudré, N., Reichstein, M., Reissell, A., Rosenfeld, D., Seneviratne, S., 2015. The significance of land-atmosphere interactions in the Earth system-iLEAPS achievements and perspectives. *Anthropocene* 12, 69–84. doi:10.1016/j.ancene.2015.12.001.

- Teuling, A.J., Seneviratne, S.I., Stöckli, R., Reichstein, M., Moors, E., Ciais, P., Luysaert, S., Van Den Hurk, B., Ammann, C., Bernhofer, C., et al., 2010. Contrasting response of European forest and grassland energy exchange to heatwaves. *Nature Geoscience* 3, 722–727. doi:10.1038/ngeo950.
- Teuling, A.J., Taylor, C.M., Meirink, J.F., Melsen, L.A., Miralles, D.G., Van Heerwaarden, C.C., Vautard, R., Stegehuis, A.I., Nabuurs, G.J., de Arellano, J.V.G., 2017. Observational evidence for cloud cover enhancement over western European forests. *Nature communications* 8, 1–7. doi:10.1038/ncomms14065.
- Theeuwes, N.E., Barlow, J.F., Teuling, A.J., Grimmond, C.S.B., Kotthaus, S., 2019. Persistent cloud cover over mega-cities linked to surface heat release. *npj Climate and Atmospheric Science* 2, 1–6. doi:10.1038/s41612-019-0072-x.
- Toledo, F., Haeffelin, M., Wærsted, E., Dupont, J.C., 2021. A new conceptual model for adiabatic fog. *Atmospheric Chemistry and Physics* 21, 13099–13117. doi:10.5194/acp-21-13099-2021.
- Torregrosa, A., Combs, C., Peters, J., 2016. GOES-derived fog and low cloud indices for coastal north and central California ecological analyses. *Earth and Space Science* 3, 46–67. doi:10.1002/2015EA000119.
- Trigo, I.F., Dacamara, C.C., Viterbo, P., Roujean, J.L., Olesen, F., Barroso, C., de Coca, F.C., Carrer, D., Freitas, S.C., García-Haro, J., Geiger, B., Gellens-Meulenberghs, F., Ghilain, N., Meliá, J., Pessanha, L., Siljamo, N., Arboleda, A., 2011. The Satellite Application Facility for Land Surface Analysis. *International Journal of Remote Sensing* 32, 2725–2744. doi:10.1080/01431161003743199.
- Tunved, P., Hansson, H.C., Kerminen, V.M., Ström, J., Maso, M.D., Lihavainen, H., Viisanen, Y., Aalto, P.P., Komppula, M., Kulmala, M., 2006. High Natural Aerosol Loading over Boreal Forests. *Science* 312, 261–263. doi:10.1126/science.1123052.
- Turner, J., Allam, R., Maine, D., 1986. A case-study of the detection of fog at night using channels 3 and 4 on the Advanced Very High-Resolution Radiometer (AVHRR). *Meteorological Magazine* 115, 285–290.
- Twomey, S., 1977. The Influence of Pollution on the Shortwave Albedo of Clouds. *Journal of Atmospheric Sciences* 34, 1149–1152. doi:10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2.

- Underwood, S.J., Ellrod, G.P., Kuhnert, A.L., 2004. A Multiple-Case Analysis of Nocturnal Radiation-Fog Development in the Central Valley of California Utilizing the GOES Nighttime Fog Product. *Journal of Applied Meteorology* 43, 297–311. doi:10.1175/1520-0450(2004)043<0297:AMAONR>2.0.CO;2.
- Valiente, J.A., Estrela, M.J., Corell, D., Fuentes, D., Valdecantos, A., Baeza, M.J., 2011. Fog water collection and reforestation at a mountain location in a western mediterranean basin region: Air-mass origins and synoptic analysis. *Erdkunde* 65, 277–290.
- Vaughan, M.A., Powell, K.A., Winker, D.M., Hostetler, C.A., Kuehn, R.E., Hunt, W.H., Getzewich, B.J., Young, S.A., Liu, Z., McGill, M.J., 2009. Fully Automated Detection of Cloud and Aerosol Layers in the CALIPSO Lidar Measurements. *Journal of Atmospheric and Oceanic Technology* 26, 2034–2050. doi:10.1175/2009JTECHA1228.1.
- Vautard, R., Yiou, P., Van Oldenborgh, G.J., 2009. Decline of fog, mist and haze in Europe over the past 30 years. *Nature Geoscience* 2, 115–119. doi:10.1038/ngeo414.
- Wærsted, E.G., Haeffelin, M., Dupont, J.C., Delanoë, J., Dubuisson, P., 2017. Radiation in fog: quantification of the impact on fog liquid water based on ground-based remote sensing. *Atmospheric Chemistry and Physics* 17, 10811–10835. doi:10.5194/acp-17-10811-2017.
- Wærsted, E.G., Haeffelin, M., Steeneveld, G.J., Dupont, J.C., 2019. Understanding the dissipation of continental fog by analysing the LWP budget using idealized LES and in situ observations. *Quarterly Journal of the Royal Meteorological Society* 145, 784–804. doi:10.1002/qj.3465.
- Wang, J., Chagnon, F.J.F., Williams, E.R., Betts, A.K., Renno, N.O., Machado, L.A.T., Bisht, G., Knox, R., Bras, R.L., 2009. Impact of deforestation in the Amazon basin on cloud climatology. *Proceedings of the National Academy of Sciences* 106, 3670–3674. doi:10.1073/pnas.0810156106.
- Weathers, K.C., Ponette-González, A.G., Dawson, T.E., 2019. Medium, vector, and connector: Fog and the maintenance of ecosystems. *Ecosystems* 23, 217–229. doi:10.1007/s10021-019-00388-4.
- Williams, A.P., Schwartz, R.E., Iacobellis, S., Seager, R., Cook, B.I., Still, C.J., Husak, G., Michaelsen, J., 2015. Urbanization causes increased cloud base height and

- decreased fog in coastal Southern California. *Geophysical Research Letters* 42, 1527–1536. doi:10.1002/2015GL063266.
- Winkler, K., Fuchs, R., Rounsevell, M., Herold, M., 2021. Global land use changes are four times greater than previously estimated. *Nature communications* 12, 1–10. doi:10.1038/s41467-021-22702-2.
- WMO, 2018. Volume I – Measurement of Meteorological Variables, in: *Guide to Instruments and Methods of Observation*. World Meteorological Organization Geneva.
- Wrzesinsky, T., Klemm, O., 2000. Summertime fog chemistry at a mountainous site in central Europe. *Atmospheric Environment* 34, 1487–1496. doi:10.1016/S1352-2310(99)00348-9.
- Wu, W., Zhang, L., 2013. Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in north-eastern Puerto Rico. *Applied Geography* 37, 52–62. doi:10.1016/j.apgeog.2012.10.012.
- Xu, R., Li, Y., Teuling, A.J., Zhao, L., Spracklen, D.V., Garcia-Carreras, L., Meier, R., Chen, L., Zheng, Y., Lin, H., Fu, B., 2022. Contrasting impacts of forests on cloud cover based on satellite observations. *Nature Communications* 13, 670. doi:10.1038/s41467-022-28161-7.
- Yan, S., Zhu, B., Huang, Y., Zhu, J., Kang, H., Lu, C., Zhu, T., 2020. To what extents do urbanization and air pollution affect fog? *Atmospheric Chemistry and Physics* 20, 5559–5572. doi:10.5194/acp-20-5559-2020.
- Yan, S., Zhu, B., Zhu, T., Shi, C., Liu, D., Kang, H., Lu, W., Lu, C., 2021. The Effect of Aerosols on Fog Lifetime: Observational Evidence and Model Simulations. *Geophysical Research Letters* 48, e2020GL61803. doi:10.1029/2020GL091156.
- Ye, H., 2009. The influence of air temperature and atmospheric circulation on winter fog frequency over Northern Eurasia. *International Journal of Climatology* 29, 729–734. doi:10.1002/joc.1741.
- Zuidema, P.A., Babst, F., Groenendijk, P., Trouet, V., Abiyu, A., Acuña-Soto, R., Adenesky-Filho, E., Alfaro-Sánchez, R., Aragão, J.R.V., Assis-Pereira, G., Bai, X., Barbosa, A.C., Battipaglia, G., Beeckman, H., Botosso, P.C., Bradley, T., Bräuning, A., Brienen, R., Buckley, B.M., Camarero, J.J., Carvalho, A., Ceccantini, G., Centeno-Erguera, L.R., Cerano-Paredes, J., Chávez-Durán, Á.A., Cintra, B.B.L.,

Cleaveland, M.K., Couralet, C., D'Arrigo, R., del Valle, J.I., Dünisch, O., Enquist, B.J., Esemann-Quadros, K., Eshetu, Z., Fan, Z.X., Ferrero, M.E., Fichtler, E., Fontana, C., Francisco, K.S., Gebrekirstos, A., Gloor, E., Granato-Souza, D., Haneca, K., Harley, G.L., Heinrich, I., Helle, G., Inga, J.G., Islam, M., Jiang, Y.m., Kaib, M., Khamisi, Z.H., Koprowski, M., Kruijt, B., Layme, E., Leemans, R., Leffler, A.J., Lisi, C.S., Loader, N.J., Locosselli, G.M., Lopez, L., López-Hernández, M.I., Lousada, J.L.P.C., Mendivelso, H.A., Mokria, M., Montóia, V.R., Moors, E., Nabais, C., Ngoma, J., Nogueira Júnior, F.d.C., Oliveira, J.M., Olmedo, G.M., Pagotto, M.A., Panthi, S., Pérez-De-Lis, G., Pucha-Cofrep, D., Pumijumnong, N., Rahman, M., Ramirez, J.A., Requena-Rojas, E.J., Ribeiro, A.d.S., Robertson, I., Roig, F.A., Rubio-Camacho, E.A., Sass-Klaassen, U., Schöngart, J., Sheppard, P.R., Slotta, F., Speer, J.H., Therrell, M.D., Toirambe, B., Tomazello-Filho, M., Torbenson, M.C.A., Touchan, R., Venegas-González, A., Villalba, R., Villanueva-Diaz, J., Vinya, R., Vlam, M., Wils, T., Zhou, Z.K., 2022. Tropical tree growth driven by dry-season climate variability. *Nature Geoscience* 15, 269–276. doi:10.1038/s41561-022-00911-8.

Further appendices

Original publication: Pauli et al. (2020)

Pauli, E., Andersen, H., Bendix, J., Cermak, J., Egli, S., 2020. Determinants of fog and low stratus occurrence in continental central Europe - a quantitative satellite-based evaluation. *Journal of Hydrology* 591, 125451. doi:10.1016/j.jhydrol.2020.125451.

Contributions:

Eva Pauli conceived the study design, conducted the investigation and developed the methodology including coding. Eva Pauli visualized and interpreted the results, wrote the original draft and incorporated the reviewer comments.

Peer-Review:

The manuscript was submitted to the *Journal of Hydrology* on June 22nd, 2020. After a peer-review with two anonymous reviewers, the manuscript was accepted on August 21st, 2020 and published on August 30th, 2020.

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 1.2. The numbering of figures was changed and section titles were renamed.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Determinants of fog and low stratus occurrence in continental central Europe – a quantitative satellite-based evaluation

Eva Pauli^{a,b,*}, Hendrik Andersen^{a,b}, Jörg Bendix^c, Jan Cermak^{a,b}, Sebastian Egli^c^a Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany^b Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany^c Laboratory for Climatology and Remote Sensing, Philipps-University of Marburg, Germany

ARTICLE INFO

This manuscript was handled by marco borga,
Editor-in-Chief

Keywords:

Fog
Low stratus
Machine learning
Land surface
Atmosphere-land surface interactions
Europe

ABSTRACT

The formation and development of fog and low stratus clouds (FLS) depend on meteorological and land surface conditions and their interactions with each other. While analyses of temporal and spatial patterns of FLS in Europe exist, the interactions between FLS determinants underlying them have not been studied explicitly and quantitatively at a continental scale yet. In this study, a state-of-the-art machine learning technique is applied to model FLS occurrence over continental Europe, using meteorological and land surface parameters from geostationary satellite and reanalysis data. Spatially explicit model units are created to test for spatial and seasonal differences in model performance and FLS sensitivities to changes in predictors, and effects of different data preprocessing procedures are evaluated. The statistical models show good performance in predicting FLS occurrence during validation, with $R^2 > 0.9$ especially in winter high pressure situations. The predictive skill of the models seems to be dependent on data availability, data preprocessing, time period, and geographic characteristics. It is shown that atmospheric proxies are more important determinants of FLS presence than surface characteristics, in particular mean sea level pressure, near-surface wind speed and evapotranspiration are crucial, together with FLS occurrence on the previous day. Higher wind speeds, higher land surface temperatures and higher evapotranspiration tend to be negatively related to FLS. Spatial patterns of feature importance show the dominant influence of mean sea level pressure on FLS occurrence throughout the central European domain. When only high pressure situations are considered, wind speed (in the western study region) and evapotranspiration (in the eastern study region) gain importance, highlighting the influence of moisture advection on FLS occurrence in the western parts of the central European domain. This study shows that FLS occurrence can be accurately modeled using machine learning techniques in large spatial domains based on meteorological and land surface predictors. The statistical models used in this study provide a novel analysis tool for investigating empirical relationships in the FLS – land surface system and possibly infer processes.

1. Introduction

Fog influences several anthropogenic and natural systems: it affects traffic at land, sea and in the air (Leigh et al., 1998; Pagowski et al., 2004), it plays an important role in climate processes (Vautard et al., 2009; Egli et al., 2017) and is often vital for ecological systems since it can supply water and nutrients to ecosystems (Bruijnzeel et al., 2006; Gottlieb et al., 2019; Mitchell et al., 2020). While fog is generally defined as a suspension of water droplets leading to visibilities <1000 m (Glickman, 2000), it can be further classified based on its formation mechanisms. In central Europe, radiation fog is the most frequent fog type (Fuzzi et al., 1992; Gulpepe et al., 2007) and especially prevalent during winter (Cermak et al., 2009; Egli et al., 2017). Radiation fog

forms due to radiative cooling at the surface or at the near-surface layer of the atmosphere (Price, 2019) under clear skies typically encountered during anticyclonic conditions (Gulpepe et al., 2007). The formation of radiation fog is positively influenced by low temperatures, low wind speeds and a stable boundary layer (Cuxart and Jiménez, 2012; Pérez-Díaz et al., 2017; Price, 2019). Fog is a particular case of a low stratus cloud positioned directly at the Earth surface. From the satellite perspective, (radiation) fog and other low stratus clouds are frequently treated together as a single category (fog and low stratus: FLS) (Cermak and Bendix, 2011).

In addition to the contribution of atmospheric processes (van Oldenborgh et al., 2010), FLS occurrence is also heavily linked to surface-atmosphere exchanges (Bergot et al., 2005). These exchanges are

* Corresponding author.

E-mail address: eva.pauli@kit.edu (E. Pauli).<https://doi.org/10.1016/j.jhydrol.2020.125451>

Received in revised form 7 August 2020; Accepted 21 August 2020

Available online 30 August 2020

0022-1694/ © 2020 Elsevier B.V. All rights reserved.

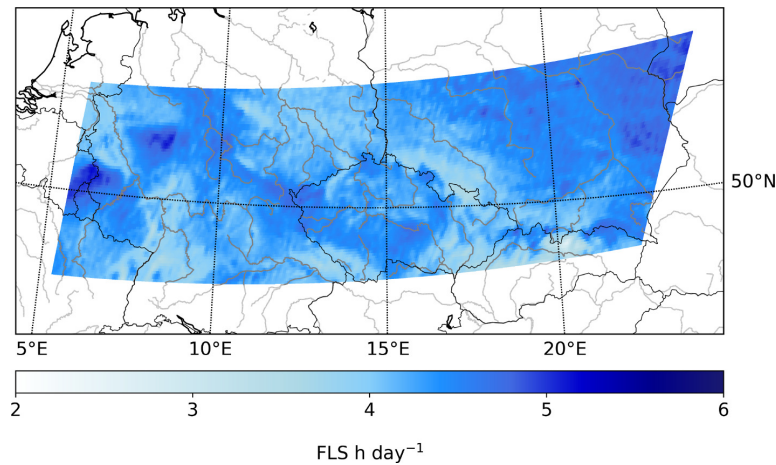


Fig. 1. Average fog and low stratus hours per day from 2006–2015 in the central European study area based on the product by Egli et al. (2017).

classified here as two pathways of opposite directions: 1) Fog and low stratus influence on the land surface, especially vegetation and 2) land surface influences on the formation, properties and distribution of FLS.

Pathway 1) is mainly expressed via the supply of water through fog and by influencing the amount of nutrients reaching the vegetation. This has been investigated in various ecosystems around the world such as coastal deserts (Eckardt et al., 2013; Ebner et al., 2011; Roth-Nebelsick et al., 2012; Mitchell et al., 2020; Lehnert et al., 2018), tropical montane cloud forests (Berry et al., 2015) and in the Californian redwood forests (Dawson, 1998). Furthermore, fog can influence the productivity in cultivated crops by modifying radiation, e.g. leading to a higher water use efficiency in strawberries on foggy days (Baguskas et al., 2018). For this study, pathway 2), the influences of the land surface on FLS formation, properties and distribution is in the focus. Generally, topography and landform can modify the influence of atmospheric dynamics on FLS occurrence (van Oldenborgh et al., 2010). Advected FLS can accumulate in steep windward slopes (Ball and Tzanopoulos, 2020) and lower advected fog densities are found in leeward locations (Ball and Tzanopoulos, 2020; Torregrosa et al., 2016). The formation and persistence of radiation fog can also be influenced by topographically induced drainage flows and turbulence (Price, 2019). Furthermore, the growing phase of radiation fog is controlled by advection slightly above the fog layer, which is strongly linked to the topography (Cuxart and Jiménez, 2012). Besides topography, turbulent mixing, soil temperature and soil moisture fluxes influence fog formation and dissipation (Maronga and Bosveld, 2017). Low wind speeds are necessary to generate turbulence for FLS development (Klemm and Wrzesinsky, 2007) whereas modestly stronger wind above the nocturnal boundary layer can inhibit the development of radiation fog (Bergot and Lestringant, 2019). Furthermore, wind direction can also influence FLS occurrence but is location-specific (Blas et al., 2002; Klemm and Wrzesinsky, 2007; Wrzesinsky and Klemm, 2000). FLS occurrence is also strongly impacted by the availability of aerosols that can act as cloud condensation nuclei (CCN) (Ramanathan et al., 2001). Biovolatile organic compounds (BVOC's), emitted by vegetation can form secondary organic aerosols (SOA) and act as CCN and thereby favor cloud formation (Pöschl et al., 2010; Shrivastava et al., 2017). Improvements in air quality have been observed to lead to a reduction in fog (Klemm and Lin, 2016; Gray et al., 2019). Furthermore the land cover type influences FLS, with lower FLS frequency over urban areas compared to rural areas, e.g. in Milan (Bendix, 1994), in Munich (Sachweh and Koepke, 1995; Sachweh and Koepke, 1997), in Delhi (Gautam and Singh, 2018), in multiple cities in the Netherlands

(Izett et al., 2019) or California (Williams et al., 2015), likely in parts due to the urban heat island effect.

The general patterns of FLS occurrence in Europe have been investigated using ground-based observations and satellite data in the last few years. Egli et al. (2017) found FLS occurrence to be highest in winter with higher FLS occurrence in the northeast of Europe than in the Mediterranean region. Using a hybrid approach with ground truth data and data from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) on board Meteosat Second Generation (MSG), Egli et al. (2018) derived a 10-year climatology of ground fog showing high fog occurrences in mountainous areas in spring and summer and high values in lowlands and plains in winter (Egli et al., 2018). Cermak (2018) investigated fog and low cloud frequency using LiDAR (Light detection and ranging) aboard CALIPSO (Cloud-Aerosol LiDAR and Pathfinder Satellite Observations) data and found similar patterns, with a general decrease of FLS frequency from north to south and high values in the Baltic sea regions.

FLS processes take place across a large range of different scales, from small scales of aerosol activation (10^{-7} m) to synoptic scales (10^6 m). While investigations about temporal and spatial FLS patterns over Europe exist, as well as many insights on specific local and non-local processes determining FLS formation and development, no explicit satellite-based investigation of the determinants of FLS development spanning multiple years and a large spatially coherent region exists to date. The aim of this study is to quantitatively determine the relative importance of several land-cover and meteorological parameters for FLS occurrence in a large central European domain. This is done by using a machine learning technique, investigating its predictive performance concerning FLS occurrence as well as model sensitivities. The high occurrence frequencies of FLS especially in central Europe during winter provide a good database in conducting this analysis. In this paper, we would like to unravel the question what the main drivers of large-scale spatial and temporal FLS patterns in central Europe are, and how these drivers influence FLS patterns.

2. Data and methods

2.1. Study area

The study is conducted in a large domain in continental Europe, from 48° to 53°N and 5° to 15°E including parts of the countries Austria, Belarus, Belgium, Czech Republic, France, Germany, Luxembourg, Netherlands, Poland, Slovakia and Ukraine (see Fig. 1). This continental

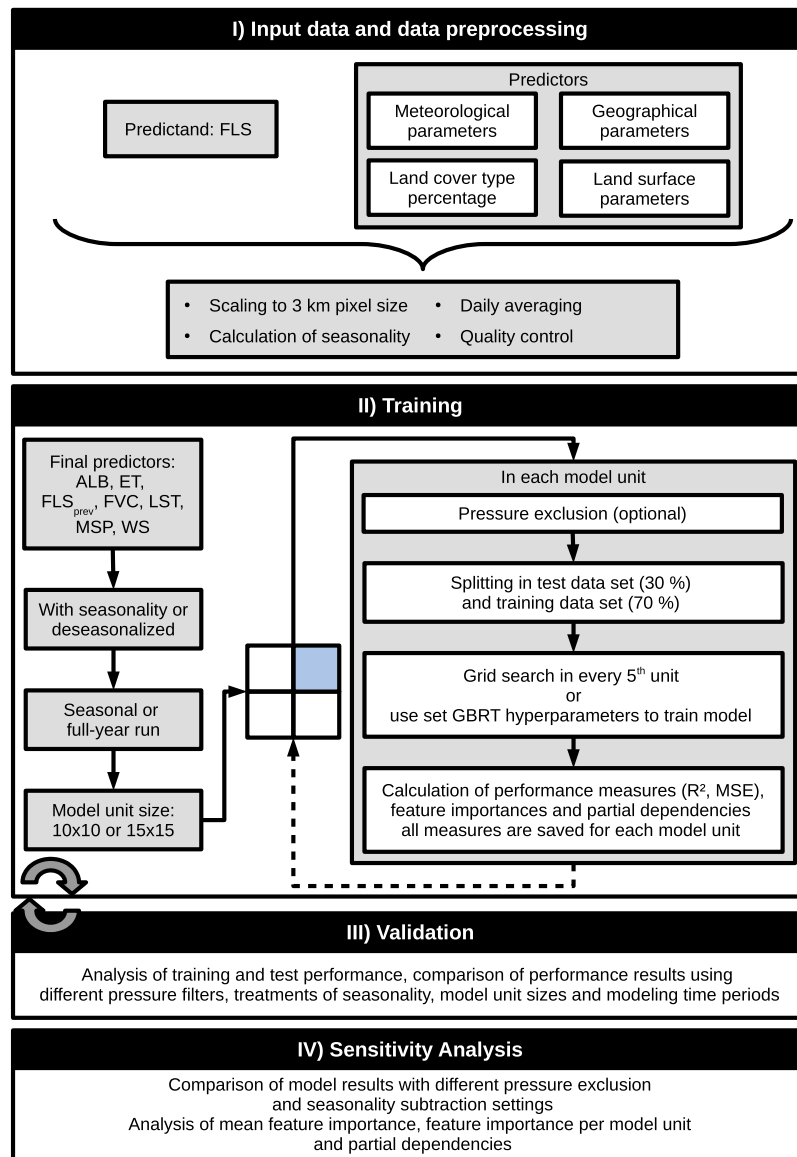


Fig. 2. Schematic of the model workflow. Step II and III are conducted multiple times for different settings concerning pressure filters, treatments of seasonality, model unit sizes and modeling time periods.

study area was chosen to allow for a focus on continental situations unimpacted by the effects of local circulations such as land-sea and mountain breeze systems. Thus, large mountain ranges such as the Alps and areas close to the coast were excluded. Nonetheless, small local topographic differences and smaller topography-induced circulations still exist in the presented study area. The effects of such local-scale modulations on FLS occurrence and FLS – land surface interactions are reduced by subdividing the study area into a number of smaller units. In this study, two model unit sizes are tested (10×10 and 15×15 SEVIRI pixels), which are further described below.

2.2. Data

The FLS product used in this study presented by Egli et al. (2017) is based on geostationary satellite data from the Meteosat Second Generation platforms Meteosat 8, 9 and 10. The SEVIRI system on board of the satellites scans the full hemisphere every 15 min. The sub-satellite resolution is 3 km in 11 spectral bands with an additional high-resolution visible channel with 1 km resolution. The FLS product is a modification of the Satellite-based Operational Fog Observation Scheme (SOFOS) by Cermak (2006) and covers the years 2006 to 2015. To produce a data set of daily FLS occurrence, the original 15-min product is locally averaged into FLS hours day^{-1} , which serves as the target quantity in this study. The mean FLS distribution for the chosen

study area over the complete time period can be seen in Fig. 1.

To describe the physical state of the land surface and its possible influence on FLS distribution, several land surface features are included in this study, based on products from the EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA-SAF) (Trigo et al., 2011). All of these products are created using data from Meteosat SEVIRI and thus have the same spatial resolution as the FLS product. To describe temperature and moisture availability of the surface, land surface temperature (LST) and evapotranspiration (ET) are included. Additional land surface characteristics relevant for energy and moisture fluxes are described by the fraction of vegetation cover (FVC) and the bi-hemispherical albedo for the total shortwave range (ALB).

To describe atmospheric conditions, ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) is used. ERA5 is the follow-up of ERA-interim (Dee et al., 2011) and provides higher spatial resolution (0.25°) (Hersbach, 2016) which is especially useful for investigating land-atmosphere interactions. In this study, mean sea level pressure (MSP) and the u and v component of wind at 10 meters are included. The u and v component of the wind are combined to calculate the near surface wind speed (WS). As a measure for the persistence of meteorological conditions that can lead to FLS, the FLS value of the previous day (FLS_{prev}) is included.

Additional information on the height above sea level, geographical position or land cover type were tested but are not included in the final model, as they did not lead to marked improvements in model performance and, equally to a location ID, could lead to overfitting (Meyer et al., 2018; Meyer et al., 2019). Meyer et al. (2019) further assume that using predictor variables that describe the spatial location prevent the model from making reliable spatial predictions. As such, the final models are based on 7 predictors (ALB, ET, FLS_{prev}, FVC, LST, MSP, WS) to predict the duration of FLS cover.

2.3. Methods

Interactions between the atmosphere and the land surface are often highly dimensional. With machine learning, non-linear and complex multivariate problems can be handled (Lary et al., 2016; Olden et al., 2008). Especially for problems where theoretical knowledge is limited but the amount of available observational data is large, machine learning can be an ideal tool for addressing these problems (Lary et al., 2016). In this study, a model was constructed using the machine learning technique Gradient Boosting Regression Trees (GBRT), aiming at predicting FLS hours per day on the basis of the meteorological and surface predictors described above. GBRTs use an ensemble of weak base learners (decision trees) which follow the negative gradient of the loss function (Friedman, 2001; Natekin and Knoll, 2013). They are flexible when using categorical and numerical data and have large predictive power, being able to represent nonlinear relationships between predictors and the predictand (Hastie et al., 2001). Furthermore, one of the key strengths of this machine learning technique is its high level of interpretability, since a number of interpretation methods exist for GBRTs (Elith et al., 2008). These properties have been exploited in recent studies to analyze complex atmospheric systems with GBRTs (Fuchs et al., 2018; Stirnberg et al., 2020). For this study, the GBRT implementation of the scikit-learn library in python was used (Pedregosa et al., 2011).

2.3.1. Data preprocessing

In a first step (Fig. 2I), the ERA5 data sets are rescaled to the SEVIRI pixel resolution of 3×3 km at nadir of the FLS data set using a nearest-neighbor approach. As described above, wind speed at 10 m height is calculated using the u and v components of the wind at 10 m height. Furthermore, daily averages are calculated for FLS, LST, ET, MSP and WS. FVC and ALB are already in the form of daily means. Daily averages are used, as the focus lies on the day-to-day variability of FLS occurrence. Poor quality data and data with high error values, e.g. due to

adjacent cloud-covered areas is excluded using the following data quality control steps. For FVC, only data that is free of snow and water and with reliable input ranges for the FVC algorithm is included as well as data with an absolute error smaller than 0.2. For ALB, data with an absolute error greater than 0.2 is excluded. For ET, data where input variables for the ET algorithm have insufficient quality or are missing, is excluded. For LST, cloud-filled pixels or pixels partly contaminated by clouds, snow or ice are excluded. Since some cloud-covered or cloud adjacent pixels with either implausible high or low surface temperatures are not caught by the quality flag, they are removed by setting the valid LST range depending on the month. For the full year, the valid data range is set to -60 °C to $+60$ °C. This range is then adjusted depending on the time of year: the maximum is adjusted to $+40$ °C from October to March and the minimum to -50 °C from March to September. With this, plausible LST data at the upper or lower ends of the temperature range can be kept in each season. The data excluded in the quality control as described above are marked as invalid in each data set. Any such flagging leads to the exclusion of the data point, as the model only considers data points with valid data in all data sets.

Since seasonal influence on FLS as well as on numerous other parameters is large, the models are run with two seasonality settings: 1) The seasonality is kept in the data and 2) the seasonal cycle is subtracted from the data. These two seasonality settings are used to investigate to what extent model performance and sensitivities are dependent on seasonal effects. When removing the seasonality, only the day-to-day variability is modeled, whereas keeping the seasonality in the data also models the seasonal interactions of FLS and its predictors. Here, the seasonality is defined as the mean over the investigated time period for every pixel for every day of the year and then smoothing this time series using a Savitzky-Golay filter (Savitzky and Golay, 1964). The seasonality is then subtracted from the data sets of each specific feature.

2.3.2. Model construction

The study area is subdivided into spatially explicit model units, in which data is treated summarily at the SEVIRI resolution for training, testing and sensitivity analyses (compare Fig. 2II). Two spatial sizes of the model units are tested, with model units containing either 10×10 or 15×15 SEVIRI pixels. Due to the high latitudinal position and the rather shallow satellite viewing angle of 45° in the study area, one SEVIRI pixel corresponds to about a width of 4 km and a height of 7 km (EUMETSAT, 2013). For the 10×10 model unit, this leads to a mean model unit area of roughly 2800 km² and 6300 km² for the 15×15 unit respectively. This model unit set-up is chosen to account for regional differences in the study area, while still choosing a unit size where some variability in land cover is present. Furthermore this set-up allows for the evaluation of spatial differences in model performance and sensitivities. Thus, the 15×15 pixel units contain more data for training and testing, whereas the 10×10 units provide smaller and thus more homogeneous model units.

To account for seasonal differences, seasonal model runs are conducted as well as full-year runs. These seasonal runs are winter (December, January and February: DJF), spring (March, April and May: MAM), summer (June, July and August: JJA) and fall (September, October and November: SON). In the considered time periods (full-year or seasons), separate models are trained using either all data or only the 20% highest pressure situations in each model unit to focus on situations where FLS is favored (see Section 1). The pressure threshold is calculated from the distribution of pressure values for every model unit and model time period separately.

In each model unit, data is split into separate training and test data sets, containing 70% and 30% of the data respectively. To ensure that training and test data contain comparable FLS distributions, the test-training split is coupled with a stratification of the FLS data set. In general, the number of available data points varies from model unit to model unit e.g. due to cloud contamination e.g. for LST and FVC data,

Table 1
Parameter grid for the grid search and final hyperparameters used for model training.

	Number of estimators	Learning rate	Maximum depth	Minimum samples per leaf
parameter grid	500, 1000, 1500, 2000, 2500, 3000	0.1, 0.05, 0.03, 0.02, 0.01	2, 3	10, 14, 18
final settings	3000	0.1	3	10

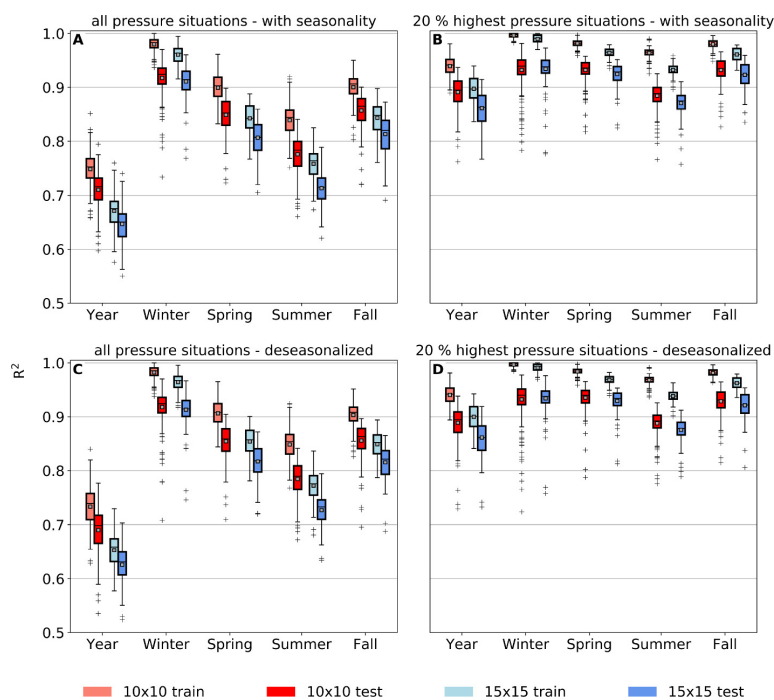


Fig. 3. R^2 of training and test set over all grid sizes and seasons using either all pressure situations (left) or only high pressure situations (right). The top row shows the results using the data with seasonality, while the bottom row displays the model results using deseasonalized data.

but can be higher than 400,000 (full-year, all pressure situations, 15×15 pixel model unit size), but also as low as 3800 (seasonal model, with pressure filter in a 10×10 pixel model unit).

Considering different seasonality and pressure filters, model unit sizes and modeling time periods (seasons or full-year), 30 different model settings are investigated in this study.

2.3.3. Hyperparameter tuning

Hyperparameter tuning is conducted using grid search with a small range of possible values to avoid overfitting, e.g. by using only low values for the maximum depth (Table 1). Since grid search is computationally expensive, it is conducted in every 5th model unit over a range of feature combinations and model settings to investigate the influence on the hyperparameters. The hyperparameters proposed in the grid search are similar over model units, feature combinations and model settings and the most common hyperparameters are used as the final settings.

2.3.4. Validation and sensitivity analysis

To test the influence of different model settings on model performance and sensitivities, the validation (Fig. 2III) and sensitivity analysis (Fig. 2IV) are conducted for all model settings described above. Model performance is evaluated using the coefficient of determination (R^2) and the mean squared error (MSE) between observed and predicted FLS in each model unit. The validation results are then evaluated

concerning the differences between settings for each unit separately and over all units. Potential overfitting is analyzed by comparing R^2 and MSE for the test and training data sets. Correlations between variables are quantified using the Spearman's rho coefficient to account for effects of outliers and non-linearity in the investigated correlation.

To investigate the most important features for model performance, the permutation feature importance is calculated for all model units. The permutation feature importance measures the increase in prediction error after permuting (randomly shuffling) one feature (Breiman, 2001). This breaks the relationship between the feature and the target, thus indicating how much the model depends on that feature (Strobl et al., 2007; Molnar, 2019). Compared to the often used impurity based feature importance (gini importance) the permutation importance is not as strongly biased towards variables with many categories and the continuous variable (Strobl et al., 2007). Thus the term feature importance relates to permutation importance in this paper. The mean feature importance is calculated over all model units but is also investigated for each model unit separately. To improve comparability, the permutation feature importance is scaled so the sum of permutation importance over all features equals 1.

While the feature importance only displays the relative importance of a feature, the sign and nonlinearity of the predictand response relative to changes of each predictor variable can be analyzed using the partial dependence. In the context of this paper, the partial dependencies quantify the average change of the predicted FLS values

relative to either one or two features (one-variable partial dependence and two-variable partial dependence) while accounting for the average effects of all other variables (Elith et al., 2008; Friedman, 2001; Fuchs et al., 2018). The partial dependence is calculated by gridding the investigated feature and calculating the corresponding average FLS prediction while the complement features are varied over their distribution (Molnar, 2019).

3. Results

3.1. Model performance

As a first overview of model performance, GBRT performance during training and validation with independent data is analyzed for all different model unit sizes as well as seasonality and pressure settings. The results of this validation are shown in Fig. 3. R^2 values are generally higher when filtering for high pressure situations, but are similar for runs containing seasonality versus runs that use deseasonalized data. The mean R^2 values of the model units in the 15×15 runs tend to be lower than in the 10×10 runs. The highest mean R^2 (averaged over all model units) is found in the pressure-filtered, deseasonalized 10×10 spring run (0.94), but is similar high (~ 0.93) in other runs (e.g. in the pressure-filtered 15×15 winter run with seasonality). A clear seasonal pattern in model performance exists, with skill generally lowest in summer and highest in winter, albeit this seasonal pattern is weakened when only high pressure situations are considered. The difference in R^2 between the model runs with all pressure values (Fig. 3A and C) versus those considering only high pressure situations (Fig. 3B and D) is therefore lowest in winter. The full-year models feature the lowest R^2 of runs without a pressure filter (mean R^2 0.6–0.7), but also reach values ~ 0.9 in the pressure-filtered runs. The regional variability of model performance, expressed as the spread in R^2 values, is highest in winter and for the full-year run, and lowest in spring and fall. The training R^2 is significantly ($p < 0.01$) higher than the test R^2 indicating some level of overfitting in the statistical models. Depending on the season, the training R^2 is on average slightly (4–8%) higher over all model units, and smaller than that in the 15×15 and full-year models. Still the highest test R^2 shows the ability of the models to generalize relationships between the predictors and FLS patterns learned during training. The difference between test and train R^2 is higher in the 10×10 model units and lowest in the full-year runs. The model predictions do not feature a noticeable bias (mean bias = 0.0, median bias ~ -0.3). MSE patterns are similar to R^2 , with higher MSE values (thus worse performance) in the 15×15 runs and a larger difference of MSE between seasons than between model runs with different seasonality subtraction settings (Fig. 10). This means that model performance is more dependent on the modeled season than on seasonality subtraction settings. As the MSE is dependent on the absolute values, MSE patterns are influenced by the seasonal FLS distribution, thus leading to high MSE values in winter and fall, and lower MSE in spring and summer.

To investigate the reasons for overfitting, the difference between training and test R^2 and MSE is correlated with the number of data points available for training and testing. For the correlation with the absolute difference between training and test R^2 the Spearman's rho is approximately -0.5 ($p < 0.01$) especially in the full-year and winter run. For the MSE correlation, the Spearman's rho is -0.5 ($p < 0.01$) in the full-year, spring and fall runs. Thus, overfitting is partly controlled by the number of available data points, and models generalize better when more data points are available for training and testing.

Due to the good overall performance in the 10×10 model units and the higher spatial resolution, only the results of the 10×10 model units are presented in the following sections. Where not specifically pointed out, results of the 15×15 model units are similar, though.

Fig. 4 shows spatial patterns of R^2 relative to the domain average R^2 for two different model settings (full-year models including seasonality with (Fig. 4A) and without (Fig. 4B) pressure filter). This highlights

model units that either have distinctly better or worse performance than the overall mean. Additionally, in Fig. 4C a topographic map is shown for the study area, together with two scatterplots of the R^2 deviation with the mean height per model unit (Fig. 4D) for values west and east of 10°E separately. The corresponding Spearman's rho, p-value and R^2 for the complete correlations (negative and positive R^2 deviation values together) are added in the top right corner of the Fig. 4D plot.

The domain average R^2 for the model run including all pressure situations is 0.71, for the model run using only high pressure situations it is 0.89. While the spatial patterns of skill are similar for both model settings, the models using only high pressure situations feature a lower deviation from the domain average R^2 . Generally, in the north-eastern part of the model region, e.g. in Poland and the central Czech Republic and the south-eastern and north-eastern parts of Germany, the R^2 in the model units shows a positive deviation from the domain average. In the western part of the model region (western Germany), as well as in parts of the south western part of the model region (northern parts of the Vosges, the upper Rhine valley, Black Forest) the R^2 is lower than the domain average. This is also the case for the mountain ranges surrounding the Czech Republic as well as in the south eastern part of the model region (High Tatras). The relationship between R^2 deviation and topography is visible in a) the visual relationship between the two (e.g. compare Fig. 4A and C) and b) the correlation of the two (Fig. 4D). With a Spearman's rho of -0.36 the latter is especially strong for values east of 10°E . A correlation of A and B with the distance to the coastline produced a Spearman's rho close to 0 with a p-value > 0.05 . Small differences between pressure settings are visible in the north eastern and south western part of the model region, as well as the border triangle between Germany, Poland and the Czech Republic at 15°E . The apparent spatial patterns of skill are similar during all seasons, and independent of skill measure (R^2 or MSE) and training or test data set.

3.2. Feature importance

In Fig. 5 the domain-average feature importance is shown for all seasons and specific model settings as stacked barplots. The left-hand panel (Fig. 5A) shows results for the model run that contains seasonality, while the right-hand panel shows the results of the deseasonalized model (Fig. 5B). In these subplots, the darker-colored bars on the left display the results of the model runs containing all pressure situations, while the lighter-colored bars on the right show the results of the model using only high pressure situations. Since results in the 15×15 model unit runs are similar, only 10×10 model unit results are shown and described here.

In general, clear similarities but also some distinct differences in mean feature importance are apparent: MSP, FLS_{prev}, WS and ET are among the most important features in all model runs. During winter and summer, MSP and WS are the most important features, while in the spring, fall and full-year run, ET and FLS_{prev} gain importance. When filtering for high pressure situations, ET becomes more important while the importance of MSP decreases. This is especially apparent in the spring, fall and full-year runs. LST reaches its highest feature importance in summer, but overall has little importance for model performance. ALB and FVC have little importance in all model runs however, this is slightly increased when filtering for high pressure situations in summer.

Differences between the model runs containing seasonality and the deseasonalized model runs are present in the exact values of feature importance but general patterns are similar. The largest differences in feature importance are apparent between the different seasons and the pressure exclusion runs. Model unit size and subtraction of seasonality only produce small differences in the mean permutation feature importance over all model units.

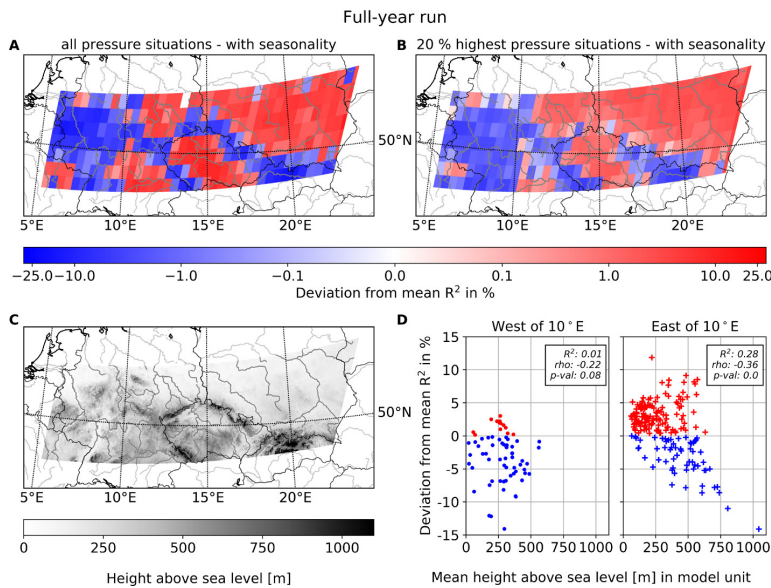


Fig. 4. Map of deviation from the mean R^2 over all model units (domain average) in %, a topographic map (based on Hijmans et al. (2005)) of the study area and a correlation of these two. The left upper panel (A) shows the results from the model run using all pressure situations, the right upper panel (B) shows the results from the high pressure model run. Both include the seasonality. The domain average R^2 is computed for all model settings separately. For better visibility the colormap is scaled logarithmic. In (C) the topographic map for the study area is shown. In (D) the results from (A) are correlated with the mean height per model unit for model units values west and east of $10^\circ E$ separately. The corresponding Spearman's rho, p-value and R^2 of the correlation of the combined negative (blue) and positive (red) R^2 deviation values are plotted in each subplot. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Spatial patterns of relevant features

In Fig. 6 the most important feature per model unit of the full-year run is shown for four different model settings. In general, the spatial patterns of the most important feature show clear differences between the model runs using all pressure situations (Fig. 6A and C) versus those using the 20% highest pressure values (Fig. 6B and D). When all pressure situations are used, MSP is the most important feature in most model units. This pattern is especially prevalent in the deseasonalized model run. Here, MSP is the most important feature in all model units except for some units in the eastern part of the study area, where ET is the most important feature. In the model run that contains seasonality (Fig. 6A), ET is the most important feature in numerous model units east of $10^\circ E$, in parts of eastern Germany, most of the Czech Republic and parts of Poland. In three model units in the Czech Republic, FLS_{prev} is the most important feature.

As expected, in the high pressure filtered model runs MSP is much less important. Instead, ET is the most important feature in large parts

of the eastern study area. Other than ET, MSP, FLS_{prev} and WS are the most important features in numerous model units across both seasonality settings (Fig. 6B and D). MSP is the dominant feature in most parts of western Germany and eastern France and in some units in the Czech Republic and Poland. FLS_{prev} is dominant in the north western part of the Czech Republic, western Germany as well as parts of France and Poland. WS is the most important feature in some scattered model units of the model run including seasonality. In the deseasonalized model run, WS is dominant in more model units mostly in the western and the north western part of the study area. The seasonal patterns of the most important feature (Fig. 7) are similar to the patterns observed in the full-year run. In the model runs using all pressure values (Fig. 7A), MSP is dominant over all seasons, with only some exceptions in winter and fall. In high pressure situations, though, seasonal differences are more pronounced (Fig. 7B): in winter, FLS_{prev} is dominant in the western part of the study area (Germany and parts of France) and WS is dominant in the eastern part of the study area. MSP is the most important feature in between WS and FLS_{prev} in central and north western parts of the study

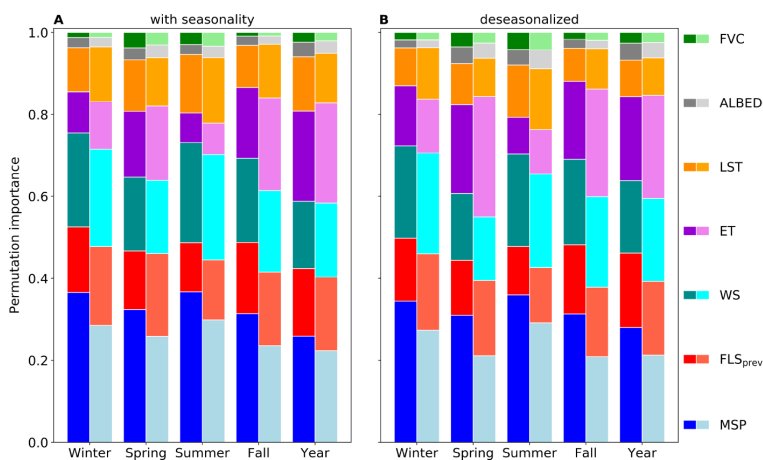


Fig. 5. Mean 10×10 feature importance over all model units for all features and all seasons. The darker, left bars display the feature importance of the model run with all pressure situations, the right, brighter bars display the feature importance of the model run with high pressure situations. The left subplot (A) shows the feature importance from the model run including the seasonality, the right subplot (B) the one of the deseasonalized model run.

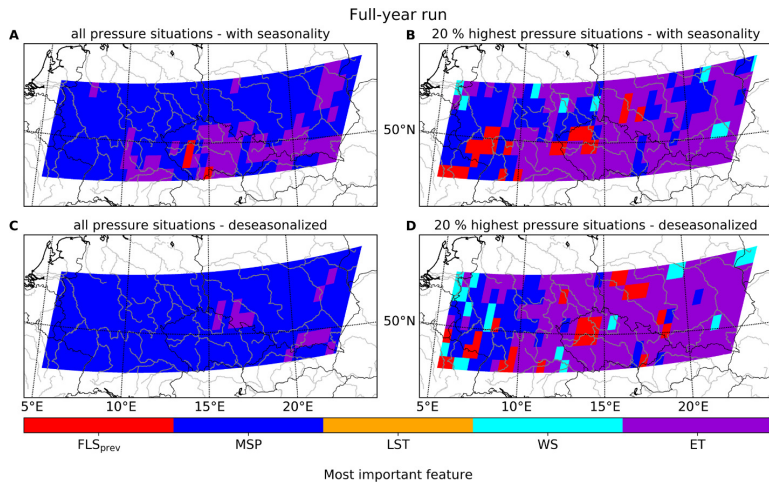


Fig. 6. Most important feature per 10×10 model unit in the full-year run over all model settings. The left column shows the model runs including all pressure situations (A and C) while the right column shows the model runs using only the 20% highest pressure situations (B and D). In the model runs of the top row, seasonality is included, in the bottom row it is subtracted.

area. Other than these three, ET and LST are the most important features in some scattered model units as well. In spring, the most important features are FLS_{prev} in south western and ET in north eastern parts of the study area. Other than that, MSP is dominant in most of the study region, with some scattered model units where LST and WS are the most dominant features. In summer, MSP is the most important feature in most model units, with WS being dominant on the border of Germany and France, in Poland, in the Czech Republic and in some scattered model units in central Germany. Apart from that only a few model units are dominated by FLS_{prev} and LST. In fall, MSP is dominant in the western part of the study area and ET is dominant in the eastern part of the study area (east of $10^\circ E$). WS is dominant in parts of eastern Germany and Poland and FLS_{prev} is the most important feature in some scattered model units in Germany, Poland and the Czech Republic.

3.4. One-variable partial dependence

In Fig. 8 the predictand responses relative to changes in all predictors are displayed as partial dependencies (see Section 2.3.4) for all considered time periods. Since the partial dependencies are calculated for each model unit separately, the mean and the interquartile ranges of the partial dependencies of all model units are presented here. Additionally, the distribution of values of each predictor is plotted below the corresponding partial dependence as a qualitative assessment of the

representativeness of the obtained partial dependencies. The partial dependencies presented here are from the 10×10 model unit run where seasonality and all pressure values are included, since the partial dependencies are similar over model unit sizes, seasonality and pressure-exclusion settings. The effect of the high pressure filter on the partial dependence of MSP is shown in the bottom right panel (all other partial dependencies of this run are similar to the ones shown). The partial dependence can be understood as the deviation from the mean of the predictand (here FLS occurrence) that can be attributed to a specific predictor. Thus, a positive partial dependence at a given feature value indicates a positive influence on modeled FLS occurrence. Similarly, a negative partial dependence at a given feature value indicates a negative influence on modeled FLS occurrence. The magnitude of the partial dependence determines the magnitude of the influence. A partial dependence close to zero indicates very little or no influence on modeled FLS occurrence at the given feature value.

As a sensitivity measure, the range of the mean partial dependence over all model units for the full-year run is given in the text, together with the corresponding slope direction of the partial dependence curve. This provides a measure of influence on modeled FLS occurrence for each feature over the corresponding value distribution with a greater sensitivity measure indicating a large positive or negative influence on modeled FLS occurrence. Seasonal differences of sensitivity are shortly described.

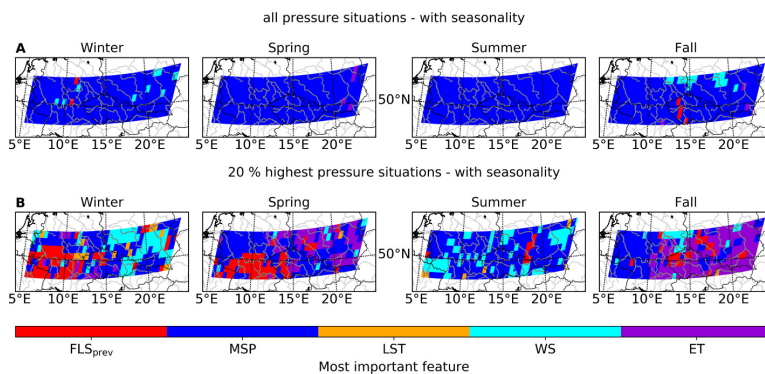


Fig. 7. Most important feature per 10×10 model unit for different seasons. The top row shows the model run using all pressure situations while the bottom row shows high pressure situations. Both model runs use the data that includes the seasonality.

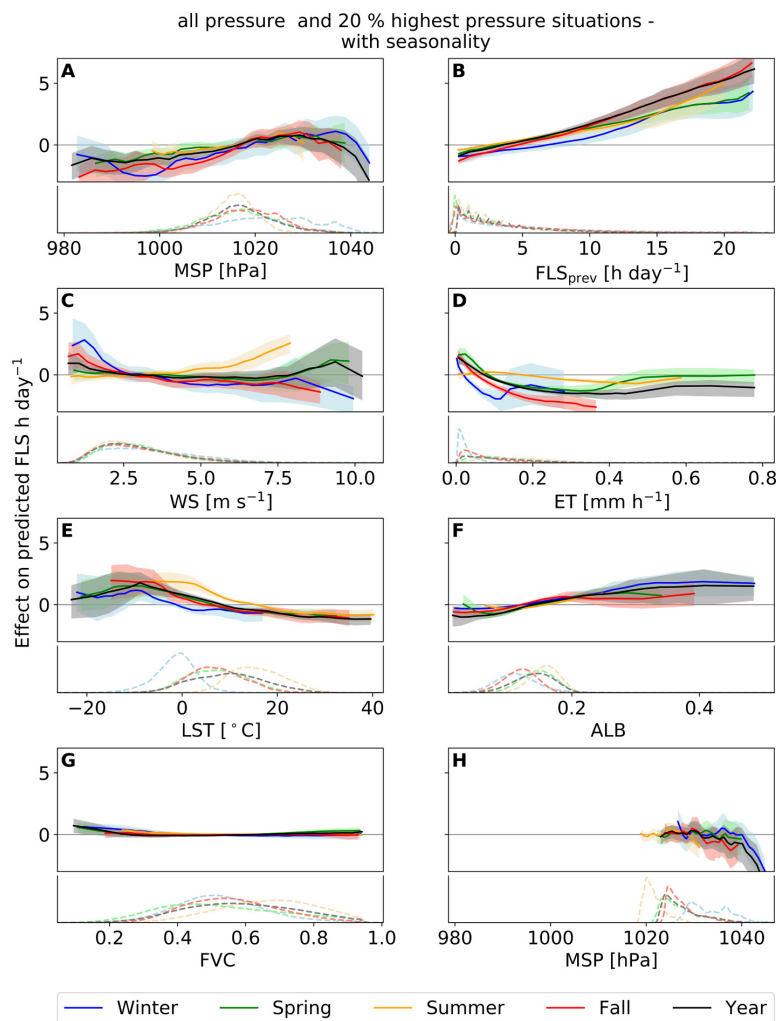


Fig. 8. Partial dependence plot showing the mean response in modeled FLS occurrence to changes in all input features over all seasons, for the 10×10 model unit run using data with seasonality. The top plot displays the partial dependence, the bottom plot shows a kernel density estimation of the data distribution. The shown features are: MSP in the model run using all pressure values (A), FLS_{prev} (B), WS (C), ET (D), LST (E), ALB (F) and FVC (G). In the additional (H) plot, the mean model response to changes in MSP in high pressure situations is shown.

The partial dependence of MSP (Fig. 8A) is nonlinear, following a clear pattern: in low pressure situations (980–1005 hPa) the partial dependence is low, however, not many cases fall into this value range. In situations with $1005 \text{ hPa} < \text{MSP} < 1030 \text{ hPa}$, where most of the cases occur, the partial dependence increases substantially, reaching its maximum at $\sim 1030 \text{ hPa}$ or even higher pressure situations during winter, before it decreases again for pressure situations around 1040 hPa. However, only a few cases (most of them in winter) exist with such high pressure of $\approx 1040 \text{ hPa}$. The full-year sensitivity of MSP is $+3.64$ and the seasonal sensitivity is highest in winter and fall ($+3.63$) and lowest in summer ($+1.73$). FLS on the previous day (Fig. 8B) shows the most pronounced partial dependence pattern over all seasons and the full-year run with increasing values over the complete value distribution. The minimum partial dependence is thus reached for small FLS_{prev} values and the maximum partial dependence for high FLS_{prev} values. The sensitivity of FLS_{prev} also shows the strong influence of FLS_{prev} on modeled FLS occurrence and lies at $+7.04$ for

the full-year run. The FLS_{prev} sensitivity reaches its seasonal maximum in fall ($+7.94$) and its minimum in spring ($+4.95$). The pattern of partial dependence of FLS on WS (Fig. 8C) is strongly dependent on season and the distribution of values. At low wind speeds ($0\text{--}2 \text{ m s}^{-1}$) the partial dependence decreases in winter, fall and the full-year run. In spring and summer it remains more or less constant at such low WS. In summer, it increases slowly, in spring and the full-year run it increases around 8 m s^{-1} and for the winter and fall run it remains more or less constant. Still it has to be considered that any WS values greater than 6 m s^{-1} are rare and the corresponding partial dependence pattern may not be as reliable and due to the low occurrence frequency not that relevant. The full-year sensitivity of WS is, compared to MSP and FLS_{prev} relatively low at -1.33 but reaches higher values in winter (-4.75). The seasonal minimum of WS sensitivity is reached in spring ($+1.65$).

The partial dependencies of the ET and LST are also both strongly dependent on the investigated season. In summer, the partial

dependence of FLS on ET (Fig. 8D) is more or less constant at a partial dependence of 0 predicted FLS h day^{-1} over the complete value distribution. In all other seasons and the full-year run the partial dependence decreases at ET values from 0.0 to 0.1 mm h^{-1} . In spring and the full-year run it then slightly increases again at 0.4 mm h^{-1} . Due to the low distribution of ET values greater than 0.3 mm h^{-1} the corresponding partial dependence values might be less reliable. The full-year sensitivity for ET is at -2.90 , the seasonal maximum of sensitivity is found in fall (-4.09) and the minimum is found in summer (-0.94). The partial dependence of FLS on LST (Fig. 8E) is highest when LSTs are below the freezing point, reaching a maximum at around $-10 \text{ }^\circ\text{C}$. At higher LSTs, the partial dependence of FLS on LST decreases. In summer, the LST value range and also these partial dependence patterns are shifted to higher temperatures. The full-year sensitivity of LST is similar to the ET sensitivity at -2.93 . The seasonal maximum of LST sensitivity is found in fall (-3.06), the seasonal minimum is found in winter (-1.85).

Similar to their low feature importance, ALB (Fig. 8F) and FVC (Fig. 8G) both have small partial dependence values. The partial dependence of ALB slightly increases over the complete value distribution of all seasons and the full-year run. The partial dependence of FVC is low and constant over all seasons. Interestingly, the full-year sensitivity of ALB reaches $+2.56$, a similar high value to the sensitivity of ET and LST. The seasonal maximum is found in winter ($+2.19$) and the minimum is found in summer ($+1.24$). The full-year sensitivity of FVC is at $+0.87$ and thus the lowest of all features. The seasonal FVC sensitivity is highest in winter ($+0.73$) and lowest in fall (-0.19).

In the pressure-filtered model run, all partial dependencies show a similar pattern, except for the partial dependencies of MSP (Fig. 8H). This is due to the reduced value range of the MSP values, thus only the decrease of the partial dependencies starting from around 1035 hPa is visible. The MSP sensitivity changes direction in the pressure filtered model and lies at -4.33 for the full-year run and is strongest in winter (-4.03) and lowest in spring (-0.74).

All in all, the response in modeled FLS occurrence is distinctive for the features FLS_{prev} , ET, LST, WS and MSP, with the latter showing a clear dependency on the application of the pressure filter. For changes in FVC and ALB, the response in modeled FLS occurrence is small. These partial dependence patterns confirm the feature importance patterns discussed above, with FLS_{prev} , MSP and ET being important features for model performance.

3.5. Two-variable partial dependence

While the isolated impacts of individual features can be investigated with the one-variable partial dependencies, two-variable partial dependence plots can show the combined effects of two variables on the response of modeled FLS occurrence, allowing for the analysis of possible co-dependencies. Based on the one-variable partial dependencies described above, six combinations of the most important features are presented in Fig. 9. These are MSP-WS (A), FLS_{prev} -MSP (B), FLS_{prev} -ET (C), ET-MSP (D), FLS_{prev} -WS (E) and ET-LST (F). In Fig. 9, the two-variable partial dependencies of the full-year run using all pressure situations including the seasonality are shown. Similar to the one-variable partial dependencies the exact data preprocessing pertaining to seasonality treatment or pressure filtering does not substantially influence the two-variable partial dependencies. To outline characteristics of the distribution of the considered features and as an assessment for the representativeness of the results, two distribution borders are marked in the plots below: inside the outer dashed line 99% of the data can be found, inside the inner dashed line 75% of the data is present. Additionally, on the equivalent axes of the feature combinations, the 1st, 50th and 99th percentiles are marked as short lines.

The two-variable partial dependence of MSP and WS is displayed in Fig. 9A. The values of the partial dependence are low over the complete value range of both features. A diagonal pattern, where low wind

speeds and high pressure values lead to higher predicted FLS values is indicated. The combined effects of FLS_{prev} and MSP in Fig. 9B are stronger, with FLS_{prev} being the more dominant variable. Still a weak combined effect of higher MSP values and higher FLS_{prev} values leading to higher predicted FLS is visible. In Fig. 9C the combined effects of FLS_{prev} and ET are visible. While low FLS_{prev} values in combination with high ET values have a negative effect on predicted FLS, a positive effect on predicted FLS is visible for high FLS_{prev} and low ET values. The two-variable partial dependence of ET and MSP (Fig. 9D) also shows a diagonal pattern of the combined effect of high MSP values with low ET values leading to higher predicted FLS. The combined effects of FLS_{prev} and WS (Fig. 9E) are similar to FLS_{prev} and MSP: FLS_{prev} dominates the combined effects, with stronger influence of WS when FLS_{prev} values are higher, indicated by the more diagonal patterns of the two-variable partial dependence at FLS_{prev} values larger than 5 h day^{-1} . For the two-variable partial dependencies of ET and LST (Fig. 9F) low ET values with low LST values lead to high predicted FLS values.

In general, the seasonal differences of the two-variable partial dependence are governed by the seasonal differences in FLS occurrence as well as the range of values of the features. Since FLS occurrence is higher in winter and fall, the magnitude of the two-variable partial dependence is higher in winter and fall. Furthermore, the combined effects of MSP and WS are stronger in fall and winter thus confirm the combination of high MSP – low WS leading to higher predicted FLS indicated in the full-year plot. In summer and spring, two-variable partial dependencies are low. Considering the patterns seen in the one-variable and two-variable partial dependence plots FLS_{prev} shows the most pronounced and clear effect on modeled FLS occurrence and dominates the combined effects in the two-variable partial dependence plots. This is especially interesting since MSP is the most important feature in most model runs (compare Fig. 5).

4. Discussion

4.1. Model performance

In general, the statistical models are capable of accurately predicting FLS using meteorological and land-cover predictors. While subtracting seasonality does not influence model performance substantially, filtering for high pressure situations boosts model performance significantly. Filtering for high pressure situations also leads to a higher relative frequency of radiation fog situation, which helps the model learn the relevant patterns, and thus improves its performance. Performance is best in winter and worst in summer. These patterns might both be due to more frequent high pressure situations in winter (e.g. Bartoszek, 2017), which can also be seen by significantly higher MSP values in winter compared to the other seasons (winter mean MSP being 5 hPa higher) in the ERA5 data used in this study. Additionally, filtering for high pressure values makes it easier for the model to generalize and predict FLS. In summer and when using all pressure values, FLS occurrence may be due to local processes more frequently, such as orographic lifting (Egli et al., 2019) and not specifically tied to large-scale weather systems. This is in line with the results shown in van Oldenborgh et al. (2010) who found that atmospheric dynamics contribute up to 40% to the variability of fog occurrences in winter, whereas in summer, the contribution of circulations on fog occurrence is more dependent on local factors.

The spatial differences of R^2 show lower model performance in units with high topographic variability such as parts of the mountain ranges surrounding the Czech Republic and the High Tatras. Furthermore the deviation from the mean R^2 (Fig. 4) and MSE per model unit is greater in areas with greater height above sea level and with greater standard deviation of height in that model unit. The correlation between the mean height above sea level and deviation from the mean R^2 per model unit is especially prevalent east of 10°E , with a stronger negative deviation from the mean R^2 at larger heights (compare Fig. 4D). This

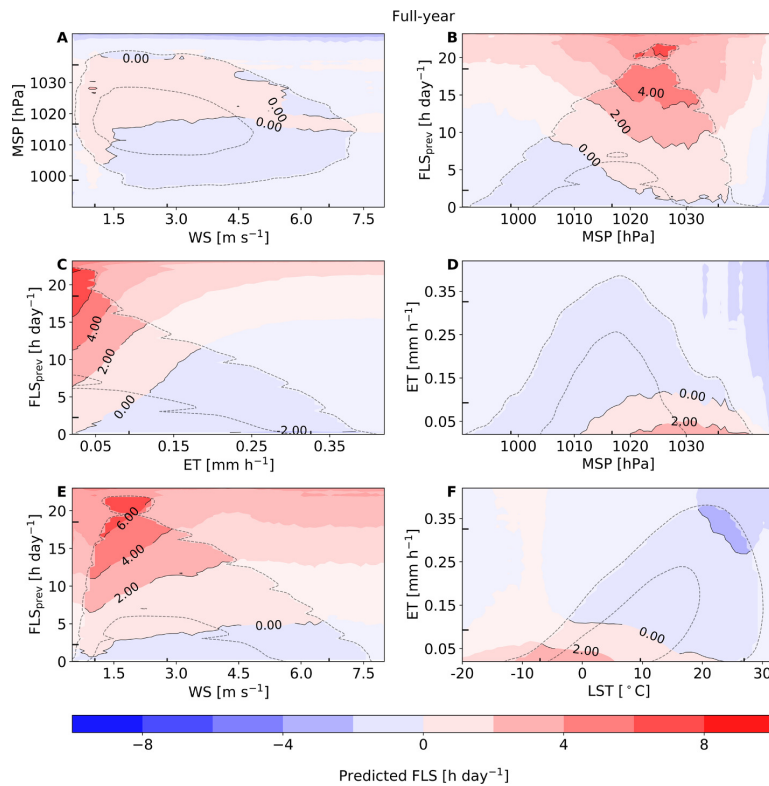


Fig. 9. Two-variable partial dependence on FLS (hours day⁻¹) for six feature combinations for the 10 × 10 full-year run including the seasonality. The feature combinations are MSP and WS (A), FLS_{prev} and MSP (B), FLS_{prev} and ET (C), ET and MSP (D), FLS_{prev} and WS (E) and ET and LST (F). The outer dashed circle contains 99% of the data, the inner dashed circle contains 75% of the data. On the x and y axes of the feature combinations, the 1st, 50th and 99th percentile are marked as short lines.

correlation is strongest in spring and fall. The decreasing performance in these either high altitude or topographically highly variable model units could be due to more pronounced small-scale, local processes that the model is not capable of reproducing. This could be also due to the missing terrain information in the selected predictors, e.g. according to Cuxart and Jiménez (2012) the advection flows generated by the topography are crucial for the growing phase of radiation fog. As stated in the Methods, using static geolocation variables such as height above sea level or geographic position can lead to overfitting and might be useful in training but not in making spatial predictions, as stated in Meyer et al. (2018) and Meyer et al. (2019). Terrain information (as height above sea level) is left out for this reason.

Besides the missing terrain information, the low resolution of the ERA5 data might lead to worse performance in topographically highly variable model units. In these units, the ERA5 data might be less accurate in general due to grid averaging. Here, the quality of the ERA5 may be improved by using a different interpolation technique like kriging or bicubic interpolation during data preprocessing. However, such a change is not expected to markedly change the results. Generally the model performance in the eastern parts of the study area is high, likely due to rather low variability in terrain and a large influence of large-scale circulation patterns on fog occurrence leading to mostly radiation or advection fog. These processes are well captured by the chosen predictors and therefore well represented in the models.

Interestingly, the high pressure filter can also lead to lower performance in some model units in winter, such as near the Harz and in parts of Poland. This is not the case at any other time of year, though. Two things might be responsible for this pattern: First, in the high pressure situations in winter, available data for training and testing has been reduced strongly. Thus, false predictions have a stronger influence on the R² of the model unit when less predictions are made in general.

In model units with more available data points, false predictions have less influence on the model unit R². Secondly, the lower performance in these units may indicate that in these regions, fog occurrence during winter may be caused by processes that are not inherently captured by the set of predictors. For example, terrain-induced processes may play a role as stated above, such as on the Mt. Brocken at 1142 m asl which is immersed in clouds about 50% of the time (Acker et al., 2002).

A thorough investigation of the influence of the FLS distribution on model performance shows that in model units with higher FLS occurrence, model performance is better. The relationship of model performance and mean FLS occurrence for each model unit over all seasons is strong (Spearman's rho = 0.8) and significant (p < 0.01) especially in the model runs using all pressure situations. This confirms the previously stated relationship of better performance in winter and worse performance in summer. These findings underline the influence of data availability and distribution on model performance.

Overfitting is apparent in all models, with larger differences between test and training performance in the 10 × 10 units than in the 15 × 15 units. This is likely due to the higher number of available data points in the 15 × 15 units for training and testing, shown with the decreasing difference between training and test performance with increasing number of data available in the model unit. More data thus makes it possible to learn more FLS situations and the subsequent behavior of the predictor variables. Furthermore, in a larger domain, a model needs to be able to represent rather general FLS patterns instead of specific local FLS patterns. A highly localized model might be also prone to overfitting. This could be investigated in future studies, by varying the model unit size while keeping the number of available training points constant. The different model unit sizes show the conflict between sufficient data for training and testing and small enough regional models to model smaller scale, regional FLS patterns. This

conflict between data availability and spatial resolution is frequently encountered in observational studies. Conducting hyperparameter tuning in all model units for all model unit sizes in combination with regularization would lead to higher computation times, but could reduce the observed overfitting.

4.2. FLS drivers and their temporal and spatial patterns

Generally, fog formation and presence depend on multiple factors, such as the synoptic situation, smaller-scale atmospheric circulations and local orographic factors (Pérez-Díaz et al., 2017). The important role of the synoptic situation is underscored by the results of this study, with the high feature importance of MSP and WS, which are used as proxies for atmospheric dynamics. In high pressure situations, often related to anticyclonic conditions, an inversion can develop during nighttime which is a prerequisite for the formation of radiation fog. The results obtained from the sensitivity analysis indicate that lower wind speed together with high pressure lead to a stronger stability and thus persisting FLS. This interplay of MSP and WS is also visible in Fig. 9. The influence of low wind speeds on turbulence generation during FLS development is not visible here as such phenomena occur below the spatial resolution of the ERA5 data. Wind direction has not been included in this study, since creating a general FLS – land surface model with a low number of predictors and minimum amount of rescaling has been a priority in this study, to first create a general model framework which can be further developed based on its first results in the future. For example more detailed treatments of wind speed and direction in different atmospheric layers will be investigated in further studies. In the investigated study area, especially humid air masses from the west are important for fog occurrence (Blas et al., 2002; Wrzesinsky and Klemm, 2000; Klemm and Wrzesinsky, 2007). The importance of atmospheric dynamics and air masses is also visible by the higher frequency of model units with WS and MSP as the most important feature in the western parts of the study region, which possibly indicates a decreasing relevance of westerly (moist) air masses with increasing continentality. As noted above, wind direction should thus be added as a predictor in further studies.

The positive relationship of FLS cover with FLS_{prev} is most likely due to persisting FLS over several days due to persisting atmospheric conditions. Additionally, feedback processes between FLS and the surface exist: in the presence of FLS, surface heating due to solar radiation is reduced during daytime which leads to lower near-surface temperatures in the subsequent night and a reduction of the difference between air and dew point temperature. This relationship might be stronger in basins and valleys where FLS tends to persist longer due to lower wind speeds inside the FLS layer and inversions or cold pool situations that can last for several days (Cuxart and Jiménez, 2012; Scherrer and Appenzeller, 2014). To investigate processes and sensitivities in larger river valleys, model units could be delineated specifically considering topographic information in further studies. In these larger model units, SHAP (SHapley Additive exPlanations) values, introduced by Lundberg and Lee (2017) could be investigated, which allow for the investigation of individual predictions leading to a global understanding of the model (Lundberg et al., 2020), providing a tool to distinguish and individually analyze different FLS regimes.

Two land surface based predictors that are closely related to each other due to the dependence of the saturation vapor pressure on temperature, described by the Clausius–Clapeyron equation, are LST and ET. This means that higher land surface temperatures lead to higher evapotranspiration, with the latter being constrained by moisture availability (especially in summer), but also by solar radiation as well as the vapor pressure deficit. Concerning FLS processes, one would expect LST to have a negative relationship with FLS occurrence and ET to have a positive influence on FLS occurrence due to the moisture input into the boundary layer. However, in this study, both LST and ET show a negative relationship with modeled FLS, the latter one likely

confounded by temperature, but also winds. FLS occurrence might be lower in high ET situations due to the larger vapor pressure deficit and thus drier air. Generally, ET is strongly influenced by the variability of atmospheric conditions (Teuling et al., 2010; Seneviratne et al., 2012) e.g. with anticyclonic circulation patterns leading to high ET values in Poland (Bogawski and Bednorz, 2016). At the scales considered in this study, the influence of atmospheric conditions on ET – FLS patterns is likely larger than the effects of moisture input via ET on FLS formation.

In general, temperature and humidity during FLS formation are connected via the process of condensation: Cooling leads to a rise in relative humidity, a reduction of the atmospheric saturation vapor pressure, and, once saturation is reached, condensation of water vapor to available cloud condensation nuclei and fog formation (Bergot and Lestringant, 2019; Steeneveld and de Bode, 2018). This effect is especially prevalent in winter, where fog formation is favored by a higher frequency of low temperatures together with high relative humidity conditions (Wrzesinsky and Klemm, 2000). Hunová et al. (2018) list relative humidity and air pollutants as the most important features when modeling fog in the Czech Republic. As the presented statistical model presents a general working ground, a proxy for CCN has not been included yet but will be in further studies. The negative influence of higher LST on FLS cover in urban areas (Williams et al., 2015; Izett et al., 2019) can be suspected here, but is not visible in the results due to the size of the model units.

Still, compared to LST, ET plays a stronger role in this study, especially in spring and fall, where the combined effects of low temperatures favorable for FLS and a higher soil moisture leading to moderate evapotranspiration are prevalent. This is mainly visible in the eastern part of the study area, where ET is the most important feature in most model units in spring and fall. In these units, the moisture input by maritime westerly air masses might be limited which could lead to a higher importance of local moisture sources. According to van Oldenborgh et al. (2010) fog during summer in the Netherlands only forms when enough moisture is present, with higher summer fog occurrences when moisture input in the preceding early summer or spring months has been larger. Here, in high pressure situations, day to day variations in ET are more important for predicting FLS occurrence than day to day variations in MSP, which is visible in the feature importance of ET and MSP in the deseasonalized model run.

In this study, the influence of radiative properties of the surface on FLS occurrence is only indicated by the small positive influence of increasing ALB values on FLS occurrence. Still, the feature importance and partial dependence of ALB is rather low and does not feature a clear pattern. The small effect of ALB on FLS occurrence might also be due to the multiple directions in which ALB might influence FLS occurrence. While surfaces with high ALB absorb less radiation during daytime and thus might lead to a lower air temperature and higher relative humidity above these surfaces, during nighttime, when radiation fog usually forms, darker surfaces might provide more moisture initially (such as water bodies or forests). Furthermore, FLS reduces the incoming solar radiation thus masking some expected radiative effects of ALB on FLS occurrence. In further studies, the effects of radiation on FLS occurrence could be further investigated by adding radiation fluxes as predictors to the model.

Similar to ALB, the influence of FVC on FLS occurrence is also not clearly visible in the results obtained from the study. Duynderkerke (1991) lists vegetation as a driver for FLS occurrence but this might not be visible here due to the spatial resolution and the stronger influence of meteorological drivers that mask such influences of the land surface and cover. Still, with the exclusion of low pressure situations and subsequent decreasing the influence of atmospheric dynamics can help the investigation of the influence of land surface parameters on FLS distribution. Filtering for specific weather situations, as in Egli et al. (2019) could decrease the variability and influence of the considered atmospheric predictors and may help isolate the influence of land surface parameters on FLS distribution under specific conditions.

As already stated, certainly not all factors influencing FLS distribution are considered in this study. In part, this study is limited by the spatial resolution of the data sets used, as local and regional morphology are important drivers for FLS occurrence (e.g. Blas et al., 2002). Furthermore, the landform, turbulent boundary layer mixing, above-cloud humidity and winds, soil conductivity, availability of CCN and the distance to the coast are listed as influential FLS drivers (e.g. Blas et al., 2002; Blas et al., 2010; Steeneveld and de Bode, 2018; Hunová et al., 2018). In further studies, the presented baseline model can be expanded by integrating these predictors to investigate their influence on model performance and sensitivities.

5. Conclusion

In this study, a machine learning technique, gradient boosting regression trees, was used to predict observed FLS occurrence over continental Europe, using meteorological and land surface parameters as predictors. To analyze spatial patterns of model skill and sensitivities, spatially explicit 10×10 and 15×15 SEVIRI pixel model units were created covering the entire study area. Additionally, models were applied for different model setups and data preprocessing procedures. The models were then applied to the entire, full-year data, as well as in different seasons.

In general, the statistical models were able to accurately predict FLS occurrence in all regions of the study area, with R^2 values between 0.6 and 0.94 during validation with independent data. Model skill was observed to be highest in winter and lowest in the full-year run. Model performance increases when only high pressure situations are considered and when smaller (10×10 SEVIRI pixels) model units are used. Some overfitting is apparent in all models and depends on model unit size, with larger modeling domains featuring less overfitting. Using deseasonalized data only has a small effect on model performance.

Analyses of feature importance reveal that features pertaining to atmospheric dynamics are more relevant to predict FLS than surface characteristics. The most important features for FLS prediction are MSP, WS, FLS on the previous day, ET and to some extent LST. Albedo and FVC are less important in the statistical models. When only considering high pressure situations, MSP becomes less important, while ET and WS gain importance, the latter especially in spring and fall. Spatial patterns of the most important feature show the dominance of MSP in most of the model units when using all pressure situations. When only high pressure situations are considered, ET is dominant in central and eastern parts of the study area while WS, FLS_{prev} and MSP are dominant in western parts of the study area.

A sensitivity analysis was conducted with the statistical models, using the partial dependence technique. While there are some differences in partial dependencies when excluding low pressure situations, seasonality and model unit size settings do not seem to have a marked influence. A positive influence of FLS_{prev} and MSP and a negative influence of ET, LST and WS on modeled FLS were found. These patterns are also confirmed in analyses of two-variable partial dependencies, which were used to study the combined effects of MSP, WS, ET, LST and FLS_{prev} . This analysis showed that especially the combination of high FLS_{prev} , high MSP and low WS values leads to high predicted FLS values. This is also the case for low LST and low ET values.

Considering the modeling framework, there are several limitations that should be noted. First of all, using spatially explicit model units

reduces the available number of data points for training and testing. Thus, the available data puts a constraint on the spatial resolution at which such an analysis can be conducted. Still, there is potential for future improvements: Creating larger model units could provide more data that could also be used for validation and reduce the risk of overfitting. Grid search or a different methodology for hyperparameter tuning could be applied in all units separately and with fewer model units, less computing time would be consumed. While the generalization of hyperparameters is useful in this model set-up, conducting a thorough grid search in all model units would lead to a more complete picture of ideal hyperparameters for the model but still may have little influence on model skill and overfitting. It has to be noted that FLS is influenced by a large number of parameters, some of which are not represented in the model, such as topography, humidity, wind direction, soil moisture and aerosol loading. Additionally, FLS formation is also influenced by the state of higher atmospheric layers for example through moisture advection and mixing. While the model set-up presented here provides a general working ground, these features may be integrated in more complex statistical frameworks in the future. Furthermore, FLS processes take place from small scales of aerosol activation (10^{-7} m) to synoptic scales (10^6 m). Accurately addressing the issues concerning scales will thus be a critical aspect going forward, specifically as processes that are not important on one scale might be crucial on a different scale.

To gain further insights into the relationship between FLS and the land surface, FLS properties such as cloud top height and liquid water path can also be analyzed with the presented modeling framework. While the model units can be varied in size, they can also be grouped into areas of similar topography or land cover which could help in further analyzing FLS – land cover dependencies.

CRedit authorship contribution statement

Eva Pauli: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Hendrik Andersen:** Methodology, Writing - review & editing. **Jörg Bendix:** Writing - review & editing. **Jan Cermak:** Conceptualization, Methodology, Writing - review & editing. **Sebastian Egli:** Resources, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

LSA-SAF data were obtained from the LSA-SAF data catalogue (<https://landsaf.ipma.pt/en/data/catalogue/>, last access 09 April 2019). ERA5 data were obtained from the Copernicus Climate Data Store (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>, last access: 02 April 2020). Eva Pauli has been financially supported by the Graduate Funding from the German States. The valuable comments of two anonymous reviewers and the editor helped improve the original manuscript.

Appendix A. Mean squared error (MSE) of all models

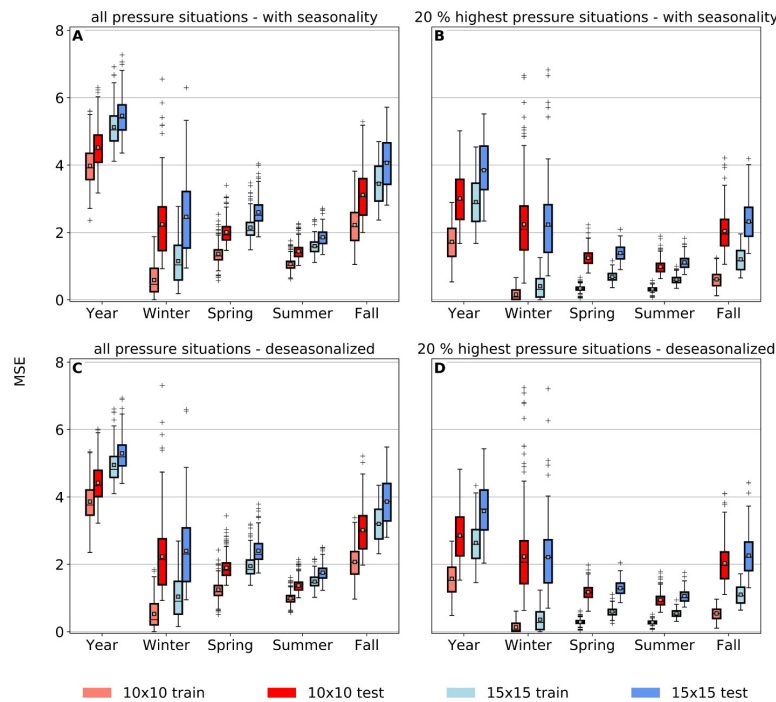


Fig. 10. MSE of training and test set over all grid sizes and seasons using either all pressure situations (left) or only high pressure situations (right). The top row shows the results using the data with seasonality, while the bottom row displays the model results using deseasonalized data.

References

- Acker, K., Mertes, S., Möller, D., Wiprecht, W., Auel, R., Kalaß, D., 2002. Case study of cloud physical and chemical processes in low clouds at Mt. Brocken. *Atmos. Res.* 64, 41–51. [https://doi.org/10.1016/S0169-8095\(02\)00078-9](https://doi.org/10.1016/S0169-8095(02)00078-9).
- Baguskas, S.A., Clemesha, R.E., Loik, M.E., 2018. Coastal low cloudiness and fog enhance crop water use efficiency in a California agricultural system. *Agric. Forest Meteorol.* 252, 109–120. <https://doi.org/10.1016/j.agrformet.2018.01.015>.
- Ball, L., Tzanopoulos, J., 2020. Interplay between topography, fog and vegetation in the central South Arabian mountains revealed using a novel Landsat fog detection technique. *Remote Sens. Ecol. Conserv.* <https://doi.org/10.1002/rse2.151>.
- Bartoszek, K., 2017. The main characteristics of atmospheric circulation over East-Central Europe from 1871 to 2010. *Meteorol. Atmos. Phys.* 129, 113–129. <https://doi.org/10.1007/s00703-016-0455-z>.
- Bendix, J., 1994. Fog climatology of the Po Valley. *Rivista di meteorologia aeronautica* 54, 25–36.
- Bergot, T., Carrer, D., Noilhan, J., Bougeault, P., 2005. Improved site-specific numerical prediction of fog and low clouds: a feasibility study. *Weather Forecast.* 20, 627–646. <https://doi.org/10.1175/WAF873.1>.
- Bergot, T., Lestringant, R., 2019. On the predictability of radiation fog formation in a mesoscale model: a case study in heterogeneous Terrain. *Atmosphere* 10. <https://doi.org/10.3390/atmos10040165>.
- Berry, Z.C., Johnson, D.M., Reinhardt, K., 2015. Vegetation-zonation patterns across a temperate mountain cloud forest ecotone are not explained by variation in hydraulic functioning or water relations. *Tree Physiol.* 35, 925–935. <https://doi.org/10.1093/treephys/tpv062>.
- Bogawski, P., Bednorz, E., 2016. Atmospheric conditions controlling extreme summertime evapotranspiration in Poland (central Europe). *Nat. Haz.* 81, 55–69. <https://doi.org/10.1007/s11069-015-2066-2>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brujinzeel, L.A., Eugster, W., Burkard, R., 2006. Fog as a hydrologic input. In: Anderson, M., McDonnell, J. (Eds.), *Encyclopedia of Hydrological Sciences*. doi: 10.1002/0470848944.hsa041.
- Blas, M., Polkowska, Zaneta, Sobik, M., Klimaszewska, K., Nowinski, K., Namiesnik, J., 2010. Fog water chemical composition in different geographic regions of Poland. *Atmos. Res.* 95, 455–469. <https://doi.org/10.1016/j.atmosres.2009.11.008>.
- Blas, M., Sobik, M., Quiel, F., Netzel, P., 2002. Temporal and spatial variations of fog in the Western Sudety Mts., Poland. *Atmos. Res.* 64, 19–28. [https://doi.org/10.1016/S0169-8095\(02\)00076-5](https://doi.org/10.1016/S0169-8095(02)00076-5).
- Cermak, J., 2006. SOfOS – a new Satellite-based Operational Fog Observation Scheme. Ph.D. thesis, Philipps-Universität Marburg. doi: 10.17192/z2006.0149.
- Cermak, J., 2018. Fog and low cloud frequency and properties from active-sensor satellite data. *Remote Sens.* 10, 1209. <https://doi.org/10.3390/rs10081209>.
- Cermak, J., Bendix, J., 2011. Detecting ground fog from space – a microphysics-based approach. *Int. J. Remote Sens.* 32, 3345–3371. <https://doi.org/10.1080/01431161003747505>.
- Cermak, J., Eastman, R.M., Bendix, J., Warren, S.G., 2009. European climatology of fog and low stratus based on geostationary satellite observations. *Quart. J. Roy. Meteorol. Soc.* 135, 2125–2130. <https://doi.org/10.1002/qj.503>.
- Cuxart, J., Jiménez, M., 2012. Deep radiation fog in a wide closed valley: study by numerical modeling and remote sensing. *Pure Appl. Geophys.* 169, 911–926. <https://doi.org/10.1007/s00024-011-0365-4>.
- Dawson, T.E., 1998. Fog in the California redwood forest: ecosystem inputs and use by plants. *Oecologia* 117, 476–485. <https://doi.org/10.1007/s004420050683>.
- 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.M., 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, I., 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. *Quart. J. Roy. Meteorol. Soc.* 137, 553–597. <https://doi.org/10.1002/qj.828>.
- Duynkerke, P.G., 1991. Radiation fog: a comparison of model simulation with detailed observations. *Mon. Weather Rev.* 119, 324–341. [https://doi.org/10.1175/1520-0493\(1991\)119<0324:RFACOM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<0324:RFACOM>2.0.CO;2).
- Ebner, M., Miranda, T., Roth-Nebelsick, A., 2011. Efficient fog harvesting by *Stipagrostis sabulicola* (Namib dune bushman grass). *J. Arid Environ.* 75, 524–531. <https://doi.org/10.1016/j.jaridenv.2011.01.004>.
- Eckardt, F., Soderberg, K., Coop, L., Muller, A., Vickery, K., Grandin, R., Jack, C., Kapalanga, T., Henschel, J., 2013. The nature of moisture at Gobabeb, in the central Namib Desert. *J. Arid Environ.* 93, 7–19. <https://doi.org/10.1016/j.jaridenv.2012.01.011>.
- 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. <https://doi.org/10.3390/rs10040628>.
- Egli, S., Thies, B., Bendix, J., 2019. A spatially explicit and temporally highly resolved analysis of variations in fog occurrence over Europe. *Quart. J. Roy. Meteorol. Soc.* 145, 1721–1740. <https://doi.org/10.1002/qj.3522>.
- Egli, S., Thies, B., Dröner, J., Cermak, J., Bendix, J., 2017. A 10 year fog and low stratus climatology for Europe based on Meteosat Second Generation data. *Quart. J. Roy. Meteorol. Soc.* 143, 530–541. <https://doi.org/10.1002/qj.2941>.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>.
- EUMETSAT, 2013. MSG level 1.5 image data format description. Technical Report

- European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) Darmstadt, Germany.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232.
- 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.* 18, 16537–16552. <https://doi.org/10.5194/acp-18-16537-2018>.
- Fuzzi, S., Facchini, M.C., Orsi, G., Lind, J.A., Wobrock, W., Kessel, M., Maser, R., Jaeschke, W., Enderle, K.H., Arends, B.G., Berner, A., Solly, I., Krusiz, C., Reischl, G., Pahl, S., Kaminski, U., Winkler, P., Ogren, J.A., Noone, K.J., Hallberg, A., Fierlinger-Oberlinninger, H., Puxbaum, H., Marzorati, A., Hansson, H.-C., Wiedensohler, A., Svenningsson, I.B., Martinsson, B.G., Schell, D., Georgii, H.W., 1992. The Po Valley fog experiment 1989. *Tellus B* 44, 448–468. <https://doi.org/10.1034/j.1600-0889.1992.r01.4.00002.x>.
- Gautam, R., Singh, M.K., 2018. Urban heat island over Delhi punches holes in widespread fog in the Indo-Gangetic Plains. *Geophys. Res. Lett.* 45, 1114–1121. <https://doi.org/10.1002/2017GL076794>.
- Glickman, T., 2000. *Glossary of Meteorology*, second ed. American Meteorological Society, Boston.
- Gottlieb, T.R., Eckardt, F.D., Venter, Z.S., Cramer, M.D., 2019. The contribution of fog to water and nutrient supply to *Arthroa leubnitziae* in the central Namib Desert, Namibia. *J. Arid Environ.* 161, 35–46. <https://doi.org/10.1016/j.jaridenv.2018.11.002>.
- Gray, E., Gilardoni, S., Baldocchi, D., McDonald, B.C., Facchini, M.C., Goldstein, A.H., 2019. Impact of air pollution controls on radiation fog frequency in the Central Valley of California. *J. Geophys. Res. Atmos.* 124, 5889–5905. <https://doi.org/10.1029/2018JD029419>.
- Gultepe, I., Tardif, R., Michaelides, S., Cermak, J., Bott, A., Bendix, J., Müller, M.D., Pagowski, M., Hansen, B., Ellrod, G., et al., 2007. Fog research: a review of past achievements and future perspectives. *Pure Appl. Geophys.* 164, 1121–1159. <https://doi.org/10.1007/s00024-007-0211-x>.
- Hastie, T., Tibshirani, R., Friedman, J., 2001. *The Elements of Statistical Learning – Data Mining, Inference, and Prediction*. Springer Series in Statistics New York, NY, USA. doi: 10.1007/978-0-387-84858-7.
- Hersbach, H., 2016. The ERA5 atmospheric reanalysis. In: *AGU Fall Meeting Abstracts*, 12–16 December 2016, San Francisco.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. <https://doi.org/10.1002/joc.1276>.
- Hunová, I., Brabec, M., Malý, M., Valeriánová, A., 2018. Revisiting fog as an important constituent of the atmosphere. *Sci. Total Environ.* 636, 1490–1499. <https://doi.org/10.1016/j.scitotenv.2018.04.322>.
- Izett, J.G., van de Wiel, B.J.H., Baas, P., van Hooft, J.A., Schulte, R.B., 2019. Dutch fog: on the observed spatio-temporal variability of fog in the Netherlands. *Quart. J. Roy. Meteorol. Soc.* 145, 2817–2834. <https://doi.org/10.1002/qj.3597>.
- Klemm, O., Lin, N.-H., 2016. What causes observed fog trends: air quality or climate change? *Aerosols Air Qual. Res.* <https://doi.org/10.4209/aaqr.2015.05.0353>.
- Klemm, O., Wrzesinsky, T., 2007. Fog deposition fluxes of water and ions to a mountainous site in Central Europe. *Tellus B: Chem. Phys. Meteorol.* 59, 705–714. <https://doi.org/10.1111/j.1600-0889.2007.00287.x>.
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geosci. Front.* 7, 3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>.
- Lehnert, L.W., Thies, B., Trachte, K., Achilles, S., Osses, P., Baumann, K., Bendix, J., Schmidt, J., Samolov, E., Jung, P., Leinweber, P., Karsten, U., Büdel, B., 2018. A case study on fog/low stratus occurrence at Las Lomitas, Atacama Desert (Chile) as a water source for biological soil crusts. *Aerosol Air Qual. Res.* 18, 254–269. <https://doi.org/10.4209/aaqr.2017.01.0021>.
- Leigh, R.J., Drake, L., Thampapillai, D.J., 1998. An economic analysis of terminal aerodrome forecasts with special reference to Sydney Airport. *J. Transport Econ. Policy* 32, 377–392.
- 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, 2522–2539. <https://doi.org/10.1038/s42256-019-0138-9>.
- Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30, 4765–4774.
- Maronga, B., Bosveld, F.C., 2017. Key parameters for the life cycle of nocturnal radiation fog: a comprehensive large-eddy simulation study. *Quart. J. Roy. Meteorol. Soc.* 143, 2463–2480. <https://doi.org/10.1002/qj.3100>.
- 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. Software* 101, 1–9. <https://doi.org/10.1016/j.envsoft.2017.12.001>.
- Meyer, H., Reudenbach, C., Willauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction. *Ecol. Model.* 411, 108815. <https://doi.org/10.1016/j.ecolmodel.2019.108815>.
- Mitchell, D., Henschel, J.R., Hetem, R.S., Wassenaar, T.D., Strauss, W.M., Hanrahan, S.A., Seely, M.K., 2020. Fog and fauna of the Namib Desert: past and future. *Ecosphere* 11, e02996. <https://doi.org/10.1002/ecs2.2996>.
- Molnar, C., 2019. *Interpretable Machine Learning*. URL: <https://christophm.github.io/interpretable-ml-book/>.
- Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. *Front. Neurobot.* 7, 21. <https://doi.org/10.3389/fnbot.2013.00021>.
- Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. *Quart. Rev. Biol.* 83, 171–193. <https://doi.org/10.1086/587826>.
- van Oldenborgh, G.J., Yiou, P., Vautard, R., 2010. On the roles of circulation and aerosols in the decline of mist and dense fog in Europe over the last 30 years. *Atmos. Chem. Phys.* 10, 4597–4609. <https://doi.org/10.5194/acp-10-4597-2010>.
- Pagowski, M., Gultepe, I., King, P., 2004. Analysis and modeling of an extremely dense fog event in Southern Ontario. *J. Appl. Meteorol.* 43, 3–16. [https://doi.org/10.1175/1520-0450\(2004\)043<0003:AAMOAE>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0003:AAMOAE>2.0.CO;2).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pöschl, U., Martin, S.T., Sinha, B., Chen, Q., Gunthe, S.S., Huffman, J.A., Borrmann, S., Farmer, D.K., Garland, R.M., Helas, G., Jimenez, J.L., King, S.M., Manzi, A., Mikhailov, E., Pauliquevis, T., Petters, M.D., Prenni, A.J., Roldin, P., Rose, D., Schneider, J., Su, H., Zorn, S.R., Artaxo, P., Andreae, M.O., 2010. Rainforest aerosols as biogenic nuclei of clouds and precipitation in the Amazon. *Science* 329, 1513–1516. <https://doi.org/10.1126/science.1191056>.
- Price, J., 2019. On the formation and development of radiation fog: an observational study. *Bound.-Layer Meteorol.* 172, 167–197. <https://doi.org/10.1007/s10546-019-00444-5>.
- Pérez-Díaz, J., Ivanov, O., Peshev, Z., Álvarez Valenzuela, M., Valiente-Blanco, I., Evgenieva, T., Dreischuh, T., Gueorguiev, O., Todorov, P., Vaseashta, A., 2017. Fogs: physical basis, characteristic properties, and impacts on the environment and human health. *Water* 9, 807. <https://doi.org/10.3390/w9100807>.
- Ramanathan, V., Crutzen, P.J., Kiehl, J.T., Rosenfeld, D., 2001. Aerosols, climate, and the hydrological cycle. *Science* 294, 2119–2124. <https://doi.org/10.1126/science.1064034>.
- Roth-Nebelsick, A., Ebner, M., Miranda, T., Gottschalk, V., Voigt, D., Gorb, S., Stegmaier, T., Sarsour, J., Linke, M., Konrad, W., 2012. Leaf surface structures enable the endemic Namib desert grass *Stipagrostis sabulicola* to irrigate itself with fog water. *J. Roy. Soc. Interface* 9, 1965–1974. <https://doi.org/10.1098/rsif.2011.0847>.
- Sachveh, M., Koepke, P., 1995. Radiation fog and urban climate. *Geophys. Res. Lett.* 22, 1073–1076. <https://doi.org/10.1029/95GL00907>.
- Sachveh, M., Koepke, P., 1997. Fog dynamics in an urbanized area. *Theor. Appl. Climatol.* 58, 87–93. <https://doi.org/10.1007/BF00867435>.
- Savitzky, A., Golay, M.J., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36, 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Scherrer, S.C., Appenzeller, C., 2014. Fog and low stratus over the Swiss Plateau – a climatological study. *Int. J. Climatol.* 34, 678–686. <https://doi.org/10.1002/joc.3714>.
- Seneviratne, S.I., Lehner, I., Gurtz, J., Teuling, A.J., Lang, H., Moser, U., Grebner, D., Menzel, L., Schroff, K., Vitvar, T., Zappa, M., 2012. Swiss prealpine Rietholzbad research catchment and lysimeter: 32 year time series and 2003 drought event. *Water Resour. Res.* 48. <https://doi.org/10.1029/2011WR011749>.
- Shrivastava, M., Cappa, C.D., Fan, J., Goldstein, A.H., Guenther, A.B., Jimenez, J.L., Kuang, C., Laskin, A., Martin, S.T., Ng, N.L., Petaja, T., Pierce, J.R., Rasch, P.J., Roldin, P., Seinfeld, J.H., Shilling, J., Smith, J.N., Thornton, J.A., Volkamer, R., Wang, J., Worsnop, D.R., Zaveri, R.A., Zelenyuk, A., Zhang, Q., 2017. Recent advances in understanding secondary organic aerosol: Implications for global climate forcing. *Rev. Geophys.* 55, 509–559. <https://doi.org/10.1002/2016RG000540>.
- Steenefeld, G.-J., de Bode, M., 2018. Unravelling the relative roles of physical processes in modelling the life cycle of a warm radiation fog. *Quart. J. Roy. Meteorol. Soc.* 144, 1539–1554. <https://doi.org/10.1002/qj.3300>.
- Stimberg, 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. <https://doi.org/10.1029/2019JD031380>.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinf.* 8, 25. <https://doi.org/10.1186/1471-2105-8-25>.
- Teuling, A.J., Seneviratne, S.I., Stöckli, R., Reichstein, M., Moors, E., Clais, P., Luysaert, S., Van Den Hurk, B., Ammann, C., Bernhofer, C., et al., 2010. Contrasting response of European forest and grassland energy exchange to heatwaves. *Nat. Geosci.* 3, 722–727. <https://doi.org/10.1038/ngeo950>.
- Torregrosa, A., Combs, C., Peters, J., 2016. GOES-derived fog and low cloud indices for coastal north and central California ecological analyses. *Earth Space Sci.* 3, 46–67. <https://doi.org/10.1002/2015EA000119>.
- Trigo, I.F., Dacamara, C.C., Viterbo, P., Roujean, J.-L., Olesen, F., Barroso, C., de Coca, F.C., Carrer, D., Freitas, S.C., García-Haro, J., Geiger, B., Gellens-Meulenberghs, F., Ghilain, N., Meliá, J., Pessanha, L., Siljamo, N., Arboleda, A., 2011. The satellite application facility for land surface analysis. *Int. J. Remote Sens.* 32, 2725–2744. <https://doi.org/10.1080/01431161003743199>.
- Vautard, R., Yiou, P., Van Oldenborgh, G.J., 2009. Decline of fog, mist and haze in Europe over the past 30 years. *Nat. Geosci.* 2, 115–119. <https://doi.org/10.1038/ngeo414>.
- Williams, A.P., Schwartz, R.E., Iacobellis, S., Seager, R., Cook, B.I., Stille, C.J., Husak, G., Michaelsen, J., 2015. Urbanization causes increased cloud base height and decreased fog in coastal Southern California. *Geophys. Res. Lett.* 42, 1527–1536. <https://doi.org/10.1002/2015GL063266>.
- Wrzesinsky, T., Klemm, O., 2000. Summertime fog chemistry at a mountainous site in central Europe. *Atmos. Environ.* 34, 1487–1496. [https://doi.org/10.1016/S1352-2310\(99\)00348-9](https://doi.org/10.1016/S1352-2310(99)00348-9).

Original publication: Pauli et al. (2022a)

Pauli, E., Cermak, J., & Teuling, A. J. (2022). Enhanced nighttime fog and low stratus occurrence over the Landes forest, France. *Geophysical Research Letters*, 49, e2021GL097058. doi:10.1029/2021GL097058

Contributions:

Eva Pauli conceived the study design, conducted the investigation and developed the methodology including coding. Eva Pauli visualized and interpreted the results, wrote the original draft and incorporated the reviewer comments.

Peer-Review:

The manuscript was submitted to *Geophysical Research Letters* on November 15th, 2021. After a peer-review with two anonymous reviewers, the manuscript was accepted on February 17th, 2022 and published on February 23rd, 2022.

Usage of text passages and modifications:

The publication is included in abbreviated form in chapter 3. Parts of the method section and of the introduction of the original manuscript were moved to chapter 1.2 and 1.3. The numbering of figures was changed and section titles were renamed.



Geophysical Research Letters®



RESEARCH LETTER

10.1029/2021GL097058

Key Points:

- Fog and low stratus (FLS) cloud cover is enhanced over the French Landes forest at night
- FLS cloud cover enhancement is most pronounced in summer and fall
- Low wind speed and low temperatures over the forest are potential drivers of FLS enhancement

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

E. Pauli,
eva.pauli@kit.edu

Citation:

Pauli, E., Cermak, J., & Teuling, A. J. (2022). Enhanced nighttime fog and low stratus occurrence over the Landes forest, France. *Geophysical Research Letters*, 49, e2021GL097058. <https://doi.org/10.1029/2021GL097058>

Received 15 NOV 2021

Accepted 17 FEB 2022

Author Contributions:

Conceptualization: Eva Pauli, Jan Cermak, Adriaan J. Teuling
Formal analysis: Eva Pauli
Investigation: Eva Pauli
Methodology: Eva Pauli, Jan Cermak, Adriaan J. Teuling
Visualization: Eva Pauli
Writing – original draft: Eva Pauli
Writing – review & editing: Eva Pauli, Jan Cermak, Adriaan J. Teuling

© 2022. The Authors.
 This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Enhanced Nighttime Fog and Low Stratus Occurrence Over the Landes Forest, France

Eva Pauli^{1,2} , Jan Cermak^{1,2} , and Adriaan J. Teuling³

¹Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, ²Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, ³Hydrology and Quantitative Water Management Group, Wageningen University, Wageningen, The Netherlands

Abstract Understanding the drivers of fog and low stratus (FLS) cloud occurrence is important for traffic, ecosystems, and climate models, but it is challenging to analyze due to the complex interactions between meteorological factors and land cover. Here, we use active and passive satellite data, as well as reanalysis data to investigate nighttime FLS occurrence over the expansive Landes forest in France from 2006 to 2015. We find significant FLS enhancement over the forest compared to surrounding areas, especially in summer and fall. Lower wind speed and lower temperatures are found over the forest at night, which can enhance FLS development over the forest. Still, other drivers, such as biogenic volatile organic compounds acting as cloud condensation nuclei, are most likely important as well. The results show that the influence of forests on boundary layer clouds is not limited to convective daytime conditions.

Plain Language Summary Fog and low stratus clouds (FLS) are influenced by various drivers. Their relationship to land cover, specifically forest, is thus difficult to investigate. In this study, we analyze nighttime FLS cover over a large forest area in south-western France using a mix of different types of satellite data. We find higher FLS occurrence over the forest area compared to its surroundings, especially in summer and fall. Lower temperatures and wind speed over the forest could contribute to this enhancement. These results can help when predicting FLS for traffic and underline the importance of different land cover types for weather and climate.

1. Introduction

Interactions between the land surface and clouds are manifold and highly dependent on cloud type as well as geographical region. The intensity and direction of these interactions are still uncertain. Considering the influence of surface characteristics on energy fluxes on the surface and the boundary layer (Pielke, 2001), specifically the effects of land cover on boundary layer clouds has been the subject of research in the past (Ray et al., 2003; Teuling et al., 2017; Theeuwes et al., 2019; Wang et al., 2009). An enhancement of convective cumulus cloud cover compared to the surroundings has been reported over forests, including the large Landes forest in southern France (Teuling et al., 2017), megacities (Theeuwes et al., 2019) and natural bushland (Ray et al., 2003). However, an increase in shallow cloud cover has also been found over deforested areas in eastern Amazonia (Wang et al., 2009). This is in line with findings from Xu et al. (2022), who found enhanced cloud cover over most temperate and boreal forests but decreased cloud cover over forests in Amazonia, Central Africa, and the Southeast US. Possible reasons for shallow cumulus enhancement over forests are higher evaporation (Gentine et al., 2013) and higher sensible heat flux (Bosman et al., 2019; Gambill & Mecikalski, 2011) over forests, as well as higher aerodynamic roughness, leading to the development of a forest breeze (Mahrt & Ek, 1993). Besides the physical mechanisms for low cloud cover enhancement, biogeochemical processes can contribute to cloud development. Biogenic volatile organic compounds (BVOCs), emitted by forests can form secondary organic aerosols (SOA), which can act as cloud condensation nuclei (CCNs) and thereby favor cloud formation (Pöschl et al., 2010; Shrivastava et al., 2017). Enhanced emission of BVOCs has been found over the Landes forest (Kammer et al., 2018) while enhanced cloud cover due to the release of BVOCs has been found over boreal forests (Spracklen et al., 2008).

While the mechanisms for convective cloud enhancement over forests are understood relatively well, this is less in the case for fog and low stratus clouds (FLS). Topography is known to strongly influence FLS, leading to a higher FLS cover in valleys compared to mountainous areas (Bendix, 1994; Scherrer & Appenzeller, 2014). Higher temperatures, lower saturation of air with water vapor, and an increase in air quality lead to reduction of

fog (Gray et al., 2019; Klemm & Lin, 2016; Yan et al., 2020), and have been observed to induce fog holes over urban areas (Gautam & Singh, 2018; Williams et al., 2015). An earlier onset of fog formation has been observed over fields compared to bare soil (Roach, 1995) and over a homogeneous grass surface compared to a surface with trees (Mazoyer et al., 2017). The influence of fog on vegetation has been well-studied for several specific ecosystems around the world, including tropical-montane cloud forests (TMCFs) (Weathers et al., 2019), in the Namib desert (Gottlieb et al., 2019) or in the Californian redwood forests (Dawson, 1998). Here, fog plays a major role for water and nutrient input. Still, temporally and spatially extensive studies looking at the influence of land cover on fog and low clouds are rare.

Satellite data can potentially provide information on fog and low stratus clouds over larger areas not covered by weather stations. Passive sensors, such as the geostationary Meteosat Spinning Enhanced Visible and Infrared Imager (SEVIRI), provide high temporal resolution and cover a large area, and have been used successfully for the detection of FLS in Europe (Cermak et al., 2009; Cermak & Bendix, 2011; Egli et al., 2017) and the Namib desert (Andersen & Cermak, 2018). Still, classification errors, small-scale FLS features and multiple cloud layers can lead to misclassifications (Cermak & Bendix, 2008; Cermak, 2018). These classification errors can be minimized when using active satellite data such as LiDAR data from Cloud-Aerosol LiDAR and Infrared Pathfinder Satellite Observations (CALIPSO) (Cermak, 2018; Vaughan et al., 2009). Despite its low temporal sampling rate, CALIPSO data is still highly valuable for the study of fog and low cloud patterns, especially when combined with passive satellite data.

To isolate potential local effects of forests on fog and low cloud occurrence, the strong influence of topography on FLS occurrence discussed above should be minimized. A forested area with low topographic variability (cf. Figure S1 in the Supporting Information S1) and large spatial extent is the Landes forest in southern France. Here, we analyze nighttime (0–6 UTC) fog and low stratus cloud cover over the Landes forest over a period of 10 years. We compare the FLS detection based on passive and active satellite data, and thus minimize the influence of potential misclassifications on the results. Furthermore, we investigate seasonal differences of FLS cover over the area and analyze the influence of wind and temperature on the observed patterns. With this long-term analysis, we can discuss potential interactions of FLS and forests, independent of satellite sensor and short-term fluctuations in FLS cover.

2. Data and Methods

2.1. Data

The Landes forest covers an area of about 12,000 km² in southern France directly at the Atlantic coast. The forest is mainly composed of maritime pine (*Pinus pinaster*) (Kammer et al., 2018) and shows a distinct contrast to its surrounding land cover types (compare Figure 2d). The study site has previously been used for the investigation of daytime cumulus clouds using satellite data, showing its potential for the investigation of land-atmosphere interactions (Teuling et al., 2017).

The primary fog and low cloud data set used in this study was created by Egli et al. (2017). It uses passive satellite data from Meteosat SEVIRI following the Satellite-based Operational Fog Observation Scheme (SOFOS) by Cermak (2006) using the approach presented in Cermak and Bendix (2007) for nighttime data. The resulting data set provides a binary FLS mask for every 15-min time step covering the entire central European land mass and the years 2006–2015. The data set has been validated against ground observations, showing that 80%–90% of FLS and no-FLS situations are classified correctly (Egli et al., 2017). The binary FLS mask is available for daytime and nighttime hours but not during, as well as shortly before and after twilight due to sensor and algorithm constraints. For the purpose of this study, nighttime observations (0–6 UTC) of FLS are used. This is in line with BVOC emissions observed during night over the Landes forest (Kammer et al., 2018), which can potentially serve as CCN. The time frame of 0–6 UTC was chosen to minimize the amount of missing observations during twilight and to select a time frame with nighttime observations across all seasons. As a plausibility check for the FLS data set by the passive satellite sensor, an additional FLS classification based on active satellite data from CALIPSO was created. For this, the CALIPSO level 2 1-km cloud-layer product (Version 4.20) (NASA Langley Atmospheric Science Data Center DAAC, 2018) was used. The cloud and aerosol discrimination (CAD) algorithm classifies 90% of layers correctly (Liu et al., 2009). Over the 10 years, 186 nighttime overpasses over the

Landes region were available. Due to missing data in both data sets, 179 of those overpasses could be used for comparison.

For the surface (10 m) winds, ERA5-land reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF) was used (Muñoz Sabater, 2019). For temperature and wind data at different pressure levels ERA5 was used as well (Hersbach et al., 2018). The land cover data plotted in the results is taken from HILDA (Winkler et al., 2021). The Fraction of Vegetation Cover data serving as the background for the binary FLS maps are based on data from the EUMETSAT Satellite Application Facility on Land Surface Analysis (LSA-SAF) (Trigo et al., 2011).

2.2. Methods

The CALIPSO overpasses usually took place between 02:15 and 02:30 in the night. The CALIPSO FLS mask was derived similarly to the approach presented in Cermak (2018). First, the cloud layer altitude was calculated by subtracting the terrain altitude from the observed feature altitude. Then all cloud layers with a cloud top height equal to or smaller than 2.5 km and a cloud base height equal to or smaller than 2 km were defined as FLS. The thresholds differ slightly to those used in Cermak (2018) to include not only fog but also low stratus clouds.

To compare the SEVIRI based data set with the FLS mask derived from CALIPSO, the FLS observations of both data sets at the location of the CALIPSO overpass were contrasted by creating a confusion matrix. The time step of the SEVIRI-based data set used for comparison was 02:15 UTC. To take into account the larger pixel size of SEVIRI and the small mismatch of observation time of both data sets, the comparison was done for the complete CALIPSO swath area across the forest area and not pixel-by-pixel. For example, an observation was marked as true positive, when any pixel along the CALIPSO swath for both data sets showed FLS. If there was no FLS pixel in both data sets, the observation was marked as true negative.

This comparison was followed by creating nighttime averages of FLS using the data from 0 to 6 UTC of the SEVIRI based data set. The respective days belonging to one of the categories of the confusion matrix were flagged accordingly.

3. Results

3.1. Cross-Validation of FLS Products

Out of the 179 used CALIPSO observations, 50 were identified as true positive and 83 as true negative. On 12 observations, no FLS was identified by CALIPSO but FLS was identified by the SEVIRI based data set (false positive). On 34 observations, FLS was identified by CALIPSO, but no FLS was present in the SEVIRI data set (false negative). Possible reasons for this could be multilayer cloud situations or classification errors in the FLS data set (cf. Cermak and Bendix (2008); Cermak (2018)). Further reasons are also described in the discussion. The confusion matrix can be found in Table S1 in the Supporting Information S1.

Two true positive cases are displayed in Figure 1. On 2008-07-14 (Figure 1a) most of the forest is covered by a large FLS patch, which is also visible in the CALIPSO profile, where the cloud top is situated at approximately 2.5 km and cloud base at or below 2 km. A slight mismatch of the two products is visible at the CALIPSO swath at 44°N, where no FLS is present in the SEVIRI-based product, but FLS is present in the CALIPSO based product. On 2015-09-20 (Figure 1b) both FLS patches over the forest and south of the forest are present in both data sets.

3.2. Climatological Means

To decrease and visualize the influence of potential misclassifications on the climatological mean of nighttime FLS cover, three types of climatological means were constructed using the FLS data set:

1. Climatological mean over all days of the SEVIRI based data set by Egli et al. (2017) (3,652 days)
2. Climatological mean over days with CALIPSO overpasses (179 days)
3. Climatological mean over days where the FLS observations of the SEVIRI and CALIPSO based data set match (true positives and true negatives) (133 days)

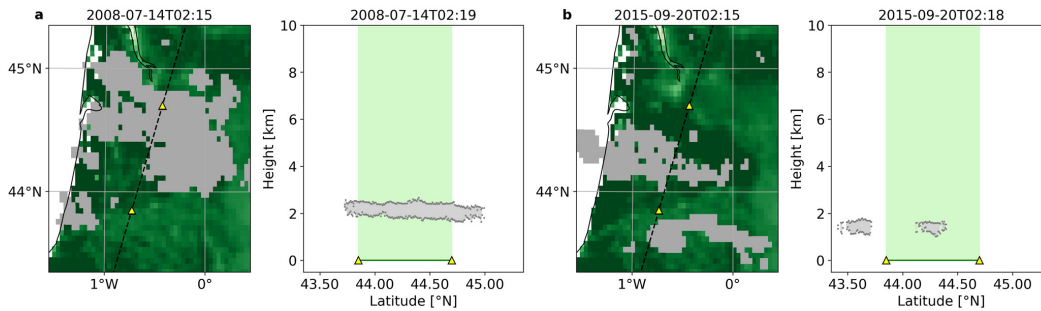


Figure 1. Example validation of Spinning Enhanced Visible and Infrared Imager (SEVIRI)-based fields with the corresponding Cloud-Aerosol LiDAR and Infrared Pathfinder Satellite Observations (CALIPSO) profiles. Shown are two days, 2008-07-14 (a) and 2015-09-20 (b), with the SEVIRI based Fog and low stratus (FLS) maps on the left and the corresponding CALIPSO profiles on the right. The gray pixels in the maps of the SEVIRI-based data set display FLS, the background is a map of mean fraction of vegetation cover of the study area. For orientation, the CALIPSO swath is plotted as a dashed line in the maps. In the CALIPSO profiles, the cloud layers are plotted as a gray area at their corresponding height and with their latitudinal extent. The green line, as well as the light green background in the profiles mark the forest area. The yellow triangles in both the maps and the profiles mark the beginning of the forest area.

The corresponding maps, together with a land cover map of the Landes region are shown in Figure 2. In all three climatology maps, mean FLS cover from 0 to 6 UTC is higher over the forest, with differences in mean FLS cover between forest and non-forest most distinct in the northern forest area. Nighttime FLS cover is up to 1 hr longer over the forest than over the surrounding areas and local patches of enhanced FLS cover over the forest are visible (e.g., south of 44°N). The difference in nighttime FLS cover over forest versus other land cover types is significant (two sample *t*-test, $P < 0.05$, Table S2 in the Supporting Information S1) for all three climatologies, and strongest for the true positive and negative days (Figure 2c). This significance independent of calculated climatology, shows the robustness of the results.

3.3. Seasonal Analysis

A seasonal analysis of FLS cover (only true positive and true negative observational days) shows that nighttime FLS enhancement over the forest is significant ($P < 0.05$) for all months except January, February, June, and October (see Table S3 in the Supporting Information S1 for all *t*-test results and Figures S1 and S2 in the Supporting Information S1 for all monthly maps). The difference in mean FLS cover between forest and non-forest areas is shown here for the months of May, July, August, and September (Figure 3a), when FLS is most likely the result of more localized processes, as opposed to the winter months. Nonetheless, differences in FLS cover between forest and non-forest areas can still be significant in winter but are most likely due to higher FLS cover over the study area but lower FLS cover in the Pyrenees (south of 43.5°N) (cf. Egli et al. (2017)). In May, especially the central Landes forest shows enhanced nighttime FLS compared to its surroundings. In July, enhanced FLS cover extends to the south of the forest, toward the Pyrenees. In August, the shape of the forest is quite well replicated by the pattern of enhanced FLS cover over the forest. In September, nighttime FLS cover is enhanced mostly in the central and western parts of the forest. Similar to the patterns depicted in the full-year means in Figure 2, local patches of enhanced FLS cover are apparent inside the forest area in all of those monthly plots.

To investigate the potential reasons for high FLS cover over the Landes region, maps of mean wind speed are created. They show lower wind speed over the forest compared to the surrounding area (around 0.5 m s^{-1}) (Figure 3b) and wind speed is highest directly at the coastline. The differences in mean wind speed above forest and non-forest pixels are significant (two sample *t*-test, $P < 0.05$, Table S3 in the Supporting Information S1) in March, April, June, July, and August. The main wind direction is west, with variations over the different months and over the study area. Patterns of wind direction also seem to be split into two subpatterns, with westerly winds prevailing near the Pyrenees (south of 44°N) and changing wind directions north of 44°N, for example, with northerly winds in July and August.

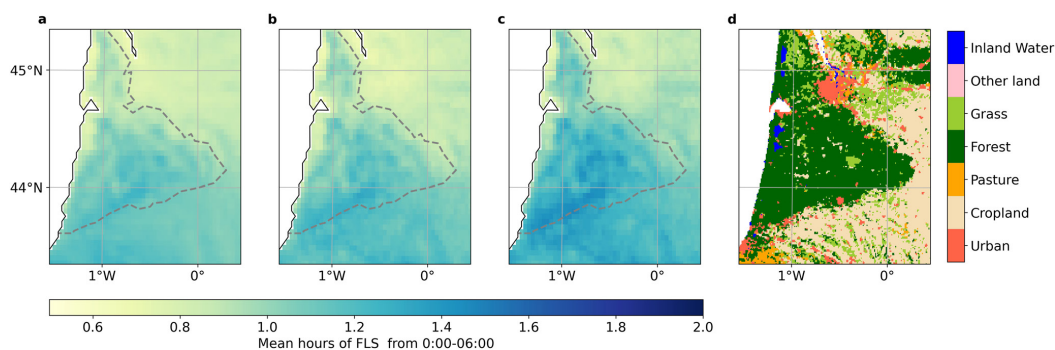


Figure 2. Spatial distribution of climatological mean Fog and low stratus (FLS) cover over the study region. In (a) all observations from 2006 to 2015 are used to calculate mean FLS cover (3,652 days), in (b) only days with Cloud-Aerosol LiDAR and Infrared Pathfinder Satellite Observations overpasses (179 days), in (c) only true positive and true negative observations (133 days), and (d) shows HILDA land cover for the year 2006. The gray dashed line approximately marks the forest border.

3.4. Vertical Temperature Profiles

To better understand boundary layer effects on enhanced FLS cover over the Landes region, vertical profiles of temperature gradient in combination with wind vectors on the different pressure levels are analyzed. For this study, the temperature gradient is calculated by subtracting the temperature of a pressure level from the temperature of its overlying layer, that is, positive values indicate a temperature inversion. In Figure 4, hourly means over the respective true positive and true negative days in August (Figures 4a and 4b) and September (Figures 4c and 4d) are presented. Both longitudinal (Figure 4a) and latitudinal (Figure 4b) vertical profiles show a temperature inversion (positive temperature gradient) over the forest compared to the surrounding areas, especially at

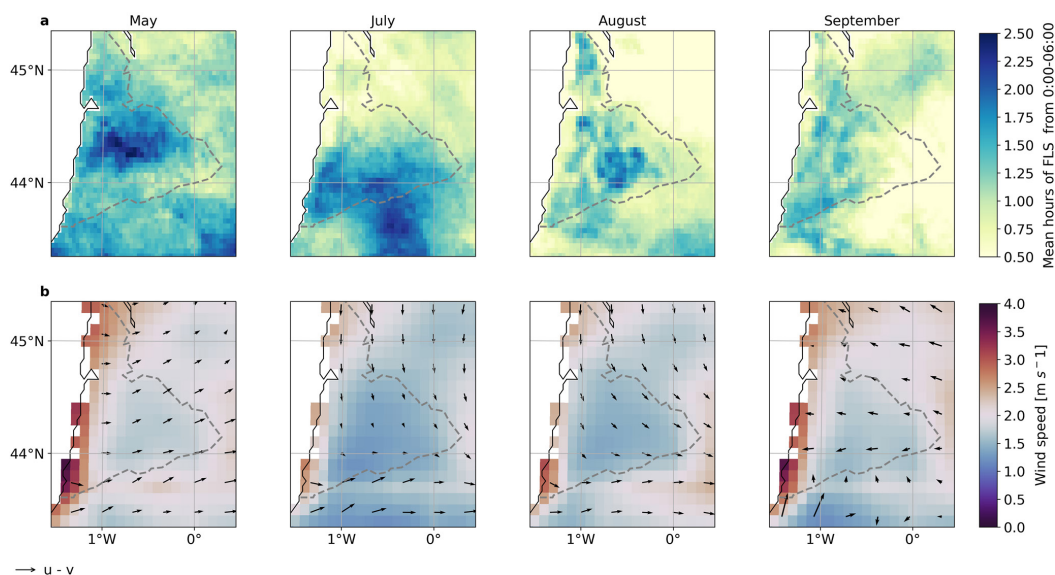


Figure 3. Climatological mean fog and low stratus hours by month based on true positive and true negative observations (a), and the corresponding ERA-5 land wind speed and wind direction (u and v wind components) on the respective days (mean from 0–6UTC) (b). The number of true negative and true positive days available to calculate the climatologies is as follows: 9 (May), 11 (July), 12 (August), and 13 (September). The gray dashed line approximately marks the forest border.

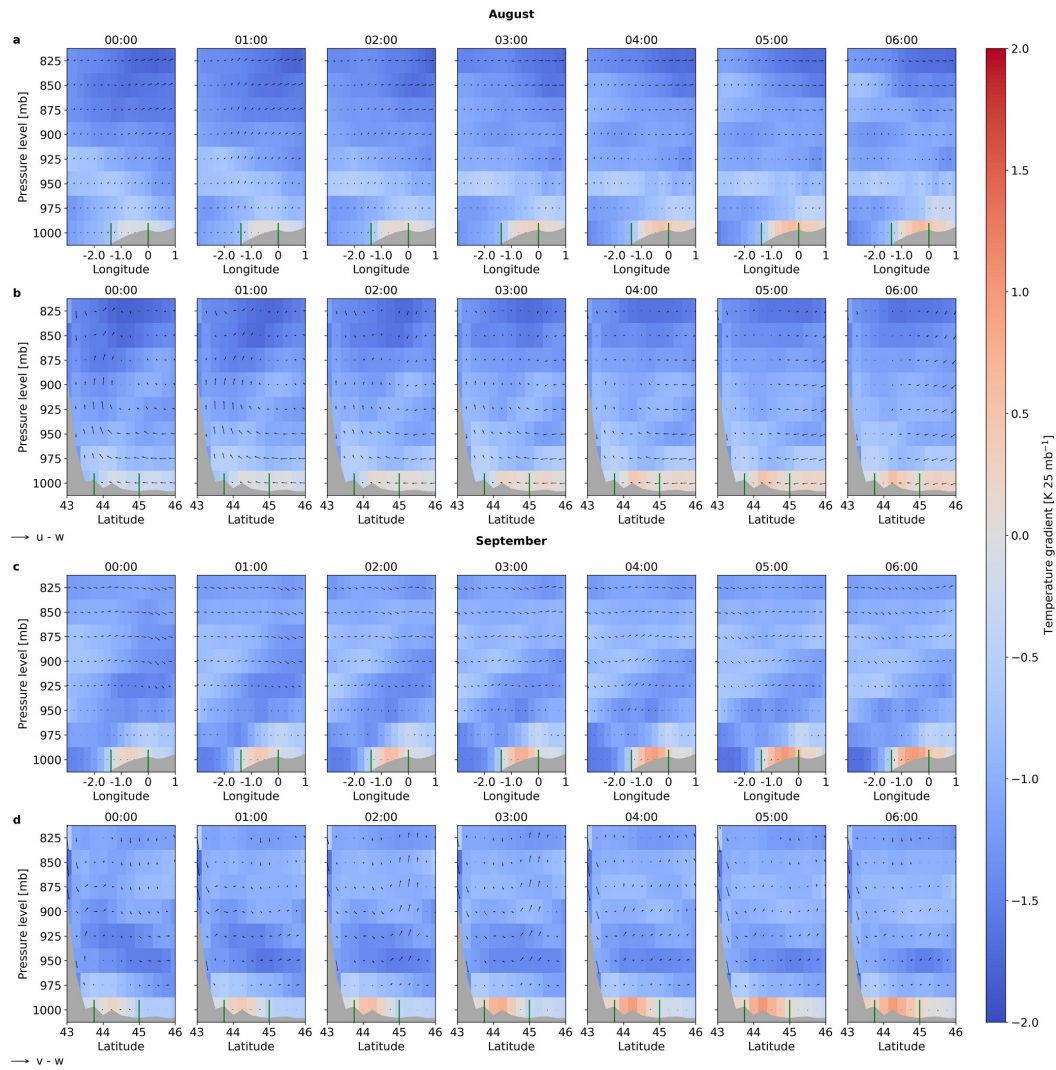


Figure 4. Profiles of mean hourly temperature gradient in $\text{K } 25 \text{ mb}^{-1}$ for the true positive and true negative days in August (a and b) and September (c and d), from 825 to 1000 mb along 44.25°N (a and c) and -0.25°W (b and d). In the longitudinal profile (a and c) u-w wind vectors are plotted, in the latitudinal profile (b and d) v-w wind vectors are plotted. For visibility reasons, the w vector is enhanced by a factor of 20. The location of the Landes forest is marked in both plots as green vertical lines. The topography is plotted in gray.

05:00 and 06:00 UTC. In August, the temperature gradient reaches values up to $+0.67 \text{ K } 25 \text{ mb}^{-1}$ at -0.25°W in the longitudinal profile and $+0.57 \text{ K } 25 \text{ mb}^{-1}$ at 44.25°N in the latitudinal profile. In September, the temperature gradient reaches values up to $+1 \text{ K } 25 \text{ mb}^{-1}$ at 44.25°N and -0.5°W (both in the latitudinal and longitudinal profile). The u-w component vectors plotted in the longitudinal plots are very weak, whereas the v-w component vectors (latitudinal plot) show wind from the north in the pressure levels up to 850mb during the night. Especially at 00:00 and 01:00 UTC air rises at around 44°N , just before the Pyrenees. Over the night the v-w wind component weakens.

4. Discussion

The passive and active satellite data used in this study reveal enhanced FLS cover over the Landes forest compared to the surroundings. Both data sets agree well in most cases (74%), disagreement is potentially due to misclassifications in the FLS data set (cf. Cermak (2018)) or due to difficult FLS detection in the transitional zone (“twilight zone”) between aerosols and clouds (Koren et al., 2007). Local patterns of higher FLS cover over the forest area are visible, especially in the southern part of the Landes forest. This is similar to the findings by Teuling et al. (2017), who found enhanced daytime cumulus cover over the Landes forest with local maxima in the southern part of the forest. FLS enhancement over the forest compared to the surrounding areas is primarily visible in summer and fall, when local processes are potentially more important than in winter.

Various drivers are important for the development of FLS (Pauli et al., 2020). Here, the roles of wind speed and temperature are investigated. Based on the results of the *t*-test we identify lower wind speeds over the forest as a potential driver of higher FLS occurrence, especially in the summer months. While some turbulence is required for the formation of a stable fog layer (Haefelin et al., 2010), lower wind speeds are generally beneficial for FLS development (Bergot, 2016; Bergot & Lestringant, 2019; Gradstein et al., 2011; Pauli et al., 2020; Roach, 1995). Similar to the processes described in Gradstein et al. (2011) for lowland cloud forests, low wind speeds combined with nighttime cooling and saturation of air are a potential pathway leading to the enhanced FLS cover described in this study.

The vertical profiles of temperature gradient show lower temperatures and a positive temperature gradient in the forest area at night, which is unusual for temperate forests since they usually have higher nighttime temperatures than their surroundings due to turbulence and the storage of heat (Li et al., 2015; Schultz et al., 2017). Still, in both of these studies, the Landes region seems to be an exception, showing lower temperatures over the forest compared to unforested areas. A potential explanation is a strong nighttime cooling through evapotranspiration, similar to nighttime cooling over forests in tropical regions (Li et al., 2015). In combination with the observed temperature inversion, nighttime cooling increases the relative humidity over the area, supporting the development of fog and low stratus.

Further likely reasons for enhanced FLS cover over the Landes forest are the interplay between BVOC emissions and high evapotranspiration over the forest area. High loadings of natural aerosols from late spring to early fall have been found for boreal forests (Tunved et al., 2006). It has been shown that secondary organic aerosol emissions together with evapotranspiration over forests can lead to an increase of liquid water path and cloud droplet number concentration in low-level liquid clouds (Petäjä et al., 2022). This could also be a potential pathway in the Landes forest, where BVOC emissions have been measured in summer (Kammer et al., 2018). A higher number of CCN and therefore potentially a higher number of small cloud droplets could lead to more FLS identified by both FLS detection algorithms. Measurements of BVOC emissions and fog and low cloud occurrences in the Landes forest could test the interactions between BVOCs and FLS in the future.

The patterns and drivers of higher FLS occurrence over the forest area is further modified by the general synoptic situation and geographic position of the Landes forest, with the Atlantic Ocean to the west and the Pyrenees to the south. While the former is a source of moisture, the latter might enhance stationarity of air masses in the region, preventing high wind speeds and supporting the build-up of atmospheric moisture.

5. Conclusion

In this study, we have analyzed nighttime fog and low stratus cloud cover over the Landes forest in southwestern France using active and passive remote sensing products. We have found significantly higher FLS cover over the forest compared to non-forest areas and identified lower wind speed and a temperature inversion over the forest as potential drivers. As these parameters only partially explain the enhanced FLS cover over the forest, further atmospheric and biophysical drivers should be included into the analysis in the future, such as soil moisture, evapotranspiration, and BVOC emissions. For future work we propose a systematic approach combining modeling and sensitivity studies to further quantify the role of forests for fog and low stratus cloud formation over varying geographic and synoptic backgrounds.

Data Availability Statement

The fog and low stratus cloud data set can be downloaded from http://vhrz669.hrz.uni-marburg.de/lcrs/data_pre.do?citid=291. The CALIPSO level 2 1-km cloud-layer product (Version 4.20) is distributed by the Atmospheric Science Data Center (ASDC) at https://doi.org/10.5067/CALIPSO/CALIPSO/LID_L2_01KMCLAY-STANDARD-V4-20. The era5-land data is available at <https://doi.org/10.24381/cds.e2161bac>. The era5 data on different pressure levels can be downloaded from <https://doi.org/10.24381/cds.bd0915c6>. FVC data is provided by the EUMETSAT Satellite Application Facility on Land Surface Analysis (Trigo et al., 2011) and can be downloaded from <https://landsaf.ipma.pt/en/products/vegetation/fvc/>.

Acknowledgments

The authors declare that they have no conflict of interest. Eva Pauli has been financially supported by the Graduate Funding of the German States. We thank two anonymous reviewers for their careful and constructive reviews which have helped improve the manuscript. Open access funding enabled and organized by Projekt DEAL.

References

- Andersen, H., & Cermak, J. (2018). First fully diurnal fog and low cloud satellite detection reveals life cycle in the namib. *Atmospheric Measurement Techniques*, *11*(10), 5461–5470. <https://doi.org/10.5194/amt-11-5461-2018>
- Bendix, J. (1994). Fog climatology of the Po valley. *Rivista di Meteorologia Aeronautica*, *54*(3–4), 25–36.
- Bergot, T. (2016). Large-eddy simulation study of the dissipation of radiation fog. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 1029–1040. <https://doi.org/10.1002/qj.2706>
- Bergot, T., & Lestringant, R. (2019). On the predictability of radiation fog formation in a mesoscale model: A case study in heterogeneous terrain. *Atmosphere*, *10*(4), 165. <https://doi.org/10.3390/atmos10040165>
- Bosman, P. J. M., van Heerwaarden, C. C., & Teuling, A. J. (2019). Sensible heating as a potential mechanism for enhanced cloud formation over temperate forest. *Quarterly Journal of the Royal Meteorological Society*, *145*(719), 450–468. <https://doi.org/10.1002/qj.3441>
- Cermak, J. (2006). *SOFOS - a new satellite-based operational fog observation Scheme*. (Doctoral dissertation, Philipps-Universität Marburg). <https://doi.org/10.17192/z2006.0149>
- Cermak, J. (2018). Fog and low cloud frequency and properties from active-sensor satellite data. *Remote Sensing*, *10*(8), 1209. <https://doi.org/10.3390/rs10081209>
- Cermak, J., & Bendix, J. (2007). Dynamical nighttime fog/low stratus detection based on meteosat seviri data: A feasibility study. *Pure and Applied Geophysics*, *164*(6–7), 1179–1192. <https://doi.org/10.1007/s00024-007-0213-8>
- Cermak, J., & Bendix, J. (2008). A novel approach to fog/low stratus detection using Meteosat 8 data. *Atmospheric Research*, *87*(3), 279–292. <https://doi.org/10.1016/j.atmosres.2007.11.009>
- Cermak, J., & Bendix, J. (2011). Detecting ground fog from space – A microphysics-based approach. *International Journal of Remote Sensing*, *32*(12), 3345–3371. <https://doi.org/10.1080/01431161003747505>
- Cermak, J., Eastman, R. M., Bendix, J., & Warren, S. G. (2009). European climatology of fog and low stratus based on geostationary satellite observations. *Quarterly Journal of the Royal Meteorological Society*, *135*(645), 2125–2130. <https://doi.org/10.1002/qj.503>
- Dawson, T. E. (1998). Fog in the California redwood forest: Ecosystem inputs and use by plants. *Oecologia*, *117*(4), 476–485. <https://doi.org/10.1007/s004420050683>
- Egli, S., Thies, B., Dröner, J., Cermak, J., & Bendix, J. (2017). A 10 year fog and low stratus climatology for Europe based on Meteosat Second Generation data. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 530–541. <https://doi.org/10.1002/qj.2941>
- Gambill, L. D., & Mecikalski, J. R. (2011). A satellite-based summer convective cloud frequency analysis over the southeastern United States. *Journal of Applied Meteorology and Climatology*, *50*(8), 1756–1769. <https://doi.org/10.1175/2010JAMC2559.1>
- Gautam, R., & Singh, M. K. (2018). Urban heat island over Delhi punches holes in widespread fog in the Indo-Gangetic Plains. *Geophysical Research Letters*, *45*(2), 1114–1121. <https://doi.org/10.1002/2017GL076794>
- Gentine, P., Ferguson, C. R., & Holtslag, A. A. M. (2013). Diagnosing evaporative fraction over land from boundary-layer clouds. *Journal of Geophysical Research: Atmospheres*, *118*(15), 8185–8196. <https://doi.org/10.1002/jgrd.50416>
- Gottlieb, T. R., Eckardt, F. D., Venter, Z. S., & Cramer, M. D. (2019). The contribution of fog to water and nutrient supply to *Arthroa leubnitziae* in the central Namib Desert, Namibia. *Journal of Arid Environments*, *161*, 35–46. <https://doi.org/10.1016/j.jaridenv.2018.11.002>
- Gradstein, S. R., Obregon, A., Gehrig, C., & Bendix, J. (2011). Tropical lowland cloud forest: A neglected forest type. In L. A. Bruijnzeel, F. N. Scatena, & L. S. Hamilton (Eds.), *Tropical montane cloud forests: Science for conservation and management* (p. 130). Cambridge University Press.
- Gray, E., Gilardoni, S., Baldocchi, D., McDonald, B. C., Facchini, M. C., & Goldstein, A. H. (2019). Impact of air pollution controls on radiation fog frequency in the central valley of California. *Journal of Geophysical Research: Atmospheres*, *124*(11), 5889–5905. <https://doi.org/10.1029/2018JD029419>
- Haefelin, M., Bergot, T., Elias, T., Tardif, R., Carrer, D., Chazette, P., et al. (2010). Parisfog: Shedding new light on fog physical processes. *Bulletin of the American Meteorological Society*, *91*(6), 767–783. <https://doi.org/10.1175/2009BAMS2671.1>
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2018). *ERA5 hourly data on pressure levels from 1979 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). <https://doi.org/10.24381/cds.bd0915c6>
- Kammer, J., Perraudin, E., Flaud, P.-M., Lamaud, E., Bonfond, J., & Villenave, E. (2018). Observation of nighttime new particle formation over the French Landes forest. *The Science of the Total Environment*, *621*, 1084–1092. <https://doi.org/10.1016/j.scitotenv.2017.10.118>
- Klemm, O., & Lin, N.-H. (2016). What causes observed fog trends: Air quality or climate change? *Aerosol and Air Quality Research*, *16*(5), 1131–1142. <https://doi.org/10.4209/aaqr.2015.05.0353>
- 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*(8). <https://doi.org/10.1029/2007GL029253>
- Li, Y., Zhao, M., Motesharrei, S., Mu, Q., Kalnay, E., & Li, S. (2015). Local cooling and warming effects of forests based on satellite observations. *Nature Communications*, *6*(1), 1–8. <https://doi.org/10.1038/ncomms7603>
- Liu, Z., Vaughan, M., Winker, D., Kittaka, C., Getzewich, B., Kuehn, R., & Hostetler, C. (2009). The CALIPSO lidar cloud and aerosol discrimination: Version 2 algorithm and initial assessment of performance. *Journal of Atmospheric and Oceanic Technology*, *26*(7), 1198–1213. <https://doi.org/10.1175/2009JTECHA1229.1>
- Mahr, L., & Ek, M. (1993). Spatial variability of turbulent fluxes and roughness lengths in HAPEX-MOBILHY. *Boundary-Layer Meteorology*, *65*(4), 381–400. <https://doi.org/10.1007/bf00707034>

- Mazoyer, M., Lac, C., Thouron, O., Bergot, T., Masson, V., & Musson-Genon, L. (2017). Large eddy simulation of radiation fog: Impact of dynamics on the fog life cycle. *Atmospheric Chemistry and Physics*, 17(21), 13017–13035. <https://doi.org/10.5194/acp-17-13017-2017>
- Muñoz Sabater, J. (2019). *ERA5-Land hourly data from 1981 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [Dataset]*. <https://doi.org/10.24381/cds.e2161bac>
- NASA Langley Atmospheric Science Data Center DAAC. (2018). *CALIPSO lidar level 2 1 km cloud layer, V4-20*. NASA Langley atmospheric science data center DAAC. https://doi.org/10.5067/CALIPSO/CALIPSO/LID_L2_01KMCLAY-STANDARD-V4-20
- Pauli, E., Andersen, H., Bendix, J., Cermak, J., & Egli, S. (2020). Determinants of fog and low stratus occurrence in continental central Europe - A quantitative satellite-based evaluation. *Journal of Hydrology*, 591. <https://doi.org/10.1016/j.jhydrol.2020.125451>
- Petäjä, T., Tabakova, K., Manninen, A., Ezhova, E., O'Connor, E., Moisseev, D., & Kerminen, V.-M. (2022). Influence of biogenic emissions from boreal forests on aerosol–cloud interactions. *Nature Geoscience*, 15(1), 42–47. <https://doi.org/10.1038/s41561-021-00876-0>
- Pielke, R. A. (2001). Influence of the spatial distribution of vegetation and soils on the prediction of cumulus convective rainfall. *Reviews of Geophysics*, 39(2), 151–177. <https://doi.org/10.1029/1999RG000072>
- Pöschl, U., Martin, S. T., Sinha, B., Chen, Q., Gunthe, S. S., Huffman, J. A., & Andreae, M. O. (2010). Rainforest aerosols as biogenic nuclei of clouds and precipitation in the amazon. *Science*, 329(5998), 1513–1516. <https://doi.org/10.1126/science.1191056>
- Ray, D. K., Nair, U. S., Welch, R. M., Han, Q., Zeng, J., Su, W., & Lyons, T. J. (2003). Effects of land use in southwest Australia: 1. Observations of cumulus cloudiness and energy fluxes. *Journal of Geophysical Research: Atmospheres*, 108(D14). <https://doi.org/10.1029/2002JD002654>
- Roach, W. (1995). Back to basics: Fog: Part 2—The formation and dissipation of land fog. *Weather*, 50(1), 7–11. <https://doi.org/10.1002/j.1477-8696.1995.tb06053.x>
- Scherrer, S. C., & Appenzeller, C. (2014). Fog and low stratus over the Swiss Plateau - A climatological study. *International Journal of Climatology*, 34(3), 678–686. <https://doi.org/10.1002/joc.3714>
- Schultz, N. M., Lawrence, P. J., & Lee, X. (2017). Global satellite data highlights the diurnal asymmetry of the surface temperature response to deforestation. *Journal of Geophysical Research: Biogeosciences*, 122(4), 903–917. <https://doi.org/10.1002/2016JG003653>
- Shrivastava, M., Cappa, C. D., Fan, J., Goldstein, A. H., Guenther, A. B., Jimenez, J. L., & Zhang, Q. (2017). Recent advances in understanding secondary organic aerosol: Implications for global climate forcing. *Reviews of Geophysics*, 55(2), 509–559. <https://doi.org/10.1002/2016RG000540>
- Spracklen, D. V., Bonn, B., & Carslaw, K. S. (2008). Boreal forests, aerosols and the impacts on clouds and climate. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 366(1885), 4613–4626. <https://doi.org/10.1098/rsta.2008.0201>
- Teuling, A. J., Taylor, C. M., Meirink, J. F., Melsen, L. A., Miralles, D. G., Van Heerwaarden, C. C., & de Arellano, J. V.-G. (2017). Observational evidence for cloud cover enhancement over western European forests. *Nature Communications*, 8(1), 1–7. <https://doi.org/10.1038/ncomms14065>
- Theeuwes, N. E., Barlow, J. F., Teuling, A. J., Grimmond, C. S. B., & Kotthaus, S. (2019). Persistent cloud cover over mega-cities linked to surface heat release. *npj Climate and Atmospheric Science*, 2(1), 1–6. <https://doi.org/10.1038/s41612-019-0072-x>
- Trigo, I. F., Dacamara, C. C., Viterbo, P., Roujean, J.-L., Olesen, F., Barroso, C., & Arboleda, A. (2011). The Satellite Application Facility for Land Surface Analysis. *International Journal of Remote Sensing*, 32(10), 2725–2744. <https://doi.org/10.1080/01431161003743199>
- Tunved, P., Hansson, H.-C., Kerminen, V.-M., Ström, J., Maso, M. D., Lihavainen, H., & Kulmala, M. (2006). High natural aerosol loading over boreal forests. *Science*, 312(5771), 261–263. <https://doi.org/10.1126/science.1123052>
- Vaughan, M. A., Powell, K. A., Winker, D. M., Hostetler, C. A., Kuehn, R. E., Hunt, W. H., & McGill, M. J. (2009). Fully automated detection of cloud and aerosol layers in the CALIPSO lidar measurements. *Journal of Atmospheric and Oceanic Technology*, 26(10), 2034–2050. <https://doi.org/10.1175/2009JTECHA1228.1>
- Wang, J., Chagnon, F. J. F., Williams, E. R., Betts, A. K., Renno, N. O., Machado, L. A. T., & Bras, R. L. (2009). Impact of deforestation in the Amazon basin on cloud climatology. *Proceedings of the National Academy of Sciences*, 106(10), 3670–3674. <https://doi.org/10.1073/pnas.0810156106>
- Weathers, K. C., Ponette-González, A. G., & Dawson, T. E. (2019). Medium, vector, and connector: Fog and the maintenance of ecosystems. *Ecosystems*, 23, 217–229. <https://doi.org/10.1007/s10021-019-00388-4>
- Williams, A. P., Schwartz, R. E., Iacobellis, S., Seager, R., Cook, B. I., Still, C. J., & Michaelsen, J. (2015). Urbanization causes increased cloud base height and decreased fog in coastal Southern California. *Geophysical Research Letters*, 42(5), 1527–1536. <https://doi.org/10.1002/2015GL063266>
- Winkler, K., Fuchs, R., Rounsevell, M., & Herold, M. (2021). Global land use changes are four times greater than previously estimated. *Nature Communications*, 12(1), 1–10. <https://doi.org/10.1038/s41467-021-22702-2>
- Xu, R., Li, Y., Teuling, A. J., Zhao, L., Spracklen, D. V., Garcia-Carreras, L., & Bojje, F. (2022). Contrasting impacts of forests on cloud cover based on satellite observations. *Nature Communications*, 13. <https://doi.org/10.1038/s41467-022-28161-7>
- Yan, S., Zhu, B., Huang, Y., Zhu, J., Kang, H., Lu, C., & Zhu, T. (2020). To what extents do urbanization and air pollution affect fog? *Atmospheric Chemistry and Physics*, 20(9), 5559–5572. <https://doi.org/10.5194/acp-20-5559-2020>

Original publication: Pauli et al. (2022b)

Pauli, E., Cermak, J. & Andersen, H. (2022) A satellite-based climatology of fog and low stratus formation and dissipation times in central Europe. *Quarterly Journal of the Royal Meteorological Society*, 148, 1439-1454. doi: 10.1002/qj.4272

Contributions:

Eva Pauli conceived the study design, conducted the investigation and developed the methodology including coding. Eva Pauli visualized and interpreted the results, wrote the original draft and incorporated the reviewer comments.

Peer-Review:




The manuscript was submitted to the *Quarterly Journal of the Royal Meteorological Society* on December 17th, 2021. After a peer-review with three anonymous reviewers, the manuscript was accepted on March 19th, 2022 and published on March 23rd, 2022.

Usage of text passages and modifications:

The publication is included in abbreviated form in chapter 4. Parts of the method section and of the introduction of the original manuscript were moved to chapter 1.1 and 1.2. The numbering of figures was changed and section titles were renamed.

RESEARCH ARTICLE

A satellite-based climatology of fog and low stratus formation and dissipation times in central Europe

Eva Pauli^{1,2}  | Jan Cermak^{1,2}  | Hendrik Andersen^{1,2} 

¹Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

²Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Correspondence

Eva Pauli, Institute of Meteorology and Climate Research and Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.
Email: eva.pauli@kit.edu

Funding information

Graduate Funding of German States

Abstract

Knowledge of fog and low stratus (FLS) cloud patterns and life cycles is important for traffic safety, for the production of solar energy and for the analysis of cloud processes in the climate system. While meteorological stations provide information on FLS, a data set describing FLS formation and dissipation times on large spatial and temporal scales does not yet exist. In this study, we use logistic regression to extract FLS formation and dissipation times from a satellite-based 10-year FLS data set covering central Europe. The resulting data set is the first to provide a geographic perspective on FLS formation and dissipation at a continental scale. The patterns found show a clear dependency of FLS formation and dissipation times on topography. In mountainous areas, FLS forms in the night and dissipates in the morning. In river valleys, the typical FLS life cycle shifts to formation after sunrise and dissipation in the afternoon. Seasonal patterns of FLS formation and dissipation show similar FLS formation and dissipation times in winter and autumn, and in spring and summer, with longer events in the former two seasons.

KEYWORDS

Europe, fog, fog dissipation, fog formation, logistic regression, low stratus

1 | INTRODUCTION

Fog and the timing of its formation and dissipation is important for traffic on land, at sea and in the air (Leigh *et al.*, 1998; Pagowski *et al.*, 2004), and is critical for the quality of the prediction of solar power production (Köhler *et al.*, 2017). Fog is also a crucial water source for various ecosystems around the world (Bruijnzeel *et al.*, 2006; Gottlieb *et al.*, 2019) and plays an important role in climate processes (Vautard *et al.*, 2009).

In general, fog can be defined as a cloud touching the ground with a horizontal visibility less than 1 km

(American Meteorological Society, 2012). It can be further classified into different types based on its formation mechanisms. Radiation fog is the most frequent fog type in central Europe (Fuzzi *et al.*, 1992; Gultepe *et al.*, 2007a). It forms through radiative cooling of the ground, a build-up of a temperature inversion and subsequent condensation of the excess water vapour, and is sustained by radiative heat loss at the fog top (Roach, 1995). Geographically, radiation fog occurs more frequently in valleys, such as large river valleys, compared to mountainous areas, as shown by Scherrer and Appenzeller (2014) for the Swiss Plateau and Fuzzi *et al.* (1992) and Bendix (1994) for the

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Quarterly Journal of the Royal Meteorological Society* published by John Wiley & Sons Ltd on behalf of the Royal Meteorological Society.

Po Valley mostly in autumn and winter, when the conditions described above are present (Cermak *et al.*, 2009; Egli *et al.*, 2017).

Dissipation of radiation fog can take place through various pathways. After sunrise, the increasing solar radiation leads to a ‘burn-off’ of fog due to warming of air and mixing of the boundary layer (Roach, 1995; Haeffelin *et al.*, 2010; Maalick *et al.*, 2016). Short-wave radiative warming can also lead to a transition of fog into a low stratus cloud (Maalick *et al.*, 2016). Turbulent heat fluxes can contribute strongly to the loss of liquid water content in the fog layer and thus to fog dissipation (Wærsted *et al.*, 2019). Furthermore, higher wind speeds can lead to the erosion of fog at its top (Bergot, 2016), and high clouds moving over the fog can reduce radiative cooling of the cloud tops, resulting in the dissipation of the fog layer (Roach, 1995). Land surface characteristics can also modulate fog patterns, for example, over cities that often feature ‘fog holes’ over urban areas (Yan *et al.*, 2020).

The knowledge of formation and dissipation time of fog relies strongly on ground-based observational data and localized process studies with numerical models such as large-eddy simulations. These have been conducted for example in France (e.g., Haeffelin *et al.*, 2010; Dupont *et al.*, 2012; Wærsted *et al.*, 2019; Karimi, 2020) or in the Netherlands (Duynderkerke, 1991; Steeneveld and de Bode, 2018) over time-scales ranging from 6 days (Dupont *et al.*, 2012) up to 7 years (Wærsted *et al.*, 2019). According to these studies, radiation fog usually forms during the night through nocturnal cooling (Roach, 1995) and dissipates a few hours after sunrise (Haeffelin *et al.*, 2010; Bergot, 2016; Steeneveld and de Bode, 2018). While these studies provide insights on the small-scale processes of fog formation and dissipation, large-scale processes play a major role as well. Thus, data at large spatial scales are necessary to obtain knowledge on fog formation and dissipation processes across different landscapes.

Geostationary satellite data have proven suitable for the observation of fog over large spatial scales, for example, over Europe (Cermak *et al.*, 2009; Cermak and Bendix, 2011; Egli *et al.*, 2017), North America (Ellrod and Gultepe, 2007; Gultepe *et al.*, 2007b; Torregrosa *et al.*, 2016) and the Namib desert (Andersen and Cermak, 2018). They provide information about fog and low stratus clouds over a large spatial area and with a high temporal resolution. When no ground information is available, fog and low stratus (FLS) are typically treated as a single category (Cermak and Bendix, 2011). Based on Cermak and Bendix (2007; 2008) a FLS dataset has been created by Egli *et al.* (2017) which provides information about FLS cover over Europe for each 15-min interval over 10 years. This dataset has been used previously for the study of large-scale drivers of FLS using machine learning (Pauli *et al.*, 2020). Still, it has

to be considered that not all FLS situations can be detected accurately by geostationary satellite data, as small-scale FLS features, classification errors and multi-layer cloud situations can lead to misclassifications (Cermak and Bendix, 2008; Cermak, 2018).

The goal of this study is to analyse and discuss the spatial patterns of continental FLS formation and dissipation times over central Europe. Both formation and dissipation times are extracted from an existing, well-validated satellite-based binary FLS data set by Egli *et al.* (2017) with a logistic regression. The seasonal differences of the extracted formation and dissipation times are analysed across Europe and the regional applicability and detail are shown in a regional study in southern Germany. The resulting novel data set gives information on FLS formation and dissipation times over continental Europe and over 10 years, on spatial and temporal scales not previously investigated. The guiding hypothesis of this study is as follows. The timing of fog and low stratus formation and dissipation in central Europe is mainly dependent on topography. Its specific patterns are further impacted by the distance to the coast and local- to regional-scale modulations of the relevant meteorological drivers of FLS (Pauli *et al.*, 2020).

2 | DATA & METHODS

2.1 | Data

The basis of this study is an FLS data set by Egli *et al.* (2017) which covers Central Europe over the years 2006–2015. The FLS data set is created using data from the Meteosat Spinning Enhanced Visible and Infrared Imager (SEVIRI) and a number of threshold tests that are based on the Satellite-based Operational Fog Observation Scheme (SOFOS) by Cermak (2006). FLS in the Egli *et al.* (2017) data set is defined as a low stratiform cloud with liquid droplets not exceeding a size of 20 μm (Egli *et al.*, 2017) but radiation fog and other stratiform clouds are not distinguished.

The FLS data set contains binary information, that is, 1 if FLS is present and 0 if FLS is not present. The data set is available for every 15 min time slot of SEVIRI at its native resolution (3×3 km at nadir). At twilight, no FLS detection is possible, leading to ϵ FLS not present (0) values in the original data set during about four to five 15-min time steps during each twilight episode. After exploring different treatments of those twilight values (keeping them at 0, setting them to NAN (not a number), temporal interpolation), setting them to NAN has been chosen as the most suitable method in this study, as this avoids the introduction of false information. However, on the basis of the Egli

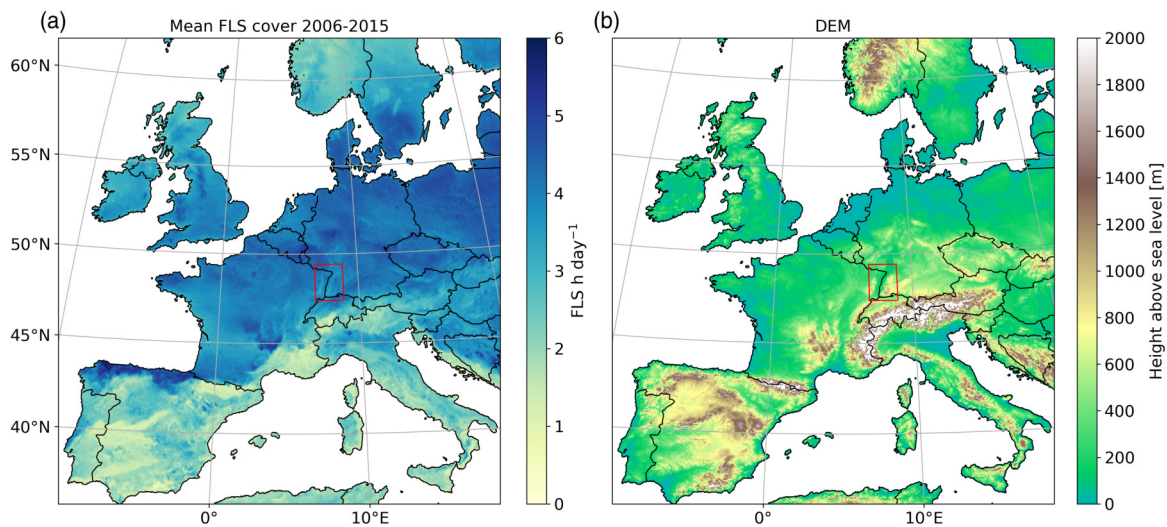


FIGURE 1 (a) Mean fog and low stratus cover in Central Europe from 2006 to 2015 (data set by Egli *et al.* (2017)) and (b) mean height above sea level. The Upper Rhine Valley, used as a regional study area for the demonstration of the algorithm and to show the spatial detail of the results, is indicated by the rectangle [Colour figure can be viewed at wileyonlinelibrary.com]

et al. (2017) data set, we are unaware of formation and dissipation time at twilight.

To provide an overview of the data used for the identification of FLS formation and dissipation time, we show the mean FLS cover in Figure 1a. The mean FLS cover over the study domain of Central Europe generally shows higher FLS frequencies over the northern latitudes and in large-scale river valleys (Figure 1a). In mountain ranges as well as in the Mediterranean, FLS frequencies are low. The dependency of FLS cover on topography can also be seen by looking at both mean FLS cover (Figure 1a) and mean height above sea level over the study area (Figure 1b). The data set can be used in various research applications, such as the investigation of large-scale drivers of FLS (Pauli *et al.*, 2020).

2.1.1 | Validation of FLS data used

In the work presented here, we use the existing and thoroughly validated FLS data set by Egli *et al.* (2017) to extract FLS formation and dissipation times by applying logistic regression. While we do create a novel FLS formation and dissipation times data set, the technique we use is not an FLS detection technique but a statistical interpretation of an already existing, thoroughly validated, satellite-based FLS data set. In this section, we provide an overview of the validation results of studies using the SOFOS algorithm for the detection of fog and low clouds (Cermak, 2006; Cermak and Bendix, 2007; 2008; Egli *et al.*, 2017). Details

on the validation procedures can be found in the relevant publications.

For the validation of the SOFOS algorithm, Meteorological Aviation Routine Weather Reports (METARs) were used, as these are available with a high temporal frequency (at least every hour) and well distributed over continental Central Europe. In general the SOFOS algorithm detects FLS very accurately (Cermak, 2006; Cermak and Bendix, 2007; 2008; Egli *et al.*, 2017). Cermak (2006) and Cermak and Bendix (2008) found that 70 to 85% of FLS situations are detected by the scheme. Situations where FLS is not detected can to a large degree be explained by overlying clouds, which are present above FLS about 25% of the time in the study domain (Cermak, 2018). False alarms are rare, that is, a pixel classified as FLS is almost never either clear or covered by a different cloud type (Cermak and Bendix, 2008). Similar to the other studies using the SOFOS algorithm, about 80–90% of FLS and no-FLS situations are classified correctly in the Egli *et al.* (2017) data set used for the extraction of FLS formation and dissipation time. The geographic patterns of validation scores show higher validation scores in continental areas with radiative FLS events as opposed to coastal areas, where advective FLS events prevail (Egli *et al.*, 2017).

We acknowledge the constraints that a satellite-based FLS climatology has when compared to ground observations. Nevertheless the good validation results show that this FLS data set can be used to derive an FLS formation and dissipation climatology as shown in the study at hand.

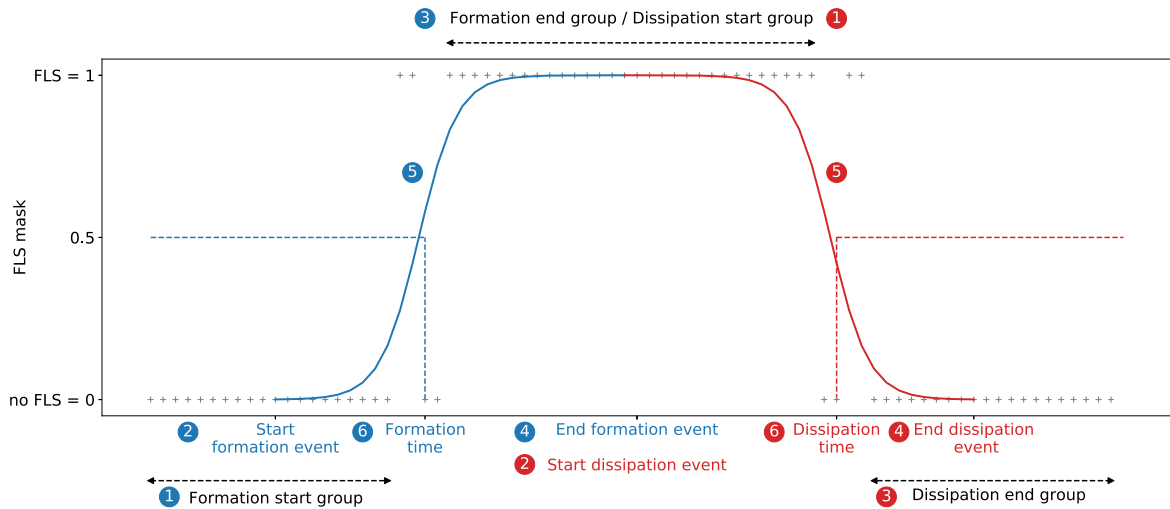


FIGURE 2 Workflow for the identification of FLS formation and dissipation times with logistic regression using an artificial example. Steps for the identification of formation time are plotted in blue, and the corresponding steps for the identification of dissipation time are plotted in red. Black crosses mark the binary FLS values [Colour figure can be viewed at wileyonlinelibrary.com]

2.2 | Methods

In this study, we apply logistic regression to statistically interpret the binary FLS data set by Egli *et al.* (2017) to determine FLS formation and dissipation times. After the algorithm is completed, a novel FLS formation and dissipation data set exists, which is derived from the original binary FLS data set by Egli *et al.* (2017). We then analyse this new data set of formation and dissipation time in the results section. In the following, we give a general overview on logistic regression and then describe the algorithm which applies logistic regression to the binary FLS time series described above.

Logistic regression is used for binary or categorical data and in this study predicts the probability of a data point belonging to one of the binary classes (Lever *et al.*, 2016; Bisong, 2019). Mathematically, the probability can be expressed as:

$$p(t) = \frac{1}{1 + \exp(-b_0 - b_t * t)}, \quad (1)$$

where p is the probability, t the time, b_0 the intercept and b_t the slope. While multiple predictor variables can be used in logistic regression, in this study, time is the only predictor. The transition from one state to another and its associated predictor value is called the “decision boundary” (Bisong, 2019). Here, this decision boundary is used to determine the time of FLS formation and dissipation.

In the atmospheric sciences, logistic regression has previously been used for the detection of hail (López and

Sánchez, 2009) or to forecast the probability of extreme precipitation events (Appelquist *et al.*, 2002; Herman and Schumacher, 2018). It has also been applied to predict the occurrence of orographic cloud cover (Wu and Zhang, 2013) or to forecast the probabilities of low visibility conditions at an airport site (Kneringer *et al.*, 2019).

2.2.1 | Algorithm

In the following we describe the algorithm which identifies FLS formation and dissipation time based on the binary cloud mask. The algorithm consists of six steps. These steps are similar for formation and dissipation time but are conducted separately. The algorithm is applied to each valid FLS event (defined below) and to each pixel over the years 2006–2015. The steps described below are marked with the corresponding numbers in Figure 2.

To apply the logistic regression to the data set, at first an FLS formation/dissipation event has to be identified (step 1). In the case of FLS formation, this means that over a period of at least ten consecutive 15-min time steps (2.5 hr) no FLS has to be present (cloud mask = 0). In the case of FLS dissipation, the requirement is similar (FLS present (cloud mask = 1) for ten consecutive 15-min time steps). By choosing a minimum value of 2.5 hr we are likely to exclude random misclassifications in the original FLS data set as well as advective FLS events with fast changes between FLS and no-FLS. One should note that the results can thus only be interpreted with respect

to such longer-lasting FLS events, and may not be representative of FLS events with a shorter life cycle. After identification of such a start group, its temporal centre point is then used as the start of the FLS formation or dissipation event (step 2). Looking at the artificial example in Figure 2, the start group of the formation event (blue curve) has 20 consecutive 15-min time steps of no-FLS. The starting point is then defined as the centre, in this case at the tenth time step of no-FLS.

The end group of a given FLS formation or dissipation event is defined similarly to the start group. In the case of FLS formation, this means that for ten consecutive 15-min time steps FLS has to be present (cloud mask = 1), and in case of dissipation FLS has to be absent for ten consecutive 15-min time steps (step 3). The end group has to be present within 30 hr from the start of the formation or dissipation event. By choosing a time period of 30 hr, it is possible to capture FLS events that persist for longer than a day. If no end group can be identified within 30 hr of the start of the formation/dissipation event (step 2), the formation/dissipation event is omitted, and no time is determined by the algorithm.

Similar to the determination of the start of the formation/dissipation event, the end of such an event is then set to be the centre of the corresponding end group (step 4). In the artificial example in Figure 2, the end group of the formation event contains 30 15-min time steps with FLS, with the end point of formation defined in the centre of the group at time step 15. During twilight, the groups of consecutive FLS or no-FLS are interrupted by NAN values as outlined in Section 2.1. If the remaining groups then do not fulfill the requirement of ten consecutive FLS or no-FLS values, this leads to a potential loss of start and end groups.

Following the identification of the start and end points of the FLS formation/dissipation events, logistic regression is applied to the binary time series between the start and end points to model the probability of FLS presence (step 5). Then, the decision boundary is used to depict the formation and dissipation times (step 6). For formation, the first 15-min time period in which the probability modelled by the logistic regression exceeds 0.5 is defined as the formation time. For dissipation, the first 15-min time period where the probability falls below 0.5 is defined as the dissipation time.

After the calculation of the formation/dissipation time, these six steps are repeated for each identified start group for each pixel separately. While the formation and dissipation situation displayed in Figure 2 can be considered as an ideal example, with a dissipation event following a formation event right away, there is also the possibility that a dissipation event is not identified right after a formation event. This occurs when an end of the dissipation event cannot be identified within 30 hr, due to alternating FLS

and no-FLS values. If the requirements discussed above are not fulfilled by either a potential formation or dissipation event, it is possible that multiple FLS formation or dissipation events are identified one after another. This leads to slight differences in the number of formation and dissipation events identified by the algorithm, which is discussed in Section 3.1.

2.2.2 | Discussion of algorithm uncertainties

It is clear that the quality of the derived FLS formation and dissipation time products is dependent on the quality of the underlying FLS data set. Therefore, systematic and random errors in the FLS data set will propagate to our classification. However, by focusing on temporally persistent FLS events, we reduce effects of random misclassifications in the data set. Another uncertainty of the derived FLS formation and dissipation time products is introduced by differences in the characteristics of the binary variability of each specific FLS event (Figure 3). An abrupt, temporally coherent change between the binary classes (little alternation between FLS and no-FLS values), leads to a higher absolute value of the slope and a steep curve fit. This leads to either the classic sigmoid shape (Figure 3a, d) or a very steep transition (Figure 3b, e), both of which are easy to interpret with respect to FLS formation and dissipation time. However, in the case of frequent changes between FLS and no-FLS values of an FLS event considered, the fitted logistic curve is flat and the slope value approaches zero (Figure 3c, f). These flat curves are difficult to interpret with respect to an FLS formation/dissipation time, and are thus excluded from the data set. We define these valid situations to feature a slope of $> |0.1|$. This threshold was defined by conducting a thorough visual analysis of different events and their corresponding curve shapes. The exclusion of these high-uncertainty events leads to an average reduction of 9% of FLS events per pixel.

2.3 | Case illustration: 7 February 2011

To illustrate how formation and dissipation time are derived from the binary FLS masks, a dissipation case from 7 February 2011 over the Upper Rhine Valley in Germany is presented in the following. The mean FLS cover and topography of the region are shown in Figure 4.

Radiation fog is a frequent phenomenon in the Upper Rhine Valley, especially in the colder months (Kalthoff *et al.*, 1998; Bendix, 2002; Bendix *et al.*, 2006; Egli *et al.*, 2017). In this rift valley between the Vosges mountains to the west and the Black Forest mountains to the east (visible in Figure 4b and roughly depicted in Figure 4 with the

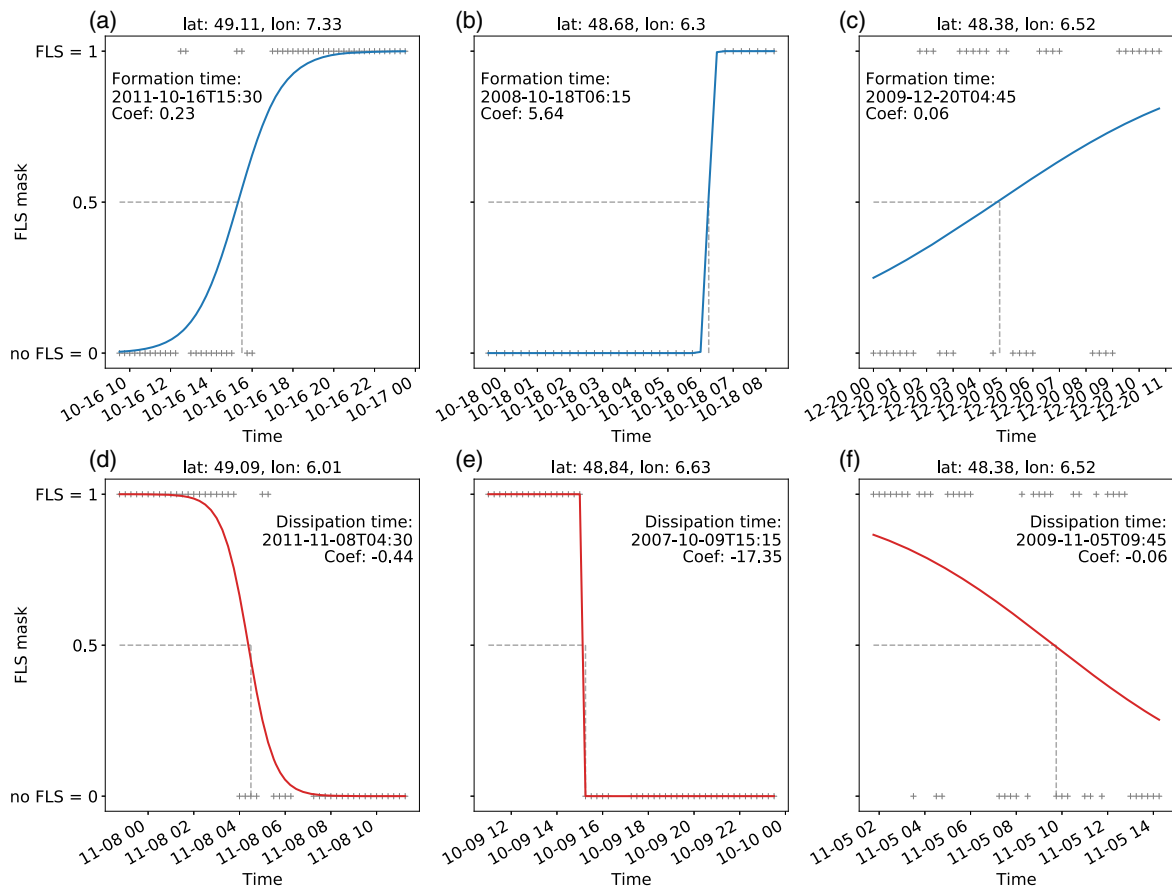


FIGURE 3 Logistic regression curves for three formation situations (a–c) and three dissipation situations (d–f). The situations in (a, d) have a classic sigmoid shape and the situations in (b, e) have a sharp transition. The curves in (c, f) do not fulfill the slope requirements (coefficient > |0.1|). The Coef parameter denotes the slope of the logistic regression curve [Colour figure can be viewed at wileyonlinelibrary.com]

500 m contour), FLS occurs more frequently in the valley than on the mountain ridges (Figure 4a).

The case illustration of 7 February 2011 focuses on three pixels in the Upper Rhine Valley (at locations A, B, C), all of them located in France. As can be seen in the binary FLS maps (Figure 5), FLS is present mainly in the northern part of the valley on the morning of 7 February and then dissipates during the day. To illustrate the derivation of the dissipation time, the binary time series at the three locations is extracted (Figure 6). A logistic regression is calculated for all time series and the time stamp where the probability of the binary value being equal to 1 falls below 0.5 is assigned to be the dissipation time. In line with the binary FLS masks in Figure 5, FLS at location B dissipates first (0930 UTC) followed by location A (dissipation at 1145 UTC). FLS at location C is most stable and dissipates in the afternoon (1430 UTC).

The approach presented here for the identification of dissipation time at the three locations is then applied to each pixel of the study area, for each identified FLS formation and dissipation event.

3 | RESULTS

3.1 | Number and duration of FLS events

As a first view of the data set, the number of formation events for which a formation time was calculated is shown (Figure 7a). The number of formation events is representative of the number of dissipation events, which is very similar with a difference around ± 10 events per pixel. As discussed above, this slight difference is present since not every formation event is followed by a dissipation

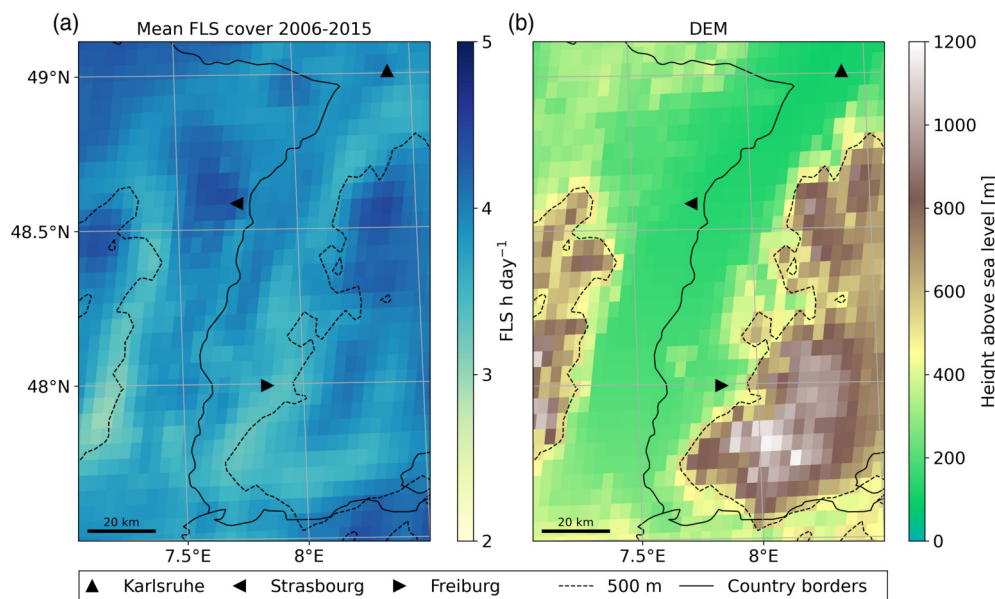


FIGURE 4 (a) Mean fog and low stratus cover in the Upper Rhine Valley from 2006 to 2015 (data set by Egli *et al.* (2017)) and (b) mean height above sea level. The country border between France and Germany also marks the course of the river Rhine. The black dashed line depicts 500 m height above mean sea level [Colour figure can be viewed at wileyonlinelibrary.com]

event which fulfills the requirements discussed above, and *vice versa*. The geographic patterns of the number of detected formation events are similar to the geographic patterns of mean FLS cover (Figure 1). A high number of formation events is identified in the northeastern part of the study area (1,000–1,200 events over 10 years), as well as in the Po Valley and at the north coast of Spain. Moreover, a higher number of formation events is detected over cities, especially visible over Milan and Paris. This is likely a systematic error of the FLS data set used, as higher FLS cover over cities can also be seen in Figure 1a. The overestimation of FLS cover over cities is likely due to their high reflectance in the $3.9 \mu\text{m}$ channel due to solar contamination during daytime, which is likely to lead to a false identification of clouds during this time (EUMETSAT, 2019). Higher daytime FLS cover over urban pixels compared to the rural surroundings can be observed in the mean diurnal course of FLS cover over London, Paris and Milan (Figure S1).

A low number of formation events (300–400 events over 10 years) is detected over the Mediterranean and over mountain ranges (Alps, Pyrenees). The same geographic patterns described here for the number of formation events apply to the number of dissipation events.

The FLS duration is calculated as the difference between dissipation time and formation time for one FLS event for each pixel. To decrease the influence of

outliers, the median instead of the mean is shown here. The geographic patterns of duration are similar to the FLS mean map and the number of dissipation events. FLS events tend to be longer in the northeastern part of the study area (400–600 min) and shorter in the Mediterranean (200–300 min). The geographic patterns of median duration are also similar to those of the mean FLS cover and the number of formation events. This is also visible when normalizing the three quantities (Figure S2) to enhance comparability. Still, the spread of values is lower for the median duration than the mean FLS cover and the number of formation events.

3.2 | Most frequent formation and dissipation times

For further analysis, the timestamps of formation and dissipation times are converted into % values as a function of day length (sunrise–sunset) for formation or dissipation during the day or as a function of night length (sunset–sunrise) for nighttime formation and dissipation. This is done to make formation and dissipation time comparable across seasons and latitudes.

To facilitate interpretation, these % values are then assigned to different classes of day- or nighttime (Table 1). To produce climatological maps of the most frequent

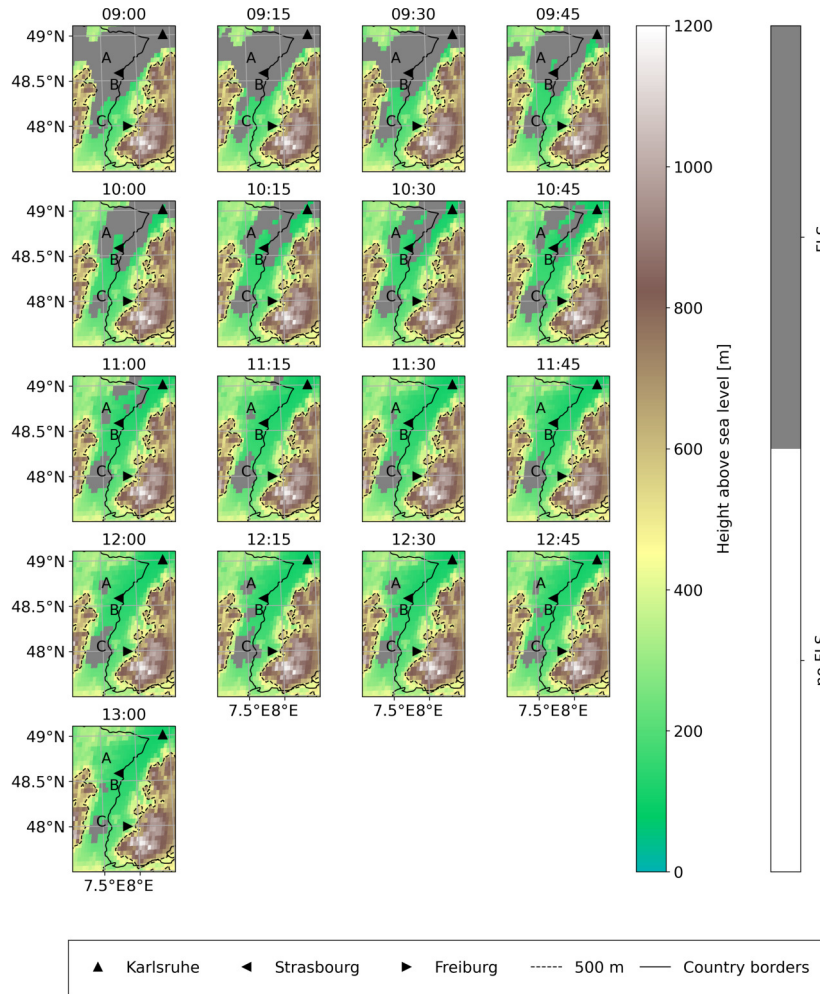


FIGURE 5 Binary 15-min FLS masks (data set by Egli *et al.* (2017)) from 0900 to 1300 UTC on 7 February 2011. Pixels with FLS cover are grey, pixels without FLS are transparent. The background is a topographic map of the region. The country border between France and Germany also marks the course of the river Rhine. The black dashed line depicts 500 m height above mean sea level [Colour figure can be viewed at wileyonlinelibrary.com]

formation and dissipation time, the mode of these classes is used for each pixel over the entire time period (Figure 8) and depending on the season (Figure 9).

When considering the entire study period (Figure 8), FLS forms most frequently around midnight in large parts of the study area such as areas of France, Germany and Italy, where secondary mountain ranges as well as rather flat areas, such as the Danube valley, are present. In the high-altitude mountain ranges (Alps, Pyrenees) and on the north coast of Spain, formation is most frequently around sunset. In the inner plateau of central Spain, western France and in the Po Valley, FLS forms most frequently around sunrise and in the morning.

The geographic distribution of most frequent dissipation time shows similar spatial patterns of equal dissipation time as was seen for formation time. On the north coast of Spain, and in high-altitude mountains such as the Pyrenees and the Alps, FLS dissipates most frequently

around sunrise or in the morning. In secondary mountain ranges such as the Massif Central in southern France and the pre-alpine areas of southern Germany, dissipation occurs mainly in the morning or around midday. In the low-altitude areas of central Europe, such as northern France, northern Germany and Poland but also in the inner plateau of Central Spain, dissipation occurs most frequently in the afternoon.

The seasonal patterns of formation and dissipation times tend to be similar in winter (DJF) and autumn (SON) and in spring (MAM) and summer (JJA) (Figure 9). In winter and autumn, FLS formation occurs most frequently in the evening or during the night and dissipation around midday or in the afternoon. In spring and summer FLS typically forms later, most frequently around sunrise, but does not last as long as during winter and dissipates in the morning. The detailed geographic patterns for each season are described below.

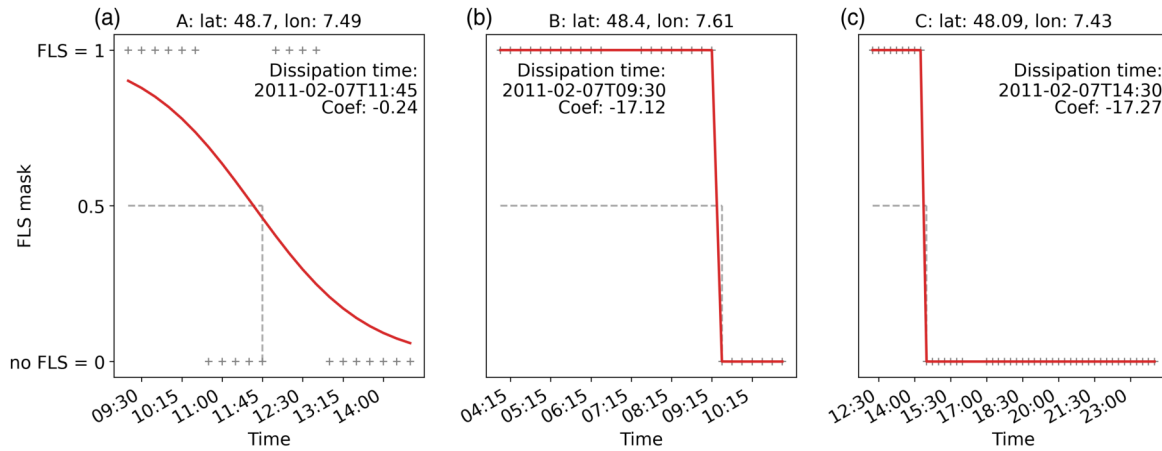


FIGURE 6 Logistic regression curves for the dissipation events on 7 February 2011 at the three locations A, B, C marked in Figure 5. The Coef parameter denotes the slope of the logistic regression curve [Colour figure can be viewed at wileyonlinelibrary.com]

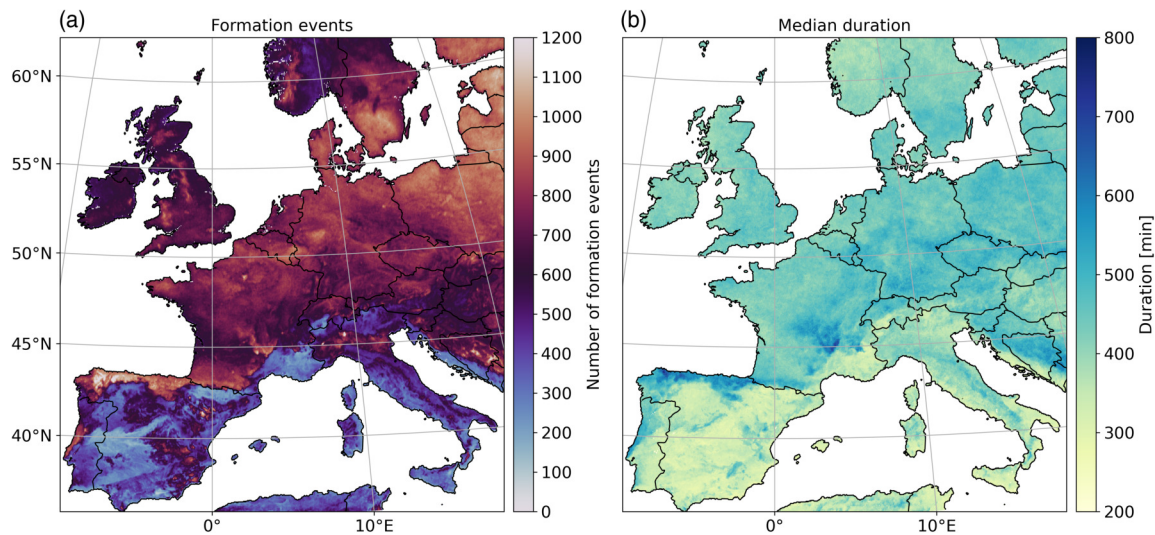


FIGURE 7 (a) Number of formation events identified by the algorithm, and (b) the median of the duration of all FLS events in which both formation and dissipation times could be determined. Both are calculated over the entire period (2006–2015) [Colour figure can be viewed at wileyonlinelibrary.com]

In winter FLS forms most frequently in the evening or around midnight, except for large areas in western France and the inner plateau of central Spain, where formation most frequently takes place in the morning. During this time, FLS dissipates most frequently in the afternoon in most regions, and earlier (in the morning or around midday) in parts of the Alps, in the secondary mountain ranges of southern Germany or southern Italy.

In spring the most frequent formation time shifts to sunrise or to the morning hours in large parts of France, central Spain and in the Po Valley. On the north coast of

Spain, southern France, in the Pyrenees and in the Alps, FLS forms most frequently at sunset. Dissipation most often takes place in the morning or around midday in most parts of the study area. In the Po Valley and in large parts of Poland, FLS is more persistent and most frequently dissipates in the afternoon.

In summer FLS formation patterns are spatially diverse, with formation at sunset (central Germany), around sunrise (Po Valley) or in the morning (France). This could be influenced by to the lower number of formation events in summer compared to other seasons in most

TABLE 1 Day- and nighttime classes with respective % ranges

Daytime			Nighttime		
Class number	Class name	% Range	Class number	Class name	% Range
1	Sunrise	0–12.4	1	Sunrise	87.5–100
2	Morning	12.5–37.4	8	Night	62.5–87.4
3	Midday	37.5–62.4	7	Midnight	37.5–62.4
4	Afternoon	62.5–87.4	6	Evening	12.5–37.4
5	Sunset	87.5–100	5	Sunset	0–12.4

parts of the study area (compare Figure S3). A clear pattern of formation at sunrise is visible at the coast of the Mediterranean Sea. Dissipation in summer is most frequently in the morning in large parts of the central study area. FLS dissipates earlier (around sunrise) at the northern coast of Spain and in the Alps, and later (in the afternoon) in western France. Still, these patterns should be interpreted with care, as FLS occurrence is low in most of the Mediterranean in spring and summer (compare hatched areas of Figure 9).

In autumn FLS formation shifts again towards nighttime hours in most of the study area, except for parts of France and Spain, where FLS forms most frequently

in the morning. FLS dissipation also shifts to afternoon hours in most parts of the study area. In the Alps and the Pyrenees, FLS most frequently dissipates at sunrise, in southern Germany around midday.

3.3 | Regional study: Upper Rhine Valley, Germany

While the formation and dissipation times data set can be used to produce climatological maps for Europe, it can also be used for a more thorough investigation of formation and dissipation patterns in regional studies. Such a regional study is presented here by looking closely at formation and dissipation patterns in the Upper Rhine Valley (red rectangle in Figure 1 and introduced in Section 2.3) in southern Germany. The FLS patterns present in that region are likely to be influenced by local to regional modulations of the synoptic-scale weather patterns. Thus the Upper Rhine Valley is an ideal region to showcase the level of spatial detail provided by the novel data set.

The most frequent formation and dissipation time over the complete period (Figure 10a) shows two distinct patches of formation and dissipation anomalies (at $\sim 7.5^\circ\text{E}$, 48°N , and at around 48.7°N). At these locations, formation is observed to be in the morning and dissipation in the

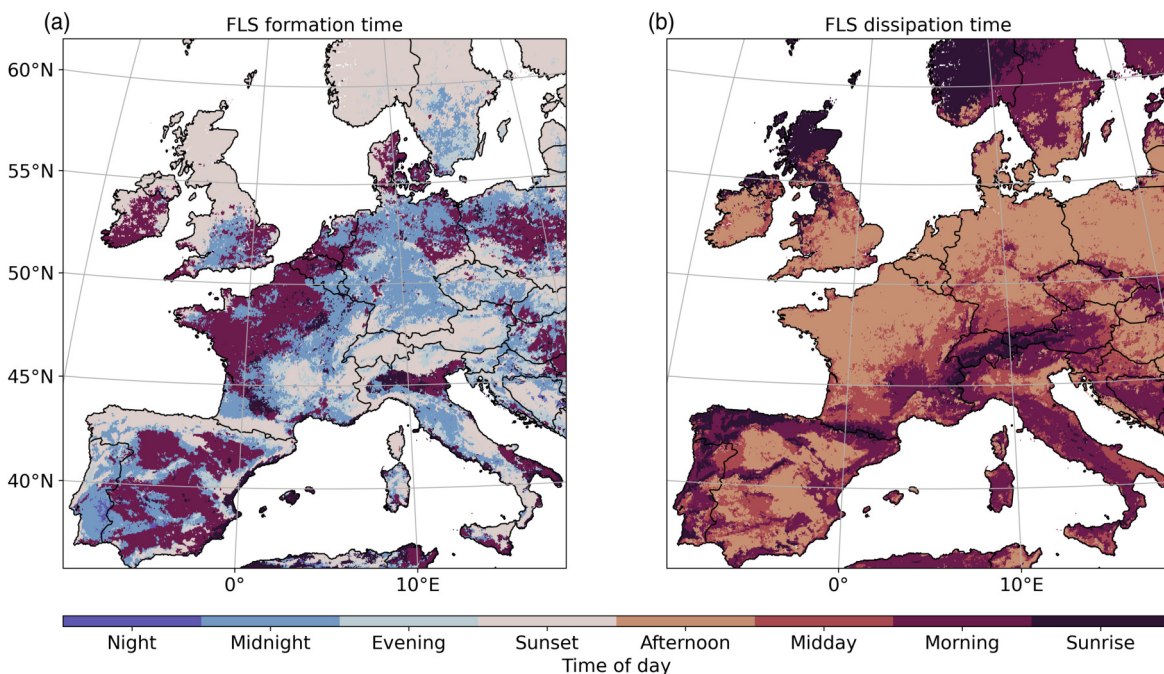


FIGURE 8 Most frequent (a) formation and (b) dissipation times over the entire study period (2006–2015) [Colour figure can be viewed at wileyonlinelibrary.com]

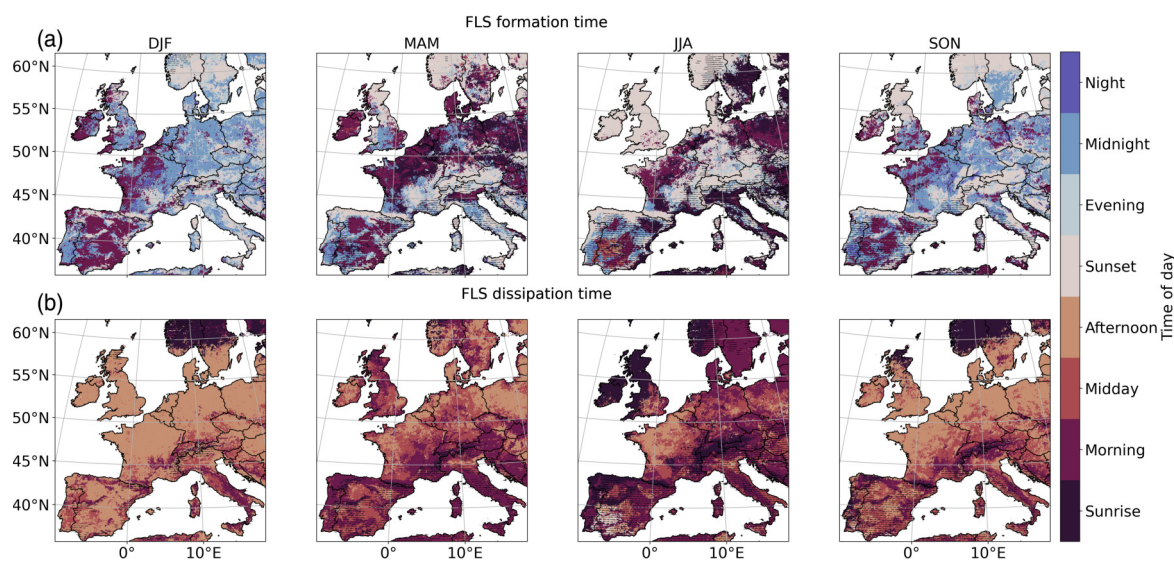


FIGURE 9 Most frequent (a) formation and (b) dissipation times for DJF, MAM, JJA and SON from 2006 to 2015. Pixels with a mean seasonal FLS occurrence of less than 2 hr-day^{-1} are marked with a horizontal line [Colour figure can be viewed at wileyonlinelibrary.com]

afternoon (Figure 10a, e). Thus FLS formation and dissipation at these patches are much later than in other places of the Upper Rhine Valley, where FLS usually forms around sunset or around midnight and dissipates in the morning or around midday. While the anomalies are clearly visible in the annual averages, they are especially pronounced in MAM, with formation in the morning and dissipation in the afternoon extending over the eastern slopes of the Vosges (Figure 10b, f).

The percentage of values that fall into the most frequent formation and dissipation time classes relative to all formation and dissipation situations provides a measure of the representativeness of the mode as a proxy for the typical formation and dissipation time for each pixel (Figure 10c, d, g, h). The percentage of values in the most frequent class lies around 15–20% for the formation over the complete year and rises to values around 20–25% in MAM and is highest in the anomalous patches described above. Considering the dissipation, the geographic distribution of % values is similar but on average 10% higher than formation, showing that the temporal variability in dissipation time is lower than the formation time.

This case-study provides two important insights into the novel formation and dissipation data set. First of all, the relatively high spatial and temporal resolution (native SEVIRI resolution; Section 2.1) make it a useful product to study regional formation and dissipation patterns. Secondly, the dissipation time features a lower temporal variability than the formation time, as evidenced by

the systematically higher fraction of events in the most frequent class. Over most of the study area, more than two thirds of dissipation events occur during the day, whereas formation time is equally distributed between daytime and nighttime in large parts of the study area (Figure S4). This is likely due to a higher number of possible formation pathways and thus formation times, whereas dissipation is strongly influenced by solar radiation. We evaluate this further in the discussion below.

4 | DISCUSSION

The geographic patterns of formation and dissipation time clearly underline the role of topography for the occurrence and development of FLS events. In large mountain ranges such as the Alps or the Pyrenees, FLS forms earlier (around sunset) than in lower terrain but also dissipates earlier (around sunrise). These geographic patterns may be interpreted as the signature of regionally characteristic processes influencing FLS formation and dissipation. It is likely that formation at sunset in those mountain ranges is due to advected FLS layers, especially on the windward slopes of those ranges, for example on the northern slopes of the Pyrenees. After sunrise, these FLS layers are likely to be ‘burned-off’ as the sun reaches the mountain tops first. In the Po Valley in Italy or in the inner plateau of Central Spain, FLS formation is likely to be due to nocturnal cooling and a subsequent transition into a low stratus cloud

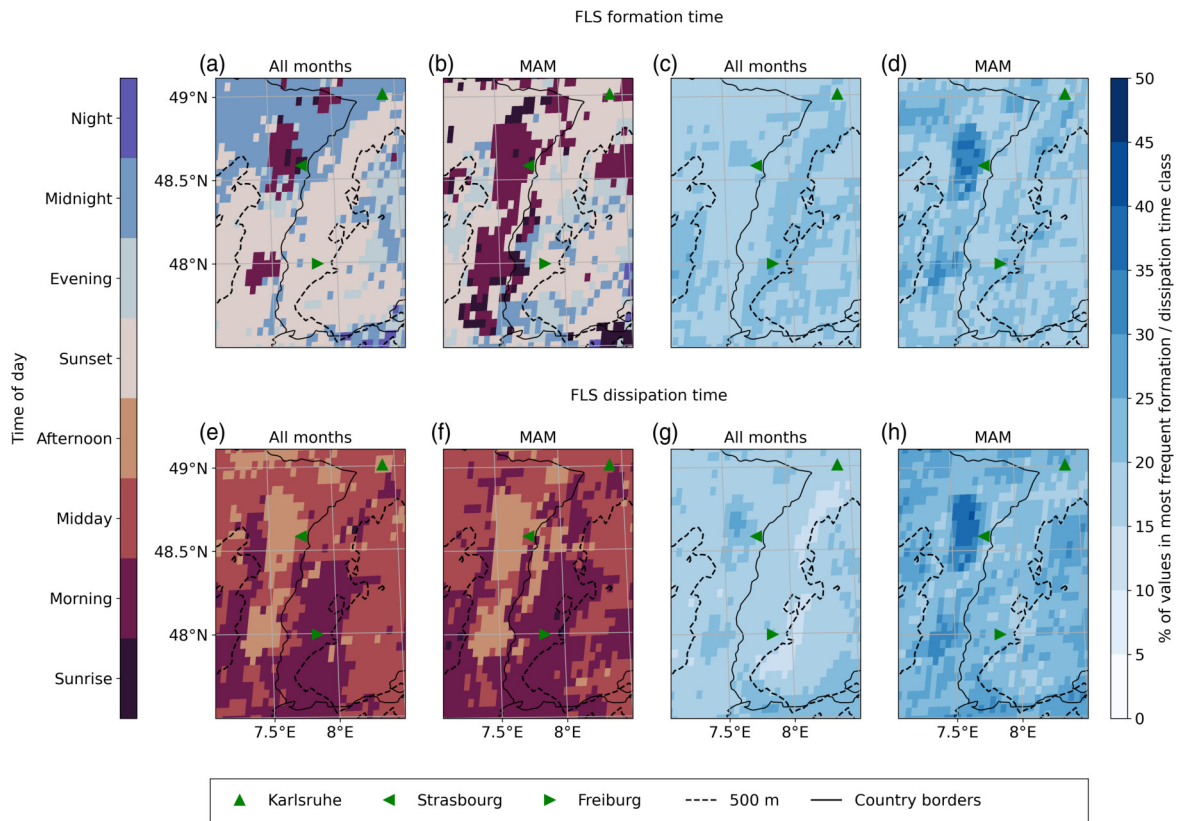


FIGURE 10 Most frequent (a, b) formation and (e, f) dissipation times in the Upper Rhine Valley and the corresponding % of values contained in the most frequent (c, d) formation and (g, h) dissipation class for the complete time period (all months) and in spring (MAM). The country border between France and Germany also marks the course of the river Rhine. The black dashed line depicts 500 m height above mean sea level [Colour figure can be viewed at wileyonlinelibrary.com]

(Roach, 1995). Other potential processes involved could be topography-induced drainage flows and turbulence (Price, 2019).

In smaller mountain ranges (e.g., the Black Forest), the dependence of formation and dissipation time on topography is not always clear. A potential explanation for this could be that the local modulation of the meteorological parameters that determine the FLS life cycle (e.g., relative humidity, wind) scales with the topographic features. In areas with moderate topography, the topography-induced local modulation of the meteorological setting would then have a weaker effect on the resulting geographic patterns of formation and dissipation times than in mountainous areas. In general, topographic features are only modulating the meteorological drivers responsible for FLS formation and dissipation. As has been shown in previous studies (Pauli *et al.*, 2020) meteorological factors are the main drivers of FLS occurrence over central Europe. In addition, the presented formation and dissipation times should be

interpreted with care over areas with a low sample size or high topographic variability, as the pixel size of the product (3–5 km, depending on the exact position) is too large to be able to depict small-scale variations in FLS.

Besides topography, the proximity to the sea is also important for the timing of FLS formation and dissipation. In general, coastal fog is strongly influenced by the meteorological conditions and ocean–atmosphere interactions (Gultepe *et al.*, 2021). The results presented here show a clear pattern of FLS formation at sunrise at the coast of the Mediterranean Sea in summer (Figure 9). According to Azorin-Molina *et al.* (2009) humid winds from the Mediterranean Sea in combination with mountain ranges close to the coastline can lead to FLS occurrence at the Mediterranean coast. Therefore, the diurnal cycle of the coastal circulation is likely a main driver of the observed patterns in FLS formation and dissipation times at the Mediterranean coast. At the Atlantic coast of northern Spain, blocking of FLS by the Cantabrian mountains (Egli

et al., 2017) and upwelling (Alvarez *et al.*, 2010) might be important for FLS formation, as the latter plays an important role in other FLS-prone regions such as at the southwestern African coast (Olivier and Stockton, 1989; Cermak, 2012; Andersen *et al.*, 2020). Close to the French Atlantic coast over the Landes forest south of 45°N, FLS forms earlier over the forest (around midnight) than over its surroundings (FLS formation in the morning), particularly in spring and summer. This is in line with enhanced nighttime FLS occurrence over this forest compared to its surroundings shown in Pauli *et al.* (2022), potentially due to enhanced emissions of biogenic volatile organic compounds over the forest, which can serve as cloud condensation nuclei (Spracklen *et al.*, 2008; Kammer *et al.*, 2018).

Another driving factor for the observed patterns is the solar radiation. As stated above, this is especially true for the dissipation time, where processes related to downwelling solar radiation (absorption inside the FLS layer, sensible heat flux) more strongly influence the dissipation of FLS compared to other, more subtle, drivers (Wærsted *et al.*, 2019). Formation can occur through various pathways during the day or during the night with formation patterns showing a much higher variability. To focus on specific FLS events and to unravel different formation and dissipation pathways, one could filter for stationary FLS events using image detection techniques or filter for meteorological conditions.

The higher number of formation and dissipation events over cities does not transfer to a difference of formation and dissipation time between cities and the surrounding land. In contrast to the literature (cf. Yan *et al.* (2020)) earlier dissipation over cities is not visible in the shown climatological means. On the one hand, this may be due to the assignment of the dissipation and formation times in % to different daytime and nighttime classes (Table 1) and the subsequent calculation of the mode, which could be investigated in more detail by looking at the raw formation and dissipation times over cities and surrounding areas. On the other hand, as discussed above, the FLS data set is likely to be flawed over cities during daytime, and thus misclassifications are likely to superimpose the actual patterns in the first place. Applying the proposed logistic regression algorithm to a robust high-resolution cloud mask over cities (Fuchs *et al.*, 2022) could add to our knowledge on the FLS life cycle over urban areas.

As these discussions on the possible processes underlying the geographic formation and dissipation patterns remain speculative, explicit regional analyses on the drivers of FLS formation and dissipation time are necessary. In addition, the potential influence of multilayer cloud situations and misclassifications on the presented FLS formation and dissipation time have to be considered

for the interpretation of the results. In a regional study, the FLS data set by Egli *et al.* (2017) used as a basis for this study has been shown to agree well with active sensor data (Pauli *et al.*, 2022). The formation and dissipation time patterns shown with this novel FLS formation and dissipation data set are also in line with LES and modelling studies over Europe (cf. Roach, 1995; Haeffelin *et al.*, 2010; Bergot, 2016; Steeneveld and de Bode, 2018), showing FLS formation in the night and dissipation after sunrise. While modelling, LES studies and local measurements display the FLS life cycle with high temporal resolution at a specific site and can also include the vertical component of an FLS event, the data set presented here provides a geographic perspective on formation and dissipation time over the whole of Central Europe.

5 | CONCLUSIONS

The central aim of this study was to investigate spatial patterns of FLS formation and dissipation times over Central Europe. For this purpose, an algorithm was designed which applies logistic regression to a binary satellite-based FLS cloud mask. With the novel data set, FLS formation and dissipation times were investigated, largely confirming known patterns of formation during the night and dissipation in the morning or in the afternoon. In general, FLS occurrence, formation and dissipation are dependent on various drivers (cf. Roach, 1995; Gultepe *et al.*, 2007a; Price, 2019; Pauli *et al.*, 2020). The results presented here underline the importance of topography-induced modulation of meteorology for FLS formation and dissipation. In mountain ranges, FLS forms most frequently at sunset and dissipates in the morning. At lower altitudes, such as in large-scale river valleys, FLS most commonly forms around sunrise and dissipates in the afternoon. Furthermore, we find a higher variability in formation times than in dissipation times, with the latter being much more dependent on solar radiation. The data set adds a geographic component to our knowledge of FLS formation and dissipation and provides a basis for future studies.

In the future, a preprocessing step could be implemented by adding a filter to study more specifically stationary FLS situations which may be indicative of radiation fog, for example using image detection techniques or focusing on specific meteorological conditions. The new algorithm can also be applied to other existing satellite-based FLS data sets with a high temporal resolution over regions where FLS are an important component of the climate and environment and station measurements are lacking, such as in the Namib Desert (Andersen and Cermak, 2018), and can be compared to diurnal patterns identified there so far (Andersen *et al.*, 2019). Furthermore

this data set holds promise to help better understand the drivers of FLS formation and dissipation at continental scales.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers and Stephanie West-erhuis for their careful reviews which have helped improve the manuscript. E.P. has been financially supported by the Graduate Funding of the German States Program.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Eva Pauli: Conceptualization, formal analysis, investigation, methodology, visualization, writing (original draft), writing (review and editing). **Jan Cermak:** Conceptualization, methodology, writing (review and editing). **Hendrik Andersen:** Methodology, writing (review and editing).

DATA AVAILABILITY STATEMENT

The fog and low stratus cloud data set serving as a basis for the algorithm can be downloaded from http://vhrz669.hrz.uni-marburg.de/lcrs/data_pre.do?citid=291. The formation and dissipation time data set is available at <https://doi.org/10.5445/IR/1000141293> (both accessed 25 March 2022).

ORCID

Eva Pauli  <https://orcid.org/0000-0002-5506-4736>

Jan Cermak  <https://orcid.org/0000-0002-4240-595X>

Hendrik Andersen  <https://orcid.org/0000-0003-2983-8838>

REFERENCES

- Alvarez, I., Gomez-Gesteira, M., deCastro, M., Gomez-Gesteira, J. and Dias, J. (2010) Summer upwelling frequency along the western Cantabrian coast from 1967 to 2007. *Journal of Marine Systems*, 79(1), 218–226. <https://doi.org/10.1016/j.jmarsys.2009.09.004>.
- American Meteorological Society (2012). Fog, in Glossary of Meteorology. Boston, MA: American Meteorological Society. <http://glossary.ametsoc.org/wiki/fog>.
- Andersen, H. and Cermak, J. (2018) First fully diurnal fog and low cloud satellite detection reveals life cycle in the Namib. *Atmospheric Measurement Techniques*, 11(10), 5461–5470. <https://doi.org/10.5194/amt-11-5461-2018>.
- Andersen, H., Cermak, J., Solodovnik, I., Lelli, L. and Vogt, R. (2019) Spatiotemporal dynamics of fog and low clouds in the Namib unveiled with ground- and space-based observations. *Atmospheric Chemistry and Physics*, 19(7), 4383–4392. <https://doi.org/10.5194/acp-19-4383-2019>.
- Andersen, H., Cermak, J., Fuchs, J., Knippertz, P., Gaetani, M., Quinting, J., Sippel, S. and Vogt, R. (2020) Synoptic-scale controls of fog and low-cloud variability in the Namib Desert. *Atmospheric Chemistry and Physics*, 20(6), 3415–3438. <https://doi.org/10.5194/acp-20-3415-2020>.
- Applequist, S., Gahrs, G.E., Pfeffer, R.L. and Niu, X.-F. (2002) Comparison of methodologies for probabilistic quantitative precipitation forecasting. *Weather and Forecasting*, 17(4), 783–799. [https://doi.org/10.1175/1520-0434\(2002\)017<0783:COMFPQ>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0783:COMFPQ>2.0.CO;2).
- Azorin-Molina, C., Sanchez-Lorenzo, A. and Calbo, J. (2009) A climatological study of sea breeze clouds in the southeast of the Iberian Peninsula (Alicante, Spain). *Atmosfera*, 22(1), 33–49.
- Bendix, J. (1994) Fog climatology of the Po Valley. *Rivista di meteorologia aeronautica*, 54(3-4), 25–36.
- Bendix, J. (2002) A satellite-based climatology of fog and low-level stratus in Germany and adjacent areas. *Atmospheric Research*, 64(1), 3–18. [https://doi.org/10.1016/S0169-8095\(02\)00075-3](https://doi.org/10.1016/S0169-8095(02)00075-3).
- Bendix, J., Thies, B., Nauss, T. and Cermak, J. (2006) A feasibility study of daytime fog and low stratus detection with TERRA/AQUA-MODIS over land. *Meteorological Applications*, 13(2), 111–125. <https://doi.org/10.1017/S1350482706002180>.
- Bergot, T. (2016) Large-eddy simulation study of the dissipation of radiation fog. *Quarterly Journal of the Royal Meteorological Society*, 142, 1029–1040. <https://doi.org/10.1002/qj.2706>
- Bisong, E. (2019). Logistic Regression, pp 243–250 in Building Machine Learning and Deep Learning Models on Google Cloud Platform. Berkeley, CA: Apress, DOI 10.1007/978-1-4842-4470-8_20, (to appear in print).
- Bruijnzeel, L.A., Eugster, W. and Burkard, R. (2006). Fog as a hydrologic input, in Encyclopedia of Hydrological Sciences, Anderson, M., McDonnell, J. (eds), DOI 10.1002/0470848944.hsa041, (to appear in print).
- Cermak, J. (2006). SOFOS – A new satellite-based operational fog observation scheme. PhD thesis, Philipps-Universität Marburg, Germany.
- Cermak, J. (2012) Low clouds and fog along the southwestern African coast – Satellite-based retrieval and spatial patterns. *Atmospheric Research*, 116, 15–21. <https://doi.org/10.1016/j.atmosres.2011.02.012>.
- Cermak, J. (2018) Fog and low cloud frequency and properties from active-sensor satellite data. *Remote Sensing*, 10(8), 1209. <https://doi.org/10.3390/rs10081209>.
- Cermak, J. and Bendix, J. (2007) Dynamical nighttime fog/low stratus detection based on Meteosat SEVIRI data: a feasibility study. *Pure and Applied Geophysics*, 164(6-7), 1179–1192. <https://doi.org/10.1007/s00024-007-0213-8>.
- Cermak, J. and Bendix, J. (2008) A novel approach to fog/low stratus detection using Meteosat 8 data. *Atmospheric Research*, 87(3), 279–292. <https://doi.org/10.1016/j.atmosres.2007.11.009>.
- Cermak, J. and Bendix, J. (2011) Detecting ground fog from space – a microphysics-based approach. *International Journal of Remote Sensing*, 32(12), 3345–3371. <https://doi.org/10.1080/01431161003747505>.
- Cermak, J., Eastman, R.M., Bendix, J. and Warren, S.G. (2009) European climatology of fog and low stratus based on geostationary satellite observations. *Quarterly Journal of the Royal*

- Meteorological Society*, 135, 2125–2130. <https://doi.org/10.1002/qj.503>.
- Dupont, J.-C., Haeffelin, M., Protat, A., Bouniol, D., Boyouk, N. and Morille, Y. (2012) Stratus-fog formation and dissipation: a 6-day case study. *Boundary-Layer Meteorology*, 143(1), 207–225. <https://doi.org/10.1007/s10546-012-9699-4>.
- Duynkerke, P.G. (1991) Radiation fog: a comparison of model simulation with detailed observations. *Monthly Weather Review*, 119(2), 324–341. [https://doi.org/10.1175/1520-0493\(1991\)119<0324:RFACOM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<0324:RFACOM>2.0.CO;2).
- Egli, S., Thies, B., Dröner, J., Cermak, J. and Bendix, J. (2017) A 10-year fog and low stratus climatology for Europe based on Meteosat Second Generation data. *Quarterly Journal of the Royal Meteorological Society*, 143, 530–541. <https://doi.org/10.1002/qj.2941>.
- Ellrod, G.P. and Gultepe, I. (2007). Inferring low cloud-base heights at night for aviation using satellite infrared and surface temperature data, pp. 1193–1205 in *Fog and Boundary Layer Clouds: Fog Visibility and Forecasting*, Gultepe, I. (ed.) Basel, Switzerland: Birkhäuser.
- EUMETSAT (2019). Algorithm Theoretical Basis Document (ATBD) Meteosat Solar Surface Radiation and effective Cloud Albedo Climate Data Records – Heliosat (SARAH-2), Darmstadt, Germany. Technical Report.
- Fuchs, J., Andersen, H., Cermak, J., Pauli, E. and Roebeling, R. (2022) High-resolution satellite-based cloud detection for the analysis of land surface effects on boundary-layer clouds. *Atmospheric Measurement Techniques Discussions*, 2022, 1–22. <https://doi.org/10.5194/amt-2022-36>.
- Fuzzi, S., Facchini, M.C., Orsi, G., Lind, J.A., Wobrock, W., Kessel, M., Maser, R., Jaeschke, W., Enderle, K.H., Arends, B.G., Berner, A., Solly, I., Krusiz, C., Reischl, G., Pahl, S., Kaminski, U., Winkler, P., Ogren, J.A., Noone, K.J., Hallberg, A., Fierlinger-Oberlininger, H., Puxbaum, H., Marzorati, A., Hansson, H.-C., Wiedensohler, A., Svenningsson, I.B., Martinsson, B.G., Schell, D. and Georgii, H.W. (1992) The Po Valley fog experiment 1989. *Tellus B*, 44(5), 448–468. <https://doi.org/10.1034/j.1600-0889.1992.t01-4-00002.x>.
- Gottlieb, T.R., Eckardt, F.D., Venter, Z.S. and Cramer, M.D. (2019) The contribution of fog to water and nutrient supply to *Arthroa leubnitziae* in the central Namib Desert, Namibia. *Journal of Arid Environments*, 161, 35–46. <https://doi.org/10.1016/j.jaridenv.2018.11.002>.
- Gultepe, I., Tardif, R., Michaelides, S., Cermak, J., Bott, A., Bendix, J., Müller, M.D., Pagowski, M., Hansen, B., Ellrod, G., Jacobs, W., Toth, G. and Cober, S.G. (2007a) Fog research: a review of past achievements and future perspectives. *Pure and Applied Geophysics*, 164, 1121–1159. <https://doi.org/10.1007/s00024-007-0211-x>.
- Gultepe, I., Pagowski, M. and Reid, J. (2007b) A satellite-based fog detection scheme using screen air temperature. *Weather and Forecasting*, 22(3), 444–456. <https://doi.org/10.1175/WAF1011.1>.
- Gultepe, I., Heymsfield, A.J., Fernando, H.J.S., Pardyjak, E., Dorman, C.E., Wang, Q., Creegan, E., Hoch, S.W., Flagg, D.D., Yamaguchi, R., Krishnamurthy, R., Gaberšek, S., Perrie, W., Perelet, A., Singh, D.K., Chang, R., Nagare, B., Wagh, S. and Wang, S. (2021) A review of coastal fog microphysics during C-FOG. *Boundary-Layer Meteorology*, 181(2), 227–265. <https://doi.org/10.1007/s10546-021-00659-5>.
- Haeffelin, M., Bergot, T., Elias, T., Tardif, R., Carrer, D., Chazette, P., Colomb, M., Drobinski, P., Dupont, E., Dupont, J.-C., Gomes, L., Musson-Genon, L., Pietras, C., Plana-Fattori, A., Protat, A., Ranguogno, J., Raut, J.-C., Rmy, S., Richard, D., Sciare, J. and Zhang, X. (2010) Parisfog: shedding new light on fog physical processes. *Bulletin of the American Meteorological Society*, 91(6), 767–783. <https://doi.org/10.1175/2009BAMS2671.1>.
- Herman, G.R. and Schumacher, R.S. (2018) Dendrology in numerical weather prediction: what random forests and logistic regression tell us about forecasting extreme precipitation. *Monthly Weather Review*, 146(6), 1785–1812. <https://doi.org/10.1175/MWR-D-17-0307.1>.
- Kalthoff, N., Binder, H.-J., Kossmann, M., Vöglin, R., Corsmeier, U., Fiedler, F. and Schlager, H. (1998) Temporal evolution and spatial variation of the boundary layer over complex terrain. *Atmospheric Environment*, 32(7), 1179–1194. [https://doi.org/10.1016/S1352-2310\(97\)00193-3](https://doi.org/10.1016/S1352-2310(97)00193-3).
- Kammer, J., Perraudin, E., Flaud, P.-M., Lamaud, E., Bonnefond, J. and Villenave, E. (2018) Observation of nighttime new particle formation over the French Landes forest. *Science of The Total Environment*, 621, 1084–1092. <https://doi.org/10.1016/j.scitotenv.2017.10.118>.
- Karimi, M. (2020) Direct numerical simulation of fog: the sensitivity of a dissipation phase to environmental conditions. *Atmosphere*, 11(1), 12. <https://doi.org/10.3390/atmos11010012>.
- Köhler, C., Steiner, A., Saint-Drenan, Y.-M., Ernst, D., Bergmann-Dick, A., Zirkelbach, M., Bouallègue, Z.B., Metzinger, I. and Ritter, B. (2017) Critical weather situations for renewable energies – Part B: low stratus risk for solar power. *Renewable Energy*, 101, 794–803. <https://doi.org/10.1016/j.renene.2016.09.002>.
- Kneringer, P., Dietz, S.J., Mayr, G.J. and Zeileis, A. (2019) Probabilistic nowcasting of low-visibility procedure states at Vienna International Airport during the cold season. *Pure and Applied Geophysics*, 176(5), 2165–2177. <https://doi.org/10.1007/s00024-018-1863-4>.
- Leigh, R.J., Drake, L. and Thampapillai, D.J. (1998) An economic analysis of terminal aerodrome forecasts with special reference to Sydney Airport. *Journal of Transport Economics and Policy*, 32(3), 377–392.
- Lever, J., Krzywinski, M. and Altman, N. (2016) Logistic regression. *Nature Methods*, 13, 541–542. <https://doi.org/10.1038/nmeth.3904>.
- López, L. and Sánchez, J. (2009) Discriminant methods for radar detection of hail. *Atmospheric Research*, 93(1), 358–368. <https://doi.org/10.1016/j.atmosres.2008.09.028>.
- Maalick, Z., Kühn, T., Korhonen, H., Kokkola, H., Laaksonen, A. and Romakkaniemi, S. (2016) Effect of aerosol concentration and absorbing aerosol on the radiation fog life cycle. *Atmospheric Environment*, 133, 26–33. <https://doi.org/10.1016/j.atmosenv.2016.03.018>.
- Olivier, J. and Stockton, P.L. (1989) The influence of upwelling extent upon fog incidence at Lüderitz, southern Africa. *International Journal of Climatology*, 9(1), 69–75. <https://doi.org/10.1002/joc.3370090106>.
- Pagowski, M., Gultepe, I. and King, P. (2004) Analysis and modeling of an extremely dense fog event in southern Ontario. *Journal of Applied Meteorology*, 43(1), 3–16. [https://doi.org/10.1175/1520-0450\(2004\)043<0003:AAMOAEO>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0003:AAMOAEO>2.0.CO;2).

- Pauli, E., Andersen, H., Bendix, J., Cermak, J. and Egli, S. (2020) Determinants of fog and low stratus occurrence in continental central Europe – a quantitative satellite-based evaluation. *Journal of Hydrology*, 591. <https://doi.org/10.1016/j.jhydrol.2020.125451>.
- Pauli, E., Cermak, J. and Teuling, A.J. (2022) Enhanced nighttime fog and low stratus occurrence over the Landes forest, France. *Geophysical Research Letters*, 49(5). <https://doi.org/10.1029/2021GL097058>.
- Price, J. (2019) On the formation and development of radiation fog: an observational study. *Boundary-Layer Meteorology*, 172(2), 167–197. <https://doi.org/10.1007/s10546-019-00444-5>.
- Roach, W.T. (1995) Back to basics: Fog: Part 2 – the formation and dissipation of land fog. *Weather*, 50(1), 7–11.
- Scherrer, S.C. and Appenzeller, C. (2014) Fog and low stratus over the Swiss Plateau – a climatological study. *International Journal of Climatology*, 34(3), 678–686. <https://doi.org/10.1002/joc.3714>.
- Spracklen, D.V., Bonn, B. and Carslaw, K.S. (2008) Boreal forests, aerosols and the impacts on clouds and climate. *Philosophical Transactions of the Royal Society A*, 366(1885), 4613–4626. <https://doi.org/10.1098/rsta.2008.0201>.
- Steenefeld, G.-J. and de Bode, M. (2018) Unravelling the relative roles of physical processes in modelling the life cycle of a warm radiation fog. *Quarterly Journal of the Royal Meteorological Society*, 144, 1539–1554. <https://doi.org/10.1002/qj.3300>.
- Torregrosa, A., Combs, C. and Peters, J. (2016) GOES-derived fog and low cloud indices for coastal north and central California ecological analyses. *Earth and Space Science*, 3(2), 46–67. <https://doi.org/10.1002/2015EA000119>.
- Vautard, R., Yiou, P. and Van Oldenborgh, G.J. (2009) Decline of fog, mist and haze in Europe over the past 30 years. *Nature Geoscience*, 2(2), 115–119. <https://doi.org/10.1038/ngeo414>.
- Wærsted, E.G., Haeffelin, M., Steeneveld, G.-J. and Dupont J.-C. (2019) Understanding the dissipation of continental fog by analysing the LWP budget using idealized LES and *in situ* observations. *Quarterly Journal of the Royal Meteorological Society*, 145, 784–804. <https://doi.org/10.1002/qj.3465>.
- Wu, W. and Zhang, L. (2013) Comparison of spatial and non-spatial logistic regression models for modeling the occurrence of cloud cover in northeastern Puerto Rico. *Applied Geography*, 37, 52–62. <https://doi.org/10.1016/j.apgeog.2012.10.012>.
- Yan, S., Zhu, B., Huang, Y., Zhu, J., Kang, H., Lu, C. and Zhu, T. (2020) To what extents do urbanization and air pollution affect fog?. *Atmospheric Chemistry and Physics*, 20(9), 5559–5572. <https://doi.org/10.5194/acp-20-5559-2020>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Pauli, E., Cermak, J. & Andersen, H. (2022) A satellite-based climatology of fog and low stratus formation and dissipation times in central Europe. *Quarterly Journal of the Royal Meteorological Society*, 1–16. Available from: <https://doi.org/10.1002/qj.4272>