# Correction of meteorological vehicle-based measurements for road weather monitoring in pursue of enabling safe automated driving

Zur Erlangung des akademischen Grades einer

DOKTORIN DER INGENIEURWISSENSCHAFTEN
(Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des

Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

**M.Sc. Meike Rietdorf**

Tag der mündlichen Prüfung: 04.11.2021

First reviewer: Prof. Dr.-Ing. Christoph Stiller
Second reviewer: Prof. Dr. rer. nat. Felix Ament

# Zusammenfassung

Um zukünftig punktgenaue Vorhersagen und somit verlässliche Warnungen vor wetterbedingten und potenziell gefährlichen lokalen Straßenbedingungen zu erstellen, werden zeitlich und räumlich hochaufgelöste meteorologische Daten benötigt. Die vorliegende Arbeit prüft die Verwendbarkeit von fahrzeugbasierten Messungen basierend auf der derzeit in Serie verbauten Sensorik. Ziel dieser Arbeit ist es zu untersuchen, ob und inwiefern eine Korrektur der fahrzeugbasierten Daten eine Steigerung des Potentials zur Verbesserung der räumlichen und zeitlichen Auflösung von meteorologischen Daten aufweist.

Die Rohdaten der Fahrzeugmessungen unterliegen starken Abweichungen zu den verwendeten Referenzdaten, hervorgerufen sowohl durch stationäre Effekte wie Messungenauigkeit und Verbauort des Sensors, als auch durch bewegungsbedingte Effekte, wie beispielsweise den Einfluss der Motorabwärme bei geringen Geschwindigkeiten.

Um die genannten Einflüsse zu untersuchen, wird zunächst ein weltweit einzigartiger Datensatz mit parallelen Daten von Serienfahrzeugen und Referenzen im Rahmen von Messkampagnen erstellt. Anschließend führt die vorliegende Arbeit eine Qualitätskontrolle und Korrektur für die vier meteorologischen Parameter Luftdruck, Lufttemperatur, relative Feuchte und Globalstrahlung durch.

Die Rohdaten verfügen für meteorologische Anwendungen über eine zu geringe Qualität. Die entwickelten und implementierten Korrekturverfahren, sowohl physikalischer Natur als auch basierend auf Machine Learning, erreichen sowohl für den Luftdruck als auch für die Lufttemperatur und die relative Feuchte signifikante Verbesserungen der vorliegenden Daten. Für die Lufttemperatur erreichen alle getesteten Modelle vergleichbar gute Ergebnisse, wohingegen bei der relativen Feuchte die Machine Learning basierten Modelle qualitativ hochwertigere Ergebnisse erzielen als das physikalische Modell. Die Machine Learning Modelle erreichen für diesen Parameter einen Anteil von

über 95 % an Daten innerhalb der einfachen Messunsicherheit. Eine allgemein gültige Aussage bezüglich der Übertragbarkeit und Wirksamkeit auf anderen Fahrzeugen und anderen als den hier getesteten Szenarien kann auf Basis der zur Verfügung stehenden Datengrundlage nicht getroffen werden.
Die Korrektur der Globalstrahlung erreicht in stationären Situationen bereits eine signifikante Verbesserung der Qualität der Ergebnisse. Für die Korrektur von mobilen Daten während der Fahrt ist das Potential der Qualitätssteigerung noch nicht ausgereizt.

Die vorliegende Arbeit verdeutlicht die Notwendigkeit der Korrektur der fahrzeugbasierten Rohdaten und zeigt das Potential der hiermit verbundenen Qualitätssteigerung auf. Weitere Untersuchungen, vor allem bezüglich der Übertragbarkeit auf Flottendaten, sowie eine größere Datengrundlage sind notwendig, um eine allgemein gültige Aussage über die Qualitätssteigerung treffen und die Korrekturen weiter in Richtung Serienreife entwickeln zu können.

# Abstract

To provide pinpoint forecasts and thus reliable local warnings of weather-related potentially hazardous road conditions, meteorological data with high temporal and spatial resolution is required. The presented work examines the usability of vehicle-based measurements based on the sensor system currently installed in series production. The objective of this work is to investigate whether and to what extent a correction of vehicle-based data increases the potential to improve the spatial and temporal resolution of meteorological data.

The raw data of the vehicle measurements is subject to strong deviations from the reference data used, caused both by stationary effects such as the measurement uncertainty or mounting position of the sensor, and by driving-related effects such as the influence of engine waste heat at low vehicle speeds.

Within the scope of measurement campaigns, a globally unique data set with concurrent data from series production vehicles and references is created in order to investigate the influences mentioned above. Subsequently, a quality control and correction for the four meteorological parameters air pressure, air temperature, relative humidity and global radiation is performed.

The quality of the raw data measured on board of a vehicle is too low to be used for meteorological applications. The developed and implemented correction methods, both physical in nature and machine learning based, achieve significant improvements of the available data for air pressure as well as air temperature and relative humidity. For air temperature, all tested models achieve comparably good results, whereas for relative humidity, the machine learning based models achieve higher quality results than the physical model. The machine learning models achieve over 95 % of data within the single measurement uncertainty for this parameter. A generally valid statement regarding the transferability and effectiveness to other vehicles and scenarios other than those tested here cannot be made on the basis of the available data.

Correction of global radiation already achieves significant improvement in the quality of results in stationary situations. For the correction of mobile data while driving, the potential of quality improvement is not yet exhausted.

The presented thesis illustrates the necessity of correcting the vehicle-based raw data and shows the potential of the associated quality improvement. Further investigations, especially with regard to the transferability to fleet data, as well as a larger data basis are necessary in order to allow for a generally valid statement about the quality improvement and to further develop the corrections to series maturity.

# Preface

This thesis has been developed within the framework of the research project FloWKar. FloWKar is a project funded by the German Federal Ministry of Infrastructure and Transport (BMVI). In the project, the AUDI AG and the German Weather Service (DWD) are cooperating, with support from this doctorate at the Karlsruhe Institute of Technology (KIT).

Recent extreme weather events, such as the heavy snowfall event in February 2021 and heavy rainfall in July 2021, show the increasing impact of weather on the transportation sector and road traffic. In addition to extreme weather events, even wintry weather conditions are enough to worsen road conditions and thus trigger potentially dangerous situations on the road.

A precise forecast of local hazards such as aquaplaning, freezing wetness and slippery conditions cannot be realized based on the currently available data basis. To increase the temporal and spatial resolution of meteorological data, vehicles are therefore being tested as "mobile weather stations" in this research project.

This work focuses on the quality control of the vehicle-based measurements and their corrections.

# Table of contents

# Abbreviations and symbols

## Abbreviations

| | |
|---|---|
| **AA** | **A**mbient **A**ir |
| **ANN** | **A**rtificial **N**eural **N**etwork |
| **BASt** | Federal Highway Research Institute - German: **B**undes**a**nstalt für **St**raßenwesen |
| **BMVI** | German Federal Ministry of Infrastructure and Transport - german: **B**undes**m**inisterium für **V**erkehr und **I**nfrastruktur |
| **CAN bus** | **C**ontroller **A**rea **N**etwork: Standard-bus-system in automotive sector |
| **DWD** | National weather service of the Federal Republic of Germany - German: **D**eutscher **W**etter**d**ienst |
| **DNN** | **D**eep **N**eural **N**etwork |
| **EIT** | **E**ngine **I**mpact **T**imer |
| **FMI** | **F**innish **M**eteorological **I**nstitute |
| **FloWKar** | Project Fleet Weather Map - German: **Flo**tten-**W**etter-**Kar**te) |
| **GDPR** | **G**eneral **D**ata **P**rotection **R**egulation |
| **GNSS** | **G**lobal **N**avigation **S**atellite **S**ystem |
| **GPS** | **G**lobal **P**ositioning **S**ystem |
| **KIT** | **K**arlsruhe **I**nstitute of **T**echnology |

| | |
|---|---|
| **KNMI** | **K**oninklijk **N**ederlands **M**eteorological **I**nstituut |
| **LHW** | **L**ocal **H**azardous **W**arning |
| **ML** | **M**achine **L**earning |
| **MMU** | **M**obile **M**easuring **U**nit |
| **MSE** | **M**ean **S**quarred **E**rror |
| **MWS** | **M**obile **W**eather **S**tation |
| **NN** | **N**eural **N**etwork |
| **OAT** | **O**utside **A**ir **T**emperature |
| **OEM** | **O**riginal **E**quipment **M**anufacturer |
| **PD1MU** | **P**ortion of **D**ata within **1** **M**easurement **U**ncertainty |
| **PD2MU** | **P**ortion of **D**ata within **2** **M**easurement **U**ncertainties |
| **PHY** | **Phy**sical model |
| **RF** | **R**andom **F**orest |
| **RH** | **R**elative **H**umidity |
| **RWS** | **R**oad **W**eather **S**tation |
| **SL** | **S**now **L**ine |
| **synop** | **Synop**tical |
| **TC** | **T**echnical **C**arrier |
| **UH** | **U**nder **H**ood |
| **WB** | **W**eather **B**ox |

# Absolute terms

| | |
|---|---|
| **$R_m$** | Universal gas constant $R_m$: 8.314462 $\frac{kgm^2}{s^2 molK}$ |
| **C1** | Absolute term 1 in Magnus formula above water C1: 6.112 hPa, above ice C1: 6.112 hPa |
| **C2** | Absolute term 2 in Magnus formula above water C2: 17.62, above ice C2: 22.46 |
| **C3** | Absolute term 3 in Magnus formula above water C3: 243.12 °C, above ice C3: 272.62 °C |
| **$M_V$** | Molar mass of water vapor $M_V$: 18.01528 $\frac{g}{mol}$ |
| **$M_A$** | Molar mass of dry air $M_A$: 28.96 $\frac{g}{mol}$ |

# Latin symbols and variables

## Lower case letters

| | |
|---|---|
| **m** | Mass |
| **p** | Air pressure |
| **$p_A$** | Partial pressure of dry air |
| **$p_V$** | Partial pressure of vapor |
| **$p_{V,sat}$** | Saturation vapor pressure |
| **s** | Specific humidity |
| **t** | time |
| **x** | Water content of an air parcel |

### Upper case letters

| | |
|---|---|
| **AoS** | Horizontal angle of incidence of the sun |
| **G** | Global radiation |
| **GT** | Ground truth |
| **M** | Molar mass |
| **RH** | Relative humidity |
| **T** | Air temperature |
| **Td** | Dewpoint temperature |
| **V** | Volume |

## Greek symbols and variables

| | |
|---|---|
| $\alpha$ | Horizontal angle of incidence of sun on vehicle |
| $\tau$ | Threshold |

## General indices

### Low indices

| | |
|---|---|
| $A$ | Dry air |
| $AA$ | Ambient air |
| $corr, NN$ | Corrected by neural network |
| $corr, PHY$ | Corrected by physical model |
| $corr, RF$ | Corrected by random forest |
| $meas$ | Measured vehicle-based signal |

| | |
|---|---|
| $ref$ | Reference signal |
| $UH$ | Under hood |
| $V$ | Vapor |
| $veh$ | Vehicle-based |
| $W$ | Water |

## High indices

| | |
|---|---|
| $i$ | Current time step |
| $i-1$ | Previous time step |

# 1 Introduction

## 1.1 Motivation

Enabling safe automated driving functions requires precise information about the surrounding of the vehicle with all its details. Since most research projects focus on influences of objects [WFSF18] and interference with living beings in traffic [RWLS18], environmental influences by weather phenomena are mostly understudied in this field [WBASA20].
In a situation with snowdrifts on traffic lanes, the vehicle needs to adjust its speed to ensure the safety of all passengers. This is only achievable if the vehicle is aware of the snowdrifts. In addition, snowdrifts hide lane markings, and therefore weaken lane keeping. Fog can be named as a second example for the necessity to include meteorological data. In heavy and dense fog, the camera systems of a vehicle are blind, and therefore the amount of information about its surrounding are significantly reduced. In general, adverse weather causes numerous accidents and delays around the world and is thus emphasizing its impact on the road and transport sector [Com04, Com10].

Today, road weather forecasts are based on regional weather models. These models use measurements from monitoring stations, including synoptical (in the following: synop) stations and road weather stations, to collect their data base. Figure 1a) gives an overview of available meteorological data from synop stations and road weather stations in Germany. Zooming in reveals the poor spatial coverage (cf. Figure 1b)). Besides a test segment in the lower part of the map section, the spatial coverage, especially in rural areas, is down to tenths of kilometers. The total of approx. 500 full-time and part-time ground measuring stations of the DWD in Germany [Deu21] results in an average coverage of 715 $km^2$ per station. The average distance between the individual stations is therefore approx. 27 km. Based on this network, reliable pinpoint forecasts are not feasible, and thereby operation of safe automated driving functions cannot be guaranteed. The situation is similar in other countries. Exemplary,

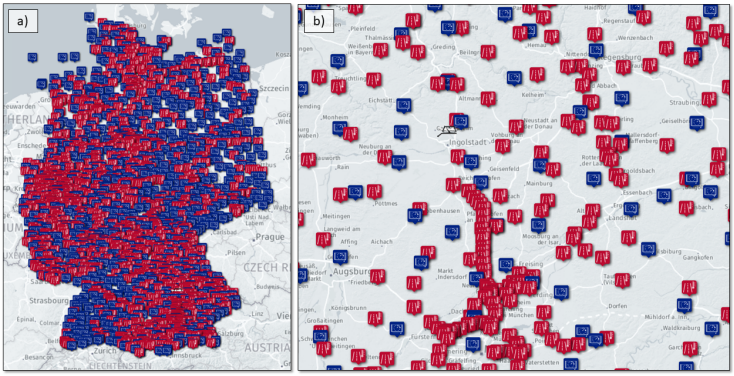


Figure 1: Current database set up by synop stations (blue) and road weather stations (red) a) Overview of currently available data base provided by national weather service of Germany b) Zoom in to visualize test track in lower left region of map, but otherwise poor spatial coverage.

in Canada the spatial coverage of road weather stations is not sufficient in all areas for deriving reliable information for operative decision making [BKF18].

To increase the precision and reliability of forecast models, needed for enabling safe automated driving functions, a denser network, both in time and space is crucial. Basically, there are two different approaches for achieving a higher resolving network, besides increasing model resolution. Either the amount of stationary monitoring stations is increased by a significant amount or mobile measuring devices are established. Increasing the number of stationary monitoring stations is costly and still only partly expedient since stationary measurements are limited to represent conditions at this specific location only. Therefore, no statement about the spatial course of a measured signal is possible. Extending the network by mobile measuring methods, allowing for measuring the spatial course of the signals and additionally increasing the density of measurements significantly, promises benefits [KNSM17]. By using mobile measuring systems the measurements are not locally limited anymore, but can perform measurements on the entire road network [BJ16]. The most obvious while arguably cost-effective approach thereby is to incorporate the environment sensors found onboard the millions of vehicles themselves.
However, since the sensor technology on board of production vehicles is not

intended for usage as a weather station, the quality of the data remains to be tested. The reliability of individual sensor systems is strongly influenced by adverse weather conditions such as rain, fog, and snow [YSYA19]. If the quality is satisfactory, the vehicle-based data has the advantage of higher spatial coverage compared to the stationary data, while particularly covering the road network and thus the area of interest for autonomous driving functions. Additionally, data collected by vehicles has a higher temporal resolution than the currently established stationary measurement stations [Deu21].
This work focuses on the quality control of the vehicle-based data for the purpose of a subsequent integration of vehicle data with established assimilation prognosis methods of the national weather services.

## 1.2 State of the art

The idea of using mobile measuring stations to increase both the spatial and temporal resolution of meteorological data is not new [MTS12, MDPO10, MIO13]. Musiak, Tillotson and Spinelli refer to vehicles in a more general term, focusing on ships and aircrafts. Cars are not explicitly mentioned. The other two papers relate to passenger and fleet vehicles and list both the enormous potential and the challenges resulting from this type of data. Mahoney and O'Sullivan emphasize that vehicle-based meteorological data provides a significant benefit to road weather accuracy but is tied to previously required quality control. However, additional research is necessary to exploit the full potential of vehicle-based data [MIO13]. In order to better title both the possibilities and the problems with fleet data, an experiment with 60 vehicles was conducted in Rotterdam (NL) with a focus on travel times [THST00]. The field test was mainly aimed at testing the communication and handling with fleet data, no meteorological data was used.

**State of the art - Projects**
Similar projects prove the topicality and the importance of solving the underlying problem of low coverage of meteorological data. One example is the cooperation between the Koninklijk Nederlands Meteorologisch Instituut Datalab (KNMI Datalab) and the Royal Netherlands Meteorological Institute. The project is aware of the low coverage of Road Weather Stations (RWS),

especially beyond highways. Although the authors identify difficulties like different CAN-bus standards, they see a significant potential of measuring meteorological data with vehicles [KR17]. However, this project focuses exclusively on the measured air temperature and precipitation information based on the operation frequency of the wipers.
At the Finnish Meteorological Institute (FMI), another similar project is carried out. A public bus is equipped with an additional road sensor [KN16] . The additional optical instrument is based on spectral analysis and is capable of measuring road surface temperature, friction and amount of water on the road surface [HNN12]. The results of two runs of a weather model, one including and one excluding measured values on board of the public bus, are compared. Although the number of data points added is comparatively small, an increase in prediction accuracy can be observed. The project solely uses additional data from the calibrated road sensor mounted on the public bus. Any influence of using vehicle's own instrumentation is still to be tested.
eHorizon.Weather is a French project to improve forecasts of road weather. For this purpose, the company Continental cooperates with the public weather service of France "Meteo France" [Met16] . The project uses selected sensor signals, for example, air temperature, wiper frequency, and air pressure, of a test fleet of about 200 vehicles. Using the on-board sensor data, they aim for developing a road weather model forecasting a preview of the road situation ahead and thereby increase safety for automated driving.
Already available on the market is a product of the start-up RoadCloud. The product is based on vehicular data to balance the low density of RWS and thereby decrease likelihood of misinterpretation of road conditions [TNK21].
All project have in common that they focus on air temperature or road conditions. Furthermore, there are projects concentrating solely on precipitation, exemplary mobileVIEW [Bra17] and a project at the University of Michigan [BPZ+19]. Nevertheless, since this work does not focus on precipitation, these projects are not further described in detail.
The FloWKar project, within the framework of which this work was developed, is co-funded by the Ministry of Transportation of the Federal Republic of Germany and presents a cooperation project between the national weather service of Germany (DWD) and the AUDI AG, a German car manufacturer. The project FloWKar has two major points of differentiation compared to the presented similar projects. First, it is the only project with an established cooperation between a national weather service and a car manufacturer. This is of great advantage, since direct information on the state of the vehicle, e.g.,

its speed and engine temperature, are known during data collection. Secondly, FloWKar is the only project where a quality check comprising several stages is conducted before the data is used for further applications. Therefore, supposedly no falsified data is used in the models, which might otherwise lead to an increase in the model error rather than mitigating the overall error and actually improving the model.

**State of the art - Air temperature**
First and foremost for virtually all meteorological applications is temperature. Besides currently ongoing research projects, publications regarding the correction of vehicle-based temperature measurements exist. However, no scientific papers have been published, but numerous patents of well-known automobile manufacturers or companies focusing on electronics can be found instead.
All considered patents focus on periods where the vehicle is either at a standstill or accelerating after a recent standstill. During these periods, the measured air temperature signal has a high chance to be affected by non-ambient heat effects. The approaches of the published patents for determining an outside air temperature (OAT) can be sorted into three different groups, distinguishable by their approaches: constant, temporal and mathematical.

The first approach is the most basic. For periods with a high chance for a flawed air temperature measurement, the previous OAT is kept constant. Thereby the current fluctuation of the signal caused by motion or indirect effects of the vehicle is filtered. However, this approach flattens out not only vehicle-induced but also potential local effects. Nevertheless, the implementation is relatively straightforward, and the results fulfill most requirements for many further applications, for example displaying the OAT in the cockpit. Typically, this approach is used for correcting the measured temperature signal for short stops [KB09,COW10] or periods with an increase in the measured temperature signal during congestions [KH09]. Super et al. keep the OAT constant, if the engine workload exceeds a certain threshold [SAL05]. Also Poublon [Pou00] uses this approach, but still differs significantly from the previous patents. The described first approach of a constant OAT is applied for standstill. To determine the first OAT before it is kept constant, is based on multiple signals, for example, the engine coolant temperature and the induction system temperature, but excludes the measured air temperature itself.

The second approach uses a temporal dependency [DBW04, Gao08]. Often the constant approach is combined with the temporal approach. Based on the

"idle counter" DeRonne decides whether the air temperature measurement is still influenced by a previous standstill or whether the temperature determined for steady run conditions is used as OAT [DBW04]. Gao proceeds in a similar way by introducing a "heat build up", which diminishes to zero if the vehicle is not situated in a heat soak mode and thereby it is not influenced by earlier standstills [Gao08]. Beyond that, prior patents already make use of a temporal dependency [RHDJ95, Wal98]. Rudzewicz calls it a time delay that is needed for the measurement to be reliable after the vehicle accelerates again and its speed exceeds a certain threshold [RHDJ95]. Wallrafen introduces a more complex "damping time", which depends on the vehicle speed and the engine temperature. The bigger the two affecting factors, the higher the damping time during which the measured temperature is considered as not trustworthy [Wal98].

The third group of patents contrasts with the previously introduced two approaches since it is based on mathematics and determines a best fit based on the preceding measurements to calculate an OAT [KBB13, Mat14, Hub16]. Ford Global Technologies patent uses Newton's Cooling Law as the base of his approach and creates a second-order polynomial function based on the decreasing air temperature measurements after a standstill [KBB13]. Thereby the theoretical final temperature value after a fully completed decline is determined and used as OAT [KBB13]. Honda Motor Co. pursue a similar approach [Mat14]. Based on the measured air temperature signal a second-order polynomial function is modeled. By comparing this function to previously determined correlations between a second-order polynomial fit and the measured air temperature, the current OAT is determined. The downside of this approach is the individual correlation between the second-order polynomial and the air temperature, since this has to be established for every vehicle type and air temperature prior to application of this correction method. Another challenge is overcoming initial difficulties for curve fitting. Hence, Hubert introduces an evolved curve fitting approach by employing trigonometry to a few early data points before the signal becomes unstable, instead of all measured values [Hub16]. Based on a mathematical model, the coefficients of the best fit for the data are determined and used for estimating an OAT.

Besides these three groups of approaches, there are a some patents using approaches not directly linked to the previously described strategies [Wue99, MS03, HZ11]. Instead of a timer, Wurtenberger, calculates a “heat resistance”

and uses it to decide whether the measured air temperature signal is falsified and if so, by how much [Wue99]. The approach resembles the approach of the temporal group without using a direct temporal dependency. Manakkal's strategy, for example, is similar to the constant approach, since the new OAT is based on the previous OAT but is adjusted depending on the duration of the stop [MS03]. Yamada's approach uses well-known signals like vehicle speed and water temperature, a representative of the engine temperature. An implemented correction function uses these signals to determine the extent of the deviation of the measured to the actual air temperature. The higher the vehicle speed and the cooler the engine temperature, the smaller is the correction function, which is explained by neglectable influences during steady run conditions.
Further notable is the approach of Schürmanns [SKF18], which is the only approach considering the global radiation as an influencing factor for the measured air temperature signal. For short standstills with an increasing measured air temperature signal, a filter for compensating heating effects of the vehicle and the sun is applied. The filter uses a heat absorption coefficient of the vehicle's surface as well as the measured global radiation and the increase in the measured air temperature signal. Considering a further heating effect in addition to the well-known waste heat of the engine is so far unique.
Nevertheless, all presented approaches aim for correcting the complete increase in the measured air temperature signal, without considering local effects that might also affect the temperature measurement. Especially the local effects are of high interest for a detailed weather map with high spatial resolution. In the end, local effects decide between critical or harmless situations for edge-cases, wherefore the awareness of local effects is crucial. Therefore, none of the presented approaches is satisfactorily. Furthermore, it is notable that all presented approaches are based on physical and empirically found correlations, no approach uses artificial intelligence for determining the OAT.

**State of the art - Other quantities**
For other meteorological quantities measured on board, such as air pressure and relative humidity, no publications in the form of scientific papers or patents are known. This is uncharted territory for research.

However, there is an implemented warning of local hazards at some OEMs already. Since the vehicle-based signals considered for triggering warnings have not been corrected beforehand, their trustworthiness is limited. Due to the

quality of the installed sensors, the AUDI AG, for example, currently relies not only on the measured sensor values, but also considers driver activity (wiper frequency, rear fog light) to assess the situation and trigger a warning. Hence, the warnings are only triggered when several vehicles confirm the situation and, therefore, have already been in this hazardous situation. The goal is to trigger this warning before the first vehicle enters this situation. This can only be achieved by road weather forecasts with higher spatial and temporal resolutions than the forecasts available today. Based on the currently available data these types of predictive warnings are not feasible.

## 1.3 Contributions and structure

The previous section illustrates the need for research regarding vehicle-based meteorological quantities to be used for further applications. Currently, only a low-resolution measurement network is available, which makes it impossible to make precise statements about road weather, especially in edge-cases with regards to passenger safety. Therefore, additional data sources are needed, ideally with a higher temporal and spatial resolution. In this work, the approach of using vehicle-based data is investigated.
The advantage of vehicle-based data is the high spatial coverage that can be achieved by using fleet vehicles. However, vehicle-based data has the disadvantage that without verification and processing their data quality is lower compared to stationary data (e.g., from synop stations). The present work examines to what extent corrections of the vehicle-based data can increase the quality in order to eliminate the disadvantage of this data source.

Since the air temperature is a prominent parameter displayed in the cockpit, models for correction are available in the form of patents (cf. Chapter 1.2). However, none of these patents test a correction via deep learning approaches. Deep Learning offers the advantage that a wide range of possible influencing parameters and their possible interdependencies can be considered.
For the other meteorological parameters, such as air pressure and relative humidity, neither scientific papers nor patents are available. Instead, there are multiple projects researching this topic. This illustrates the topicality and the need for research in the processing of vehicle-based meteorological data.

Many of the mentioned projects use raw data derived from the vehicles. This work, however, investigates the quality improvement that can be achieved by a quality check with subsequent correction of the vehicle-based signals, which sets this work apart from previous research. Cooperating with an Original equipment manufacturer (OEM) enables a holistic view of the situation which is not given in the other projects.
Before the data is anonymized and provided to a service provider or weather service, the quality of every raw signal used for meteorological applications, must be assured. For raw signals with insufficient data quality, correction methods need to be implemented and tested. This results in the first two contributions that this work accomplishes.

1. **The thesis introduces physical models for correcting vehicle-based raw data and presents the limitations.**

For every considered meteorological quantity, a model is developed, which corrects effects induced by the motion of the vehicle while effects induced by local influences remain unchanged. Therefore, physical correlations between individual parameters are considered and used as the base for the model. To what extent a physically based model can detect falsified measurements and compensate them, is evaluated.

2. **As an alternative approach, the work examines the use of machine learning models to correct the raw vehicle-based data.**

Based on identical data, this thesis elaborates different machine learning models. It needs to be established whether machine learning algorithms interpret the context correctly and are hence able to generalize the correction methods to subsequently apply it successfully to a different data set. The performance of different machine learning approaches is analyzed and evaluated.

3. **The thesis performs a comparative evaluation of the two different approaches using physical and machine learning models.**

First, the quality of the two previous approaches is examined. The strengths and weaknesses of the respective approaches are elaborated on. It is veri-

fied, whether all or how many of the processed and corrected meteorological signals reach an adequate quality level to be used for further meteorological applications. Subsequently, the results of both approaches for all considered meteorological signals are compared to identify the best suited approach for each signal in the available data set.

4. **The thesis points out an application example of an improved warning of local danger situations based on the corrected vehicle-based data.**

Based on the corrected vehicle-based data, a local danger warning system is exemplary implemented to warn the user of potential road weather hazards. This includes warnings for fog, icy conditions, and snowfall. The improved accuracy of the warnings when using corrected data compared to raw data is elaborated.

Before outlining the listed contributions of the thesis, Chapter 2 briefly describes the basics necessary to provide content to this thesis in meteorological terms and machine learning methods. Chapter 3 lists the sensor technology used in today's production vehicles, the project-specific technology carriers, and potential reference stations. The methodological approach of the necessary data acquisition, as well as the data generation itself, are also elaborated in Chapter 3. Due to the degree of novelty of this work, no preexisting data set can be used. Instead, an extensive data set is recorded within the scope of the thesis in the form of targeted measurement campaigns with a focus on real-world references to capture highly local effects that could otherwise not be constructed in laboratory environments. Chapter 4 lists the various approaches for correcting the raw data and, therefore, forms the basis for the first two contributions. Quantities considered in this thesis are air pressure, air temperature, relative humidity and global radiation. In the following Chapter 5, results of both approaches are presented, compared and quality statements are made (Contribution 3). The application example of local danger warnings, and thus Contribution 4, is found in Chapter 6. Finally, Chapter 7 summarizes the findings of the thesis and gives an outlook on further research goals.

# 2 Foundations

This work combines thematically two different and less related areas, hence a brief introduction is given to both. Chapter 2.1 introduces the basic meteorological quantities, which are helpful for understanding the following elaborations. Chapter 2.2 first gives general information on data analysis using machine learning. Subsequently, the Subchapters 2.2.1 and 2.2.2 briefly explain the individual methods of a neural net and a random forest.
The following descriptions and definitions of the fundamentals do not serve as a complete elaboration of the respective topics but provide an insight into the respective subject matter to provide the background knowledge necessary for understanding the presented thesis.

## 2.1 Meteorological quantities

**Atmospheric pressure**
In general, meteorology distinguishes between hydrostatic and dynamic pressure. The hydrostatic pressure is caused by gravity, whereas the dynamic pressure is caused by the velocity of the air flow [LC79]. At the earth's surface the influence of the flow of air is insignificant, except for some exceptional cases, wherefore the air pressure can be equated with the hydrostatic pressure of the atmosphere [LC79]. Therefore, the air pressure $p$ at a given altitude is defined as the weight of a vertical column of air with base area 1 $m^2$ extending from the altitude under consideration up to the upper boundary of the atmosphere. The weight of the air column is the force that the air column exerts on the base due to gravity [LC79]. The unit of air pressure or atmospheric pressure is Pascal. However, because of numerical values ranging above 1000 Pa often, air pressure is usually expressed in hPa.

**Air temperature**

The air temperature $T$ is the expression of the subordinate molecular motion in the air [LC79]. The greater the mean kinetic energy of the air molecules and thus their velocity, the greater the air temperature [Spe21]. To minimize influences of the environment and thus make measured values comparable, the air temperature is specified metrologically as a temperature of a radiation-protected and ventilated thermometer. Radiation shielding reduces radiation error due to solar radiation and radiation exchange with the atmosphere and environment [LC79]. Ventilation contributes to heat exchange between the air and the thermometer, preventing heat buildup on the thermometer.

In addition to the listed falsifying influences, local phenomena can also influence the air temperature measurement. However, these are not falsifying influences since a representative measurement can only be made by taking into account the local conditions and influences. Exemplary, local deviations are caused by the relief of the landscape. One possible effect of this is the formation of a reservoir for cold air in a depression or valley during the night or in winter [LC79]. In addition, the air temperature is influenced by various other parameters, such as advection and clouds [LC79].

Temperature differences between two values are specified in Kelvin [K] according to DIN 1345, even if the temperature values themselves have the unit °C [Deu93]. The step size of both units is identical.

**Humidity parameters**

Various humidity parameters are used in meteorology. In the following, the three most important ones for the understanding of this work are briefly explained.

The best-known measure of humidity is *relative humidity RH*. It is typically expressed in percent % and a measure to represent the degree of saturation of the air. Thus, it is not a direct measure of the amount of water vapor in the air. Mathematically, the relative humidity (RH) is described as the ratio of the vapor pressure $p_V$ to the saturation vapor pressure $p_{V,sat}(T)$ [WH06]:

$$RH = 100 \cdot \frac{p_V}{p_{V,sat}(T)} \tag{1}$$

The vapor pressure $p_V$ is the partial pressure of the water vapor. If the vapor pressure is less than the saturation vapor pressure $p_{v,sat}$, the vapor, i.e. the moist air, is not saturated and can absorb more water molecules. When the vapor pressure $p_V$ is equal to the saturation vapor pressure $p_{V,sat}(T)$,

the number of molecules that condense and those that evaporate are equal, therefore the air is saturated. Thus, the saturation vapor pressure describes a state of equilibrium. Since the saturation vapor pressure increases with rising temperature, the relative humidity is also temperature dependent. Warm air can therefore hold more water molecules than cold air. As a result, two air parcels A and B with temperatures $T_A$ and $T_B$, where $T_A > T_B$, can have the same relative humidity $RH_A = RH_B$, although the amount of water contained in A is greater than in B.
Closely related to relative humidity is the *dew point temperature* $T_d$. The dew point indicates the temperature to which the moist air parcel must be cooled until it is saturated [LC79]. The relative humidity is therefore 100 % at the dew point temperature. As it cools further, the contained moisture condenses [Hau10].
The third moisture measure used in the further course is the *specific humidity s*. If no water molecules are added or removed from an air parcel either by condensation or by evaporation, the specific humidity $s$ is a constant humidity measure [Wet19b]. This property is useful for the comparability of two air parcels at different locations under the assumption that no humidity was added or removed.

**Global radiation**
Global radiation $G$ is defined as the sum of direct and scattered or diffusely reflected solar radiation [LC79]:

$$G = I \cdot sin(h) + D \tag{2}$$

With $I$ intensity of direct solar radiation (solar radiation energy per time and unit area perpendicular to the direction of radiation), $h$ solar altitude and $D$ diffuse radiation from clouds and sky [LC79]. When the sky is overcast, the global radiation $G$ is reduced because the high reflectivity (albedo) of the cloud cover reflects about 75 % of the incident radiation directly back into space, and thus only 25 % of the original incident radiation penetrates the cloud cover. An exception to the described scenario, are snow landscapes. The albedo of snow is higher than 80 %, therefore radiation can be reflected several times between clouds and snow [LC79]. The global radiation is given in $W/m^2$.

## 2.2 Machine learning

One speaks of machine learning (ML), if the machine learns certain skills by itself, without being directly programmed by a user [ZN18]. Hence, the term "machine learning" refers to multiple procedures that automatically identify patterns in data and can subsequently apply these patterns to previously unseen data to predict the outcome [Mur12].

### 2.2.1 Neural networks

A neural network or a neural net (NN) is just one of many possible applications of machine learning. Since a NN depicts a certain type of nonlinear statistical model or simply a nonlinear function $f : \mathbb{R}^p \rightarrow \mathbb{R}^r$, the relationship between a pattern and a real-value can be quantitfied by it [HTF09, AB09, Run10]. Each NN is specified by so-called weights $w$ of the individual values or connections [Run10]. These are trained during a suitable learning procedure, e.g., with input/output pairs $Z = (X, Y) \in \mathbb{R}^{p+r}$. Subsequently, the network is able to approximately reproduce or predict an associated output vector $y \in Y$ for an input vector $x \in X$, in a way that $y \approx f(x)$ holds true [Run10].
To understand how a NN is constructed and to what extent it maps a nonlinear function, a few basics are necessary first.

In their original form, artificial NN (ANN) form a simplified model of the biological NN [AB09]. This means that each ANN has neurons or units (comparable to biological neurons), which are structured in layers and partially connected with each other.
Each neural network has an input layer with $n$ input variables, also called features, inputs, or predictors [HTF09], and an output layer with $m$ outputs, also called responses [HTF09]. Both $n$ and $m$ are integers and represent the number of levels of the input and output vector $x$ and $y$, respectively. There can be any number of hidden layers between the input layer and the output layer. The larger the number of layers, the higher the complexity and the computational effort for the model. If the number of hidden layers is too low, the model cannot capture the nonlinearities in the data due to missing flexibility [HTF09]. Nevertheless, there are use cases where so-called shallow [Wer21] or traditional [Mat21] networks are used. They are characterized by only having one

hidden layer and thus a total of three layers. Networks with a higher number of hidden layers are called deep neural networks (DNN). The greater the number of layers, the deeper is the NN. The NNs in this work are in the shallow region of DNNs, as they contain a maximum of six layers.
The exchange of information between the individual layers and neurons takes place, both in biological and artificial networks, via the connections between them. Between which neurons a connection exists is decided by the network topology. In the following, only the topology of a fully connected NN will be explained since alternative approaches such as recurrent NN do not find application in this work. A NN is fully connected when all neurons of the intermediate and output layers are connected to all neurons of the previous layer [Wer21]. However, direct feedbacks between neurons, i.e., the connection of a neuron with itself, or recurrent connections are excluded in fully connected NNs [Kri07].
What information is exchanged across the existing connections, in both a biological and an artificial neuron, depends on the activation state of the neuron. A biological neuron starts firing when the intensity of activation exceeds a threshold value. In an ANN, each neuron processes information based on its activation function and its previous activation state into new information, which is emitted to the following neuron [Kri07]. The activation function is normally identical for all neurons, partly except for the output layer. Depending on the underlying task, there are both stepwise and derivable activation functions.
Figure 2 illustrates the structure and possible connections of a flat fully connected DNN example, comparable to the NNs implemented for this work.

In their further structure, realization and implementation of different ANN types differ fundamentally. First, networks can be divided into the following two groups based on their problem definition: Classification and Regression. Classification networks are those that have a qualitative output, whereas regression networks refer to quantitative outputs [HTF09]. More specifically, regression networks estimate the correlation of different features [Run10] and output a continuous response, rather than a class or a binary response as it is the case with classification networks [Böh09]. All networks considered in this work are based on regression models, since a continuous instead of a stepwise or classified output is wanted.
Another fundamental difference between different networks is their learning strategy. Generally speaking, a learning strategy is an algorithm that evolves

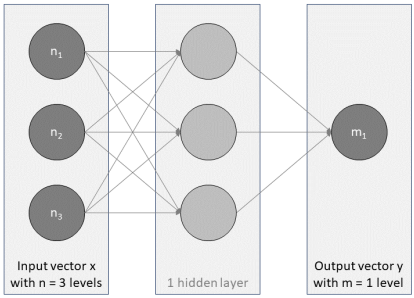


Figure 2: Structure of an exemplary neural network with connections for a fully connected NN shown only for first neuron of hidden layer and first, in this case also only, output neuron $m_1$. In this example the NN can be described by an input vector $x$ with n = 3 levels, one hidden layer and an output vector $y$ with m = 1 level.

the NN and thereby trains the model to refine the approximated output for a given input [Kri07]. The two best-known learning methods are supervised and unsupervised learning. In case of unsupervised learning, the network only receives input vectors and must use them to identify and classify similar patterns on its own [Kri07]. It does not receive feedback on the accuracy of the model throughout the learning process. In case of supervised learning, however, the model is provided with the correct output vectors in addition to the input vectors. Thereby, the model obtains information about the true relationship between input and output vector [AB09]. This process can also be called "learning by example" [HTF09]. First, the features are fed into the current state of the NN. In response, the model produces outputs. These generated outputs can be compared to target results and thereby a deviation of the current model outputs can be expressed. During its training phase, the NN continuously adjusts its model and thus its weights of the individual connections based on the deviations in order to achieve a gradually smaller deviation and thus create a more accurate network function $f$ [HTF09]. After completion of the training phase, the artificial outputs of the NN should be close enough to the target results so that the NN can be applied to further, unknown input data while providing reliable results [HTF09]. Murphy summarizes this as: "The goal is to learn a mapping from inputs to outputs." [Mur12]. In this work, supervised learning is used exclusively.

### 2.2.2 Random forests

To understand how a random forest (RF) algorithm works, which, as the name suggests, is a random collection of individual trees, the concept of a single tree, a decision tree, is necessary.

A decision tree resembles a natural tree in its structure and architecture, which gives rise to the naming. The branches in a decision tree are created by distinguishing the data based on their properties [LL20]. Rules are used to assign the data to their respective classes. Each rule implements a new decision level and thus another branch in the tree.
A RF is defined as an extensive collection of these independent trees [HTF09]. All trees in the forest are initialized and trained independently [Seg04] and randomly [LL20]. The size of each tree is given by the RF algorithm or the user, but the selection of the properties or decision criteria of each decision tree and level are based on randomness and therefore differ from tree to tree [LL20]. The overall result of the RF is determined based on the set of individual decisions or results of the individual decision trees [LL20]. For regression applications the results of the trees are averaged unweighted [Seg04, HTF09].

By initializing and combining many different independent result trees with random variances of each, a RF can achieve a better result than a single tree and increase its approximation accuracy [LL20]. By the RF method, at each decision level of each tree, only a random subsection of features or characteristics is provided to the algorithm for decision making. The subset is recomposed at each decision in each tree [Fro20]. By using random sets of input variables for growing the individual trees, the accuracy can be improved [Bre01]. Therefore, a RF algorithm is successful due to its additional randomization compared to other machine learning approaches [HTF09].

To adapt a RF to the respective use case, the user has various tuning parameters at his disposal, whereby, for example, the way of creating trees can be changed [LL20]. The maximum depths of each decision tree is one example: In general, the deeper a tree is, that is, the more decision levels it has, the lower its bias [Seg04]. However, the variance behaves in the opposite way, since with increasing depth the instability of the tree increases, which becomes visible in a decreasing approximation quality of the RF [Seg04]. Therefore, the depth of the trees should be limited by a tuning parameter. In order not to specify the maximum depth of the trees directly, a minimum number of cases at a decision

point can be specified instead [Seg04]. If this number is reached, the tree does not divide further, so that no new branch or decision level is created. This indirectly limits the maximum depth of a tree.

# 3 Sensors and data acquisition

Since the quality of the underlying data set determines the overall quality of the results, great care is taken with regard to data selection. In addition to investigated vehicle-based data, measurements for reference purposes are required. Without reference data, no conclusion regarding the quality of the vehicle-based signals can be made.
This chapter categorizes the sensors and station types available to the project. First, sensors relevant to meteorological applications installed on board of production vehicles are described in Chapter 3.1. The second subchapter discusses external sources and measurement equipment that can potentially be used as reference. The different types are described including the available signals promising for the project (cf. Chapter 3.2). The established requirements for reference data, as well as the selection of a suitable reference are listed in Chapter 3.3.
As of writing, no known, pre-existing data set meets the basic requirement of a continuous reference with the same temporal resolution as the vehicle data. Anderson et al. follow an approach where vehicle data is compared to a stationary reference, while the vehicles are in motion. In addition, the data has a low temporal and spatial resolution since it is averaged over 5 min and 1.6 km [ACD+12], respectively. In another publication, a different data set with mobile vehicle-based data is recorded with the limitation that data is collected only with vehicle external auxiliary sensors rather than on-board sensors. For evaluation no reference is used, instead the relative change of the measurement data due to change of location of the vehicle is considered [KKS+19].
Therefore, in this thesis measurement campaigns (cf. Chapter 3.4 and 3.5) are designed and conducted to generate a first-hand data set that enables evaluations and statements on the quality of the vehicle-based data by ensuring almost identical environmental conditions for both vehicle and reference. Besides complete vehicular data, including CAN traces and additional on-board sensors, a suitable ground truth is recorded alongside the vehicular data set.

None of the conducted measurement campaigns take place in a laboratory. Instead, all measurement campaigns are real-world measurements, since multiple influences that cannot be recreated in a laboratory have an impact on the measurements. An example is the influence of the changing altitude on the temperature measurement or the influence of the sun on the global radiation measurement. In addition, influences that are unknown or unquantified cannot be represented in a laboratory. Measurements in a laboratory would therefore not provide representative data, so real-world measurements are required.
Due to current regulations regarding data privacy (GDPR, [Ger18]), no field data from customer vehicles can be used for this work. Instead, two company owned vehicles (in the following technical carriers, TCs) are available for data collection.

## 3.1 Sensors on board of production vehicles

On board of current production vehicles, a variety of different sensors for a range of different requirements and purposes exist. Most of these sensors ensure flawless operation of the vehicle itself as well as for passenger comfort. The second biggest group of sensors monitors the environment of the vehicle. Well known functions driven by these sensors are for example the adaptive cruise control, which processes data of detected objects in the surrounding. Although less exposed, modern vehicles are equipped with a wide range of environmental sensors by default which promise potential for meteorological applications and are the focus of this work. Due to economic constraints and purely unit-driven business cases any sensors other than to meet the minimum requirements to operate a vehicle as advertised have hardly been considered in the past. To enter series production with an additional set of sensors for purposes and use cases that exceed the basic operation of a vehicle involves both technical and entrepreneurial risks. Without well-established business cases it is unlikely for major OEMs to fit auxiliary or specialized sensors to vehicles for applications beyond this in the near future. Therefore, the currently available sensor set on board of production vehicles forms the scope for generating a real-time weather map based on fleet data. Custom fitted sensors are solely used for reference purposes and are unavailable once the project possibly enters series production. Hence, in the end, the whole process of determining ambient conditions must solely be based on these established

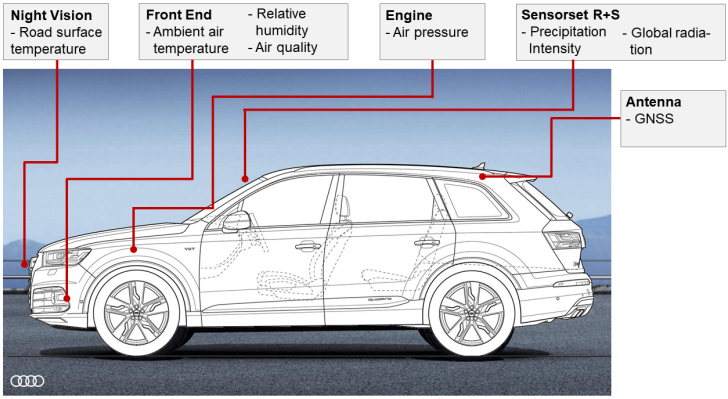


Figure 3: Overview of sensors on board of production vehicles relevant for meteorological applications.

sensors.
A major objective of this work, therefore, is to investigate how well exclusively sensors on board of production vehicles are suitable to generate a real-time weather map based on fleet data. Figure 3 gives an overview of sensors on board of production vehicles that are relevant in a meteorological context.

The foremost sensor providing data is a far-infrared nightvision camera. This camera can potentially detect the road surface temperature. However, this requires the development of a new analysis program. Up to date, the sensor is used for pedestrian and object recognition exclusively and no information about road conditions is gained.
In the front end of every vehicle, several sensors interesting for meteorological applications are located. Two air temperature sensors operating independently are mounted in different positions (cf. Figure 4). The front sensor is mounted directly behind the radiator grill in a height of approximately 0.32 m and is referred to as BCM1 sensor in the following. Its measurement uncertainty amounts to ±0.3 K (at a temperature value of 25 °C) with a system-restricted resolution of 0.5 K (cf. Appendix A.1). The second sensor that measures temperature among other parameters is located further downstream, closer to the engine block (cf. Figure 4) at the air intake for the cabin at a height of 0.80 m. In the following, this multifunctional sensor is referred to as MuFu sensor.

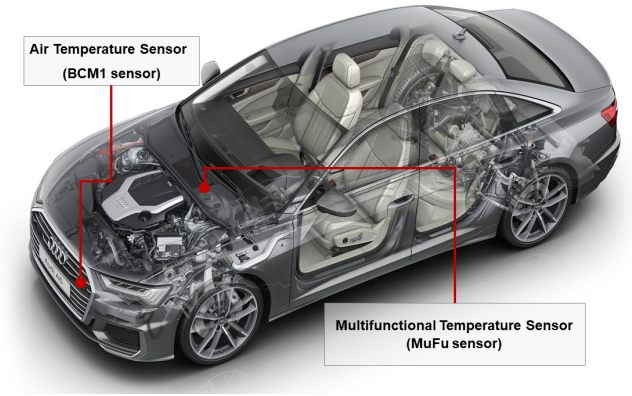


Figure 4: Two different temperature sensors located in front end of vehicle: BCM1 sensor and MuFu sensor.

For a temperature of 25 °C its measurement uncertainty likewise amounts to ±0.5 K. The measurement uncertainty for both temperature sensors increases for temperatures diverting from 25 °C. The resolution of the temperature signal of the MuFu sensor amounts to 0.5 K (cf. Appendix A.1).
The MuFu temperature sensor is a multifunctional sensor, which also measures relative humidity and dew point temperature. The relative humidity signal has a measurement uncertainty of ±3 %rH and a resolution of 1 %rH. For the dew point temperature signal neither a measurement uncertainty nor a resolution is specified, since it is an indirect signal calculated based on the measured air temperature and relative humidity.
The engine control unit itself provides the air pressure signal (measurement uncertainty: ±10 hPa, resolution: 7.9 hPa) as well as information in conjunction with the combustion process, including the engine temperature.
Another important sensor for meteorological applications is the sensor combination "Rain and Sun" (shortform: R+S), which is installed in the center of the windshield at a height of 1.38 m, in the socket of the rear mirror. The sensor has two fields of activity. Primarily, it detects precipitation and controls the wipers. In the course of the FloWKar research project the unit of the precipitation intensity was changed from an arbitrary intensity with units in % to a scale in ml/s/m$^2$. This development allows for initially usage the precipitation

intensity signal for meteorological purposes, which was previously unfeasible. Due to very limited number of data points during rain events and only temporarily available reference during these events, the rain signal is not further considered in the scope of this work. In addition to rain intensity, the sensor registers two signals for light intensity, directed to the front left and front right. The conversion to global radiation values is performed as a standard using a model from a subsidiary of the AUDI AG. The global radiation values have a measurement uncertainty of ±3-5 %.
A crucial prerequisite for integration and comparison of the sensor signals with a meteorological context is the ability to pinpoint the observations geographically as well as temporal. Therefore, the antenna module is essential since it provides the system with Global Navigation Satellite System (GNSS) coordinates (cf. Figure 3). The GPS signal is too inaccurate to control automated driving functions via this alone but is perfectly adequate for meteorological applications.
All sensors and signals relevant for answering the research questions of this thesis are listed in Table 21 in Appendix A.1 in detail. Other signals of interest in a meteorological context, such as spray indication, are available on board of production vehicles. However, they do not find any application in this work and are therefore not detailed further.

## 3.2 Sensor technology for reference purposes

Besides the presented sensors on board of production vehicles, multiple opportunities for collecting meteorological data are made use of during the project phase. Especially for reference purpose, additional external data sources are indispensable.
Table 1 gives an overview of the available additional external data sources. Characteristics describing the individual data sources, for example, mobility possibilities, are listed. In Table 2 an extract of measured signals for the respective data sources relevant for this work is illustrated. All data sources measure air pressure, air temperature and relative humidity, whereas, for example, global radiation is not included for all data sources. Detailed information, such as measurement uncertainty for each data source, are listed in Table 22 in Appendix A.1.

Synop stations are weather stations operated by the national weather service that provide meteorological data used for the numerical weather models (cf. column “Synop” in Table 1). These stations are located in largely unaffected environments. This excludes for example locations in a residential area, since buildings and objects influence the air flow and thereby the measurements might by distorted. Instead, locations amid fields are preferred. Beyond this the ground beneath the measurement equipment is standardized, too. Mowed lawn is required, any other materials, for example asphalt, are prohibited. Thereby data from different synop stations can be compared due to sufficiently identical environmental conditions. The set up often includes multiple different sensors from different manufacturers and is stationary in its nature. The individual stations are typically separated by several kilometers up to tenth of kilometers. Besides the relatively coarse spatial resolution, the temporal resolution is also relatively low. By design of the national weather service the meteorological measurement equipment of the synop stations provides one value for each quantity every 10-15 minutes. The data set contains a variety of meteorological signals. However, only the following will be considered for reference: Air pressure, air temperature, relative humidity, dew point temperature, precipitation intensity and global radiation. Air temperature is measured at two different heights of 0.05 m and 2 m simultaneously. Since the vehicle-based air temperature measurement is conducted in a height of 0.3 m above ground, neither of the two synop station temperature measurements matches perfectly. Nevertheless, the vehicle-based measurement is conducted in a height similar to the lower of the two synop station measurements. The

| Overview | Synop | RWS | MWS | WB | MMU |
|---|---|---|---|---|---|
| Short form for | Synoptical station | Road weather station | Mobile weather station | Weather box | Mobile measuring unit |
| Mobility | Stationary | Stationary | Partly mobile | Partly mobile | Mobile |
| Ground beneath measurement | Defined as gras | Road/ roadbed | Optional | Optional | Road |
| Recording rate | 10-15 minutes | 10 minutes | 1 second | 1 second | 1 second |
| Spatial resolution | Several kilometres | Several kilometres | - | - | - |
| Measurement equipment | Meteorological sensors | Meteorological sensors + road condition sensors | Meteorological sensors | Vehicle sensors | Meteorological sensors |
| Calibration | Once - before assembly | Once - before assembly | Once - before delivery | Once - before delivery | Frequently - every 24 months |

Table 1: Overview of different station types available for recording reference data with characteristics of each station type

| Measured signals | Synop | RWS | MWS | WB | MMU |
|---|---|---|---|---|---|
| Air pressure | ✓ | ✓ | ✓ | ✓ | ✓ |
| Air temperature | ✓ (0.05 m + 2 m) | ✓ | ✓ | ✓ | ✓ (0.7 m + 2 m) |
| Wet bulb temperature | - | - | - | - | ✓ (0.7 m + 2 m) |
| Relative humidity | ✓ (2 m) | ✓ | ✓ | ✓ | ✓ (2 m) |
| Dewpoint temperature | ✓ (2 m) | ✓ | - | ✓ | - |
| Global radiation | (✓) | - | ✓ | ✓ | ✓ |

Table 2: Extract of signals measurable by individual external data sources relevant for this work. Height information are included if invariable. Note, that all data sources measure additional meteorological parameters. Since these are not relevant to this work, they are therefore not listed.

relative humidity and the dew point temperature are both measured in a height of 2 m above ground. Neither the air pressure, nor the precipitation, nor the global radiation are somehow influenced by height, wherefore no heights are specified. Noteworthy is that the applicability of the global radiation measurements as a reference is limited to only a subset of synop stations as not all stations are equipped with a pyranometer.

The second station type considered for reference purpose are RWS (cf. column "RWS" in Table 1). As the name already implies, these stations are located next to roads, mainly next to highways and rural roads. As a result, the measured signals might be influenced by characteristics of the road and traffic. The proximity to the road explains that the ground beneath the measurement equipment is the road itself or the roadbed respectively. Thereby these stations can reflect the conditions on the road well but are not directly comparable to measurements from synop stations. Another downside of these stations is the commonly poor condition of the stations. Unlike with synop stations, no periodic maintenance is done after installation which results in falsifying effects due to for example shadows of the growing trees and bushes in the surrounding. Maintenance of RWS would increase costs and is therefore frequently reduced to a minimum or cut completely. Most often the stations are used for distinguishing between hazardous and safe driving conditions by road authorities though. Since RWSs are the only stations with information on road water film thickness, as well as road temperature and salinity, this unique feature provides decision support for road weather warnings. Typically,

RWS transmit data every 10 minutes. Exclusively for this project, seven RWS are adjusted to transmit data every minute. Beside meteorological sensors for air pressure, air temperature, relative humidity, dew point temperature and information about precipitation, a RWS is additionally equipped with sensors measuring road surface temperature and road surface conditions.

The mobile weather station (in the following: MWS) is a partly mobile station, which is a benefit in comparison to the two previously described station types (cf. column “MWS” in Table 1). Since the power is supplied by a car battery, this station can be set up at any desired location. For the duration of the measurement the MWS is stationary though. Due to the flexible choice of site, the ground beneath the measurement can vary. Independently of location, the MWS transmits data from meteorological sensors every second. The extent of the transmitted data is listed in Table 2. Although it is limited to five meteorological signals, all crucial parameters are covered.

The fourth type of potential reference station is the so-called Weather Box (in the following: WB). The respective information are displayed in column “WB” in Table 1. In the WB, production sensors, normally mounted in a vehicle, are installed in an industrial enclosure which results in measurements independent of the otherwise adjacent components of a vehicle. Both characterisics, "mobility" and "ground beneath measurement" are identical to the MWS, since the WB can be set up at any desired location whereby the ground beneath the measurement can change. The height of the WB is adjustable for each measurement campaign. The temporal resolution of one second is identical, too. Even the extent of the measured signals does not differ significantly (cf. Table 2). The main difference is the already mentioned use of vehicle sensors in the WB, instead of meteorological sensors for the MWS.

The last available data source is listed in the column "MMU" in Table 1. The mobile measuring unit of the national weather service (in the following: MMU) is a data source being mobile continuously. This is achieved by mounting meteorological measurement equipment on a metal structure in front of a vehicle. Thereby, the sensors are located upwind of the vehicle and distorting influences of the dynamics of the vehicle are reduced to a minimum. In general, the MMU combines the benefits of the previously presented four station types. The meteorological equipment is mobile and can thereby continuously measure close to the TCs. Hence, the spatial representativeness of the data set in comparison to data from the TCs is guaranteed. Additionally, the high temporal

resolution of one second is advantageous. Driving at a speed of 10 km/h results in a measurement location every 2.5 m, a vehicle speed of 100 km/h yield a measurement every 28 m. Finally, it is worth mentioning that this is the only station type that is calibrated frequently and on a regular basis. Regarding the measured signals most meteorological quantities are represented, except precipitation. The two signals air temperature and wet bulb temperature are even measured at two different heights, with the lower of the two (0.7 m) being close to the height of the production vehicle-based temperature measurements of 0.3 m.

In summary, each station type has its benefits regarding serving as reference for vehicle-based measurements. Being stationary or only partly mobile is yet a strong limitation, since the spatial representativeness is reduced significantly. Only data measured at timestamps when the TC was close enough to the station, can be used for referencing purposes. As a consequence, the available amount of data is reduced drastically. The composed information already implies that the MMU is most suitable for serving as a reference station. Nevertheless, due to its uniqueness, the MMU is not continuously available. Furthermore, the MMU reaches its limitations regarding precipitation. Therefore, a combination of different station types is most preferable to function as ground truth data.

## 3.3 Selecting a suitable reference

For varying measurement scenarios, reference data needs to meet different quality requirements. For a stationary measurement campaign, for instance, the source of reference data can be stationary, too. But for a mobile measurement campaign, the same data source would only provide reference data with insufficient quality since the data have a strongly limited local representativeness. In case of the TC being in the immediate vicinity of a stationary data source and being affected by the same environmental conditions (ground, air flow, irradiation, altitude, etc.), this data can be used as appropriate ground truth. For periods with some distances between the TC and the second data source, different surfaces on the ground or further factors occur, causing varying environmental conditions to both data sources. Hence, a direct comparison of both stations is not expedient. Therefore, a different data source should be

used as reference for the vehicular data, ideally a data source that can move along with the TC without measurements being disturbed by its dynamical effects. It becomes apparent that different data sources should be used for reference, depending on the measurement scenario.

In a stationary scenario, the MWS (cf. Chapter 3.2) could serve as a suitable source for ground truth data. Calibrated and off the shelf measurement equipment, as well as cost-effective purchase and flexibility regarding the installation site are benefits of this data source. Moreover, a second reference station useable in a stationary scenario is the WB (cf. Chapter 3.2). Nevertheless, the WB should not be used as the only reference, since the same sensor technology is used here as in the vehicles and thus measurement errors caused by the sensors cannot be identified. Using the WB as an additional reference station in combination with the MWS instead, is of great added value, as this allows a multi-level correction, divided into measurement uncertainties due to equipment and measurement uncertainties due to the way of installation in the vehicle.

Assuming the TC is moving, neither the MWS nor the WB are an appropriate source of reference data any longer, since data of both stationary measurement systems is not no longer continuously representative for the environment the TC is driving in. For this scenario the MMU (cf. Chapter 3.2) is the method of choice for recording reference data. Comparable to the MWS, the measuring equipment is also calibrated. Additionally, the MMU can accompany the TC continuously, whereby the representativeness through time and space of the mobile reference data is ensured.

The three reference options mentioned (MWS, WB, MMU) have the advantage over the stationary RWS and synop stations that their temporal resolution is equal to the temporal resolution of vehicular measurements, which is one second. Both RWS and synop stations have a significantly lower temporal resolution, varying between one and ten minutes depending on the station. Due to the low frequency of reference data of these two station types, this data is only sporadically accessed for evaluation.

The uniqueness of the data set generated as part of the project is based on combining different data sources, which results in a data set not only containing vehicular data from the TCs and stationary reference data from the MWS, but also including stationary reference data from the WB and mobile reference data

from the MMU. The described data set is of considerable value for all analysis in a meteorological context and thereby builds the base for the project and this work. It is a unique characteristic of the project, and thereby a strength, and a hurdle alike, since at the start of the project this data set, or a comparable one, were not existent. Therefore, a lot of time was spent on the design and setup of the experiments as well as the eventual data acquisition itself.

For generating the previously described data set including vehicular data and reference data multiple measurement campaigns with different characterizing features were necessary. The most obviously varying feature is mobility during the campaign, resulting in stationary and mobile measurement campaigns.

## 3.4 Stationary measurement campaigns

Stationary measurement campaigns have the purpose of determining the basic accuracy of the sensors mounted in vehicles, whose measured signals have to be quality assured for further applications. As mentioned in the previous subchapter, mainly two different reference stations are used. Firstly, the MWS (cf. Chapter 3.2) and secondly the WB (cf. Chapter 3.2). Combining the three data sources, namely both reference stations and a vehicle, gives information about the basic accuracy of the mounted vehicle sensors and the influence of the position of the sensor in the vehicle. Hereby, the measurement uncertainty in stationary situations can be evaluated.

During the project, two stationary measurement campaigns are executed, one in November 2020, the other one in January/ February 2021. During these stationary campaigns both TCs were located on the north side of the factory plant of the AUDI AG in Ingolstadt. The windshields were active and continuously facing north-east. Both TCs were powered off with all systems active, except the engine. Therefore, none of both TCs was moved. Southwest of the vehicles is a multi-story building of AUDI AG. The reference stations (MWS and WB) were set up directly behind the two vehicles at a distance of 1 to 2 m on a small grass strip. Due to the close spatial arrangement of the measurement campaign, the environmental conditions can be assumed to be the same, and the data can be compared directly.
The first stationary measurement campaign in November 2020 lasted more

than 3 days and thereby recorded 294431 data points for both TCs, the MWS and the WB simultaneously. The measurement has taken place during a slight cold snap, as the maximum temperature values drop daily. The maximum air temperature was about 9 °C on the first day of measurement and decreased by about 7 K in the following days. The minimum temperatures also decrease over this time period and is below 0 °C in three out of five days. On the last measuring day, the temperature values increased again.
For the second stationary measurement campaign in January/ February 2021, the measurement lasted about 7 days, resulting in 605427 data points with concurrent data for both TCs, the MWS and the WB. During this measurement campaign, milder weather prevailed than in November. The maximum temperature was about 10 °C. The minimum daytime temperature fell only occasionally to just below 0 °C. Air temperatures were therefore above freezing most of the time, but often below an artificially set mark at 5 °C. To avoid freezing of the windscreen and damage to technical equipment, a software changeover took place in the TC between the two stationary measurement campaigns. The auxiliary heater is automatically switched once the measured air temperature falls below 5 °C. The influence of the auxiliary heating on the measurements, especially on the temperature and relative humidity measurement, must be specifically investigated and evaluated before corrections compensating for measurement uncertainties induced by stationary situations can be derived based on this data.

Since the MWS suffered from sensor failures in the global radiation signal, another stationary measurement campaign was conducted. For this measurement campaign, a TC was positioned in Dürnast next to the synop station "Weihenstephan" for 4.5 days. This synop station is one of the few stations in Germany, where the global radiation is recorded by a pyranometer and has a temporal resolution of 1 minute. Due to the lower resolution compared to other reference stations, this stationary measurement campaign provides only 613 concurrent data points of the reference station and the TC, despite a duration of more than 4 days.

## 3.5 Mobile measurement campaigns

Mobile measurement campaigns contribute the remaining data to the data set. Their purpose is to identify the influence of the dynamics of the vehicle on the measured signals. For these campaigns the TC participates in traffic as a normal road user to generate data under real-world conditions. Thus, the vehicle is moving most of the time. During these campaigns data was collected for as many different scenarios as possible. This includes for example varying vehicle speeds, in scenarios downtown, out-of-town and highway of up to 100 km/h, as well as different weather situations, for example warm and sunny weather, but also rain and snowstorm. For periods, where the vehicle comes to a stop, for example due to traffic lights or stop and go traffic, the data recording continues unchanged. Out of this, a variety of scenarios is recorded during these project specific mobile measurement campaigns, covering the range of everyday usage of a vehicle. Each measurement campaign has been accompanied by the MMU providing full coverage with reference data.

The disadvantage of mobile measurement campaigns compared to stationary measurement campaigns (cf. Chapter 3.4) is the influence of extraneous traffic on the measurements, such as spray, turbulence, or shadows. These influences cannot be excluded during measurement campaigns in real road traffic. In addition, mobile measurement campaigns are only possible during working hours and with the availability of the participants, whereas stationary measurement campaigns can record data permanently. However, mobile measurement campaigns are the only possibility to collect data while driving and during real-world traffic. This database is mandatory to develop corrections for effects induced by the vehicle motion of the own vehicle.

During the project phase three multi-day mobile measurement campaigns were conducted, focusing on collecting data during winter season, since for this season road weather is most critical for applications related to vehicle and traffic safety.

The first measurement campaign took place from January $20^{th}$ to $24^{th}$, 2020. During these 5 measurement days, a total of 92849 data points with concurrent measurements for the primary TC and the reference of the MMU was recorded. During this measurement campaign, data was mostly collected while driving (country road and highway), which explains the 1405 km covered (cf. Fig-

ure 5a). In addition, data was recorded at the mountain "Hohenpeißenberg" in southern Germany. Here, significant changes in altitude occur within a few kilometers, whereby a vertical profile of meteorological variables can be approximated. Since most meteorological variables show significant and fast changes with a change in altitude, this can provide an indication of the inertia of the vehicle-based measurements. In addition, data was also recorded in Bergisch Gladbach, at the test site of the German Federal Highway Research Institute (BASt) during the January measurement campaign. Although long distances were covered during this measurement campaign, the weather situation has been comparatively uniform. In southern Germany, temperatures were mostly below the freezing point for the first three days of the measurement campaign. The last two days, which took place mainly in Offenbach and Mönchengladbach, recorded slightly warmer temperatures, with the maximum daily values above the freezing point. At night, temperatures have fallen below 0 °C for the entire period. Since the measurement campaigns took place during the day, the temperatures here were slightly higher, but still frequently below 0 °C. No precipitation occurred during the entire measurement campaign, neither in liquid nor in solid form.

The second measurement campaign was conducted from February $10^{th}$ to $12^{th}$, 2020 (3 days). A total of 54310 data points with concurrent measurements of the primary TC and MMU was recorded. The geographical focus of the measurement campaign was Offenbach (cf. Figure 5b) for logistical reasons. From there, the synop station "Wetterpark" usable as another reference possibility,

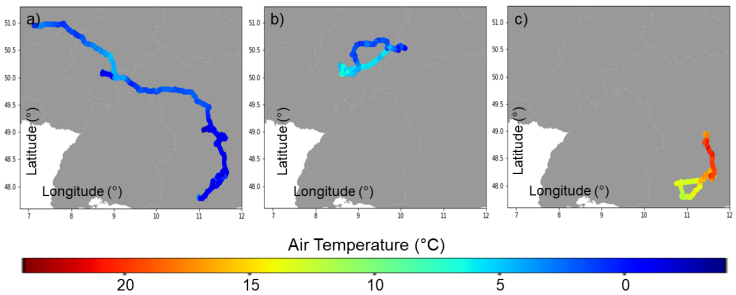


Figure 5: Reference air temperature measured by MMU plotted for all measurement campaigns to visualize route of all campaigns, a) January 2020 b) February 2020 c) September 2020 campaign.

as well as the northeast located "Wasserkuppe" due to its elevated altitude and the city of Frankfurt with dense rush hour traffic for stop and go scenarios were chosen as targets. The weather situation during these three February days was mixed, but especially on the second day characterized by blizzard conditions and heavy snowfall. Temperatures were near the freezing point, so the falling precipitation frequently alternated between rain, snow, sleet and hail. Therefore, the road situation was also variable, fluctuating between dry, wet and snow-covered roads.

From August $31^{st}$ to September $2^{nd}$, 2020, the third measurement campaign took place starting in Ingolstadt. In the three measurement days 83246 data points could be collected for which measurement data of the primary TC as well as the MMU is both available. In order to have a reference measurement for precipitation available for individual passing summer thunderstorms, the synop station "Hohenpeißenberg" in southern Germany was approached. At the same time, vertical profiles for all meteorological quantities could be recorded by trips up and down the mountain "Hohenpeißenberg". Due to the, apart from individual precipitation events, sunny weather situation, the synop station "Weihenstephan" was approached as a secondary destination, in order to cover another possibility of reference data for global radiation on the basis of this station. The weather during the three days was mostly friendly and summery with maximum daily temperature values above 17 °C. On the second day very isolated summer thunderstorms occurred.

During the three mobile measurement campaigns described above, a total of 230405 data points were collected covering 2765 km. All data recorded within the scope of these mobile measurement campaigns is characterized by concurrent measurements for the primary TC with the reference of the MMU at different weather situations, in different seasons with different driving scenarios. Additional observations for reference purposes are available for individual time periods from the synop stations mentioned above. The time window to be used depends on the distance of the TC to the station: For distances exceeding 500 m, identical environmental conditions can no longer be assumed, whereby the measurements are not suitable as a reference and are therefore not considered. Due to technical problems during all three measurement campaigns, concurrent data of the secondary TC is not continuously available. The basis for the evaluations and analyses are therefore the data from the primary TC.

# 4 Quality control and correction of measured values

Chapter 4 outlines the methodology used for quality control and correction of the individual measured meteorological quantities. Theoretically, all meteorological quantities should be tested and corrected following the same scheme (cf. Figure 6). First a input filter, including identification and elimination of logical error as well as additional outliers is conducted. Secondly and thirdly, sensor uncertainties in static situations and sensor uncertainties induced by vehicle motion are compensated. The measurement campaigns described above serve to separate and thus detail these working steps.

However, since the individual implementation of each of the three steps varies depending on the meteorological quantity investigated, this chapter is split by the meteorological quantities rather than the three steps. Subchapter 4.1 explains the working steps for correcting the air pressure, Subchapter 4.2 for air temperature, Subchapter 4.3 for relative humidty and Subchapter 4.4 for global radiation.

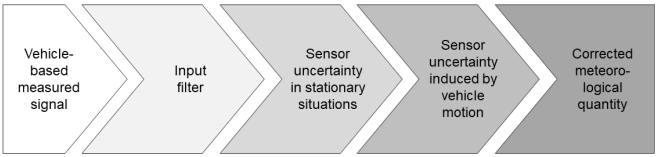


Figure 6: Schematic flowchart to test and correct meteorological quantities in the scope of this thesis.

## 4.1 Air pressure

The stages of filtering and correcting the air pressure signal are shown in Figure 7.

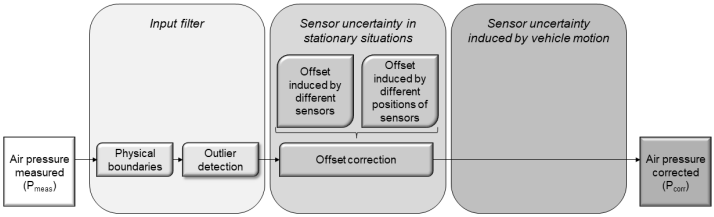


Figure 7: Flowchart for testing and correcting the air pressure signal, consisting of an input filter and compensating for sensor uncertainty in static situations.

### 4.1.1 Input filter

First, the measured air pressure signal $P_{meas}$ is subjected to an input filter. This stage consists of two working steps:

1. *Ensuring all measured values are within the measurement range.*
   In this step single raw measurements outside the physical limitations are modified. For the air pressure the only physical limitation is 0 hPa since no negative air pressures are plausible. If a measured value falls below the physical limitation, the duration of this state is checked. If it lasted shorter than a preset threshold of 5 s, the measurement of this timestep is corrected by holding the previous, plausible measurement. If the reading falls below the physical limitations for longer than the threshold allows, the erroneous measurement is excluded and replaced with a logical gap, since a failure of the measuring system is likely.

2. *Ensuring no outliers are present in the data.*
   This working step checks for sudden changes, which exceed a preset threshold of 15 hPa. This threshold seems reasonable since the air pressure is a rather inert, slowly changing quantity. Sudden changes exceeding twice the measurement resolution within 1 s can therefore

be identified as a falsified measurement. Once the outlier detection is triggered, the previously measured value is hold for no longer than 5 s.

### 4.1.2 Sensor uncertainty induced in static situations

The next stage addresses the measurement uncertainty of the vehicular sensor becoming apparent in static situations. To identify this uncertainty, static measurement campaigns as described in Chapter 3.4 were conducted. An offset between vehicle-based sensor and reference measurement is likely due to two aspects. First, due to an offset induced by different sensors for the reference station and the vehicle and secondly due to an offset induced by the mounting location of the sensor. The setup of the static measurement campaigns with two different reference stations (MWS and WB) enables the consideration of both offsets independently.
The first offset, induced by different sensors, can be determined by comparing the signal of the WB with the signal of the MWS, as illustrated by the upper grey arrow in Figure 8. Since the MWS uses a calibrated sensor, it is assumed to be the ground truth. The WB uses the same sensor mounted in the vehicle, whereby the difference between the WB signal and the MWS signal represents the offset induced by the sensor itself.
The second offset, induced by different locations of the same sensor, can be identified by comparing the WB signal with the vehicle-based signal (cf. lower grey arrow in Figure 8). The sensors in the WB and in the vehicle are the same model and only differ by their mounting position. The sensor in the WB measures in an undisturbed environment, whereas the vehicle-based sensor measures within the engine block of the vehicle.

Ideally, the total sensor uncertainty in static situations $SU_{stat,total}$ (deviation between the MWS and the vehicle-based measurement) can be described as the sum of the uncertainty based on the sensor technology $SU_{Sensortechnology}$ (difference between MWS and WB) and the uncertainty caused by the location of the sensor in the vehicle $SU_{Location}$ (difference between WB and vehicle) and can thereby be written as:

$$SU_{stat,total} = SU_{Sensortechnology} + SU_{Location} \tag{3}$$

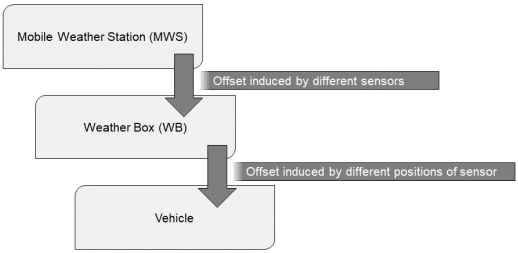


Figure 8: Compare measurements from Mobile Weather Station (MWS), Weather Box (WB) and Vehicle in two steps to identify the two offsets that build the sensor uncertainty in static situations.

If Equation 3 holds true, it can be assumed that no additional influencing factors are involved during campaigns in a static situation or are negligible. Hence, the transfer of vehicular data to stationary reference data can be modelled by the previously described two-stage correction method.
It is notable that the third stage, compensating for sensor uncertainty induced by vehicle motion, is omitted. The vehicle-based air pressure measurement represents a signal, which is independent of the motion of the vehicle. According to the ideal gas law [LC79], a change in pressure is expected when one of the factors affecting the pressure $p$ changes.

$$p \cdot V = \frac{m}{M} \cdot R_m \cdot T \tag{4}$$

Since neither the volume $V$, the mass $m$, the molar mass $M$, nor the universal gas constant $R_m$ change as a function of vehicle velocity, only a change in temperature $T$ can lead to a change in pressure. According to the manufacturer [Rob12], the pressure value output by the sensor is already corrected for fluctuations in temperature and therefore does not exhibit any further effects induced by vehicle motion. Therefore, the deviations between the vehicle-based measurement and the respective reference data are almost exclusively due to the sensor uncertainty of the production sensor system. Thus, the previously described two stages are sufficient to obtain representative data for ambient air pressure.

## 4.2 Air temperature

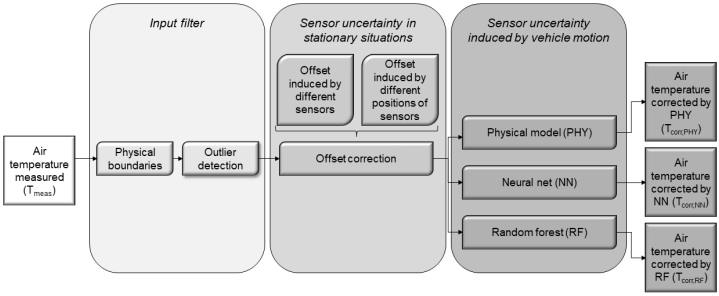


Figure 9: Flowchart for testing and correcting the air temperature signal, consisting of an input filter, compensating for sensor uncertainty in stationary situations and correcting for sensor uncertainty induced by vehicle motion. Note, that three different approaches for the third stage are implemented.

Figure 9 displays the procedure of filtering and correcting the air temperature readings to derive a representative value from the on-board measured signal. The first two stages are similar to the approach for correcting the air pressure signal, wherefore only the differences regarding these two stages will be described here.

### 4.2.1 Input filter

For the input filter of the measured air temperature signal $T_{meas}$ the two working steps are fundamentally identical to the air pressure approach. Both approaches only differ by the thresholds set for physical boundaries and outliers:

1 *Ensuring all measured values are within the measurement range.*
The physical limits for the air temperature signal are set to -40 °C as the minimum permissible temperature value and to 80 °C as the maximum permissible temperature value. The lower limit of the physical limitations is defined by the operating range for the thermometer installed in the vehicle. According to the sensor's data sheet, the upper physical limit is 155 °C [Amp15]. However, since the sensor is only indirectly

exposed to the engine block or any other heat source, the upper plausible physical limit is lowered to 80 °C. The threshold regarding the accepted duration of the measured value being outside the physical boundaries is set to 5 s, identical to the approach for the air pressure.

2 *Ensuring no outliers are present in the data.*
The pre-set threshold for accepted extents of sudden changes in the temperature value amounts to 5 K. Therefore, it is ten times the measurement resolution. Since the air temperature is a quickly, both temporal and locally, changing meteorological quantity, a proportionally higher threshold seems reasonable. The temporal threshold of 5 s is identical to the approach regarding the air pressure.

### 4.2.2 Sensor uncertainty induced in static situations

The working steps of stage two, to determine the sensor uncertainty of the vehicular sensor becoming apparent in static situations, fully align with the steps for the air pressure signal. Therefore, the static measurement campaigns are likewise used to determine the respective offsets for this meteorological quantity.
Note, that effects introduced by the motion of the vehicle cannot be specified at this time, since the vehicle remains in standstill for the whole campaign without running engine. Solely, the influence of the sensor model and the location of the sensor can be determined.

The determination of the total sensor uncertainty in static situations $SU_{stat,total}$ refers to the difference between the MWS and the vehicle-based measurement and can be calculated by Equation 3. For the statement to hold true, it is assumed that no other influences affect a measurement in a static environment.

### 4.2.3 Sensor uncertainty induced by vehicle motion

After compensating for sensor uncertainties in static situations, effects induced by the motion of the vehicle need to be corrected to get a representative air temperature signal. However, it must be taken into account that only the effects induced by the vehicle's motion and no local effects may be compensated.

Changes caused by local effects need to be identified in order to preserve them and thus obtain representative values after the correction. A prominent example of a local effect is the decrease of air temperature with an increase in altitude and vice versa [LC79]. This effect is measurable with vehicle-based sensors (positioning module) and thus represents an important local effect that must not be falsely compensated. Therefore, the physical model needs to identify the reason for a change in the measured air temperature signal first, before a correction is executed either way. If the temperature change is due to a change in altitude, the measured value is retained.

The greatest effect, which occurs due to motion of the vehicle, is caused by the engine waste heat in combination with the vehicle speed [HAVP+20]. This effect is noticeable at low speeds or during stops. Specifically, the measured air temperature increases. Even after the vehicle accelerates or starts again, the effect initially persists.
At low vehicle speeds or when the vehicle is stationary after the engine has been running, the air flow under the hood is reduced. Thus, the air heated by the engine is not replaced by fresh air, but instead remains under the hood and continues to build up as it is continuously irradiated by the engine waste heat. The measured air temperature therefore rises with increasing duration of standstill.
Immediately after the vehicle builds up speed again or continues after a stop, the measured air temperature is still distorted by the influence of the previous stop. This effect decreases with increased flow under the hood and time, until a measurement unaffected by the engine waste heat can be detected again.
This effect occurs especially in stop-and-go scenarios or in rush-hour traffic. It leads to excessively high air temperature values by the production sensor system.

To compensate for this sensor uncertainty induced by vehicle motion three different approaches are developed, implemented, and tested. The explanation for the effects of vehicle motion on the measured temperature are decisive only for the development of the physical model. However, for the data-based models, they help to select suitable features for training the models.
In the following, all three implemented models, namely the physical model, the neural network and the random forest, are described in detail.

### Correction with physical model

The first approach tested is a model based on physical connections and empirically determined thresholds. First attempts using this approach can be found in [HNK+19].

To decide whether the engine waste heat has an influence on the measured air temperature signal, the physical model introduces an "Engine Impact Timer" (EIT). The EIT is calculated based on the vehicle speed, the change in temperature since the vehicle stopped, and indirectly on the duration of the stop.
For determining the EIT at each time step a decision tree with three levels is implemented (cf. Figure 10).

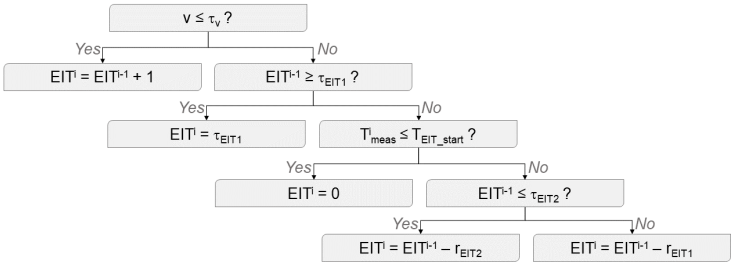


Figure 10: Flowchart for the Engine Impact Timer (EIT), referring to current timestep i and previous timestep $i-1$, vehicle speed v, threshold for vehicle speed $\tau_v = 5$ km/h, first threshold for EIT $\tau_{EIT\,1} = 600$, and second threshold for EIT $\tau_{EIT\,2} = 60$, first rate of decline $r_{EIT\,1} = 2$, and second rate of decline $r_{EIT\,2} = \frac{2}{3}$, air temperature measured at starting point of EIT rising above zero $T_{EIT\_start}$.

The EIT is triggered as soon as the speed of the vehicle falls below a predefined threshold $\tau_v$, which is empirically determined to be 5 km/h. As long as the speed of the vehicle remains below this threshold, the EIT continues to increase in a strictly monotonic manner.
For a vehicle speed above the threshold, a predefined decreasing rate ($r_{EIT\,1}$ or $r_{EIT\,2}$) is applied until the currently measured temperature $T^i_{meas}$ is smaller than the temperature at the beginning of triggering the EIT $T_{EIT\_start}$ or until the EIT has decreased to zero. Subsequently, an influence of the previous stop by the waste heat of the engine can be neglected.
The value of the synthetic EIT signal builds the starting point for the physical model PHY used for correcting the air temperature signal. The whole model

is built as a decision tree with seven levels. The detailed illustration of the decision tree is shown in Figure 11.

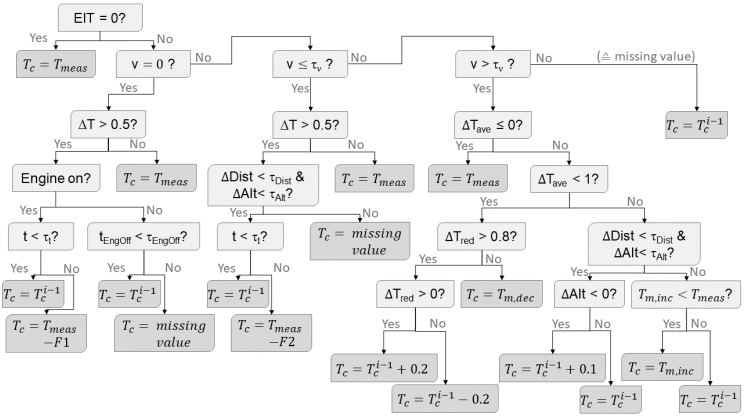


Figure 11: Flowchart of the physical model for correcting the vehicle-based air temperature signal. The corrected temperature value for the considered time step $T_{corr,PHY}$ is abbreviated by $T_c$. $T_c^{i-1}$ represents the corrected temperature value for the previous time step, $T_{meas}$ the measured temperature. A detailed explanation for all quantities can be found in Table 23 in Appendix A.2. If no time-related reference is given as superscript, the value refers to the current time step i.

As long as the EIT equals zero, an influence of waste heat of the engine can be neglected and the measured temperature signal is used as the corrected air temperature value ($T_{corr,PHY} = T_{meas}$). If the EIT is greater than zero, a possible influence of the engine waste heat has to be verified and if necessary corrected. Since the vehicle speed represents the clearest indication of a possible influence, it is used as the next decision criterion. Basically, the PHY model distinguishes between four different scenarios regarding the vehicle speed:

1. The vehicle is at standstill.
2. The vehicle travels slowly, at maximum speed of 5 km/h ($v \leq \tau_v = 5$ km/h).
3. The vehicle travels at speeds above the speed threshold of 5 km/h ($v > \tau_v = 5$ km/h).

4. No statement can be made about the speed of the vehicle, since no speed values are available.

The individual decision-making progresses for each enumerated scenario can be traced in the decision tree (cf. Figure 11). Signals used for decision-making are the engine state (on/off), the time passed as well as the distance and altitude covered since the EIT is triggered, and multiple temperature comparisons.

For the first scenario with the vehicle being at a standstill, the change in temperature since the EIT was first triggered (ΔT) is considered. If the currently measured value did not increase by more than 1 measurement resolution step of 0.5 K, an influence by the engine heat can be excluded and the measured temperature signal is adopted as a representative ambient air temperature value ($T_{corr,PHY} = T_{meas}$). However, if the measured value increases by more than 1 measurement resolution step, the state of the engine is checked (on/off).
If the engine is currently running, the time (t) passed since the last stop is taken into account. If it is below a predefined threshold ($\tau_t$) dependent on the engine temperature, the corrected temperature signal from the previous time step is retained ($T_{corr,PHY} = T^{i-1}_{corr,PHY}$). Otherwise, a correction factor F1 is determined, depending on the change in temperature since the stop and the value of EIT. Correction of the measured value $T_{meas}$ by F1 is only applied for durations of stop longer than the threshold $\tau_t$ since keeping the temperature values constant for the first time after the vehicle stops leads to more robust results.
If the engine is currently turned off and has been in this state for less than a pre-defined threshold of $\tau_{EngOff}$, the temperature signal is kept constant $T_{corr,PHY} = T^{i-1}_{corr,PHY}$. This threshold is set to 5 minutes since temperature changes due to local effects can be neglected for short time periods at one specific location (given since vehicle speed is zero). If the engine has been turned off longer than $\tau_{EngOff}$ local effects must not be neglected and therefore the output of the model becomes a missing value to avoid misleading information ($T_{corr,PHY}$ = missing value).

For the second scenario with low speeds, instead of a complete standstill, the model proceeds similarly. The only significant difference is the consideration of the altitude and distance covered since the EIT was triggered ($\Delta Alt$ and $\Delta Dist$). If the vehicle is stationary as in the previous scenario, it cannot change its position. However, in this scenario it is moving and thereby a

certain change in temperature can also be caused by a change in altitude or distance. Therefore, these changes are considered for low speeds here.

The third enumerated scenario refers to a situation where the vehicle is moving with a speed above the threshold, but it might still be influenced by the previous stop, since the EIT is greater than zero. If $\Delta T_{ave}$ ($= \overline{T_{corr,PHY}} - \overline{T_{meas}}$ with averaging referring to 20 s) is smaller or equal to zero, a measured temperature value lower or equal to the corrected temperature value is implied. Since such a reducing effect on the measured temperature value induced by vehicle motion is not known, the currently measured temperature is used as the corrected temperature ($T_{corr,PHY} = T_{meas}$). If $\Delta T_{ave}$ is positive, but still small ($\leq 1$ K) the measured temperature is only slightly higher than the previously corrected temperature. After further queries of temperature differences a variation of the previously corrected $T_{corr,PHY}^{i-1}$ or the measured $T_{meas}$ value is used as a representative ambient air temperature.
For $\Delta T_{ave}$ greater than 1 K, the measured temperature signal has to be influenced significantly by the waste heat of the engine as there is no other probable explanation. Therefore, the currently measured temperature value is not considered for determining the corrected temperature value. Instead, the previously corrected temperature value is retained ($T_{corr,PHY} = T_{corr,PHY}[i-1]$). However, for a decrease in altitude, this local effect can also cause the temperature value to grow, which is considered and accepted.

The fourth scenario applies, if no values for the vehicle speed are available. If only the vehicle speed value is not reported for any reason, but the temperature value is present, a technical error of short duration is assumed and therefore the previous corrected value is retained ($T_{corr,PHY} = T_{corr,PHY}[i-1]$).

Based on this decision tree (cf. Figure 11) the influence induced by waste heat of the engine block is corrected. The correction takes effect as soon as the vehicle stops, drives slowly or is still influenced by the previous stop and a temperature increase above one measurement resolution step is present. The local effect of temperature change due to altitude is also considered.
Compensating for effects induced by global radiation is not put into practice here, although an effect is expected, especially during summer season. However, since the data basis is not sufficient to quantify this effect and additionally the exact mathematical relationship is unknown, this effect can currently not be integrated into the physical model.

### Correction with neural network

The second approach tested is data driven, using a NN to correct the measured temperature values.

*Network architecture and training*

Since the problem at hand does not aim to perform a division into categories, but to output a continuous decimal number as the result, it is a regression problem. A fully-connected net with the softplus function as the activation function to ensure uniform activation and to prevent stepwise activation is applied. A stepwise activation decides between 0 and 1 and is therefore unsuitable for regression problems. An exception is the output layer of the network under study. Here, a linear activation function is applied, to avoid steps in the output data and thereby ensure a continuous mapping of the output data.
The Adam Optimizer (Adaptive moment estimation) [KB15] is used as the optimization algorithm using Keras in Python. Since no or only little tuning of the hyperparameters is necessary when using Adam in order to achieve robust results [KB15], it is well suited for this application.
The starting weights of the net are initialized randomly gaussian distributed.
Crucial for the network is the selection of the loss function, i.e., the quantity to be minimized during the training of the network in order to reduce the error of the network. For the present use case, the mean squared error is chosen as the loss function [SW11]:

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^{n} (y^i - \lambda(x)^i)^2 \tag{5}$$

with n being the number of all available instances in the data set, $y^i$ the value that is ammused to be true (measured by the MMU), and $\lambda(x)^i$ the predicted value. An advantage of squaring the deviation is that large deviations result in larger errors than small deviations. This penalizes the model more for large errors.

The training of the net is implemented to stop on convergence. Convergence is assumed when the MSE does not improve within 20 epochs on the validation set. A maximum of 100 epochs are granted to the net to find the best option. Training of the net is therefore terminated either by reaching the 100 epochs or by no further improvement of the net after 20 epochs. A higher number of

epochs was never necessary for this net, since the history of the training of the nets shows that the net is rarely retrained 100 times, but instead the convergence criterion is usually triggered beforehand, which leads to the termination of the training. This is partly due to the relatively small amount of data available for training the net.

The investigation of the features to be used takes place with a network structure of 2 hidden layers and 8 units, as well as the default value of $10^{-4}$ for the learning rate. After defining the ground truth and feature, the network structure is examined for its optimum for the given problem. Therefore, nets with 6, 8 and 10 units and 1-4 layers are trained. When examining the learning rate, values over 4 orders of magnitude [$10^{-2}$,$10^{-5}$] were tested.

In order to check whether the results of the network are reproducible, and the network thus has a high ability to generalize the underlying data set, the repeatability of the network is tested. For this purpose, 10 nets are trained with exactly the same basic conditions, i.e., ground truth, features, number of hidden layers and units, learning rate, and underlying data set are identical in all 10 nets. For this test the final network as described on the next pages is used.

*Input data and features*

First, all time steps at which one of the quantities has a missing value are filtered from the data, reducing the number of available data points from 210209 to 207345.
Subsequently, the data is divided into three fractions. The largest fraction contains $\frac{2}{3}$ of the data and is used for training the network. The remainder is divided equally between a validation and a test data set, each containing $\frac{1}{6}$ of the data. Since the available data comes from three different measurement campaigns at different weather conditions and times of the year, it is important that the network receives data from all three measurement campaigns for training. The first $\frac{2}{3}$ of each measurement campaign is used for training, the following $\frac{1}{6}$ for validation and the last $\frac{1}{6}$ for testing. This ensures that the three data sets train, valid, test each contain data from all three measurement campaigns.
After splitting the data, the input data of the network is scaled using the MinMaxScalar [Sci20] scaling them to the range of [0,1]. This is necessary because the input parameters have numerical values in different dimensions.

The speed of the vehicle, for example, may read up to a value of 160 km/h, whereas the measured temperature is often close to the freezing point (0 °C). Without scaling, the speed would be assigned a higher weighting based only on its numerical values, limiting the learning process of the network. By scaling the input variables, all values are in the same order of magnitude and the network chooses the weighting depending on the influence of the respective feature.

The determination of the features and the ground truth is decisive for the quality of the NN. For the present problem, various feature combinations are tested (cf. Figure 25 in Appendix A.3). A feature selection limited to the quantities measured directly by the vehicle and corrected by stage one and two only is not appropriate. The network would lack temporal correlations and other quantities to be able to train a reliable model. Therefore, additional features are provided to the network, which are not directly measured by the vehicle-based sensors but represent processed signals.
The network is provided with the already known EIT, which has been implemented and explained in detail for the physical network (cf. Chapter 4.2.3).
Further the net is provided with the temperature change since the EIT was triggered ($\Delta T$) and the duration of the triggered EIT ($D_{EIT}$). $\Delta T$ provides the network with a measure of the impact of changing vehicle operating conditions, in this case stopping or slow speeds. For each point in time when the EIT is greater than zero and therefore a falsification of the measurement data due to the vehicle dynamics is likely, the difference between the currently measured temperature and the temperature measured at the beginning of triggering the EIT, is calculated and stored as $\Delta T$. The larger $\Delta T$ is, the more the current measurement signal is distorted and the stronger the correction of the measurement signal must be. The second synthetic signal is the $D_{EIT}$, which refers to the duration the EIT has been active. This signal indirectly gives the network the possibility to cover temporal information.
In total, nets varying from 3-6 features, based on 7 different features are trained and tested (cf. Figure 24 in Appendix A.3).

*Ground truth*

Besides a suitable feature selection, the ground truth also decides about the quality of the network and the corrected values.
When selecting the ground truth, the measured reference value of the MMU is used first. However, it becomes apparent that using real measured temperatures

as ground truth weakens the generalization ability of the network. For each possible temperature value, there is an insufficient number of training examples for the network to successfully train and generalize. Since the majority of the data was collected in winter, the network learns decisively for the cold temperature spectrum. If the network is subsequently confronted with data from the summer campaign at warmer temperatures, it can be expected that the net is hardly able to correct these. To successfully train a network for this problem, while using measured temperature values as ground truth, a larger data base covering all weather conditions equally is required. However, building a data set of this size is practically out of the scope of the project's resources.
Nevertheless, in order to train a network that is more robust against seasonal temperature variations, no explicitly measured temperature value is used as ground truth. Instead, the difference between the reference value and the vehicle-based value is used:

$$GT = T_{ref} - T_{veh} \tag{6}$$

This makes the network independent of the absolute numerical value of the temperature. Instead, situations in which both signals differ more strongly, i.e., situations in which the vehicle-based signal is more distorted, are emphasized. This procedure allows the network to be trained independently of the seasons and weather situations for which the data was collected. However, it should be noted that the output of the network is therefore an offset and not a final temperature value. Based on this offset, the corrected temperature can be calculated in a final step.

*Final network*

Since the NN implemented is compared to other approaches correcting the air temperature, the best NN structure is determined in this chapter. The detailed results of the investigations described for the neural network approach can be found in Appendix A.3.

The feature combination that results in the lowest MSE is the composite of the following 6 features, sorted by impact on MSE (cf. Figure 25 and Figure 26 in Appendix A.3):

1. $\Delta T$

2. v
3. $D_{EIT}$
4. $Eng_{I/0}$ - Binary status whether engine is running (1) or is turned off (0)
5. Measured air temperature after stage two correction
6. EIT

It may be surprising at first that the influence of the temperature, i.e., the parameter that is to be corrected, on the network is in fifth and thus penultimate position. This can be explained by the fact that the difference instead of the absolute measured temperature values serve as ground truth. The front positions of the two synthetic parameters $\Delta T$ and $D_{EIT}$ underline, how crucial the influence of these quantities and the associated temporal influence on the net is.
It should be noted that altitude, which is an indication of local effects, is not considered in the network since it does not improve the results for the underlying data set.

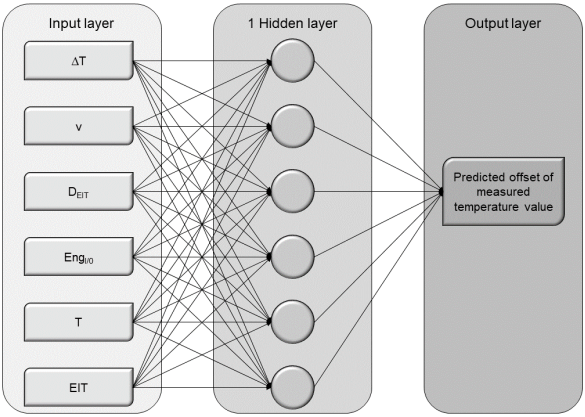


Figure 12: Overview of the implemented neural network (NN) used for correcting the vehicle-based air temperature measurement.

Testing different numbers of units and layers gives the knowledge, that a deeper (more hidden layers) and more complex (more units) network structure does not improve the results (cf. Figure 27 in Appendix A.3). Instead, reducing the size of the net can reduce the MSE and thus improve the results. The investigations show that the optimal structure for the network is a single hidden layer and 6 units. The final structure of the network including the used features is shown in Figure 12.
The MSE becomes the lowest for a learning rate of $10^{-4}$ (cf. Figure 28 in Appendix A.3). The learning curve of the network over the epochs is shown in Figure 34 in Appendix A.3.

During testing the reproducibility of the net, the MSE of the training data is identical for all 10 networks to two decimal places (cf. Figure 29 in Appendix A.3). The MSE of the validation data and the test data fluctuate by 1.94 % and 1.87 % around the MSE averaged over all 10 networks (standard deviation in percent). These small deviations of the 10 networks among each other show a high reproducibility and thus generalization ability of the network.

#### Correction with random forest

Lastly, a random forest (RF) is presented as another data driven approach. There are no known applications where a random forest has been trained on vehicle-based data or a similar problem (cf. Chapter 1.2). The idea to implement a random forest for the underlying problem is based on the structure of the physical model, which resembles a decision tree. Using a random forest (RF) being composed of multiple decision trees generated by training is therefore tested for its performance in comparison to the PHY model, which also uses a decision tree, and the NN.

*Forest architecture and training*

The random Forest is implemented in Python with the sklearn module. Since the underlying application is identical to the NN, it is also a regression problem. The Mean Squared Error (MSE) is employed as the criterion to be minimized.

In the training phase, the model trains a fixed number of trees each independent of each other. Each of these trees is split to a predetermined maximum depth. Fanning the tree too deeply can potentially compromise the tree's

ability to generalize, comparable to overfitting for the neural network. Different combinations of number of trees and maximum depth are tested, ranging from 8-100 trees with a respective depth of 4-14 levels.
The repeatability of the RF results is tested by training the model with identical boundary conditions 10 times independently.

*Input data and features*

Before the actual RF can be implemented and trained, the data must be prepared for use. The preparation of the potential input data is identical to the procedure of data preparation for the NN. For the data all times at which missing values are present are deleted (cf. Chapter 4.2.3).
Similar to the NN, the RF is a supervised learning model. Since the RF already uses the ground truth data when developing the model and dividing the data accordingly to build the branches of the individual trees, the RF does not need validation data to adjust the model afterwards. Therefore, the RF is based purely on training and test data. The available data from all three measurement campaigns is divided into two data subsets. The training set includes the first $\frac{5}{6}$ of each measurement campaign and is thus equal to the sum on the training and validation data set of the NN model. The test data set is identical for both machine learning applications and therefore also includes the last $\frac{1}{6}$ of each measurement campaign for the RF. The split of the data is intentionally very similar to the split for the NN to be able to better compare the results of the two data-based approaches. For the RF no scaling of the data is necessary.

When selecting features, first, a random forest is trained with the same input variables as the physical model. Subsequently, the model is provided with two more input variables, thus a total of 9 features (cf. model A in Table 30 in Appendix A.4). To identify the individual influence of each of the 9 features, different models are trained with only 8 of the 9 features at a time. By comparing the obtained MSE to the MSE of the model using all 9 features, the influence of the respective feature not taken into account can be determined.

*Ground truth*

The random forest uses the same ground truth as the neural network since the task is identical. The difference between the vehicle-based air temperature signal after completion of the stage two correction process and the measured value of the MMU is used as ground truth (cf. Equation 6).

*Final forest*

The best RF structure is determined in this chapter since the results of the RF are compared to the other implemented approaches correcting air temperature. A detailed overview of the different RFs tested and their structures can be found in Appendix A.4.

Figure 30 in Appendix A.4 list the different models tested and their respective MSE for the test data set. Omitting the feature "Duration of current stop" ($D_{stop}$) leads to an improvement of the result (cf. model F to model A) and is therefore excluded. The feature "Maximum EIT of current period" ($EIT_{max}$) has the lowest impact (cf. model I) and is therefore also neglected. The 7 remaining features that contribute most to the reduction of MSE are listed below according to their influence (cf. model K in Figure 30).

1. $\Delta T$
2. v
3. Measured air temperature after stage two
4. $\Delta Alt$ - Change in altitude since the vehicle triggered the EIT for the first time
5. $Eng_{I/0}$ - Binary status whether engine is running (I) or is turned off (0)
6. $D_{EngOff}$ - Duration the engine has been turned off
7. EIT

Selecting features is made with 8 trees with a maximum depth of 10 levels each. However, the MSE of the test data set varies relatively little due to the variations (cf. Table 31 in Appendix A.4). The RF with the smallest MSE of the test data set has 15 trees, each with a maximum depth of 10 levels. Due to the relatively small amount of data, no deeper or more complex forest has to be used.

The results for testing the RF for its reproducibility are displayed in Table 32 in Appendix A.4. The 10 MSE values of the test data have a percentage standard deviation of 1.71 %. The model therefore exhibits high repeatability and thus good ability to generalize the underlying data.

## 4.3 Relative humidity

Correcting the relative humidity signal using the principle of input filter, correction of sensor uncertainty in static situations and correction of sensor uncertainty induced by vehicle motion is not appropriate in this form due to the mounting position directly next to the engine block (cf. position of MuFu sensor in Figure 4) as well as the measuring method of the vehicle-based sensor. For relative humidity, four individual correction models (PHY-A, PHY-B, NN and RF) are developed, which are shown in Figure 13.

It is noticeable that the correction of the RH follows a more complex process. The modules "Input filter", "Compensating sensor uncertainty induced in static situations" and "Compensating effects induced by vehicle motion", which are known from the previous chapters, are varied in their order, and supplemented by further modules. The new modules "Calculate air temperature inside sensor" and " RH transfer to different location" are necessary due to the

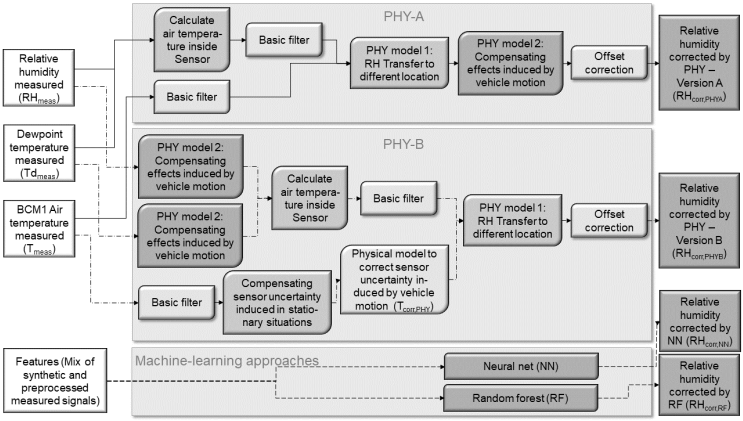


Figure 13: Flowchart for testing and correcting the relative humidity signal in three different ways. First, PHY- Version A, an approach based on physical relationships compensating for effects induced by vehicle motion last. Secondly, PHY- Version B, an approach that is built on the same physical relationships as PHY - Version A but compensates for effects induced by vehicle motion first. Lastly, two models based on machine learning are used to correct the vehicle-based raw data, namely a neuronal net (NN) and a random forest (RF).

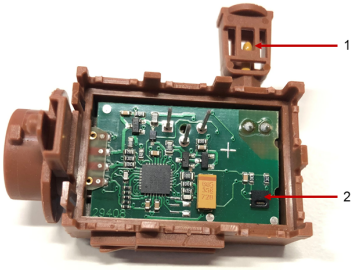


Figure 14: Multifunctional sensor with origin of different signals shown. 1: Location outside sensor to detect air temperature signal, 2: Location inside sensor case to detect relative humidity signal.

architecture of the RH sensor (cf. Figure 14). The relative humidity is measured inside the sensor case (cf. position 2 in Figure 14), which has a higher air temperature due to the waste heat generated by the sensor operation, compared to an air temperature measurement outside the sensor case (cf. position 1 in Figure 14). Since air temperature and relative humidity are directly coupled, a higher air temperature inside the sensor case simultaneously means a lower relative humidity. This explains the effect of relative humidities measured with values too low. At the same time, it illustrates the need for sensor individual correction models, since previously presented models do not take into account effects arising from the architecture of the respective sensor.
The critical components from the PHY-A and PHY-B models in Figure 13 are the calculation of the temperature inside the sensor case and subsequently the transfer of relative humidity from inside the sensor case (cf. position 2 in Figure 14) to any other position where the air temperature is known. The following section derives the physical formulas for both calculations. Subsequently, the components "input filter" and "compensating effects induced by vehicle motion" are explained. Lastly, the offset corrections as the final step for both physical models are clarified.

### 4.3.1 Air temperature inside a sensor case

To calculate the air temperature at any location, in the present case inside the sensor case, the relative humidity and the dew point temperature must be known for the approach derived here.
Both the relative humidity and the dew point temperature are determined based on the air inside the sensor, whereas the extracted air temperature signal of this sensor is measured outside the sensor. Therefore, the air temperature representative for the air inside the sensor has to be calculated. This can be done by using physical relationships exclusively.

Starting point is the water content $x$ of an air parcel, which is defined as the mass of water $m_W$ in relation to the mass of dry air $m_A$ contained by the air parcel and is therefore also called the mixing ratio [WH06]. As long as the air parcel is not saturated, the mass of the water can also be described as the mass of vapor $m_V$ since the water is in gaseous form.

$$x = \frac{m_W}{m_A} = \frac{m_V}{m_A} \tag{7}$$

The second principle needed for the derivation is the definition of the total pressure $p$. It consists of the partial pressures from the involved components. In the case of having a humid air parcel the total pressure is the sum of the partial pressure of the water vapor $p_V$ and the partial pressure of the dry air $p_A$, since according to Daltons law the total pressure is always built by the sum of the individual partial pressures [Dal21].

$$p = p_V + p_A \tag{8}$$

The ideal gas law describes the relationship between the following characteristics of an air parcel of a hypothetical ideal gas: pressure $p$, volume $V$, temperature $T$ and the amount of gas, presented by mass $m$ and molar mass $M$. $R_m$ is the universal gas constant (cf. Equation 4).
Building the ratio of the partial pressures $p_V$ and $p_A$ by using the converted ideal gas law yields the following equation:

$$\frac{p_V}{p_A} = \frac{\frac{m_V}{M_V} \cdot R_m \cdot \frac{T}{V}}{\frac{m_A}{M_A} \cdot R_m \cdot \frac{T}{V}} \tag{9}$$

Since temperature $T$, volume $V$, and the universal gas constant $R_m$ are identical for both components of the air parcel, they cancel each other out. Furthermore, the ratio of the masses of the vapor and the dry air $\frac{m_V}{m_A}$ can be replaced by $x$, compare equation 7. Solving for $x$ results in the first part of the following equation. This can be further simplified by using the known molar masses for water $M_V = 18.01528\ \frac{g}{mol}$ and for dry air $M_A = 28.96\ \frac{g}{mol}$:

$$x = \frac{M_V}{M_A} \cdot \frac{p_V}{p_A} = 0.622 \cdot \frac{p_V}{p_A} \tag{10}$$

Solving Equation 10 for the partial pressure of dry air $p_A$, substituting this into equation 8 and solving it for the partial pressure of the vapor $p_V$ results in:

$$p_V = \frac{x}{0.622 + x} \cdot p \tag{11}$$

Combining Equations 11 and 1 (definition of relative humidity in Chapter 2.1) by substituting the partial pressure of the vapor $p_V$ into the definition of the relative humidity $RH$, results in:

$$RH = \frac{x}{0.622 + x} \cdot \frac{p(T)}{p_{V,sat}(T)} \tag{12}$$

For the next step making use of the dew point temperature $T_d$ is essential. For the dew point temperature, the relative humidity $RH$ is always 1. Therefore, equation 12 equals 1, for the air temperature being equal to the dew point temperature.

$$RH = \frac{x}{0.622 + x} \cdot \frac{p(T_d)}{p_{V,sat}(T_d)} = 1 \tag{13}$$

Dividing equation 12 by equation 13 results in the following expression, which can be simplified by assuming an identical pressure ($p(T) = p(T_d)$). Since the difference in air pressure that is caused by a change in air temperature is negligible in relation to the absolute value, the air pressure terms cancel each other out:

$$RH = \frac{\frac{x}{0.622+x} \cdot \frac{p(T)}{p_{V,sat}(T)}}{\frac{x}{0.622+x} \cdot \frac{p(T_d)}{p_{V,sat}(T_d)}} = \frac{p_{V,sat}(T_d)}{p_{V,sat}(T)} \tag{14}$$

Solving this for the saturation vapor pressure $p_{V,sat}(T_d)$ results in:

$$p_{V,sat}(T_d) = RH \cdot p_{V,sat}(T) \tag{15}$$

The previously mentioned saturation vapor pressure $p_{V,sat}$ describes the state where a system is in its state of equilibrium since the exchange of molecules between different sates of aggregation is balanced. The saturation vapor pressure can be described by the Magnus formula [Son90] which represents an approximation and was derived by Heinrich G. Magnus in 1844 [Wet19a]. C1, C2, and C3 represent empirically determined constants.

$$p_{V,sat}(T) = C1 \cdot exp(\frac{C2 \cdot T}{C3 + T}) \tag{16}$$

Substituting the Magnus formula into equation 15 results in:

$$C1 \cdot exp(\frac{C2 \cdot T_d}{C3 + T_d}) = RH \cdot C1 \cdot exp(\frac{C2 \cdot T}{C3 + T}) \tag{17}$$

Solving for temperature results in a formula by which the air temperature $T$ can be calculated based on the dew point temperature $Td$ and the relative humidity $RH$, which are both determined by the multifunctional MuFu sensor based in the vehicle. Here, the relative humidity is used as a number between 0 and 1.

$$T = \frac{C3 \cdot (log(1/RH) + \frac{C2 \cdot T_d}{C3 + T_d})}{C2 - (log(1/RH) + \frac{C2 \cdot T_d}{C3 + T_d})} \tag{18}$$

### 4.3.2 Transferring relative humidity

In the following, the needed equation to transfer relative humidity from inside the sensor case to any location outside the sensor case is derived. The location to which the relative humidity is to be transferred is freely selectable as long as the air temperature at this location is known. Since the calculated relative humidity should be representative for the ambient air, it is calculated for the location of the BCM1 temperature sensor at the radiator grille (cf. Figure 4). By correcting the measured temperature signal of the BCM1 sensor beforehand, a representative temperature signal is assumed.

The physical basis for the transfer is built by the specific humidity $s$, which is defined as the ratio of vapor pressure $p_V$ to air pressure $p$, multiplied by the ratio of the molar masses of the involved components [LC79]. Here, the molar masses of interest are for water and dry air and their ratio results in 0.622, compare equation 10.

$$s = \frac{p_V}{p} \cdot 0.622 \tag{19}$$

The main advantage of the specific humidity in comparison to other indices of humidity is, that it remains constant for varying temperatures as long as the water content is not changed. This can be assumed since no significant amount of moisture is added or taken away from the air parcel while moving from one sensor location to another. Note, that the relative humidity value does not remain constant, this only applies to the specific humidity value. Therefore, the specific humidity values are equal for the BCM1 sensor, which is representative for the ambient air (AA) as well as for the multifunctional MuFu sensor, which is located under the hood (UH) of the vehicle (cf. sensor positions in Figure 4).

$$s_{UH} = s_{AA} \tag{20}$$

Substituting equation 19 in equation 20 and afterwards substituting the for vapor pressure $p_V$ solved equation 1, yields the following:

$$\frac{RH_{UH} \cdot p_{V,sat,UH}}{p_{UH}} = \frac{RH_{AA} \cdot p_{V,sat,AA}}{p_{AA}} \tag{21}$$

By assuming a constant air pressure for both locations of the sensors ($p_{UH}$ = $p_{AA}$), the air pressures cancel themselves out. Substituting the definition for the saturated vapor pressure (equation 16) into the previous equation and solving for the relative humidity of the ambient sensor, results in:

$$RH_{AA} = RH_{UH} \cdot exp(\frac{C2 \cdot T_{UH}}{C3 + T_{UH}} - \frac{C2 \cdot T_{AA}}{C3 + T_{AA}}) \tag{22}$$

By this, the relative humidity of the moving air mass can be calculated for the position of the BCM1 sensor (= $AA$), although it is only measured at the location of the multifunctional MuFu sensor (= $UH$).

### 4.3.3 Input filter

When correcting relative humidity, it is not the RH that is subject to an input filter, but instead the signals that are necessary for further calculation. Therefore, the only quantities that are subject to an input filter for relative humidity are two temperature values. One is the calculated temperature in the MuFu sensor, the other is the measured temperature at the BCM1 sensor. The procedure for the input filter for temperature values is already described in Chapter 4.2.1. Since neither physical limit values nor the extent of acceptable jumps change within a time step, the identical input filter can be applied as long as it concerns air temperature values.

### 4.3.4 Sensor uncertainty induced by vehicle motion

The influences induced by vehicle motion have a similar effect on the relative humidity as on the air temperature, only vice versa. In concrete terms, this means that the relative humidity is also distorted in the same situations as the air temperature, but to the opposite extent. This is due to the relationship between air temperature and relative humidity. An air parcel which is subjected to a temperature change without the addition or removal of water has a lower relative humidity at higher air temperatures. This is due to the fact that warm air can absorb more water particles until it is saturated. Therefore, the same absolute water content at higher temperatures leads to a lower relative humidity.

Since the triggers and justifications for the influence on the relative humidity are thus the same as those on the air temperature, only the individual situations of importance with their direct effect on the relative humidity are listed below. In situations with the vehicle traveling at low speeds or being stationary, an increased temperature is measurable (cf. Chapter 4.2.3 for detailed explanation). Simultaneously, the relative humidity drops. Since the amount of water contained in the air trapped under the hood does not increase, the relative humidity drops.
Following a previous stop, the air temperature measurement as well as the relative humidity remain distorted due to the accumulated air under the hood. With increasing time distance to the last stop, the effect vanishes, and the relative humidity returns to higher values. For a situation, when the vehicle decelerates under the set speed-related threshold again before the effect of the

previous stop has disappeared, not only the measured air temperature increases again, but also the relative humidity decreases.

The flowchart in Figure 13 shows that the compensation of the effects by vehicle motion can take place both after the transfer of RH (PHY-A) and before (PHY-B).
PHY-A corrects the effects induced by vehicle motion subsequently to the input filter and the transfer of the relative humidity. Therefore, only one quantity, the newly calculated RH has to be compensated for effects induced by vehicle motion. If compensating these effects is based solely on one quantity, the effort of the correction is less than for PHY-B but the result is more prone to errors, since individual errors in the correction have a greater impact on the result.
Therefore, in addition to PHY-A, PHY-B is implemented, which follows a different order of correction and thereby executes the correction of effects induced by vehicle motion on several input variables.
The data used for the transfer calculation of the relative humidity, is also falsified by the vehicle motion. Thus, computing the temperature inside the sensor is performed knowing that it is based on wrong input values. Approach PHY-B therefore corrects the sensor uncertainty induced by vehicle motion first and subsequently calculates the transferred relative humidity.
Both approaches are described in more detail in the following subchapters.

#### Correction with physical model A

Comparable to the physical model of the temperature correction model (cf. Chapter 4.2.3), this model is also based on the principle of an engine impact timer ($EIT_{RH}$). This $EIT_{RH}$ differs from the EIT for temperature in a way that the threshold for speed $\tau_{v,RH}$ was increased so that the $EIT_{RH}$ is triggered earlier. This is necessary since the relative humidity reacts more sensitively towards stops or slow speeds due to the mounting position of the two sensors. The temperature sensor (cf. BCM1 sensor in Figure 4) is located at a greater distance from the motor block than the RH sensor (cf. MuFu sensor in Figure 4). Due to the small distance to the heat source, the MuFu sensor reacts faster to a reduced air exchange under the hood. Apart from this, the procedure for determining $EIT_{RH}$ is identical to EIT (cf. Figure 10).

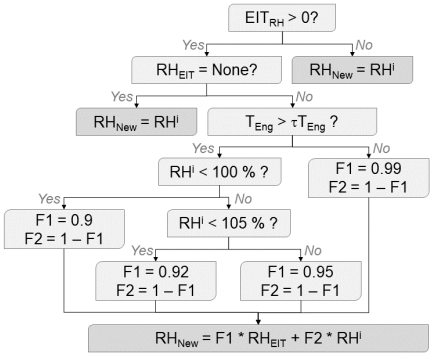


Figure 15: Flowchart for compensating the effects induced by vehicle motion for the calculated relative humidity in model PHY-A. $EIT_{RH}$ being the Engine Impact Timer for the relative humidity, $RH_{EIT}$ the relative humidity in the second the vehicle falls below the speed related threshold of 20 km/h und thereby triggers the EIT. The threshold $\tau_{Enginetemp}$ is set to 100 °C. RH denotes the previously valid correction level of the relative humidity, in this model therefore the transferred relative humidity. The designation i as a superscript refers to the current time step, i−1 to the previous one.

Figure 15 illustrates the flow of the correction of effects induced by vehicle motion.
First $EIT_{RH}$ is checked. If $EIT_{RH}$ is not triggered, meaning the vehicle is traveling at speeds above 20 km/h and is no longer affected by the previous stop, the current relative humidity value ($RH^i$) is stored as the new relative humidity $RH_{New}$. If the $EIT_{RH}$ is greater than zero, the relative humidity is therefore likely to be influenced by effects induced by vehicle motion. Therefore, the relative humidity at the time when the $EIT_{RH}$ was first triggered ($RH_{EIT}$) is checked.

If $RH_{EIT}$ is not set, i.e., there was no value of relative humidity in the second the vehicle speed fell below the velocity threshold, the current value of relative humidity is stored as the new relative humidity $RH_{New}$. However, if the relative humidity at the time of the first triggering of $EIT_{RH}$ is known, this can be considered for the further correction.
In the next step, the engine temperature is considered. The following section is used to define the two factors F1 and F2, which are used for weighting the different available values of the relative humidity. The values of both factor F1 and F2 are empirically determined based on the underlying data set. For a cold

engine the factors F1 and F2 are assigned 0.99 and 0.01 respectively. At warm engine temperatures, the current value of the relative humidity is additionally considered when F1 and F2 are defined. The more distinct the distortion of the current relative humidity value $RH^i$ is (RH values above 100 %), the greater will be F1. F2 behaves the opposite way. Subsequently, the new relative humidity $RH_{New}$ is calculated by weighting the $RH_{EIT}$ value with factor F1 and the current RH value with factor F2. It becomes apparent that the value at the beginning of $EIT_{RH}$ being triggered, is given greater importance, since it ($RH_{EIT}$) is weighted with F1. This is due to the fact that the value is assumed to be valid at this time. Therefore, the values for F1 are always at least 0.9 which corresponds to a weighting of 90 %. The current RH value ($RH^i$) is considered with the factor F2 and is of the size of the remaining weight, at maximum 10 %.

#### Correction with physical model B

Model PHY-B follows the approach of correcting the signals used to calculate the transfer of relative humidity beforehand. Hereby, it can be assumed that the transfer of the relative humidity takes place with input parameters which are no longer falsified by effects induced by vehicle motion.
The two signals for which compensation of these effects is necessary are the relative humidity inside the sensor case and the measured dew point temperature $T_d$ (cf. Figure 16).

If the vehicle is traveling at speeds lower than $\tau_{v,RH}$ or is even stationary, the values for both relative humidity and dew point temperature are kept constant. However, if the current relative humidity value is higher than the previously highest RH value, it will be adopted instead. For the dew point temperature, the opposite interpretation applies. In case of current values smaller than the previous minimum, the current value is accepted. This additional rule ensures that a possibly falsified value is not held on to for too long, but that the most reliable value, i.e., the highest relative humidity or lowest dew point temperature value respectively, is used.
As soon as the vehicle travels faster than the set threshold, it reviews how much time has passed since the last stop. If the last stop was more than a pre-set threshold $\tau_t$ ago, the current measured value is used. Otherwise, a timer is started at the time when the vehicle first drives faster than the limit

value. The timer then decreases monotonously. As long as the timer is greater than zero, the highest value for relative humidity or lowest value for dew point temperature is stored as the current value. Once, the timer decreases down to zero, the current value is accepted as the new value since an influence of the previous stop that has been more than $\tau_t$ seconds ago, can be neglected due to the ventilation under the hood induced by driving.

### Offset for both PHY models

For both models PHY-A and PHY-B an offset correction takes place in addition to all previously described steps. Regardless of the transfer (cf. Chapter 4.3.2), the multifunctional MuFu sensor, which builds the basis for the relative humidity correction, has a measurement uncertainty and deviation, both in dew point temperature and relative humidity measurement [Sen14]. The determination of the offset of the variables included in the calculations is not possible without further ado due to the lack of reference values. Therefore, the offset is subtracted from the converted relative humidity afterwards. Due to the different approaches, different empirically determined offsets result for the two models. The offset for model PHY-A is 10 %rH, for PHY-B it is 14 %rH. Both offsets

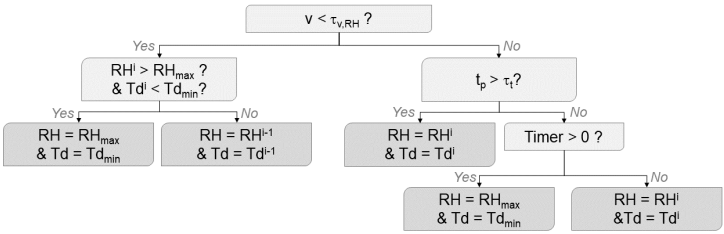


Figure 16: Flowchart for compensating the effects induced by vehicle motion for the calculated relative humidity in model PHY-B. RH being relative humidity and $T_d$ dew point temperature. $t_p$ is the time that has passed since the last stop, $\tau_t$ refers to a preset threshold of 3 hours, which is determined as a rounded-up mean value from the timer suggestions from three patents [RHDJ95, KB09, COW10]. The timer is started as soon as the vehicle exceeds the speed-related limit $\tau_v$ for the first time and decreases monotonously afterwards. The init value of the timer is determined empirically and set to 500, which corresponds to approximately 8 min. The designation i as a superscript refers to the current time step, i−1 to the previous one.

are relatively large, which is due to the significant influence of slight temperature and humidity changes, or measurement uncertainties, on the conversion during the transfer of relative humidity.
Simultaneously, it is checked whether the new value corrected by the respective offset ($RH_{New}$ - Offset) is above the maximum permissible relative humidity of 100 %. If yes, the value is capped at 100. Otherwise, the respective previously determined value $RH_{New}$ is reduced by the respective valid offset and set as $RH_{corr,PHY}$.

#### Correction with neural network

For the relative humidity also a NN is applied. First attempts using this approach are depicted in [HS20]. Most of the architecture related parameters and hyperparameters are identical to those from the neural net of air temperature (cf. Chapter 4.2.3). In the following, these properties are therefore listed, but not further specified. Deviations to the previously described network for air temperature are described in detail.

*Network architecture and training*

Identical to the temperature network, this is also a regression problem, which is implemented using a fully-connected network.
Also, the activation function, namely the softplus function, is the same. Furthermore, this network uses the Adam Optimizer of Keras in Python, which is used in the temperature net likewise. The random initialization of the starting weights of the relative humidity net is adopted from the temperature net. This is also true for the loss function. Again, the MSE is chosen (cf. Equation 5).
Training parameters, such as the convergence criterion, are identical to the temperature net.
Different architectural combinations are tested, ranging from 1-4 layers with 6-10 units. Additionally, the learning rate is varied $[10^{-5}, 2^{-2}]$.

Lastly, the repeatability of the network is tested. First, the same net is trained 6 times on the identical train data set. Afterwards, the allocation of the data is changed, resulting in a different train data set for each of the 6 additional trainings.

*Input data and features*

Also, for this network the data must first be preprocessed. However, no general correction of the input variables takes place for this network. Instead, the preprocessing refers solely to the filtering of missing values in the individual quantities. In addition, values from the February campaign are neglected for the data set to be evaluated due to the lack of robust reference data. Through this preprocessing step, the number of available data points is reduced from 149705 to 146843, corresponding to a decrease of 1.9 %. It is noticeable, that the amount of data points is almost $\frac{1}{3}$ smaller compared to the application of a neural net to the air temperature. The reason for this is the strongly limited availability for reference data regarding the relative humidity due to measuring difficulties during the February campaign.

The division of the data set into three sub data sets to train, validate and test the model, is carried out in a different way compared to the temperature net. Originally, the same proceeding of always using the first $\frac{2}{3}$ of each measurement campaign for training, the following $\frac{1}{6}$ for validation and the last $\frac{1}{6}$ for testing, was implemented. However, a pattern in the data was noticed that can neither be explained by the measurement environment nor the architecture of the network. Nevertheless, the network is significantly affected. Regardless of the rest of the network structure, it returns the lowest values of the loss function for the test data set, rather than for the train or validation data set. To find the cause for this behavior, the complete network structure was tested and varied to exclude influences of the initialization and individual parameters or hyperparameters. However, the explanation cannot be found in the network, but in the data. The behavior of this network can only be justified by a data set that is easier to learn for the model towards the end of each measurement campaign. The observed pattern of the data cannot be explained by the measurement conditions or other known environmental conditions.
To avoid this effect, the division into train, validate and test sections is done randomly. Still $\frac{2}{3}$ of the data is used for training, $\frac{1}{6}$ of the data for validation and $\frac{1}{6}$ of the data for testing, but the selection of the data points used in each of the three data subsets is not based on a temporal criterion. Each data point is used exactly once, since the test data set would otherwise contain data points that are already known to the network from the train and validation phase.
For scaling of the input variables this network also uses the MinMaxScalar to obtain values from [0,1] (cf. Chapter 4.2.3).

Limiting the features to signals directly measured by the vehicle, leads to a loss function 9 times as high as the final setting. Therefore, different combinations, also considering synthetic signals such as $EIT_{RH}$, different relative humidity signals and the change in relative humidity since the last stop ($\Delta RH$) are tested. Overall, 9 features are tested for their influence on the net. The number of features considered simultaneously ranges from 4 to 9.

*Ground truth*

Due to the experience regarding the ground truth of the temperature network, the relative humidity net is applied and tested with the difference between the reference value of the MMU and a vehicle-based value. Three different stages of the vehicle-based relative humidity signal are considered for determining the ground truth. Therefore, three different possibilities for the ground truth arise. The difference between reference signal measured by MMU and:

1. vehicle-based raw signal of relative humidity ($GT = RH_{raw} - RH_{ref}$),
2. vehicle-based relative humidity signal transferred from inside the sensor case to BCM1 sensor without compensating for effects induced by vehicle dynamics ($GT = RH_{tran} - RH_{ref}$),
3. vehicle-based relative humidity signal corrected for effects induced by vehicle motion without transferring it to a more representative location ($GT = RH_{vm} - RH_{ref}$).

*Final network*

Since the NN is implemented to enable a comparison between different approaches to correct the RH, the final setting of the network is described here. Appendix A.5 lists the detailed test results of the network.

To determine the final feature and ground truth selection, different constellations are tested and evaluated. The same selection of features yields significantly different results with a different ground truth. Three different possibilities are tested as ground truth (cf. Figure 33 in Appendix A.5), each with the same selection of features. Best results are yielded using the difference between the MMU and the vehicle-based raw signal of the relative humidity (cf. 1. in list above).
Subsequently, different constellations of features are tested (cf. Figure 34 in

Appendix A.5). Since neither the vehicle speed v nor the duration of the stop $D_{stop}$ have a significant influence on the MSE of train, validate and test data (cf. model E and F in Figure 35 in Appendix A.5), they are not considered in the final net. Best results are achieved with the following 7 features (cf. model M):

1. $RH_{tran}$ - raw signal of relative humidity transferred to BCM1 sensor without correcting for vehicle motion
2. $\Delta RH$ - Change in relative humidity since the EIT is triggered
3. $RH_{raw}$ - raw signal of relative humidity
4. $T_{sc}$ - Calculated temperature inside the sensor case corrected for effects induced by vehicle motion
5. $RH_{vm}$ - raw signal of relative humidity corrected for effects induced by vehicle motion (no transfer of signal)
6. $T_{Eng}$ - Engine temperature
7. $EIT_{RH}$

The order of the listed features corresponds to their influence on the network. Thus, the greatest influence is exerted by $RH_{trans}$, followed by the synthetic parameter $\Delta RH$, the least by $EIT_{RH}$.

The decision which features to use, is made based on a network with 2 hidden layers, 8 units and a learning rate of $10^{-4}$. Varying the number of hidden layers and units can improve the MSE. Best results are achieved for 3 hidden layers and 8 units (cf. Figure 36 in Appendix A.5), although the improvement of the MSE was down to 0.3 %.
Changing the learning rate, had a significantly higher effect on the MSE. By increasing it to $15^{-3}$, the MSE is almost halved (cf. Figure 37 in Appendix A.5). Figure 35 in Appendix A.5 shows the development of training and validation MSE for the final net.

Figure 17 displays the final structure of the network including the used features. For the six networks, each trained on the same data set, the standard deviation in percent for the train, validate, and test data set is 13.11 %, 13.34 %, and 13.42 %, respectively (cf. Figure 38a) in Appendix A.5). The standard

deviation in percent of the 6 networks with a new data assignment for each of the 6 nets, is 15.48 %, 15.29 %, and 16.10 % for the train, validate, and test data sets, respectively (cf. Figure 38b) in Appendix A.5). The deviations of the individual nets therefore deviate more strongly from each other than in the case of the air temperature. The reproducibility of the tested network for relative humidity can therefore not be assumed without performing further tests.

### Correction with random forest

A RF is implemented as a second data driven model. The structure of the RF for relative humidity is strongly based on the structure of the RF for air temperature since the underlying task is comparable. Therefore, the detailed explanations in this chapter are reduced to structural decisions that differ from those for the RF of temperature.

*Forest architecture and training*

For the initialization of the RF, as in Chapter 4.2.3, the sklearn module in Python is used to execute the underlying regression task. The MSE is used as

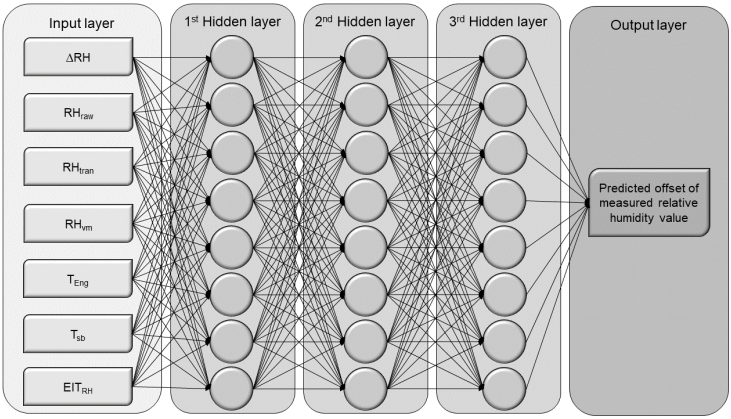


Figure 17: Overview of the implemented neural network (NN) used for correcting the vehicle-based relative humidity measurement.

the parameter to be minimized (cf. Equation 5). Different combinations of number of trees and maximum depth are tested, varying from 8-100 trees with 4-20 levels for each tree.

Lastly, the RF is tested for repeatability of results. First, 10 RFs are trained on the identical data set, then 10 more on a newly compiled data set for each RF.

*Input data and features*

Comparable to the NN, the pre-processing is limited to the filtering of missing values and excluding data from the February campaign since the reference data is not reliable. Thus, the RF has an identical data base to the NN totaling 146843 s.

The division into train and test data occurs randomly also for the RF, which is due to the structure of the data noticed in the context of the implementation of the NN (cf. Chapter 4.3.4). $\frac{5}{6}$ of the randomly chosen data is used for training the RF, the remaining $\frac{1}{6}$ for testing. To ensure comparability between the two data driven approaches, the same random partitioning is used as for the NN (cf. Chapter 4.3.4).

Different constellations are tested for the final feature selection. The starting constellation is the feature selection for the NN comprising the 7 features listed above. Afterwards, a different feature is neglected for every training phase to elaborate the influence of this specific feature on the data.

*Ground truth*

The RF uses the same ground truth as the NN, namely the difference between the reference data of the MMU and the raw data of the vehicle: $GT = RH_{raw} - RH_{ref}$.

*Final forest*

In order to be able to access the final RF for the evaluation of the vehicle-based data in the following chapter, the final structure is already presented in this section. The detailed test results for the RF for relative humidity are given in Appendix A.6.

The MSE of the test data is reduced by excluding the engine temperature $T_{Eng}$ as a feature. Neglecting any other of the tested features increased the MSE

and thereby decreases the result of the RF (cf. Figure 39 in Appendix A.6). Therefore, the feature selection for the RF is limited to the following six features, which are listed here in descending order to their influence on the RF (cf. model C):

1. $RH_{tran}$ - raw signal of relative humidity transferred to BCM1 sensor
2. $RH_{raw}$ - raw signal of relative humidity
3. $T_{sc}$ - Calculated Temperature inside the sensor case corrected for effects induced by vehicle motion
4. $EIT_{RH}$
5. $RH_{vm}$ - raw signal of relative humidity corrected for effects induced by vehicle motion
6. $\Delta RH$ - Change in relative humidity since the EIT is triggered

The feature with the most significant influence is $RH_{tran}$, since a neglection of it increases the MSE of the test data by a factor 3.9.

Testing the architecture of the RF reveals a decrease in the MSE of the test data the more trees with a higher maximum depth are used (cf. Figure 40 in Appendix A.6). The MSE is reduced by at least 30 % every time another two levels are added to each of the 15 trees. Afterwards the decrease amounts to less than 25 %, wherefore this is set as an artificial boundary and 14 is chosen as a maximum depth for the forest. The number of trees is identical to the RF of temperature (15), but the number of maximum depths allowed for RH (14) is higher than for temperature (10). This indicates the higher complexity of correcting RH compared to temperature.

All 10 RFs trained on the identical data base have a percentage standard deviation from the obtained MSE to the averaged MSE of 2.41 % and of 4.91 % for a newly compiled data set for each RF (cf. Figure 41a) and b) and in Appendix A.6). The standard deviation therefore varies more when the RF is trained on different data sets, but the mean MSE is almost identical for both scenarios. The repeatability of the RF results is thus high, ensuring a good ability to generalize the underlying data through the developed RF. It is noticeable that the mean MSE test of the RF is about 10 percentage points lower than for the NN.

## 4.4 Global radiation

The differentiation into the individual components of direct and diffuse radiation (cf. Chapter 2.1) is not possible based on the sensors installed as standard on board of a production vehicle. Therefore, the global radiation G is considered as a whole taking into account the two vehicle-based signals. Both signals are output by the light and rain sensor in the windshield. The two values differ in their orientation and therefore their range for which they are representative. One signal covers the front left, the other the front right.

Figure 18 shows the flowchart for the global radiation correction. First, both vehicle-based raw signals $G\_le_{meas}$ and $G\_ri_{meas}$ are subjected to the input filter (cf. Chapter 4.4.1).
In addition to the two raw signals of global radiation, the position (latitude and longitude), as well as date, time and heading of the vehicle are required. This information is used to establish the relationship between the solar position and the vehicle-based measurements (cf. Chapter 4.4.2). First the current position of the sun is calculated. With the addition of the orientation of the vehicle (heading), the horizontal angle of incidence of the sun in relation to the vehi-

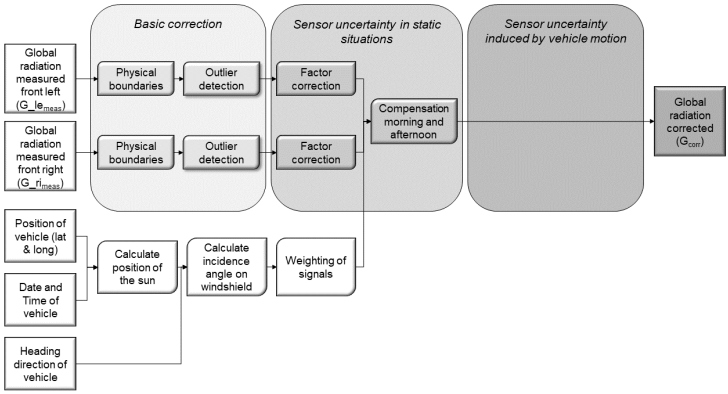


Figure 18: Flowchart for correcting the vehicle-based global radiation signals and merging them into one final signal. Note, besides the vehicle-based raw signals, also the position, date and time and heading of the vehicle need to be known.

cle is determined. Based on this angle of incidence, a weighting of the two vehicle-based global radiation signals is developed.
Chapter 4.4.2 describes the calculation of a merged global radiation signal based on the input filtered global radiation signals and the weighting of the signals developed in Chapter 4.4.2. First, a factor correction of the vehicle-based signals is performed. Second, based on the previously developed weighting, a merged global radiation signal is calculated from both individual signals. Third, this is subjected to a final correction to compensate for the varying effects between morning and afternoon.

The methodology for global radiation is developed on the basis of the stationary measurement campaign of TC2 next to the synop station "Weihenstephan". Since the vehicle was stationary during the entire campaign, the influences on the measurement can be limited to the sun and the direct environment.

### 4.4.1 Input filter

The input filter of both measured global radiations signals is content-related identical to the air pressure as well as the air temperature approach. Only the individual thresholds are chosen differently.

1. *Ensuring all measured values are within measurement range.*
   The only physical boundary for global radiation is a lower limit, which is set to zero. The value of the global radiation can only decrease to zero during nighttime. Any negative values indicate false values.

2. *Ensuring no outliers are present in the data.*
   Since the global radiation is strongly dependent on cloud cover and shading due to trees, it can change rapidly. Therefore, changes in the data of up to 200 Watt each timestep are accepted without being changed.

### 4.4.2 Relationship between position of sun and vehicle

#### Calculation of position of the sun

To establish a relationship between solar position and the vehicle-based measurements, the solar position is first calculated. Here, a distinction must be made between two different angles of sun incidence: The vertical and the horizontal angle of incidence.
The calculation of the vertical sun angle is given in Appendix A.7.
The azimuth angle or the horizontal angle of incidence of the sun $AoS$ is calculated as specified in equation 23 [Gie20]:

$$AoS = \frac{acos(y) \cdot 180}{\pi} \tag{23}$$

where y is defined as:

$$y = -\frac{sin(\frac{\pi}{180} \cdot l) \cdot sin(\frac{\pi}{180} \cdot S) - sin(\frac{\pi}{180} \cdot d)}{cos(\frac{\pi}{180} \cdot l) \cdot sin(acos(sin(\frac{\pi}{180} \cdot S)))} \tag{24}$$

where $l$ is the latitude of the vehicle and $S$ is the sunheight (cf. Equation 32). The parameter $d$ represents the declination and is based on the time. Equation 35 in Appendix A.7 displays the detailed calculation path.

#### Calculation of incidence angle on windshield

In addition to the calculation of the sun height, Appendix A.7 also contains the calculation of the vertical angle of incidence of the sun on the windshield of the vehicle.
The horizontal angle of incidence of the sun on the vehicle $\alpha$ includes the previously calculated azimuth angle of the sun AoS and the direction of travel of the vehicle H for Heading and is defined as follows:

$$\alpha = AoS - H + 90° \tag{25}$$

Since the horizontal angle of incidence on the vehicle is not related to the orientation of the vehicle, but to the orientation of the windshield where the sensor is mounted, 90° is added to the difference between the azimuth angle

of the sun AoS and the orientation of the vehicle H. This results in an angle of incidence of 90° and thus a perpendicular incidence on the windshield and the sensor when the vehicle is driving exactly towards the sun.

### Weighting of signals

Since both vehicle-based signals have their own assigned range, which partially overlaps with the range of the neighboring signal, a weighting of the individual signals depending on the angle of incidence of the sun on the vehicle seems reasonable.

First, effects induced by the horizontal angle of incidence of the sun are considered. For this, the full circle of possible horizontal angles of incidence on a vehicle is considered. The angles of incidence are divided into nine sectors depending on their value, as shown graphically in Figure 19.

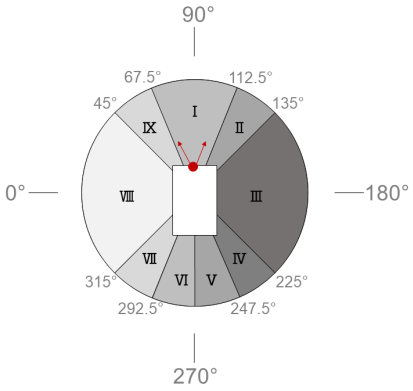


Figure 19: Schematic representation of the sector division for weighting the global radiation based on the horizontal angle of incidence of the sun on the vehicle. The car is shown as a white box, windshield facing up towards 90°. The small red circle represents the available sensor, the red arrows represent the viewing direction of the two global radiation signals from this sensor. Note, the angle of incidence of the sun on the vehicle is defined as zero when the sun shines on the vehicle from the left side. Sun shining exactly from ahead thus means an angle of incidence of 90°.

Two weights are defined for each of these sectors, one for each signal. "a" represents the sensor oriented to the front left and "b" the sensor aligned to the front right. An overview of this is provided by Table 3.

The steeper the sun shines on the respective sensor, the higher is the weighting of the respective signal. The signal, from which side the sun shines on the vehicle, makes up the main part of the weighting in each case. For example, if the sun falls on the vehicle from the left side with a horizontal incidence angle of 40°, signal "a" of the front sensor is weighted with 75 %, and signal "b" is weighted with 25 %. At an angle of incidence of 90°, which corresponds to a direct irradiation from the front, both signals "a" and "b" are equally weighted with 50 % each.

The shown weighting of the two signals with respect to the horizontal angle of incidence of the sun represents a highly simplified version. Instead, an arbitrarily more complex version can be defined, which assigns continuous activation to the signals instead of step-by-step activation.

Since the vehicle is at standstill for the whole campaign, the combination of horizontal and vertical incidence angle on the vehicle is almost identical for every day of the campaign. The influence of both angles of incidence can therefore not be evaluated simultaneously. To investigate the influence of the vertical angle of incidence, another measurement campaign is necessary, in which the vehicle is next to the same reference but has different vertical angles of inclination on the different days. This artificially varies the vertical angle of

| Sector | Incidence angle of sun on windshield | Sun from | a | b |
|---|---|---|---|---|
| I | [67.5, 112.5) | ahead | 0.5 | 0.5 |
| II | [112.5, 135.0) | front-right | 0.2 | 0.8 |
| III | [135.0, 225.0) | right | 0.25 | 0.75 |
| IV | [225.0, 247.5) | behind-right | 0.2 | 0.8 |
| V | [247.5, 270.0) | behind(-right) | 0.1 | 0.9 |
| VI | [270.0,292.5) | behind(-left) | 0.9 | 0.1 |
| VII | [292.5, 315.0) | behind-left | 0.8 | 0.2 |
| VIII | [315.0,360.0] & [0.0, 45.0) | left | 0.75 | 0.25 |
| IX | (45.0, 67.5] | front-left | 0.8 | 0.2 |

Table 3: Overview of weighting for two sensors a and b depending on horizontal sun incidence angle on the vehicle.

incidence and thus separates its influence as far as possible, whereby it can be determined.

The known effects, which exert an influence on the global radiation, are all visible in stationary situations, so that no further correction for effects induced by vehicle motion has to be made. Nevertheless, the correction of the global radiation is not completely finished yet, because the influence of the vertical incidence angle cannot be quantified due to the data basis. The weighting of the two signals of the global radiation sensor is therefore based exclusively on the horizontal angle of incidence of the sun on the vehicle.

### 4.4.3 Sensor uncertainty induced in stationary situations

#### Factor correction

After the input filter, both vehicle-based signals are subjected to a factor correction. Based on the available data, the factor $F_{SU}$ is empirically determined to be 0.7. The same factor $F_{SU}$ applies to both signals, since they are measured by the identical sensor and differ only in their alignment, but not in the measuring principle or sensor.

The factor correction only affects the times when the measured value is greater than zero. At night, when the global radiation is zero, the values therefore remain unchanged. During the day, at global radiation values greater than zero, the values are reduced by 30 %.

The factor $F_{SU}$ is set to 0.7 empirically based on the available data and therefore does not take into account any physical relationships. This factor correction is comparable to the offset correction for the effects induced by sensor technology for the air pressure and the air temperature. However, no offset but a factor is used for the global radiation.

Both vehicle-based signals corrected by a factor of 0.7 are then merged into one global radiation signal using the weighting from Chapter 4.4.2.

### Compensation of varying effects during daily course

Since the calculated global radiation signal, which is based on the two corrected individual signals as well as the weighting of these signals, shows anomalies that are dependent on the time of the day, a further correction is required.

In the morning, the calculated global radiation signal shows values too high compared to the reference. In the afternoon, the trend reverses and the vehicle-based values are too low compared to the reference.
The cause of this is not the time of day itself, but the angle of incidence of the sun on the vehicle, which changes during the course of the day. However, whether the deviations are due to the vertical or horizontal angle of incidence, cannot be determined from the present data set.

To compensate this effect, a linear function of the correction factor $MA$ for Morning/Afternoon is determined empirically. For very flat incidence angles in the morning, the global radiation value is corrected by the factor $MA = 0.8$ and thus reduced, whereas the factor $MA$ for flat incidence angles in the afternoon is 1.2 and thus increases the global radiation value. Using these two factors, a linear function is created that continuously covers the range of horizontal incidence angle from 0 to 180°. The linear equation of the factor $MA$ is:

$$MA = \frac{1}{450} \cdot \alpha + 0.8 \tag{26}$$

with $\alpha$ being the horizontal angle of incidence of the sun on the windshield of the vehicle. For the determination of the factor of the respective time, the horizontal, instead of the vertical, angle of incidence of the sun is used. Due to the stationary vehicle, the ratio between the horizontal and vertical incidence angle is almost identical over time every day in the present data of the stationary measurement campaign lasting one week. Therefore, the vertical incidence angle is also indirectly taken into account in this correction step.

# 5 Experimental evaluation

Since not all potential reference data can be assumed to be valid throughout, Chapter 5.1 discusses the limitations of the references used. Chapter 5.2 explains the quality parameters that are used in the following for evaluation and assessment. The following chapters contain the results and evaluations of the four meteorological quantities studied: Air pressure (Chapter 5.3), air temperature (Chapter 5.4), relative humidity (Chapter 5.5), and global radiation (Chapter 5.6).

## 5.1 Quality limitations of used references

During the measurement campaigns, anomalies occurred at the reference stations, which partially limit the suitability as a reference for individual quantities. In the following, the limitations of the MWS and the MMU are listed parameter-wise. For the MWS difficulties have been encountered for the following two parameters global radiation and air temperature. The project vehicle TC1 as well as the MMU show occasional difficulties with the relative humidity.

The signal of the global radiation measured by the MWS has attracted negative attention over the course of the project. No total failure occurred, instead a successive decrease of the measured quantity occurred. At night, the global radiation correctly continued to drop to zero. However, during the day, the amplitude continued to decrease daily, regardless of coverage or position of the sun. Therefore, the MWS was sent to the manufacturer for inspection and was no longer used as reference. As an alternative reference for the global radiation, the synop station "Weihenstephan" was subsequently used within the framework of an additional stationary measurement campaign. Since the global radiation is not only influenced by the position of the sun, but is also influenced by clouds, the low temporal resolution of 1 min of the synop station poses a challenge using this data as reference. For comparing vehicle-based

data to synop station data, the vehicle-based data must be averaged over 1 min to create an identical time basis for the evaluation. Statements about the quality of the second-by-second data based on the synop station is not possible.
The measuring device, which is mounted on the MMU to measure the global radiation, was originally not planned as a reference. The sensor is able to reproduce the changes of the global radiation, but the output absolute value must first be corrected by an offset before it is representative. This offset is determined as part of a comparison measurement next to the "Weihenstephan" synop station. After the offset correction, the values of the MMU averaged over 1 min agree well with those of the synop station. Therefore, it is assumed that the second-by-second data of the MMU is suitable as a reference.

The air temperature signal of the MWS has some deficiencies, too. As mentioned above, the MWS was returned to the manufacturer for inspection. During the repair, the manufacturer, according to his own information, used a soldering spray on the circuit board. It was not considered that the solder used has the property of becoming conductive if there is a high relative humidity over a longer period. This in turn has an influence on the measured signal and therefore probably causes a decaying temperature signal. This becomes clear in the second stationary measurement campaign in January/February 2021, where the air temperature of the MWS records opposite temperature trends to all other sensors for high relative humidities (cf. Figure 36 in Appendix A.8). The signal of the air temperature of the MWS is therefore not trustworthy for the period of the second stationary measurement campaign and is thus not used as a reference.

Additionally, the auxiliary heater of the TC1 was temporarily switched on during the second stationary measurement campaign to protect the additional electronics on board of the vehicle from damage caused by the cold. However, the parking heater, which is positioned next to the engine block, exerts a significant influence on the adjacent relative humidity measurement. The RH signal is therefore strongly distorted, which is not due to effects induced in static situations but only to the auxiliary heater. Thus, this data is not suitable for determining the sensor uncertainty induced in static situations. Hence, only the data from the first measurement campaign, in which no parking heater was active, is used for the evaluating the relative humidity signal during static measurement campaigns.

Normally, the MMU is used in high-pressure weather conditions to measure heat effects in urban areas. The use in bad weather or precipitation is a premiere for the measurement vehicle. Problems occurred during the mobile measurement campaign in February 2020, as the measuring instruments for relative humidity temporarily froze due to the persistent cold combined with multiple precipitation and snowfall events. This can be recognized by a strongly jagged and unnatural course of the measurement signal. However, an exact start and end time is difficult to define, since the freezing and thawing of the measuring instrument is a gradual process. For this reason, none of the MMU measurements of relative humidity for the February campaign can be used for reference purposes.

## 5.2 Quality parameters used for evaluation

Various quality parameters are used to evaluate the results, such as the MSE, data portion within single measurement uncertainty (PD1MU) and data portion within twice the measurement uncertainty (PD2MU).

The most common of the three quality control measures for forecasting and correction models is the MSE. It indicates the squared difference between estimated and true value. The estimated value is the corrected vehicle-based signal. The true value is the associated reference, either the measured value of a stationary reference (MWS or synop station) or the mobile reference (MMU).

The other two measures of quality refer to the proportion of the corrected data that can be assumed to be valid. This quality measure is strongly dependent on how large the maximum permissible deviation is defined, up to which a measurement is considered valid. The greater the maximum permissible difference between the corrected and the true value is set, the higher the proportion of data that is assumed to be valid. However, this does not increase the quality of the data, even if the percentage increases. Since the proportion automatically grows to 100 % if the maximum deviation is large enough, it is necessary to define a suitable and meaningful maximum allowable deviation. In this work, the single measurement uncertainty is defined as the limit for the maximum permissible deviation between the corrected vehicle-based and the reference value. Corrected values that deviate from the reference value by less than the single measurement uncertainty are therefore assumed to be valid. A

lower limit would incorrectly exclude correct values, since the deviation may be caused by the measurement inaccuracy of the sensor, rather than a real deviation. Since the measurement uncertainty is different for each evaluated signal, the limit value for the respective maximum permissible deviation also varies.

The measurement uncertainties for the respective meteorological quantity considered are composed of the measurement uncertainties for the reference and the vehicle-based sensor. Table 4 lists the measurement uncertainties for all sensors used sorted by meteorological quantity.

Exemplary, the measurement uncertainty for the air pressure during a stationary measurement campaign is 10.8 hPa, consisting of 10.0 hPa for the vehicle-based sensor and 0.8 hPa for the reference sensor.

The third quality measure is directly connected to the previous one. Instead of the single measurement uncertainty, twice the measurement uncertainty is defined as the permissible maximum difference between the corrected and the reference value. The proportion of the data determined in this way cannot be assumed to be valid. Rather, it is considered to be a measure of the spread of the corrected values around the reference value. The larger the proportion of data within twice the measurement uncertainty, the fewer outliers, equivalent to unambiguous false values, are present after correction.

| | **Vehicle-based measurement** | | **Reference for stationary measurement campaign** | | **Reference for mobile measurement campaign** | |
|---|---|---|---|---|---|---|
| | MU | Sensor | MU | Sensor/ Reference station | MU | Sensor/ Reference station |
| Air pressure | ± 10 hPa | Bosch SMP 480 | ± 0.8 hPa | Bosch BMP180 (MWS) | ± 0.08 % | Transcat DPI 260 Series (MMU) |
| Air temperature | ± 0.5 K | Amphenol Advandes NKS8588 (BCM1) | ± 0.3 K | Sensirion SHT25 (MWS) | ± 0.1 K | Theodor Friedrichs Frankenberger Art (MMU) |
| Relative humidity | ± 3 %rH | Sensirion SHT20 (MuFu) | ± 2 %rH | Sensirion SHT25 (MWS) | ± 2.5 %rH | E+E EE33 (MMU) |
| Global radiation | ± 7 % | Valeo | ± 3 % | Pyranometer CM11 Kipp & Zonen (Synop) | ± 2 %*1 | EG & G Heimann (MMU) |

*1: The specified MU refers to measurement in which sensor is continuously in balance. Due to dynamic effects during travel, the horizontal position of the sensor cannot be guaranteed continuously. Since there is no quantifiable statement about the increase in MU, this is not considered further.

Table 4: Measurement uncertainty of all sensors (vehicle-based and reference) relevant for evaluation.

If not stated differently the following plots all have the same color scheme. Light blue represents the vehicle-based raw signal, dark blue the PHY-corrected signal, red the NN-corrected signal, yellow the RF-corrected signal and grey the reference signal provided by either the MWS and the WB for stationary campaigns or the MMU for mobile campaigns.

## 5.3 Air pressure

### 5.3.1 Stationary measurement campaigns

A total of 887674 s of data (approx. 10 days) are available for the determination of the sensor uncertainty induced in static situations for atmospheric pressure. They are divided into 288946 s (32.6 %) obtained during the first stationary measurement campaign in November 2020 and 598728 s (67.4 %) during the second stationary measurement campaign in January/ February 2021.

However, before determining the offset necessary to compensate for the effects induced in stationary situations, an input filtering to detect and filter outliers is performed. Regarding the air pressure, values less than or equal to zero are outside the measurement range and are therefore defined as outliers. This is done by keeping the value from the previous time step constant. The outlier correction is performed for a maximum of five continuous seconds, to avoid introducing a new error (cf. Chapter 4.1.1).

Step 2 and step 3 focus on determining the sensor uncertainty induced in static situations, divided into the uncertainty caused by the sensor inaccuracy $SU_{Sensortechnology}$ and the uncertainty caused by mounting the sensor in a vehicular environment instead of an undisturbed environment $SU_{Location}$.

Table 5 lists the determined uncertainties induced in static situations. The $SU_{Sensortechnology}$ represented by the difference between the raw signal of the WB and the reference signal of the MWS is almost identical for the first and the second stationary measurement campaign. The weighted average is 2.20 hPa and thus amounts to 22 % of the measurement uncertainty. The temporal course for the WB signal corrected for the $SU_{Sensortechnology}$ is displayed in Figure 37 in Appendix A.9.

The $SU_{Location}$ is built by comparing the vehicle-based raw signal corrected by the previously determined $SU_{Sensortechnology}$ to the WB signal also corrected by the $SU_{Sensortechnology}$. $SU_{Location}$ varies by less than 9 % comparing both stationary measurement campaigns. The weighted average for the static $SU_{Location}$ amounts to 6.47 hPa. Therefore, $SU_{Location}$ is also smaller than the measurement uncertainty. Figure 38 in Appendix A.9 shows the temporal course for the vehicle-based signal corrected for both static $SU_{Sensortechnology}$ and static $SU_{Location}$ in comparison to the corrected WB signal.

As described in Chapter 4.1.2, the sum of both sensor uncertainties induced by sensor technology and location, sums up to the total static error $SU_{stat,total}$ if no further influencing static factors are involved. Therefore, the total sensor uncertainty induced in static situations amounts to 8.67 hPa for combining all data from both static measurement campaigns. Note, that combining $SU_{Sensortechnology}$ and $SU_{Location}$ still results in a value smaller than the measurement uncertainty.

To control, whether additional influencing static factors affect the vehicle-based measurement, the deviation between the vehicle-based signal corrected by $SU_{stat,total}$ and the reference signal of the MWS is determined for each time step. The temporal course for the air pressure signal of TC1 corrected for the $SU_{stat,total}$ is displayed in Figure 39 in Appendix A.9.

Without correcting the vehicle-based data, the averaged deviation amounts to 8.66 hPa, with a maximum deviation of -13.05 hPa. After correcting the vehicle-based signal, the averaged deviation is decreased to 0.01 hPa. Since the measurement uncertainty of the air pressure is 10.8 hPa (cf. Table 4), the averaged deviation of 0.01 hPa lies within this uncertainty. The manufacturers' specifications can thus be confirmed, and no significant uncertainties caused

| | Stationary measurement campaign 1 Nov 2020 | Stationary measurement campaign 2 Jan/ Feb 2021 | Both stationary measurement campaigns combined |
|---|---|---|---|
| $SU_{Sensortechnology}$ | 2.26 hPa | 2.17 hPa | 2.20 hPa |
| $SU_{Location}$ | 6.84 hPa | 6.28 hPa | 6.47 hPa |
| $SU_{stat,total}$ | 9.10 hPa | 8.45 hPa | 8.67 hPa |

Table 5: Sensor uncertainty induced in stationary situations for air pressure based on two static measurement campaigns.

by $SU_{Sensortechnology}$ or $SU_{Location}$ occur.
Since the correction of sensor uncertainties induced in static situation is a pure offset correction, the distribution of the deviations is not changed.

However, since the maximum deviation after correction accounts to 4.46 hPa and is thereby smaller than the measurement uncertainty of the vehicle-based sensor, all corrected vehicle-based measurements of the air pressure are within the measurement uncertainty for both stationary measurement campaigns. Therefore, PD1MU as well as PD2MU, both amount to 100 %.

### 5.3.2 Mobile measurement campaigns

Since the air pressure is largely independent of the state of the vehicle (cf. Chapter 4.1.2), no complex correction for effects induced by vehicle motion is needed. Therefore, the combination of input filter and correction of sensor uncertainties induced in stationary situations represents the total correction for the vehicle-based air pressure signal.

During the three mobile measurement campaigns, 207013 s of concurrent air pressure data of the TC1 and the reference (MMU) are collected. In Figure 20 (upper plot) the temporal course of the vehicle-based corrected air pressure

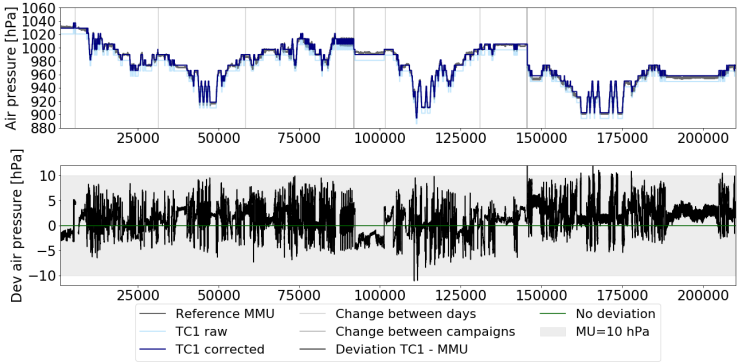


Figure 20: Time course of corrected vehicle-based air pressure signal of TC1 in comparison to reference signal (MMU) for mobile measurement campaigns.

signal in comparison with the reference for mobile campaigns (MMU) is shown. Additionally, the deviations between the corrected vehicle-based and the reference signal are shown in the lower plot.

Using the $SU_{stat,total}$ offset correction the averaged deviation of the vehicle-based signal to the reference signal from the MMU can be reduced by almost 81 % down to 1.40 hPa (cf. Table 6). The average deviation amounts to less than $\frac{1}{6}$ of the measurement uncertainty, which indicates a correction specific to this one vehicle, but does not prove it. In the next chapter, the correction developed on TC1, is transferred to TC2. This can be used to verify whether the correction is specific in a way that it provides reliable results only on TC1 or whether the empirically determined offset values are also valid on TC2.
Table 6 highlights the good agreement between the corrected and reference data, which is at least 99 % for each of the measurement campaigns (PD1MU). The PD2MU values of 100 % throughout show that there are no outliers in the available vehicle-based data. Therefore, all data are within twice the measurement uncertainty. The deviations distribute evenly with a standard deviation of 2.63 hPa (cf. Figure 40 in Appendix A.9 and Table 6).

Based on the available data of the three mobile measurement campaigns, no difference caused by the seasons can be detected. All results regarding the deviations between the signals are comparable.

| TC1 | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 2 Feb 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|---|---|
| Data | Raw | Corr | Raw | Corr | Raw | Corr | Raw | Corr |
| Data points [sec] | 89016 | | 53416 | | 64581 | | 207013 | |
| Ave Dev [hPa] | -7.28 | 1.39 | -8.79 | -0.12 | -5.99 | 2.68 | -7.27 | 1.40 |
| Max Dev (abs) [hPa] | 15.99 | 9.89 | 19.76 | 11.09 | 14.09 | 12.24 | 19.76 | 12.24 |
| Std dev [hPa] | 2.46 | | 2.63 | | 2.14 | | 2.63 | |
| PD1MU [%] | 83.64 | 100.00 | 63.66 | 99.87 | 96.99 | 99.95 | 82.65 | 99.95 |
| PD2MU [%] | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

Table 6: Overview of the results for raw data (Raw) and corrected data (Corr) of TC1 for the mobile measurement campaigns.

### 5.3.3 Transferability of models

The results presented above refer exclusively to the analysis of data from TC1. To test whether the determined correction $SU_{stat,total}$ is also valid on another vehicle, the identical correction developed on TC1 is applied to TC2.

Since TC2 experienced intermittent technical difficulties during the measurement campaigns, the amount of data from TC2 is less than from TC1 and totals 103274 s for the air pressure signal of all three mobile measurement campaigns. Thus, the data basis of TC2 amounts to slightly less than half of the existing data of TC1.

The use of the $SU_{stat,total}$, which was developed on TC1, also gives solid results on TC2. The time course of the TC2 signal, as well as the deviations from the reference, are shown in Figure 21. Due to measurement failures, TC2 does not show data for all time points. The tabular overview of the results of the individual mobile measurement campaigns is shown in Table 42 in Appendix A.9.

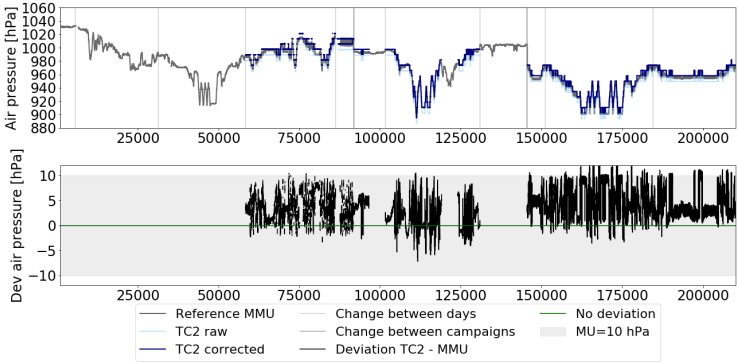


Figure 21: Time course of corrected vehicle-based air pressure signal of TC2 in comparison to reference signal (MMU) for mobile measurement campaigns. Correction applied to TC2 was developed on TC1 and therefore, appears to be vehicle non-specific. Due to technical difficulties, there are data gaps in TC2.

For the data of the three mobile measurement campaigns, the correction to TC2 achieves 98.81 % of the data within the simple measurement uncertainty (cf. Table 42) and is thus only slightly smaller than for TC1 with 99.95 %.

For TC2 data no clear influence of the individual measuring campaigns, and thus of the seasons is recognizable.

It is noticeable that the correction developed on the TC1 slightly overcorrects the TC2 data. The mean deviation of the vehicle-based data from the reference of the MMU, is changed from -4.88 hPa for the raw data to 3.79 hPa for the corrected data. An offset correction of lesser extent would give even better results on TC2, but would also represent a more specific solution. Nevertheless, it is clear that the correction of the TC1 also takes effect on the TC2 and improves the results compared to the raw data.
On the basis of the available data for TC2, the air pressure correction seems to be a generally valid solution that is not only limited to TC1 as previously assumed (cf. Chapter 5.3.2).

## 5.4 Air temperature

### 5.4.1 Stationary measurement campaigns

For the determination of the offset, which is necessary to correct the sensor uncertainty induced in stationary situations, data from both stationary measurement campaigns in November 2020 and January/February 2021 is available. However, as already explained in Chapter 5.1, the temperature measurement of the MWS in the second measurement campaign is unsuitable as a reference. Therefore, only data from the stationary November measurement campaign is considered for the following evaluation. This reduces the amount of data to 294431 s, which is about $\frac{1}{3}$ of the data of the original data set.

Before the sensor uncertainty induced in stationary situations can be determined, an input filter as described in Chapter 4.2.1 is executed to filter for outliers and values outside physical boundaries.

As described in Chapter 4.2.2, a multilevel evaluation is performed to determine the offset, which is visible in stationary situations. This offset is due to sensor

uncertainty induced by sensor technology on one hand and location of the sensor on the other hand.

In the first step, the mean deviation between the WB and the MWS is determined to isolate the uncertainty induced by sensor technology. The average deviation, and thus $SU_{sensortechnology}$ is 0.91 K. Since $SU_{Sensortechnology}$ is greater than the measurement uncertainty of 0.8 K for the temperature comparison between the WB and the MWS, there is a discrepancy between the two sensors that is not exclusive to the measurement uncertainties of both sensors. Therefore, under the assumption that the MWS measures the true temperature value, the WB sensor shows an anomaly.
Both, the WB data, as well as the data of the TC1, is corrected by this offset since both measurement systems are affected by this influence as they both use identical sensors.
In the second step, the mean deviation between TC1 and the WB is calculated. For this purpose, the used signals of both data sources, TC1 and WB, have already been corrected for the $SU_{sensortechnology}$. The subsequently determined mean deviation between TC1 and WB is -0.1 K ($SU_{Location}$ = -0.1 K).

The offset caused by the location of the sensor, $SU_{Location}$, thus counteracts the offset for compensating the inaccuracy caused by the sensor technology and dampens this offset. This results in a total sensor uncertainty induced in stationary situations of 0.81 K:

$$\begin{aligned} SU_{stat,total} &= SU_{Sensortechnology} + SU_{Location} \\ &= 0.91K - 0.1K \\ &= 0.81K \end{aligned} \tag{27}$$

Finally, the previously determined offset of 0.81 K is directly verified by comparing the corrected TC1 based temperature signal and the reference value of the MWS. If both signals are identical, no further influences disturb the stationary temperature measurement. The temporal course of the air temperature values for the stationary measurement campaign is shown in the upper plot in Figure 43 in Appendix A.10, the resulting deviations are displayed in the plot below. Table 43 in Appendix A.10 lists the associated quality parameters.

The mean deviation of 0.00 K confirms the successful correction of effects induced in stationary situations. The offset correction of 0.81 K is 0.01 K higher than the measurement uncertainty of the temperature for stationary

campaigns of 0.8 K. Assuming that the MWS measures the real temperature, the vehicle-based temperature in the uncorrected state therefore deviates by slightly more than the measurement uncertainty. Without the offset correction, the vehicle sensor therefore exhibits anomalies to the real temperature measured by the MWS.
By correcting for the offset $SU_{stat,total}$, the proportion of data within the single measurement uncertainty (PD1MU) can be improved by 36 percentage points. Reasons for deviations greater than twice the measurement uncertainty can be influences by other factors, such as global radiation. To compensate for these influences, a more complex correction of the effects induced in stationary situations might be necessary, which, however, is not tested within the scope of this work.

### 5.4.2 Mobile measurement campaigns

When applying the previously determined $SU_{stat,total}$ to the data of the mobile measurement campaigns (207347 s), the offset reduction of the raw values by $SU_{stat,total}$ of 0.81 K, has reduced both the average deviation and the maximum deviation. However, an unexpected result occurred as additionally the proportion of data that can be assumed to be valid (PD1MU) has worsened instead of improved by approximately 11 percentage points. The detailed listed numerical values are deposited in Table 44 in Appendix A.10. A decrease in the proportion of the data that has a deviation equal or less than the single measurement uncertainty to the reference (PD1MU) indicates that the previously determined offset overcompensates for the sensor uncertainty induced in stationary situations for the data of the mobile measurement campaigns.

Therefore, on the present data basis of the mobile measurement campaigns with the MMU as reference, another offset is determined empirically, which counteracts the previously determined $SU_{stat,total}$ and thus attenuates it. The necessary damping is 0.48 K. Therefore, the total offset is reduced to 0.33 K. This raises the PD1MU from 53.11 % for the raw data to 58.46 %, instead of lowering it to 41.90 % (cf. Table 44 in Appendix A.10).

The reason for the necessity of a modification of the offset correction can be found in the different reference stations for the stationary and the mobile measurement campaigns. The data indicate that the reference stations MWS

and MMU have an offset of approximately 0.5 K among themselves. The measurement uncertainty for the comparison between MWS and MMU accounts to 0.4 K, meaning the offset between both stations is outside the acceptable range. Due to the given measuring conditions for both measuring stations, the exact offset cannot be verified by a parallel long-term measurement.
Assuming that the MWS and the MMU have an offset among themselves, the vehicle-based data of the TC1 can never match both reference stations at the same time. This illustrates the problem and at the same time emphasizes the importance of choosing a suitable reference for comparative measurements. The question arises whether either the MWS or the MMU represents a valid value of the air temperature. It is highly probable that neither of the two stations consistently measures the exact temperature value. However, it is not possible to determine a valid temperature value more precisely than with the present reference equipment during this project. Since the MWS has already attracted negative attention regarding the quality of its temperature measurement and the further evaluation takes place within the scope of the mobile measurement campaigns, the MMU is assumed to be true.

### 5.4.3 Physical model

An improvement of the results by applying the physical model (PHY) is clearly shown in Figure 22 as well as in Table 7.

In Figure 22, the x-axis represents the time and thus the available data points, the y-axis represents the air temperature in °C. The three plots present one mobile measurement campaign each. The vertical subdivisions in light gray mark the breaks between individual days during each measurement campaign.

A significant improvement of the signal due to the physical correction becomes visible. Time sections with strong deviations between the raw signal and the reference, such as at about 68000 and 98000 s, are modified by the physical model in a way that they are subsequently in good approximation to the reference signal. However, time sections where the physical model overcompensates for the uncertainty induced by vehicle motion (e.g., at about 48000 s) can also be identified. Furthermore, it is noticeable that the deviations between the corrected PHY signal and the MMU signal are more pronounced for the third mobile measurement campaign than for the two previous measurement

campaigns. The fact that this behavior cannot be observed in the two previous measurement campaigns indicates to another effect influencing the temperature measurement, which is not yet considered in the physical model.

Table 7 confirms the previously described results, becoming apparent in Figure 22. The mean deviation for all data of the mobile measurement campaigns (207347 s) is reduced by a total of 94 %, with the first 19 % achieved by the offset correction performed to compensate for effects induced in stationary situations. The remaining 75 % of the improvement are obtained by the physical model. Hereby, the mean deviation decreases by 94 %.

Likewise, the maximum deviation is significantly reduced by the physical model. The offset correction reduces it by only about 1 %, whereas the physical model applied afterwards reduces it by another 59 %. The magnitude of the maximum deviation is reduced by the correction, but outliers are still present in the data. The remaining outliers are isolated errors, as shown by Figure 22 and confirmed by a low standard deviation (cf. Table 7).

The standard deviation of the corrected data decreases by more than $\frac{2}{3}$ of its original value. The corrected values are therefore less widely scattered around

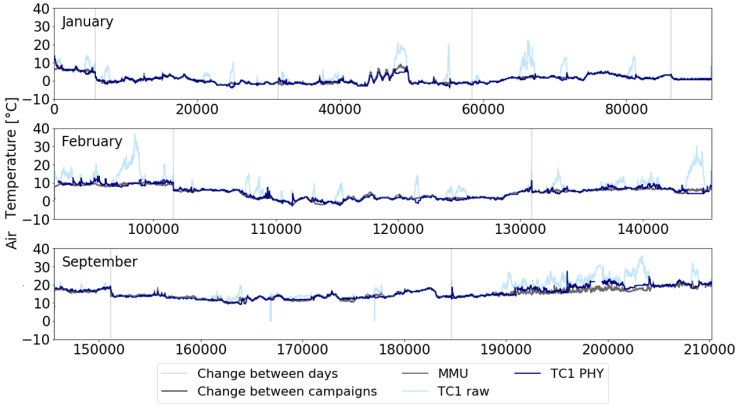


Figure 22: Time course of three air temperature signals (darkgrey: MMU, lightgrey: TC1 raw, blue: TC1 corrected based on physical model) for all three mobile measurement campaigns in January, February and September 2020 spread over three plots. Vertical gray subdivisions mark a change of day during each measurement campaign.

the mean deviation, which is shown in Figure 44 in Appendix A.10. Instead, more data points are located within a short distance from the mean deviation. Therefore, the corrected data concentrates stronger around their mean deviation of 0.1 K. For the raw data (cf. Figure 44a)), the distribution is stretched to the right, which indicates the majority of the vehicle anomalies being greater than zero. The physical model can reduce the disproportionately frequent positive anomalies, resulting nearly in a Gaussian distribution (cf. Figure 44b)).

Considering all data from the mobile measurement campaigns, the PD1MU increases by more than $\frac{1}{4}$. It should be noted that the PD1MU of the raw data of the February and September campaign is also significantly lower than for the January campaign. The percent increase in PD1MU due to the physical model is similar for the September and January campaign and largest for the February campaign. The nevertheless lower PD1MU is due to different ambient conditions compared to the January campaign. These deviating ambient conditions exert an influence on the measured temperature signal, which is not sufficiently corrected by PHY model.

One influence to be mentioned here is the global radiation. It is responsible for the long-lasting pronounced deviations between the raw data and the reference data in September (cf. Figure 22). Since the number of data where the global radiation shows significant differences from the January campaign is limited to the September campaign, the data basis is insufficient to modify the physical model in such a way that it also provides reliable results in situations with strong global radiation.

| TC1 PHY | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 2 Feb 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|---|---|
| | Raw data | PHY | Raw data | PHY | Raw data | PHY | Raw data | PHY |
| Data | 89043 seconds | | 53716 seconds | | 64588 seconds | | 207347 seconds | |
| Ø Dev [K] | 1.23 | -0.08 | 2.26 | 0.16 | 1.92 | 0.31 | 1.71 | 0.10 |
| Max Dev [K] | 21.51 | 7.58 | 27.15 | 7.03 | 18.43 | 10.97 | 27.15 | 10.97 |
| Std dev [K] | 2.53 | 0.72 | 3.55 | 0.99 | 3.02 | 1.24 | 3.00 | 0.99 |
| PD1MU [%] | 66.32 | 80.31 | 36.25 | 54.47 | 48.92 | 58.72 | 53.11 | 66.95 |
| PD2MU [%] | 79.70 | 92.68 | 57.12 | 80.91 | 61.70 | 79.31 | 67.99 | 85.50 |

Table 7: Overview of the correction results of TC1 for the physical model (PHY) for the mobile measurement campaigns.

The winter weather (snowstorm) that occurred during the mobile measurement campaign in February 2020 explains the lower quality of the data from this measurement campaign. Although the February data has the highest relative improvement in PD1MU by 50 %, the achieved PD1MU remains the lowest compared to the other campaigns due to the lowest PD1MU for the raw data. A further modification of the physical model to increase the PD1MU in scenarios with heavy snowfall, as occurred during the February campaign, is necessary. However, due to the very limited data basis in these extreme weather situations, in combination with the unknown exact influence on the temperature measurement, the development of a suitable modification of the physical model is not feasible based on the available data set.

The PD2MU behaves similar to the PD1MU. By the physical model, an increase of $\frac{1}{4}$ of the PD2MU to 85.50 % is obtained for data of all mobile measurement campaigns.

### 5.4.4 Neural network

The evaluation of the results obtained by the correction using the neural network (NN) is limited to 34422 s (cf. Table 8). This represents 16.6 % of the total amount of available data collected during all three mobile measurement campaigns. As listed in Chapter 4.2.3, the remaining $\frac{5}{6}$ of the data is needed to train and validate the network.

Figure 23 represents the data of the mobile measurement campaigns. The x-axis corresponds to the available data points, the y-axis to the air temperature. As described in Chapter 4.2.3, the last sixth of each of the three mobile measurement campaigns is used to test the network. Hence, results of the net are only available for these three sections (cf. red signals path in Figure 23). For the first two mobile measurement campaigns the NN results look trustworthy. For the third measurement campaign larger differences between the NN results and the reference signal of the MMU are noticeable.
Table 8 confirms the lower quality results for the third measurement campaign in September, which were already visually evident from Figure 23.

The mean deviation between the vehicle-based temperature values and the MMU reference are reduced on average over all three mobile measurement campaigns by almost 80 % by the NN. The mean deviation of all data is thereby

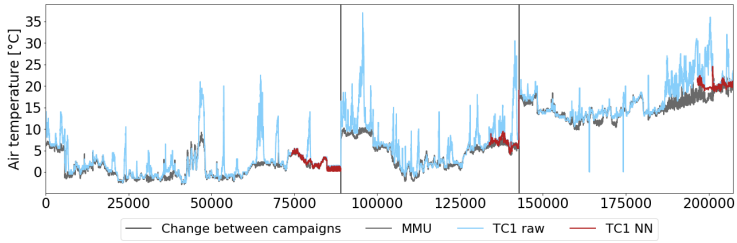


Figure 23: Temporal air temperature course of the reference signal (MMU, gray) and the vehicle-based raw signal of TC1 (lightblue). The temporarily available results of the NN for the last sixths of each mobile measurement campaign is plotted in red. The vertical lines mark a change between measurement campaigns.

in the same magnitude as the measurement uncertainty. To test whether this good agreement with the reference is due to a highly specified net or whether the net can identify the influencing correlations, the transferability of the net to a different data set is conducted in Chapter 5.4.7. Since the mean deviation for the raw data of the February and September campaign is comparable, but the mean deviation of the NN corrected values is more than twice as high for the September campaign than for the February campaign, this suggests that there are other effects affecting the data in September not included as features in the network.

The maximum deviation for all three mobile measurement campaigns is reduced by 75 % by the NN. For the September data the maximum deviation after NN correction is about twice as large (6.24 K) compared to the corrected

| TC1 NN | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 2 Feb 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|---|---|
| | Raw data | NN | Raw data | NN | Raw data | NN | Raw data | NN |
| Data | 14840 seconds | | 8818 seconds | | 10764 seconds | | 34422 seconds | |
| Ø Dev [K] | 0.70 | 0.05 | 4.03 | 0.58 | 4.33 | 1.32 | 2.69 | 0.58 |
| Max Dev [K] | 13.15 | 2.09 | 25.20 | 3.11 | 18.43 | 6.24 | 25.20 | 6.24 |
| Std dev [K] | 1.85 | 0.40 | 4.70 | 0.61 | 4.42 | 1.40 | 4.03 | 1.03 |
| PD1MU [%] | 86.52 | 91.69 | 16.91 | 57.61 | 29.01 | 35.71 | 50.71 | 65.46 |
| PD2MU [%] | 91.59 | 97.50 | 30.03 | 85.53 | 40.01 | 50.25 | 59.69 | 79.66 |

Table 8: Tabular overview of the correction result of TC1 for the neural network (NN) for the mobile measurement campaigns.

February data (3.11 K). Thus, the September measurement campaign also provides the largest maximum deviation, comparable to the results for the mean deviation.

The NN reduces the standard deviation by 74 %, from 4.03 K to 1.03 K. The largest improvement in standard deviation is seen in the February campaign, which reduces its standard deviation by 87 % from 4.70 K to 0.61 K. For the September campaign, the standard deviation of the raw data is slightly smaller than that for the February data. Nevertheless, the standard deviation after the NN correction is more than twice as large (1.40 K) as the one of the February campaign (0.61 K). The data of the September campaign is therefore in comparison more widely scattered around their mean deviation, which reinforces the thesis of further influencing parameters during the September campaign. However, an insufficient amount of data is available, especially for summer environmental conditions (comparable to the September campaign), which emphasizes an overall risk of applying the trained network to data with unknown environmental conditions or environmental conditions rarely occuring in the training data.

This is confirmed by the proportion of data lying within a single measurement uncertainty to the reference value (PD1MU). The September campaign achieves the lowest PD1MU after NN correction, although the PD1MU of the raw data of the February campaign is significantly lower than for the September campaign. Considering all data (34422 s), an increase of PD1MU by almost 30 % is achieved by the network. For the raw data 50.71 % of the data can be assumed to be valid, the NN increases the proportion to 65.46 %.

For the data within twice the measurement uncertainty (PD2MU), an increase of one third from 59.69 % to 79.66 % is achieved by the network. Regarding the PD2MU of the January campaign, it is noticeable that the PD2MU of the raw data is already above 90 %. Nevertheless, the value for this campaign can be raised further to 97.50 %. The largest percental increase in PD2MU is present for the February campaign, where the proportion increases by 85 % in relation to the raw values. The visualization for the February campaign (second section in Figure 23) depicts the reason for the significant increase in PD1MU by a factor of 3.4 and PD2MU by 90 %. In the considered time interval, the vehicle-based raw signal has multiple peaks lasting several seconds to minutes, which are well corrected by the network.

### 5.4.5 Random forest

Likewise for this machine learning-based approach, only 16.6 % of the data collected within the three mobile measurement campaigns are available for evaluation, as the remaining data is used to build the random forest model (RF).

Figure 24 represents the temperature signals over the time course of the collected data. Visually, there is good agreement between the RF results and the reference data for the January campaign. For the February campaign, the RF results show stronger fluctuations compared to the reference signal. In September, the course of the reference signal is well reproduced by the RF results, but the corrected signal contains multiple peaks over several seconds.

Considering all data available for evaluation (34423 s), the RF reduces the mean deviation by 94 % to -0.16 K. The reduction in mean deviation is smaller for the September than for the February campaign, which results in a higher mean deviation in absolute terms for the September campaign. This indicates that the RF cannot compensate for the influence of at least one foreign parameter on the data during this campaign.

In terms of maximum deviation, the RF reduces the value to less than a quarter, from 25.20 K for raw data to 5.40 K for RF-corrected data. In both the January and the February campaign, the maximum deviation is reduced by over 85 % by the RF. For the September campaign the reduction amounts to 70 %.

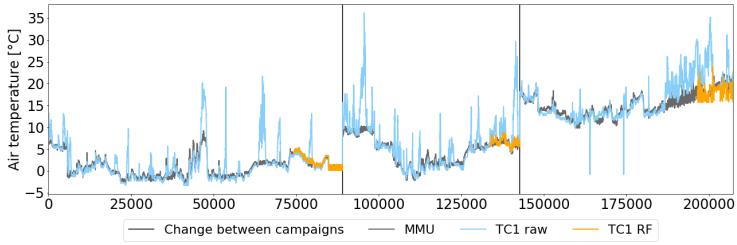


Figure 24: Temporal air temperature course of the reference signal (MMU, gray) and the vehicle-based raw signal of TC1 (lightblue). The temporarily available results of the RF for the last sixths of each mobile measurement campaign is plotted in yellow. The vertical lines mark a change between measurement campaigns.

| TC1 RF | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 2 Feb 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|---|---|
| | Raw data | RF | Raw data | RF | Raw data | RF | Raw data | RF |
| Data | 14840 seconds | | 8818 seconds | | 10765 seconds | | 34423 seconds | |
| Ø Dev [K] | 0.70 | 0.04 | 4.03 | 0.24 | 4.33 | -0.77 | 2.69 | -0.16 |
| Max Dev [K] | 13.15 | 1.87 | 25.20 | 3.07 | 18.43 | 5.40 | 25.20 | 5.40 |
| Std dev [K] | 1.85 | 0.35 | 4.70 | 0.7 | 4.42 | 1.37 | 4.03 | 0.97 |
| PD1MU [%] | 86.52 | 91.32 | 16.91 | 57.58 | 29.01 | 37.82 | 50.70 | 65.94 |
| PD2MU [%] | 91.59 | 99.47 | 30.03 | 90.58 | 40.01 | 63.02 | 59.69 | 85.79 |

Table 9: Overview of the correction results of TC1 for the random forest (RF) for the mobile measurement campaigns.

The RF correction reduces the standard deviation of the vehicle-based data over all mobile measurement campaigns by three quarters to 0.97 K. The largest percentage reduction in standard deviation is obtained for the February campaign (85 %), followed by the January campaign (81 %). The September campaign shows the least improvement with 69 % and has the largest standard deviation of the three mobile measurement campaigns with a value of 1.37 K. This strengthens the assumption that further influencing parameters are present during the September campaign.

The proportion of data that can be assumed to be valid (PD1MU) increases by 30 % to 65.94 %. For the January campaign, the PD1MU achieves 91.32 %, whereas for the September campaign only 37.82 % of the data are within the range of single measurement uncertainty around the reference value. However, for the September campaign, the PD1MU is smaller, but the increase in terms of raw data is larger (30 % increase compared to 6 % increase for January). Therefore, the RF achieves a smaller increase in quality for the data of the January campaign than for the data of the September campaign. The biggest increase in quality of the corrected data is achieved for the February campaign, where the PD1MU is risen by a factor of 3.4.

In terms of PD2MU the September campaign performs worst with a value of 63.02 % of the RF corrected data. Conversely, the January and February campaigns achieve a PD2MU over 90 %. The most frequent deviations, despite RF correction, are therefore present during data of the September campaign. Overall, the PD2MU can be increased by 44 % to 85.79 % due to the RF correction.

### 5.4.6 Comparing different model results

Figure 45 in Appendix A.10 shows the time periods of the three mobile measurement campaigns in January, February and September, where results are available for all three models. The top plot shows January data (14841 s), the middle plot February data (8952 s) and the bottom plot September data (10629 s).

The anomalies between each of the mentioned four signals to the reference of the MMU are shown in Figure 25. Each subplot includes all data points where values for all four signals are available. The horizontal gray lines mark the range of single measurement uncertainty. If the anomalies remain within this range, the associated data points can be assumed to be valid. Note the different scales of the y-axis. The y-axis of the anomalies of the raw data covers a larger range than for the corrected data and therefore makes the anomalies look

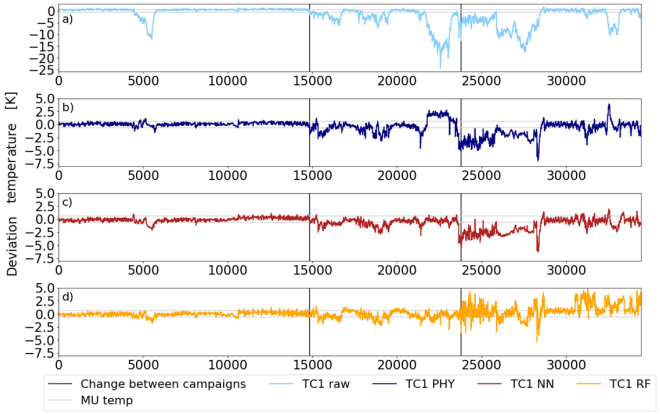


Figure 25: Time course of anomalies of vehicle-based temperature signals. Four vehicle-based temperature signals are considered: TC1 raw in lightblue, TC1 corrected based on physical model (PHY) in darkblue, TC1 corrected based on neural network (NN) in red, TC1 corrected based on random forest (RF) in yellow. Anomalies are shown for all three mobile measurement campaigns in January, February, and September 2020, separated by vertical black lines. Note, the y-axis section varies for anomalies of raw data (30 K) and corrected data (13 K). Therefore, the deviations of the raw data appear smaller than they would be if mapped at the same scale as the corrected data.

smaller. However, the deviations of the raw signal from the reference signal are several times larger than those of the corrected data.

Compared to the reference signal of the MMU, the temporarily significant deviations between raw signal and reference become clear (cf. Figure 45 and Figure 25a)). In January, the raw signals deviate by up to 12 K for a period lasting about 1000 s (approx. 15 min). In the February campaign, both the frequency and the magnitude of the deviation increase. During the September campaign, deviations occur mainly in the first half of the period considered.

It is noticeable that all three developed correction models (PHY, NN, and RF) apply a correction at times with large deviations between the raw signal and the reference (cf. Figure 45). Therefore, all three models react to the given situation and do not perform a constant offset correction. Nevertheless, all models show different patterns of behavior and therefore different anomalies during correction (cf. Figure 25).

For the January and February campaign all three models achieve comparably good results. Nevertheless, both ML models differ from the PHY model in the section from about 21000-23000 s during the February campaign. Here the PHY model shows almost constant and too low temperature values. Both, the NN and the RF, can correct this scenario better.

For the September campaign the results of the PHY and the NN approach resemble each other. Especially, this can be seen between 24000-28000 s where both models correct individual peaks in the raw signal and can represent the time course well but have an offset to the reference signal. This offset results in negative anomalies (cf. Figure 25b) and c)) and temperature values too high (cf. Figure 45). As discussed in Chapter 5.4.3, the influence of global radiation is a potential explanation for the observed offset. Wang et al. confirm the effect of the sun on temperature measurements and suggest training a neural network to compensate for it [WGMY17]. Since the global radiation is weaker during the winter campaigns (January and February) and is additionally reduced by the prevailing weather situations (e.g., snowstorm), this effect is not visible in the January and February campaign. Since neither the NN nor the PHY consider global radiation, it is a potential explanation for the observed offset. The behavior of the RF corrected data deviates significantly from the other two model results in the September campaign (especially between 24000-28000 s). The RF data is on average closer to the reference signal than the other

two models but shows stronger fluctuations (cf. Figure 25d)). The RF data provides intermittent reliable results for the September campaign but appears unstable compared to the other two models due to frequent and comparatively strong fluctuations. Nevertheless, the results do not show any offset, although the RF, just like PHY and NN, does not consider global radiation. The lack of information affects PHY and NN in an offset where RF is affected by more fluctuating results.

In addition to the visual representation, Table 10 and 45 in Appendix A.10 are used to compare the different model results. The numerical values listed for the raw signal and the PHY results differ from the results in Table 7 since Table 10 and Table 9 only consider data points for the evaluation at which results are available for all three models. The amount of data considered is therefore significantly lower.

All models improve the mean deviation of the raw data by at least 78 %. Considering all 34421 s, the mean absolute deviation for the random forest is lowest (-0.16 K), which is not surprising since the RF achieves the lowest mean deviation for all of the mobile campaigns considered individually. Considering all campaigns, the PHY model performs better (mean deviation = 0.45 K) than the NN model (mean deviation = 0.58 K).

The RF results also show the best value for the absolute maximum deviation for all mobile measurement campaigns. Nevertheless, the PHY model has the lowest maximum deviation for the January data. In terms of percentage improvement, the RF achieves the highest reduction of the maximum deviation (by 79 %), followed by the NN (by 75 %) and the PHY model (by 72 %).

| TC1 | All three mobile measurement campaigns combined | | | |
|---|---|---|---|---|
| | Raw data | PHY | NN | RF |
| Data | 34421 seconds | | | |
| Ø Dev [K] | 2.69 | 0.45 | 0.58 | -0.16 |
| Max Dev [K] | 25.20 | 7.04 | 6.24 | 5.40 |
| Std dev [K] | 4.03 | 1.23 | 1.03 | 0.97 |
| PD1MU [%] | 50.71 | 64.53 | 65.46 | 65.94 |
| PD2MU [%] | 59.69 | 75.13 | 79.66 | 85.79 |

Table 10: Tabular overview of the correction results of TC1 for all three models physical (PHY), neural network (NN) and random forest (RF) for the mobile measurement campaigns combined.

The highest standard deviation (1.23 K) and simultaneously the lowest reduction of 69 %, is found for the PHY model. The machine-learning based models, NN and RF, achieve a reduction of 74 and 76 %, respectively. For the February and September campaigns, the picture emerging over all mobile measurement campaigns is confirmed and the PHY results show the highest standard deviation. Conversely, for the January campaign data the deviations of the PHY results from the reference value show the smallest scatter around their mean deviation (Std dev = 0.29 K).

For the 34421 s considered, 50.71 % of the raw data is within the range of single measurement uncertainty around the reference value. All three models raise the PD1MU to at least 64 %. All models therefore show a significant improvement in data quality. The RF has the largest percentage of valid data (65.94 %), followed by the NN (65.46 %) and the PHY model (64.53 %). The deviations of 1 percentage point, show the comparable quality regarding the amount of valid data generated by the three models. Regarding the individual campaigns, it is noticeable that for both the January and September campaigns, the PHY model has the highest PD1MU. For both campaigns, the values of both machine-learning based approaches are at most 4.8 and 0.8 percentage points lower, respectively.
The proportion of the data that is within twice the measurement uncertainty (PD2MU) after correction is at least 75 % for all models. Three-quarters of the data corrected in the time interval under consideration is thus in the range of +/- 1.2 K around the reference value. The RF results, already highest for PD1MU, are as well highest for PD2MU with 85.79 %.

The comparable results of all three models are illustrated by the distribution of anomalies. Figure 46 in Appendix A.10 shows the histograms for the raw data, the PHY-corrected data, the NN-corrected data, and the RF-corrected data. The negative anomalies of the raw data, which are synonymous with excessively deviating temperature values, are striking. The strongly deviating raw data has already become clear in Figure 45 and Figure 25a) and is emphasized again by the histogram. The range of deviations can be significantly reduced by all three models. However, for better comparison of the histograms, the scales of the axes are retained. Due to the reduced range of anomalies, the proportion of data with lower anomalies increases, causing the histograms of the corrected data to become more concentrated. In addition, the strong leftward bias in the distribution of the raw data anomalies is reduced by all three models. However,

by the PHY model, there is an accumulation of anomalies in the positive region around 2 K. The NN results in the most compact distribution of anomalies, which is slightly skewed toward negative anomalies, but does not represent further accumulations in the positive range. The RF reduces the negative anomalies most effectively but shows more positive anomalies than the other two models.

Overall, the assessment for the best model is thus campaign dependent. For the January campaign, the PHY model has the best result with 96.16 % valid values and a PD2MU of 99.82 %. On the following two campaigns external influences were present, whose effect has not been clearly formulated so far, which is why the influences are not considered in the PHY model. This decreases the quality of the results, but still reaches an average deviation of maximum 1.27 K.
For the February and September campaign, the RF achieves the highest data quality data. However, the number-based results of the two machine-learning approaches (NN and RF) hardly differ from each other. Nevertheless, the RF is rated as the better model as it corrects the data of the September campaign the best of all three models. However, significant deviations from the reference signal can also be observed with this model, especially in the last approx. 3000 s. Still, the RF reaches the highest PD2MU for both the February and September campaigns. In addition, both the mean and maximum deviation values, as well as the standard deviation, are lowest for the RF results. The additional unknown influence of global radiation results in stronger variations in the RF results, but the results remain on average in the correct order of magnitude and do not deviate by an offset, as both the PHY model and the NN do.

Whether the good results of all three models are attributed to a very specific model solution in each case is examined in the following chapter by means of a transfer to another data set of another vehicle.

The detailed tracking for the differences between the PHY model and the machine learning models is limited since machine learning models have the property of a black box. Therefore, the individual decisions of the machine learning models are not or only very limited traceable for a developer. Nevertheless, the overall very similar results for all three models can be attributed to the similar input variables considered in the models (cf. Table 46 in Appendix A.10). The five parameters "DeltaTemp", "Vehicle speed", "Engine on", "Temperature corrected for effects induced in stationary situations" and

"EIT" are used in all three models. The weighting of the influences is identical for the two machine learning models for the first two input variables and is largest in each case.
The NN considers the smallest set of parameters with six input variables and is the only model that uses the "Duration EIT". However, no significant improvements due to this parameter can be seen in the results compared to the PHY model and the RF. In this context, it is noticeable that the NN does not consider local effects. The influence of the change of altitude on the model result improves the result minimally with the available data. Hence, this input variable is neglected in order not to generate overfitting of the network (cf. Chapter 4.2.3). However, since both the PHY model and the RF consider the same signal of change in altitude, the better results of the RF in September cannot be explained by this signal. Since the RF does not consider a single signal that is neglected in the PHY model, the differences between the two models cannot be explained by the input data but only by the processing of the data and thus to the model itself.
However, since the differences are of small magnitude and the PD1MU of all three models differs by less than 2 percentage points, the first five input variables are the most important influencing variables for the present data set. Nevertheless, local effects should be considered to allow for a change in the measured temperature signal due to local effects instead of changing them. In the available training data, the NN does not detect a direct correlation between a temperature change and a change in altitude. Still, the evaluation of the input variables to be considered may be different on other data sets, wherefore an extension of the considered input parameters by local effects is useful.
In addition, the data shows an influence due to global radiation. The influence by the global radiation is additionally dependent on the paint color of the respective vehicle. Due to a lower albedo and higher immissions for black surfaces, exemplary the effect will be stronger for vehicles with black paint than for vehicles with white paint. However, due to the limited data available, the exact influence of global radiation cannot be quantified, wherefore it cannot be considered in the PHY model. For both machine learning models, the data basis is also too small to determine and consider a clear influence by global radiation.

### 5.4.7 Transferability of models

To test the transferability of the three models, all models, which are originally developed using TC1 data, are transferred to TC2. Neither an adjustment, nor a new training on the TC2 data takes place.

Figure 26 illustrates the results of the transferability test of the models to TC2. The y-axis of each of the three subplots represents the same section of air temperature in °C. Furthermore, all subplots share the same x-axis and therefore show the identical data basis, composed of all available data for TC2 recorded during the mobile measurement campaigns. Note that there are 94944 s and thus 2.7 times as much data as in Chapter 5.4.6 for the evaluation of TC1. Since all data from TC2 can be used for evaluation and no data is needed for training or validation of the models, the amount of data is larger for TC2 than for TC1, although TC2 had intermittent failures during the mobile measurement campaigns.

The raw signal shows significant deviations from the reference signal of the MMU in multiple time periods. The deviation which lasts the longest is

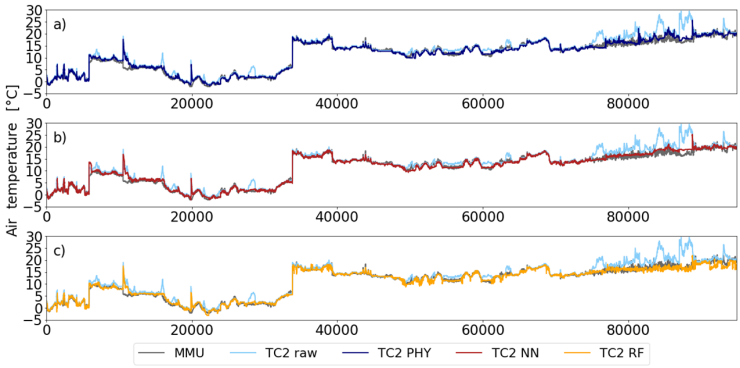


Figure 26: Time course of different model results for all data collected with TC2 during mobile measurement campaigns. All three subplots show the identical data basis with the raw signal of TC2 in light blue and reference data of MMU in dark grey. Additionally, results of the physical model (PHY) are plotted in a), results of the neural network (NN) in b) and results of the random forest (RF) in c).

recorded in September (from approx. 75000-90000 s). The previously occurring deviations are of shorter duration, but comparable magnitude.

The physical model (cf. Figure 26a)) compensates well for most deviations between the vehicle-based raw signal and the reference and runs in good approximation to the reference signal. However, isolated peaks, such as at about 10000 and 20000 s, are not sufficiently compensated by the PHY model. Towards the end of the data recording (approx. 75000-90000 s) the physically corrected signal shows stronger fluctuations than the reference signal. Nevertheless, apart from individual fluctuations, this data is also in good approximation to the reference signal.

The neural network results (cf. Figure 26b)) strongly resemble those of the physical model. This model also reproduces well both the time course and the magnitude of the reference signal, but like the PHY model exhibits difficulties at the peaks around 10000 and 2000 s. In addition, the NN provides temperature values too high for a sudden temperature rise at approx. 7000 s. In contrast to the PHY model, the NN compensates the fluctuations at the end of the data recording (approx. 75000-90000 s) with less fluctuations in the corrected signal but has a slight offset to the reference signal.

The random forest (cf. Figure 26c)) is also able to correct the raw TC2 data and provide results comparable to those of the other two models. Like the other two models the RF is not able to compensate the peaks at 10000 and 20000 s. The RF can reproduce the sudden temperature rise at approx. 7000 s comparable to the PHY without outputting excessive temperature values like the NN. Especially, the last conspicuous time interval (approx. 75000-90000 s) is best corrected by the RF. The RF corrected signal shows stronger fluctuations than the NN corrected signal but has no offset to the reference signal. However, in the subsequent final seconds the RF results show insufficient temperature values, as it overcompensates for the increase in the raw data.

Table 11 confirms the statements made on the basis of Figure 26. All three models reduce the mean deviation between the vehicle-based signal and the reference signal of the MMU by at least 88 %. The mean deviations achieved for TC2 are even lower than the ones for TC1 and therefore also significantly lower than the measurement uncertainty. In terms of maximum deviation, the results for all three models differ by only 0.79 K but are higher than the maximum deviations for TC1. The percentage reduction of the standard deviation by the

| TC2 | All three mobile measurement campaigns combined | | | |
|---|---|---|---|---|
| | Raw data | PHY | NN | RF |
| Data | 94944 seconds | | | |
| Ø Dev [K] | 1.14 | 0.13 | 0.11 | -0.11 |
| Max Dev [K] | 13.23 | 11.94 | 11.15 | 11.72 |
| Std dev [K] | 1.89 | 0.99 | 0.99 | 1.01 |
| PD1MU [%] | 53.28 | 68.85 | 67.99 | 68.78 |
| PD2MU [%] | 70.95 | 88.48 | 88.86 | 87.97 |

Table 11: Tabular overview of the correction results of TC2 for all three models physical (PHY), neural network (NN) and random forest (RF), that are developed on TC1 and transferred to TC2 for the mobile measurement campaigns combined.

models is lower for TC2 than for TC1.
The lower average deviation of the raw data of TC2 to the reference signal compared to the raw data of TC1, already indicates that the data of TC2 are closer to the reference signal. This is confirmed by the proportion of valid data (PD1MU), which is 53.28 % for TC2 (PD1MU of TC1 = 50.71 %). The proportions of valid data of the corrected signals for all models are also higher for TC2 than for TC1. The percentage increase in PD1MU comparing the corrected and raw data differs by a maximum of 2 % for TC2 from TC1 values. Thereby, all models achieve a comparable increase in quality on TC2 as on TC1.
The PD2MU is higher for the TC2 data than for the TC1 for all three models, but the percentage increase compared to the raw data is up to 8 % less than for the TC1. Thus, more of the TC2 data is in the range of twice the measurement uncertainty than for TC1, but the proportion of raw data for TC2 was already 11 percentage points higher.

In summary, the models developed on TC1 can be transferred to TC2 without a loss of quality. No adjustments to the models are necessary. The results indicate that none of the three models developed on TC1 are that specific that they cannot be applied for a different vehicle. However, in order to make this statement generally valid, a sufficient data basis is missing in this work. Both considered vehicles TC1 and TC2 are station wagons, but they differ in exterior paint color and in the installed engine. The different engines installed influence the performance of the vehicle, but more relevant for the correction are the influences on the temperature measurement due to the different waste heat and the different cooling system. In addition, the data basis for both vehicles was recorded during the same measurement campaigns. For a generally valid

statement about the transferability of the models, they must be tested on further data, on different vehicles with different engine constellations and paint colors and on further measurement campaigns.

## 5.5 Relative humidity

### 5.5.1 Stationary measurement campaigns

To correct the vehicle-based relative humidity data, only the data from the first stationary measurement campaign in November 2020 (288789 s) are used (cf. Chapter 5.1). During the second stationary measurement campaign, the parking heater of the TC1 was temporarily switched on to protect the additional electronics on board of the vehicle from damage by the cold. However, the parking heater, which is positioned next to the engine block, exerts a significant influence on the adjacent relative humidity measurement. The RH signal is therefore strongly distorted, which is not due to effects induced in stationary situations but only to the parking heater. Thus, this data is not suitable for determining the sensor uncertainty induced in stationary situations.

Chapter 4.3 already explains that a pure offset correction of the relative humidity to compensate for the effects induced in stationary situations is not useful due to the design of the sensor. The approach chosen instead to transfer the relative humidity is explained in detail in Chapter 4.3.1 and 4.3.2. To illustrate the difference between the two approaches, Figure 47 and Table 47, both in Appendix A.11, show results for these approaches.

The offset used for the pure offset correction (version I) accounts to 29.93 %rH, and thereby covers almost $\frac{1}{3}$ of the range of values accepted for the relative humidity. The second approach (version II) transfers the raw RH value by using equation 22 and then corrects it by an offset of 1.52 %rH.

The difference between both correction versions is significant. A pure offset correction (version I) cannot reproduce the course of the reference signal and thus does not provide representative results.

The transferred RH signal (version II) runs in good approximation to the reference signal and only shows deviations temporarily. Since these are individual

time windows, the deviations are not continuous and thus cannot be clearly attributed to effects induced by sensor technology or location of the vehicle-based sensor. Instead, the deviations are caused by other, external influencing parameters.

In addition to Figure 47, Table 47 reinforces the recommendation to use the approach of version II rather than version I to correct for effects induced in stationary situations. The data portion outside twice the measurement uncertainty to the reference (13 %) is due to the previously described deviations caused by external influences.

As described above, the correction of the raw signal can be separated into two steps. Table 47 in Appendix A.11 illustrates that the first correction step causes most of the quality improvement. This allows the proportion of valid data to be raised from 0.00 % to 53.44 %, while the mean deviation is reduced. The following offset correction by the mean deviation of 1.52 %rH further increases the PD1MU to 69.96 %.

### 5.5.2 Mobile measurement campaigns

Overall, less data is available for the relative humidity evaluation than for temperature, which is due to the limited use of the reference instrument at temperatures below freezing and during snowstorms (cf. Chapter 5.1). Excluding the February campaign and the first day of the January campaign, there are 147897 s available, which corresponds to a reduction of about 28 %.

Figure 27 displays only the section of the data generated during the mobile measurement campaigns where the reference values of the MMU are trustworthy and concurrent data for TC1 and the reference MMU is available (147897 s). On the y-axis the relative humidity is shown in %. The horizontal red line corresponds to an RH value of 100 % and thus indicates saturation of the air. RH values above 100 % are only achievable under laboratory conditions, since the water particles in the real atmosphere condense at a saturation of 100 %, which prevents the relative humidity from increasing further.

Temporarily, the transferred signal shows significant deviations from the reference signal of the MMU and increases to values above 175 %rH. Since values above 100 % correspond to supersaturated air and are not found to this extent

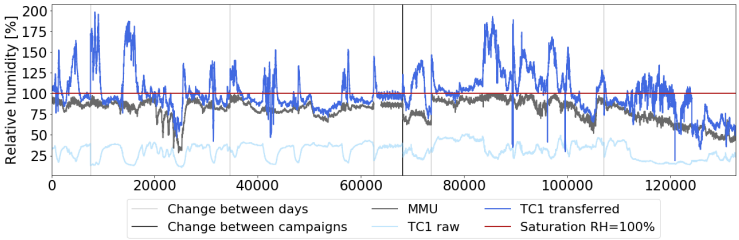


Figure 27: Time course of three different relative humidity (rH) signals. Reference signal of MMU in grey, vehicle-based raw signal of TC1 in light blue and transfer with additional offset correction in blue. Data basis built by sections of January campaign and all data from September campaign.

in the real atmosphere, both the high RH values and the significant deviations from the reference illustrate the influence of effects induced by vehicle motion. Since the transfer is developed on the data of the stationary measurement campaign, neither the calculated relative humidity, nor the input parameters are corrected for effects induced by vehicle motion. The input parameters for the transfer, the BCM1 temperature, the RH measured inside the MuFu sensor, and the dew point temperature measured inside the MuFu sensor, are each filtered only for peaks and measurement range. Therefore, the previously performed correction must be extended by a correction of the effects induced by vehicle motion.

To correct effects induced by vehicle motion, four different approaches are implemented and tested. Chapter 5.5.3 shows results of two different physical models, Chapter 5.5.4 of a neural network, and Chapter 5.5.5 of a random forest.

### 5.5.3 Physical model A and B

The main difference between the two physical models PHY-A and PHY-B is the order of the performed corrections. PHY-A first performs the transfer of the raw signal (cf. "TC1 transferred" in Figure 27) and then corrects the transferred RH signal for effects induced by vehicle motion. Conversely, PHY-B first corrects the input variables air temperature at BCM1 sensor as well as

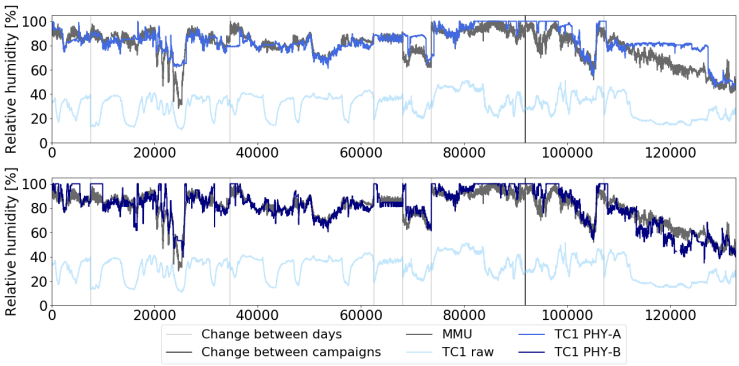


Figure 28: Time course of different relative humidity (RH) signals for data of January and September campaign. Reference signal of MMU in grey, vehicle-based raw signal of TC1 in light blue. Change between campaigns and days during a campaign are marked as vertical lines. Upper plot: Result of physical model PHY-A plotted (blue). Lower plot: Result of physical model PHY-B plotted (dark blue).

RH and dew point temperature measured inside the MuFu sensor for effects induced by vehicle motion and subsequently transfers the RH based on the corrected signals.

Results of both PHY models are shown in Figure 28.

The results of the PHY-B model (cf. Figure 28, lower plot) show a higher number of peaks and events of saturation (RH=100 %) over the complete data course. This is especially observable during the January campaign. The PHY-A model (cf. Figure 28, upper plot) reproduces the reference signal slightly better during the January campaign. Nevertheless, the PHY-A model shows weaknesses at significant changes of the RH signal in short timeframes, e.g., at approx. 35000 s and at the beginning of the third mapped measurement day of the January campaign. Here, the PHY-A model does not reflect the change strong enough or not at all. The PHY-B model tends to handle these situations better.
For the September campaign, the time course of the PHY-B signal is closer to the reference than the PHY-A signal. PHY-A has excessive RH values, especially on the first day of the September campaign, while PHY-B is a good approximation of the reference. Another weakness of the PHY-A model

| TC1 | Mobile measurement campaign 1 Jan 2020 | | | Mobile measurement campaign 1 Sep 2020 | | | All three mobile measurement campaigns combined | | |
|---|---|---|---|---|---|---|---|---|---|
| | Raw | PHY A | PHY B | Raw | PHY A | PHY B | Raw | PHY A | PHY B |
| Data | 83309 seconds | | | 64588 seconds | | | 147897 seconds | | |
| Ø Dev [K] | -50.33 | -0.69 | 0.65 | -47.30 | 7.35 | 1.41 | -49.00 | 2.82 | 0.98 |
| Max Dev [K] | 83.20 | 35.00 | 26.20 | 79.00 | 30.53 | 33.00 | 83.20 | 35.00 | 33.00 |
| Std dev [K] | 11.25 | 5.89 | 6.20 | 12.38 | 7.71 | 7.39 | 11.85 | 7.84 | 6.75 |
| PD1MU [%] | 0.00 | 78.36 | 72.97 | 0.00 | 42.04 | 57.75 | 0.00 | 62.50 | 83.34 |
| PD2MU [%] | 0.03 | 93.89 | 91.37 | 0.00 | 69.74 | 87.06 | 0.02 | 83.34 | 89.50 |

Table 12: Tabular overview of the two different physical models to correct the relative humidity signal during mobile measurement campaigns.

can be observed for the data of the last day of the September campaign. Here, the RH values of the PHY-A model remain almost constant while the reference values decrease continuously. The PHY-B model also has difficulties in accurately reproducing the trend and temporarily shows insufficient RH values, but reproduces the decreasing trend significantly better.

Table 12 lists the numerical values of the evaluation of the data of the January and the September campaign, as well as both campaigns together. Considering both campaigns (147897 s), it can be observed that the results of the PHY-B model provide higher quality than those of the PHY-A model.

The mean deviation is significantly reduced for the 147897 s by both PHY models. For the individual consideration of the two measurement campaigns, the respective percentage decreases are in the same order of magnitude. Only the results for PHY-A of the September campaign show smaller improvements due to the deviations on the last day.
The maximum deviation of all 147897 s is reduced by 58 % and 60 % for the PHY-A and PHY-B model, respectively. The reduction in maximum deviation clearly shows an increase in quality of both physical models, but outliers are still present in the corrected data. The overall standard deviation of both models is reduced by 33 % and 43 % for PHY-A and PHY-B, respectively. Particularly notable is the increase in quality by both physical models regarding the proportion of data that can be assumed to be valid (PD1MU). The PD1MU shows a significant difference between both PHY models with a difference of almost 21 percentage points. The PD2MU is also higher for PHY-B.

Overall, the PHY-B model provides better results than PHY-A. For the further evaluations and the comparison to the machine learning approaches as well as in the transferability test of the models, the PHY-B model is therefore used as the physical model.

Reasons for the better results of PHY-B are due to the two different approaches of the models. The results of correcting the input variables for the effects induced by vehicle motion (PHY-B) are more representative than correcting the calculated RH value (PHY-A). The errors that occur during the correction of the input values have a smaller consequence than the errors that occur during the correction of the output value of the RH. For more complex corrections of both input and output values, PHY-A might give better results. Based on the simplified corrections used here, the PHY-B approach is more expedient.

### 5.5.4 Neural network

In contrast to the neural network for temperature, the data for training, validation, and testing is randomly selected for the relative humidity neural network due to the underlying pattern identified in Chapter 4.3.4. Therefore, the results of the NN for relative humidity are distributed as single second values over the whole data basis of the January and September campaign. The results are shown in Figure 29.

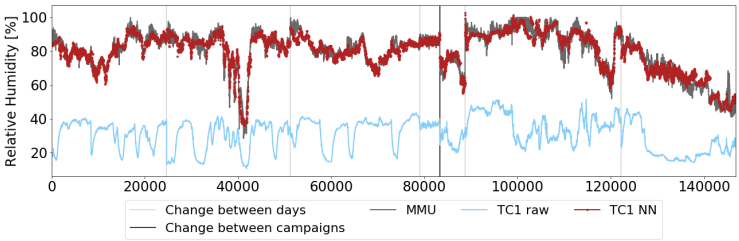


Figure 29: Time course of different relative humidity (RH) signals for data of January and September campaign. Reference signal of MMU in grey, vehicle-based raw signal of TC1 in light blue and vehicle-based signal corrected by neural network (NN) in red. Change between campaigns and days are marked as vertical lines.

| TC1 NN | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|
| | Raw data | NN | Raw data | NN | Raw data | NN |
| Data | 13833 seconds | | 10641 seconds | | 24474 seconds | |
| Ø Dev [%rH] | -50.44 | -0.30 | -47.31 | 0.08 | -49.08 | -0.13 |
| Max Dev [%rH] | 82.60 | 16.46 | 78.90 | 22.41 | 82.60 | 22.41 |
| Std dev [%rH] | 11.20 | 3.13 | 12.38 | 3.99 | 11.83 | 3.54 |
| PD1MU [%] | 0.00 | 92.47 | 0.00 | 85.33 | 0.00 | 89.36 |
| PD2MU [%] | 0.01 | 99.16 | 0.00 | 98.29 | 0.00 | 98.78 |

Table 13: Tabular overview of the neural network results (NN) to correct the relative humidity signal during mobile measurement campaigns.

Most of the NN corrected data points are in close approximation to the reference signal and reproduce well both the course and the magnitude of the reference signal. Sudden changes in RH values can be well reflected by the NN, as well as the last day of the September campaign.

The good results of the NN are confirmed by the low mean deviation as well as the low standard deviation (cf. Table 13). The mean deviation between vehicle-based and reference signal can be reduced by 99.7 % to -0.13 %rH. Since the measurement uncertainty accounts to 5.5 %rH, the mean deviation is significantly lower, which indicates a quite specific solution to the underlying task. In Chapter 5.5.7 the transferability of the net is tested and thereby it is evaluated whether the solution is specific to TC1.
No significant differences occur between the results of TC1 for the two campaigns considered.
The standard deviation over all 24474 s is reduced by 70 %. The percental decrease in standard deviation differs by 8 percentage points between the two measurement campaigns (80 % for January, 72 % for September).
After correction based on the NN, 89.36 % of the data can be accepted as valid, whereas before the PD1MU is 0.00 %. The percentage of data within twice the uncertainty increases significantly, too. The behavior is comparable for both campaigns, with the PD1MU slightly higher for the January data than for the September data (7 percentage points difference).

The correction of the relative humidity using the trained neural network thus provides a significant increase in quality compared to the raw data.

### 5.5.5 Random forest

Comparable to the neural network (NN), the data selection for the random forest (RF) for the relative humidity also takes place randomly. Figure 30 shows three relative humidity signals. The very good agreement between the RF corrected and the reference signal is noticeable. The RF results cover sudden changes in the reference signal as well as the data from the last day of the September campaign. No significant deviations from the reference signal can be detected at any time in the present data.

Table 14 lists the RF results both subdivided by measurement campaigns and both campaigns combined, confirming the results from Figure 30.

Before correction, the mean deviation takes about half of the range of values for relative humidity (-49.08 %rH). When considering the measurement campaigns individually, the average deviation does not increase to more than 0.01 %rH. The low mean deviation of RF corrected data implies a RF specific to the underlying data set. To test this, the transferability of the model is evaluated in Chapter 5.5.7.
The maximum deviation is reduced from above 82 %rH to below 9 %rH by the RF.
Averaged over all available 24474 s, the standard deviation is reduced by 91 %.
A significant difference between the campaigns cannot be identified.
The high quality of the RF results is confirmed by the PD1MU and PD2MU.
The RF raises the percentage of valid data from 0.00 % to 99.86 %. The

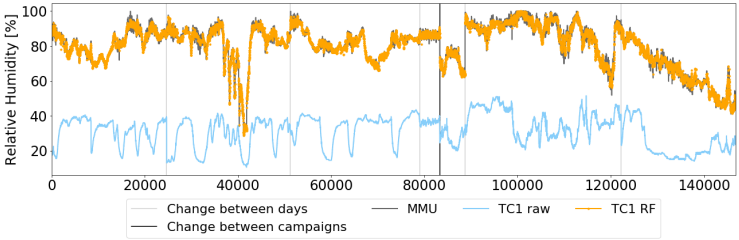


Figure 30: Time course of different relative humidity (RH) signals for data of January and September campaign. Reference signal of MMU in grey, vehicle-based raw signal of TC1 in light blue and vehicle-based signal corrected by random forest (RF) in yellow. Change between campaigns and days during a campaign are marked as vertical lines.

| TC1 RF | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|
| | Raw data | RF | Raw data | RF | Raw data | RF |
| Data | 13833 seconds | | 10641 seconds | | 24474 seconds | |
| Ø Dev [%rH] | -50.44 | -0.01 | -47.31 | 0.01 | -49.08 | 0.00 |
| Max Dev [%rH] | 82.60 | 8.24 | 78.50 | 8.12 | 82.60 | 8.24 |
| Std dev [%rH] | 11.20 | 1.06 | 12.38 | 0.97 | 11.83 | 1.02 |
| PD1MU [%] | 0.00 | 99.86 | 0.00 | 99.86 | 0.00 | 99.86 |
| PD2MU [%] | 0.01 | 100.00 | 0.00 | 100.00 | 0.01 | 100.00 |

Table 14: Tabular overview of the random forest results (RF) to correct the relative humidity signal during mobile measurement campaigns.

PD2MU is 100 %, both for all data together and for both campaigns considered separately.

The RF successfully corrects the data of the TC1 and thus leads to a significant increase in quality.

### 5.5.6 Comparing different model results

To compare the three implemented models, all three models are evaluated on the same data set. The total number of data points at which results for all three models are available is 24474 s. For the physical model, results are available at 147897 s, but only the 24474 s that are concurrent with the NN and RF data points are considered for the following evaluation.

Figure 48 in Appendix A.11 shows the results of all three models (PHY, NN, RF), as well as the vehicle-based raw data and the reference signal. Figure 31 plots the deviations of the respective signals to the reference. Note the different scales of the y-axis for a) and the other three subplots.

All three models achieve a significant improvement of the raw signal (cf. Figure 48) and thereby a significant reduction of anomalies. Nevertheless, differences between the three models and their deviations can be identified (cf. Figure 31).
The physical model (blue) shows the most pronounced deviations from the reference signal. In both January and September, the PHY signal is temporarily constantly 100 %, although the reference signal does not reach saturation. On

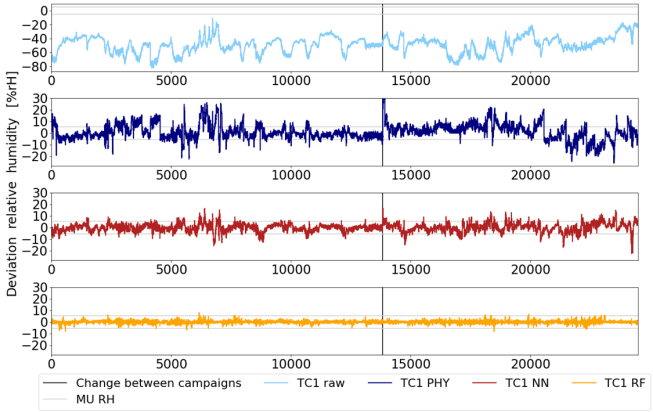


Figure 31: Time course of anomalies of vehicle-based relative humidity signals. Four vehicle-based relative humidity (RH) signals are considered: TC1 raw in lightblue, TC1 corrected based on physical model (PHY) in darkblue, TC1 corrected based on neural network (NN) in red, TC1 corrected based on random forest (RF) in yellow. Anomalies are shown for all three mobile measurement campaigns in January, February, and September 2020. Note, the y-axis section varies for anomalies of raw data (90 %rH) and corrected data (60 %rH). Therefore, the deviations of the raw data appear smaller than they would be if mapped at the same scale as the corrected data.

the last day of the September campaign, the PHY signal additionally shows significant deviations lower than the reference signal.
Both machine-learning models, the NN as well as the RF, reproduce the course of the reference signal better than the physical model. However, the reliability of the NN is reduced in time periods where the reference shows short-term and strong fluctuations. The RF is the only model that reproduces each of the mapped variations of the reference signal and does not show significant deviations from the reference at any point in time.

It is noticeable that the anomalies of the PHY model (Figure 31b)) are both larger and more frequent compared to the machine-learning models (Figure 31c) and d)). RF anomalies are consistently the smallest.

Table 15 confirms the preliminary results of both Figures. All models, considering all 24474 s, reduce the mean deviation between vehicle-based data and reference data by at least 97 %. The maximum deviation is also reduced by

the correction of all three models. However, noticeable differences do occur. In addition to the lowest mean deviation, the RF also has the lowest maximum deviation of the three models.
The same order of the model results for the consideration of the standard deviation. The PHY model reduces the standard deviation by 43 %, the NN by 70 %, and the RF by 91 %. If the two measurement campaigns are considered separately, the result is the same: The RF achieves the strongest reduction of the standard deviation.

The differences between the three models become particularly clear when evaluating the portion of valid data. The PHY model raises the PD1MU by 66.40 percentage points, the NN by another 22.96 percentage points (89.36 %). Still, with 99.86 % the RF achieves the highest percentage of valid data. The RF is the only one of the three models to achieve a PD1MU of constantly more than 99 % for both measurement campaigns. Conversely, the proportion for the PHY and NN models differs by 15 and 7 percentage points respectively in the comparison between the January and September campaign.
The PD2MU reconfirms the best quality of the RF results. For the RF corrected data, 100 % of all data is within twice the measurement uncertainty around the reference. Therefore, no outliers are present in the RF corrected data.

The strong concentration of the RF anomalies around 0 %rH and the small scattering of the anomalies is clearly shown in Figure 49 (yellow) in Ap-

| TC1 | Mobile measurement campaign 1 Jan 2020 | | | | Mobile measurement campaign 3 Sep 2020 | | | | All three mobile measurement campaigns combined | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Raw data | PHY | NN | RF | Raw data | PHY | NN | RF | Raw data | PHY | NN | RF |
| Data | 13834 seconds | | | | 10641 seconds | | | | 24474 seconds | | | |
| Ø Dev [%rH] | -50.44 | 0.64 | -0.29 | -0.01 | -47.37 | 1.49 | 0.08 | 0.01 | -49.08 | 1.01 | -0.13 | 0.00 |
| Max Dev [%rH] | 82.60 | 29.80 | 16.58 | 8.24 | 78.90 | 32.93 | 22.41 | 8.12 | 82.60 | 32.93 | 22.41 | 8.24 |
| Std dev [%rH] | 11.20 | 6.22 | 3.13 | 1.06 | 12.38 | 7.42 | 3.99 | 0.97 | 11.83 | 6.78 | 3.54 | 1.02 |
| PD1MU [%] | 0.00 | 72.89 | 92.46 | 99.89 | 0.00 | 57.95 | 85.33 | 99.86 | 0.00 | 66.40 | 89.36 | 99.86 |
| PD2MU [%] | 0.01 | 91.31 | 99.15 | 100.00 | 0.00 | 86.63 | 98.29 | 100.00 | 0.00 | 89.28 | 98.78 | 100.00 |

Table 15: Tabular overview of the results of the raw data and all three tested models in comparison. The data basis is built by points of time with results available for all three models: physical model (PHY), neural network (NN) and random forest (RF).

pendix A.11. This explains the associated high PD2MU of the RF model from Table 15. Since the scales of the axes are identical for all four subplots in Figure 49, the narrow distribution of the anomalies of the RF is particularly striking. The anomalies of the NN (cf. Figure 49 red) are also distributed around 0, but over a wider span than the RF anomalies. This explains the lower PD2MU for the NN compared to the RF. The histogram of the anomalies of the PHY model (cf. Figure 49 blue) shows not only a wider range of scatter but also an asymmetric behavior of the positive and negative anomalies.

In summary, the RF represents the best of the three implemented relative humidity correction models. It is noticeable that both machine learning models give significantly better results for this meteorological parameter than the physical model. This is due to the multiple unknowns that are approximated and assumed for the physical model or are completely disregarded. For the machine learning models, the physically explainable correlations of the influencing effects are irrelevant. However, there are also significant differences between these two approaches, making the RF the model with the greatest impact and the highest quality result for relative humidity correction.

The explanation of the qualitatively different results between NN and RF is due to the different structure of the models. The features considered by both models are identical except for one input variable. The only difference is the signal of the engine temperature. The NN considers this signal, whereas the RF does not. When neglecting the engine temperature in the NN, the results become worse (cf. Chapter 4.3.4). The better results of the RF are therefore not due to the omission of this signal. Additionally, since both models are trained on the identical data set, the data set cannot be the explanation for the different results.
For the present use case, a correction based on decision trees is therefore more powerful than a correction based on multiple layers and weighted activations. It is likely that the result would also be similar for a much larger data set. However, this is not guaranteed, since the performance of the models depends on the underlying data. However, it cannot be excluded that the result changes when the data set grows by several orders of magnitude.

For the PHY model, multiple assumptions are made, such as identical pressure and no change in the water content of the air parcel, without being able to verify them. In addition, the influence of vehicle motion and times with low velocities on the RH measurement is not precisely quantified, wherefore a simplified

correction of these influences takes place. A more detailed correction based on quantified correlations would increase the quality of the PHY results, but is not feasible within the scope of this work.

### 5.5.7 Transferability of models

To test the transferability of the three models, they are deployed to the data of technology carrier TC2 without any modifications. Since no data is needed for a new training or validation of the models, all 65864 s at which concurrent data for the TC2 and the reference is available, can be used for evaluation. Data of the February campaign is also not considered here, the explanation being identical to TC1.

The results of all three models after application to the TC2 data are shown in Figure 32.

The physical model (cf. Figure 32a)) does not give satisfactory results for the correction of the relative humidity of the TC2. Up to approx. 20000 s, the signal, apart from individual outliers at the beginning of the data, runs in

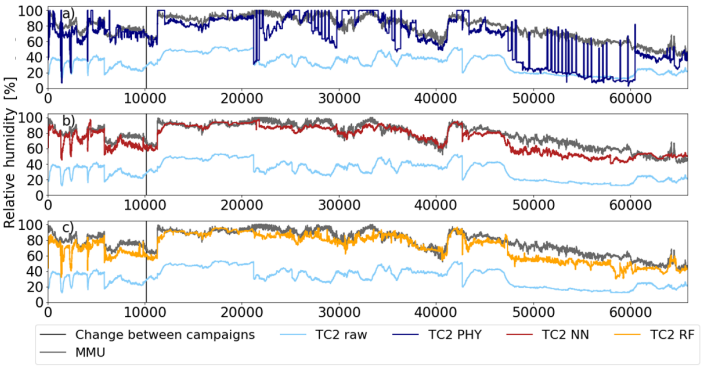


Figure 32: Time course of different model results for all data collected with TC2 during mobile measurement campaigns. All three subplots show the identical data basis with the raw signal of TC2 in light blue and reference data of MMU in dark grey. Additionally, results of the physical model (PHY) are plotted in a), results of the neural network (NN) in b) and results of the random forest (RF) in c).

good approximation to the reference. Subsequently, the PHY signal fluctuates strongly back and forth between the raw signal and the reference value. In between, the amplitudes become smaller, but increase again towards the end of the present data (from approx. 48000 s). Occasionally, the PHY signal is in good approximation to the reference, but the majority of the data shows significant fluctuations and deviations from the reference.

The neural network (cf. Figure 32b)) provides better results compared to the PHY model. The fluctuations and deviations of the NN signal are significantly lower. Nevertheless, the NN model also exhibits fluctuations at the beginning of the data. In the further course, the NN model cannot reproduce all fluctuations and tendencies of the reference but shows significant deviations from the reference only temporarily. In the section between approx. 48000 and 60000 s an offset of several %rH between the NN signal and the reference signal can be seen.

The RF results (cf. Figure 32c)) behave similarly to the NN results. At the beginning, the RF results also show isolated deviations from the reference, just like the NN. The entire course of the RF signal is close to the reference signal but fluctuates more than the NN signal. In addition, similar to the NN, the RF results deviate significantly from the reference signal occasionally, especially in the range between approx. 48000 and 60000 s. For this section, an offset between the RF signal and the reference can be observed.

Table 16 reinforces the results described earlier. It is notiveable that the mean deviation is reduced by all three models. The NN achieves the lowest mean deviation, with 5.23 %rH this is almost equal to the measurement uncertainty

| TC2 | All three mobile measurement campaigns combined | | | |
|---|---|---|---|---|
| | Raw data | PHY | NN | RF |
| Data | 65864 seconds | | | |
| Ø Dev [%rH] | -45.99 | -13.14 | -5.23 | -7.59 |
| Max Dev [%rH] | 82.90 | 73.21 | 41.84 | 50.04 |
| Std dev [%rH] | 10.70 | 19.43 | 7.09 | 7.19 |
| PD1MU [%] | 0.00 | 26.48 | 54.54 | 40.78 |
| PD2MU [%] | 0.00 | 61.08 | 76.70 | 69.68 |

Table 16: Tabular overview of the correction results of TC2 for all three models physical (PHY), neural network (NN) and random forest (RF), that are developed on TC1 and transferred to TC2 for the mobile measurement campaigns combined.

of 5.5 %rH. The highest mean deviation is present for the PHY results but is still only 2.4 times the measurement uncertainty.
In terms of maximum deviations, the NN results also reach the lowest values. In contrast to both ML models, the PHY model can reduce the maximum deviation by less than 10 percentage points.
The standard deviation illustrates the difference between the models. On one hand, the PHY model increases the standard deviation by almost 82 %. This worsening of the results is due to the strong fluctuations within the PHY signal (cf. Figure 32a)). The machine-learning based models on the other hand, achieve a reduction of 34 % and 33 % for the NN and RF, respectively.
Nevertheless, all three models show the ability to increase the proportion of valid data. The highest increase however, is achieved by the NN, for which over half of the corrected data of TC2 is valid.
For data within twice the measurement uncertainty, the ranking of the three models is the same. The NN has the most data within twice the measurement uncertainty, followed by the RF.

Overall, both machine learning models achieve better results than the physical model even after transfer to another vehicle. In comparison, the NN is more reliable for TC2 than the RF. In contrary, for TC1, the RF achieves the best results. Noteworthy, the quality of the results is significantly reduced by the transfer to the TC2 compared to the TC1. For TC1, the best model achieves a PD1MU of 99.86 %, whereas the best model for TC2 achieves only 54.54 %. On one hand, this can be caused by the underlying data set for TC2 being harder to correct for the models. On the other hand, the decreased quality of the results can also be caused by additional parameters influencing the RH of the TC2, which have not been considered for the TC1.

However, without modification and adaptation of the tested models, a transfer for relative humidity is not possible without loss of quality. Nevertheless, especially both ML models increase the quality of the raw data significantly. Even better results can be achieved for adjusting the models to this vehicle.

## 5.6 Global radiation

### 5.6.1 Stationary measurement campaigns

The basis for the evaluation of the stationary data is the measurement campaign of TC2 next to the synop station "Weihenstephan". Due to the temporal resolution of 1 min for the data of the synop station used as reference, a direct comparison between the second-by-second vehicle data and the reference data is not reasonable. A direct comparison would assign too much importance to individual peaks in the vehicle-based signal. Therefore, the vehicle data are averaged over 1 min in order to achieve an, albeit artificial, nevertheless identical temporal resolution. Therefore, 6130 min of data for the concurrent values of the synop station and the TC2 are available for the evaluation.

Figure 33 represents the time course of the different global radiation signals. The measured values are too high (cf. Figure 33a)) and are particularly noticeable in combination with the associated anomalies outside the measurement uncertainty (cf. Figure 33b)). By correcting the vehicle-based signal using the

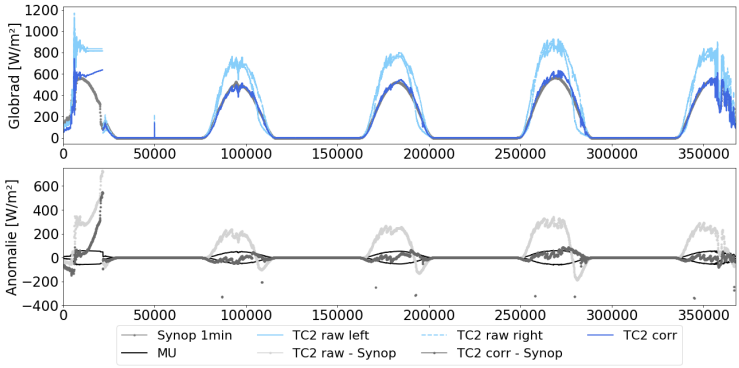


Figure 33: Time course of a) different global radiation signals for all data collected with TC2 during stationary measurement campaign next to synop station "Weihenstephan" and b) the corresponding anomalies of the vehicle-based signals to the reference. The raw signal is plotted in light blue, the corrected signal in blue and the reference in grey in a). The respective anomalies are shown in light grey for raw signal and dark grey for corrected signal in b).

| TC2 | Raw data | | Corrected data | |
|---|---|---|---|---|
| | All | Day only | All | Day only |
| Data [sec] | 6130 | 3039 | 6130 | 33039 |
| Ø Dev [W/m²] | 59.93 | 120.89 | 4.49 | 9.05 |
| Max Dev [W/m²] | 728.00 | 728.00 | 549.10 | 549.10 |
| Std dev [W/m²] | 123.25 | 152.55 | 49.20 | 69.56 |
| PD1MU [%] | 52.89 | 5.00 | 76.64 | 52.88 |
| PD2MU [%] | 55.79 | 10.86 | 84.89 | 69.56 |

Table 17: Tabular overview of results for raw data and corrected data of TC2 for the global radiation signal. The table distinguishes between all data ("All") and only data measured during the day ("Day only"), since the global radiation decreases to zero at night and thus pretends better agreement of the vehicle-based data to the reference if all data is considered.

steps described in Chapter 4.4, it can be approximated to the reference signal and the deviations reduced accordingly. The results of the intermediate steps from Chapter 4.4 are shown in Figure 50 in Appendix A.12.

For evaluating the data, it must be distinguished whether all measured data (6130 min) or only data on which the values of the synop station are greater than zero and are thus measured during the day (3309 min) are taken into account (cf. Table 17). The significant difference between the results of the two data bases is due to the fact that the global radiation drops to zero at night and thus both signals are identical. This simulates a result that is not representative for general vehicle measurements, since these take place almost exclusively during the day. Representative results are therefore only obtained if measurements during the night are excluded. The statement based on vehicle data that global radiation drops to zero at night does not provide any added value.

The "Day only" data from Table 17 illustrate the difference between the raw data and the corrected data, which is already visible in Figure 33. For example, the average deviation between the vehicle-based signal and the reference signal can be reduced by almost 93 % to 9 $W/m^2$.
It is important to emphasize the proportion of valid data, which is 5 % of the "Day only" data before the correction and is raised to 52.88 % by the correction. A likewise significant improvement is found for data within twice the measurement uncertainty (PD2MU), where an increase by a factor of 6.4 is achieved.

Although the vertical angle of incidence of the sun on the sensors is not considered in the tested correction procedure and a simplified weighting of the two sensors is implemented, the data is subjected to a significant quality improvement. Further potential for improvement lies in the two points already mentioned regarding the vertical angle of incidence and weighting.

### 5.6.2 Mobile measurement campaigns

#### Calibration of reference

When transferring the developed correction to the data of the mobile measurement campaigns, the measurement technique of the MMU to be used as a reference must first be calibrated with respect to the measured absolute value of the global radiation. For this purpose, a time section of the mobile measurement campaign in September is used, during which the MMU was located next to the synop station "Weihenstephan", which is used as a reference for the stationary measurement campaigns, for about 30 min. During this time the engine of the MMU is turned off most of the time and the distance to the station is less than 500 m. Therefore, identical environmental conditions can be assumed for both measurements.
Since the synop station outputs data with minute-by-minute resolution, the second-by-second data from the MMU are averaged over one minute to ensure comparability of the two stations.
Empirically, this database yields a factor correction of 0.76 for the MMU data. The results of the calibrated MMU measurements in comparison to the measurements of the synop station "Weihenstephan" are shown in Figure 51 in Appendix A.12.
The only deviations between the calibrated MMU and the synop data, which are larger than the measurement uncertainty, are at the beginning and end of the measurement period (cf. Figure 51b)). In these time periods, the second-by-second values of the MMU fluctuate strongly, which is not observable for the rest of the considered period. At these time points, the MMU was not at standstill, whereby the local effect of shade on the moving measurement equipment causes the fluctuations. Therefore, contrary to the original assumption, the environmental conditions of both measurements for the first and last minutes are not identical, which limits the comparability.

By calibrating the MMU values, the accuracy of the values can be increased to such an extent that they are within the measurement uncertainty and can thus be used as a reference for further measurements.

#### Vehicle-based measurements

Since the global radiation correction was developed on stationary data of the TC2, the TC2 data is also used for the evaluation of the mobile campaigns. The parameter global radiation is of particular interest in the summer months, both because of its influence on other vehicle-based parameters, which is stronger in summer, and for cross-industry applications to forecast the generated power of solar fields. Therefore, only the data from the mobile summer campaign will be considered initially. An overview of all summer campaign data is provided in Figure 53 in Appendix A.12.

It is noticeable that both the reference and the vehicle-based signal fluctuate strongly, especially when compared to the data from the stationary measurement campaign (cf. Figure 50). The large variations are due to the influence of the local effect of shade, for example, by roadside trees and other road users. Due to the shading of the sensor, the measured value drops by several orders of magnitude, only to rise again abruptly when it is directly exposed to the sun. Since the TC2 and the MMU keep a safe distance of several meters from each other during the mobile campaign, a direct comparison between the second-by-second values is not expedient, since these are strongly influenced by shade. In the case of the previously considered parameters air pressure, air temperature and relative humidity, the effect caused by shade is not decisive. Therefore, a comparison of the values every second does not have a falsifying influence on the result.

In addition to the fluctuations, it is noticeable that the vehicle-based signal is significantly below the reference signal in sections (cf. section around 86000 s in Figure 53a)). In comparison to Figure 53b) it becomes apparent that the significant deviations always occur when the sun shines on the vehicle from behind. This can be explained by the fact that the sensor signals taken into account are aligned forward. As soon as the sun shines on the vehicle from behind, the sensor in the windshield is in the shade of the vehicle itself and therefore measures values too low. For the following evaluation, therefore,

only time periods are considered in which the sun falls on the vehicle from the front (0°≤ angle of incidence ≤ 180°).

It can be deduced from both conspicuities observed that a comparison of the second-by-second values is only meaningful if both the TC2 and the MMU are exposed to the identical shade/sun conditions. This cannot be guaranteed while driving. For the September measurement campaign, three stops took place where both vehicles were exposed to identical light conditions (direct sun). Despite a measurement failure of the TC2 during one of these stops, there are 4043 s of concurrent data for the TC2 and the MMU. Comparing these second-by-second values, the mean deviation between the two signals is 27.45 $W/m^2$ with a PD1MU of 27.23 % (cf. Table 48). The percentage of valid data is very low for the fact that the conditions of the evaluated 4043 s are comparable to those of the stationary measurement campaign (cf. Table 17 in Chapter 5.6.1). One explanation could be the different seasons in which both campaigns took place and the associated vertical angle of incidence of the sun on the windshield. Since the sun is lower in February, the vertical angle of incidence on the sensor is also lower than in September.

In addition to the analysis of the second-by-second data during the three stops, data averaged over 1 min from TC2 and MMU is compared. The interval is set to 1 min, since this is the temporal resolution of the synop station and a vehicle with an average speed of 50 km/h covers a distance of about 800 m within 1 min, so that influences of individual shades are compensated for. By averaging the data, the influence of peaks is distributed over the entire period of 1 min and does not occur selectively. This creates comparability between the vehicle-based and the MMU data during driving. The disadvantage of this method is that the minute-averaged values do not represent real global radiation values, since there are lower values in the shade and higher values in the sun. For the September campaign, after averaging over 1 min, 366 min values are available for situation where the sun is shining on the vehicle from the front. The mean deviation increases to 70.12 $W/m^2$ and the proportion of valid data decreases compared to the second-by-second data during standstill (cf. Figure 53c) and Table 48). Figure 53c) shows that the behavior of both minute values is comparable, but the absolute numerical values temporarily have a significant offset from each other.

As an alternative to the values averaged over 1 min, the respective maximum (cf. Figure 53e)) and minimum (cf. Figure 53d)) global radiation values are

determined for time intervals of 1 min. The advantage of this method is that both values are representative, the max value for the section in the sun and the min value for the section in the shade. From this it can also be concluded that the closer the Min value is to the Max value, the less shading was present in the minute under consideration.
For both max and min values, the vehicle-based signal is temporarily in good approximation to the reference signal, for example between approx. 77000 and 80000 s. Before that, however, between 74000 and 76000 s, both the min and max values deviate strongly from the respective reference data. The agreement of the Min data is lower than for the Max data (cf. Table 48). One explanation for this could be the different mounting positions of the sensor on the TC2 and the MMU. In order to shadow the TC2 sensor mounted in the windshield, even smaller vehicles of the surrounding traffic can be sufficient, whereas shadowing of the sensor mounted on the roof of the MMU only occurs, for example, through large trees. The frequency of shading of the TC2 sensor is thus higher and it records shades that are not noticed by the MMU sensor, whereby the Min values differ.

Of the three synthetic 1 min data, the Max values are in best agreement with the reference with 15.03 % valid data, but there are strong deviations in sections, too. Since there is no continuous offset, it can be assumed that there is at least one other variable parameter which has a significant influence on the measurement. An influence of the vertical angle of incidence of the sun could explain, for example, why the correction of the data of the mobile campaign in September results in a significantly worse outcome than the data of the stationary campaign in February.
The fact that sunlight from behind the vehicle has a significant influence on the measurement is already mentioned at the beginning of this chapter. On the available data basis of the stationary campaign in February, no correction can be developed for this scenario since the sun was continuously shining from the front onto the vehicle. Further data is therefore necessary.
In addition, it would be helpful to perform targeted mobile measurement campaigns in which there is temporarily only sunshine and then only shade to minimize the fluctuations of the signal. However, due to time constraints, no further measurement campaigns and investigations are carried out within the scope of the present work.

# 6 Application to local danger warning

## 6.1 Preparation for local danger warning information

So far, only meteorological variables that are directly measured by the vehicle-based sensor technology have been considered. In addition, however, further parameters and thus information can be generated that are based on the measurements but cannot be measured directly by the vehicle. The parameters described below are only partially used in classical weather forecasting but are essential for road weather conditions and thus especially for safe automated driving functions. Examples of composed road weather relevant events, which are determined based on measured signals, include occurrence of fog, slippery road warnings, snowfall warnings and the likelihood of aquaplaning.

### 6.1.1 Fog detection

Fog is a potential hazard for road traffic due to the usually associated limitation of visibility. Drivers cannot command their vehicle safely and assistant functions fail due to sensors failing to operate, particularly cameras and laser scanners.
Based on the corrected vehicle measurements, warnings of fog can be generated. By networking the vehicles with each other, these warnings can be exchanged exemplary to allow for informing vehicles without necessary measurement equipment about potentially dangerous situations. This can reduce the number of vehicles at risk. For automated driving functions, an early warning can be triggered in a way that either the driver takes over or the route or the driving parameters are adapted to the situation by the system.

For the stationary determination of visibility restrictions and thus presence of fog, there are meteorological measurement principles based on the attenuation of an optical signal. However, since a visibility sensor is not installed by default, a detection principle based on meteorological correlations is being developed for the appearance of fog.

The appearance of fog means that the air can no longer hold the water it contains, resulting in condensation. The condensed water forms fine droplets, which hang in the air as fog. Since the droplets are too light, their weight force is not sufficient to let them fall to the ground. Instead, they are in a float-like state.
To cause this condition, the air under consideration must have a high moisture content and temperatures below the dew point temperature (cf. Chapter 2.1). Mathematically, this can be summarized in two conditions:

$$RH > RH_{Fog} \tag{28}$$

and

$$T \leq T_d. \tag{29}$$

For practical application, the second condition is extended to $T \leq T_d + \Delta T_{Fog}$ to prevent inaccuracies in the calculation of the dew point temperature.

Once both criteria are met, it is recommended to trigger a warning for potential fog and an associated visibility restriction. This approach does not distinguish between low clouds and fog, as this is irrelevant for road users and only the visibility restriction is decisive.

Depending on the selected limit values (e.g. $RH_{Fog}$ = 80 % instead of 90 % for RH) and safety factors (e.g. $\Delta T_{Fog}$ = +4 K instead of +2 K for the dew point temperature), the time of triggering a warning signal can be changed significantly. It is conceivable to introduce an additional parameter in addition to the actual fog warning with the values listed above, which triggers earlier. This would allow the automated driving systems or an active driver respectively to get an advanced warning whereby they are not surprised by the actual fog that might follow.

### 6.1.2 Slippery road warning

The approach currently used operationally by the German Weather Service to release slipperiness warnings, is triggered when the surface temperature is below the dew point temperature and the air temperature falls below the 0 °C limit simultaneously [Kra21a]. However, since the road surface temperature, which must be known for this approach, is currently still being developed and is not installed by default in every production vehicle, it is not possible to generate area-wide information about the slippery conditions of the road network using this approach.

Nevertheless, to predict slippery roads and trigger a warning, an approach based purely on information that each vehicle can generate independently based purely on its own measurements is necessary. To accomplish this, it is crucial to implement the wet bulb temperature, which represents a composed signal, first.
The wet bulb temperature $T_{WB}$ is measured with a damp textile cloth wrapped around a thermometer. The enclosed thermometer is then moved to accelerate the evaporation process of the moisture contained in the textile. The drier the air, the more water can evaporate, which means that more energy is extracted from the thermometer and thereby, it cools down. The wet bulb temperature displayed on the thermometer is therefore lower than the "normally" measured air temperature. The greater the difference between the wet bulb temperature and the dry bulb temperature (equivalent to the air temperature), the drier the air. If the air is saturated, no moisture can evaporate, resulting in equal wet bulb temperature and air temperature. Instead of measuring the wet bulb temperature using the complex measuring principle described above, which is not integrated as standard in the vehicles, it can alternatively be derived. The following approximation formula was established for this purpose [Stu11]:

$$\begin{aligned} T_{WB} =& T \cdot arctan(A \cdot (RH + B)^{0.5}) \\ &+ arctan(T + RH) - arctan(RH - C) \\ &+ D \cdot RH^{1.5} \cdot arctan(E \cdot RH) - F \end{aligned} \tag{30}$$

with $A$ = 0.151977, $B$ = 8.313659, $C$ = 1.676331, $D$ = 0.00391838, $E$ = 0.023101 and $F$ = 4.686035.

Based on the wet bulb temperature, a decision support for slippery road warnings can be implemented. If the wet bulb temperature falls below 0.5 °C, a slippery road warning should be triggered [Kra21b]. Since this approach is based purely on vehicle-based measurements that are already integrated as standard, it is conceivable that this approach could be used to generate area-wide information regarding slippery condition of the roads using fleet data.

If information about spray or wiper activity is available, this can additionally be used to determine the current binary wetness status of the road. This can be taken into account for the slippery road warning by exemplary, adding a likelihood for a slippery road if the warning is triggered.

### 6.1.3 Snow line and snow warning

The snow line (SL) indicates the height above N.N. above which precipitation is expected to fall in solid form. If the vehicle is located below the snow line, it can be assumed that the detected precipitation will fall in the form of rain. However, as soon as the vehicle is located above the snow line, potential precipitation will fall as snow or sleet and thus exerts a significant influence on road traffic. This parameter is of particular interest during winter and in edge cases with cold temperatures and potential precipitation events.

The calculation of the snow line is based on the wet bulb temperature [Ste83]:

$$SL = Altitude_{Vehicle} + \frac{T_{WB} - 1}{0.006} \tag{31}$$

Based on equation 31 the snow line SL can be calculated, using the current height of the vehicle ($Altitude_{Vehicle}$) and the wet bulb temperature ($T_{WB}$).
In addition to the SL, information about the current precipitation status (precipitation or no precipitation) is crucial to trigger a reliable warning of snowfall. If there is no precipitation, it cannot snow, regardless of the altitude at which the vehicle is located.
Based on the combination of the SL and the precipitation status, a distinction between three different scenarios can be made:

1. No precipitation (independent of $Altitude_{Vehicle}$)
2. Liquid precipitation (Precipitation + $Altitude_{Vehicle}$ < SL)

3. Solid precipitation (Precipitation + $Altitude_{Vehicle} \geq$ SL)

Based on this distinction, specific warnings, e.g., for snowfall in scenario three can be triggered.

## 6.2 Evaluation of application to local danger warning information

All following evaluations are based on temperature and RH values of TC1. The physically corrected values are used to calculate the wetbulb temperature and snow line. No data from the machine learning models is used because the number of data points for which contemporaneous results are available for both corrected temperature and corrected relative humidity is smaller. Since the triggering of a warning often occurs only for a short period of time, the representativeness of the results of the application increases significantly as the data basis grows. Therefore, the physically corrected data is used for which the largest common database of corrected temperature and relative humidity data is available (146843 s).

### 6.2.1 Fog detection

When evaluating fog detection and warning, it must be considered that no reference is available to confirm the presence or absence of fog. It can only be investigated whether the values of the MMU that are assumed to be true, trigger warnings of fog at the same time. This does not verify that the mathematical approach based on Equation 28 and 29 in Chapter 6.1.1 is appropriate to detect fog.

The graphical representation of the calculated warning signals can be found in Figure 54 in Appendix A.13. When a warning is triggered, the signal has the numerical value 1. A 0 therefore represents conditions for which no warning is triggered.
Table 18 lists the comparison between the MMU reference and the respective TC1 signal for all 146843 s.

| Fog warning | Portion of data with identical status | Portion of data with warning triggered only by vehicle | Portion of data with warning triggered only by reference |
|---|---|---|---|
| Data | 146843 seconds | | |
| TC1 (raw) [%] | 80.31 | 0.00 | 19.69 |
| TC1 [%] | 85.71 | 12.02 | 2.23 |

Table 18: Overview of three different portions of data for the vehicle-based raw data and the physically corrected data of TC1 for fog warning.

Without a correction of the vehicle-based data, the status regarding warning/no warning is identical for the MME and the TC1 in 80.31 % of the time. The high percentage for the raw data can be explained by the fact that few fog events occurred during the mobile measurement campaigns and the raw data does not trigger a single fog warning (cf. upper plot in Figure 54). By physically correcting the vehicle-based data, the percentage of data with identical status of TC1 and MME can be increased to 85.71 %.

Since the raw TC1 data do not trigger a warning at any point in the data set under consideration, the proportion of data for which the TC1 data incorrectly triggered a warning is zero. The percentage of erroneously triggered warnings increases to 12.02 % for the corrected TC1 data. Approximately every $8^{th}$ warning is therefore triggered, although according to the MMU data, no fog is present.

The raw TC1 data misses triggering a warning 19.69 % of the time and therefore underestimates the situation. By correcting the TC1 data, the proportion is reduced to 2.23 %.

The corrected TC1 data therefore overestimate the situation more often than they underestimate it. The latter poses the greater risk to the vehicle and its occupants. Therefore, reducing the proportion of underestimated situations is crucial. The corrections can reduce this proportion by 17 percentage points.

The fog warnings triggered based on the corrected vehicle data are in good approximation to the warnings triggered by the MMU data. However, the extent to which these warnings are representative of the occurrence of fog cannot be quantified due to the lack of a suitable reference for this purpose.

### 6.2.2 Slippery road warning

Figure 55 in Appendix A.13 shows the results for the triggered warnings regarding slippery roads. 1 means a triggered warning for this point in time, while 0 means no slippery road conditions. It is noticeable that the signal based on the raw data fluctuates more than the signals based on the corrected vehicle data and the MMU. In addition, the raw signals trigger warnings of slippery roads even during the September campaign.
The two signals based on the corrected vehicle data and the MMU data show a similar course. Both data sources almost consistently trigger a warning for slippery roads during the January campaign.

The significant improvement of the warning signal for slippery roads is confirmed by Table 19. In 64.66 % of the time the warning/no warning status matches between the raw data and the MMU. This percentage can be raised to 98.79 % by correcting the vehicle data.
Due to the triggering of warnings of slippery roads during the September campaign, the raw data shows a proportion of almost one third (29.31 %) where a warning is triggered superfluously. The corrected vehicle-based data reduce this fraction to 1.17 %, which corresponds to an improvement of 96 %.
A reduction in the proportion of data which does not trigger a warning, even though the road is slippery according to the MMU data, is also achieved by the corrected vehicle-based data. The percentage of 6.03 % for the raw data, can be reduced by 99 % to 0.04 %.

For the present data, a significant increase in the representativeness of vehicle-based road-slippery warning takes place. Potentially dangerous situations can be detected significantly more often based on the corrected vehicle data. At the same time, the proportion of warnings triggered unnecessarily is also

| Slippery road warning | Portion of data with identical status | Portion of data with warning triggered only by vehicle | Portion of data with warning triggered only by reference |
|---|---|---|---|
| Data | 146843 seconds | | |
| TC1 (raw) [%] | 64.66 | 29.31 | 6.03 |
| TC1 [%] | 98.79 | 1.17 | 0.04 |

Table 19: Overview of three different portions of data for the vehicle-based raw data and the physically corrected data of TC1 for slippery road warning.

decreasing. This effect is expected to strengthen as more vehicles measure independently of each other and thereby validate each other's measurements and warnings automatically.
However, it should be noted that in the available data, explicit road conditions are present almost throughout. A higher proportion of situations at the edge of slippery to non-slippery roads could reduce the proportion of times with the same warning status for vehicle-based and MMU based data.

### 6.2.3 Potential for snow fall

For a snowfall warning, the precipitation status must also be known in addition to the current altitude of the vehicle and the SL. Since the precipitation signal is not evaluated in this work, warnings of snowfall are not determined, but the potential for snowfall is determined. If the vehicle is above the calculated snowline, the potential for snowfall is set to 1, otherwise it is 0.

Figure 56 in Appendix A.13 shows the course of the potential for snowfall. The upper plot shows the calculated potential based on raw TC1 values, the middle and lower plots represent the signals for the corrected TC1 and MMU data, respectively.
Similar to the warnings for slippery roads (cf. Chapter 6.2.2), the raw TC1 data results in multiple triggers of snowfall potential during the September campaign. The courses of the snowfall potentials determined based on the corrected TC1 data and the MMU data are very similar. During the January campaign the possibility of snow with isolated exceptions is almost continuously present. The TC1 does not map all the exceptions.

| Potential for snow | Portion of data with identical status | Portion of data with warning triggered only by vehicle | Portion of data with warning triggered only by reference |
|---|---|---|---|
| Data | 146843 seconds | | |
| TC1 (raw) [%] | 64.91 | 30.05 | 5.04 |
| TC1 [%] | 99.17 | 0.82 | 0.02 |

Table 20: Overview of three different portions of data for the vehicle-based raw data and the physically corrected data of TC1 for snow potential.

Table 20 shows the proportions of the data for the three groups of identical status, the vehicle being over-sensitive, and the vehicle being under-sensitive for the available 146843 s.
64.91 % of the time, the potential based on the raw data of the TC1 agrees with the one based on the MMU data. By correcting the vehicle-based TC1 data, the percentage is raised to 99.17 %.
Additionally, the percentage of data in which the vehicle-based data falsely triggers is significantly reduced. For the raw data, almost $\frac{1}{3}$ of the triggered possibilities for snowfall are incorrect. For the corrected TC1 data, this fraction is reduced by 97 % and amounts to 0.82 %.
Comparable results are available for the data portion indicating the missed triggers based on the TC1. Without correction, the TC1 does not indicate a potential for snowfall 5.04 % of the time, although according to the MMU data the possibility is present. The correction reduces this percentage to 0.02 %.

The representativeness of the calculated potential for snowfall is significantly increased by correcting the vehicle-based data. Comparable to the warning of slippery roads, explicit weather conditions are present in the evaluated data. To what extent the results can be confirmed on data in which the vehicle stays around the calculated snowline cannot be evaluated on the existing data basis. Since there are neither tested vehicle-based precipitation signals nor reference signals of precipitation, a snowfall warning resulting from the potential cannot be tested.

# 7 Conclusion and Outlook

The objective of this work is the quality evaluation and correction of vehicle-based meteorologically relevant signals to allow for a beneficial use in further applications, such as weather models.

The data set generated and analyzed in this thesis with representative reference data for meteorological parameters in moving vehicle measurements, is unique worldwide. The results of this thesis emphasize that the raw vehicle-based data are of poorer quality than the stationary data currently used by DWD. Therefore, the correction of the vehicle-based data is essential for the use in a meteorological context.

Based on the collected data, different correction models are developed. The first contribution enhances physical models for the four considered quantities. Data-based machine learning models are developed for two of the considered quantities (Contribution 2). The corrections developed and implemented, through both physical and data-based models, can significantly improve the quality in a way that high-quality and high-resolution vehicle-based data are available for the majority of time (Contribution 3).

The results of the four considered quantities are briefly summarized hereafter. For the *air pressure* a physical model is implemented in this dissertation. The relatively inert quantity is corrected by the physical model to such an extent that 99 % of the available data from the mobile measurement campaigns can be assumed to be valid. Hence, there is currently no need for a data-driven approach.
In this thesis, both physical and machine learning models are implemented for *air temperature*. Each movement of the vehicle generates various influences falsifying the measurement. Comparable results regarding the proportion of valid data are achieved by the models. The machine learning models (NN and RF), however, provide better results with respect to the data portion within twice the measurement uncertainty and with respect to the mean deviation.

For the *relative humidity* both physical and machine learning models are developed in this work. Due to falsifying effects adding to the vehicular motion by the measurement taking place inside the sensor box, the relative humidity is a more complex parameter than the air temperature. In this context, differences between the physical and machine learning models become apparent. The physical model is not able to achieve the same quality as both machine learning models (NN and RF) and the data-based random forest model generates the best results.
For *global radiation*, a physical model has been implemented in this thesis, but it does not provide satisfactory results for mobile data. Only when stationary data is considered, the model can achieve good results. However, the transfer to mobile data is complicated due to rapidly changing environmental conditions, such as shading, and their large impact on global radiation, which is complicating the comparability to the reference data.

The difficulty of physical models in general is the direct dependence of how well all involved processes are known on the quality of the model. The more processes exert an influence on the measurement, the more complex the issue becomes and the more difficult it is to develop an all-encompassing physical model. This is the advantage of machine learning models, since the exact relationships do not have to be known, but the model itself searches in the data for connections and correlations. In the present application, however, the model also requires synthetic parameters in addition to the measured signals, since the situation can otherwise not be generalized. Commonly, any machine learning model relies on a representative and large data set to develop robustness. Due to the confined amount of data available, this is a limitation of the machine learning models in the context of this work. This effect is less present in the physical models, since physical relationships are independent of the data set's size.

The fourth contribution is the subsequent application of the corrected data to produce local hazard information. The increase in reliability of the warnings based on corrected vehicle data compared to the raw data is elaborated. The corrected vehicle-based data therefore offers the potential to significantly improve warnings of local hazards. As a product of the process, alerts distributed to all fleet vehicles close the loop from data collection, through data correction and processing.

The work reveals that the supply of stable, reliable data as input parameters for forecast models produced by the DWD is within reach. Nevertheless, there is a need for further research, for example to examine the transferability of the models to fleet data. This has not been possible within the scope of the present thesis, as no fleet data is available due to the currently applicable GDPR. Additionally, further data under other weather conditions need to be obtained and evaluated to address the weakness of the data-based models caused by their small data base. At the same time, this is necessary to enable the correction of the vehicle-based data in further scenarios and thus enable autonomous driving in the long term. In addition to the four quantities evaluated here, vehicle-based data shows potential for deriving further meteorological and road weather related quantities, such as precipitation intensity, precipitation type, and road surface temperature.

The application example of local hazard warning shows that the goal of increasing safety by improving the temporal and spatial resolution of the data basis is possible. This can support existing automated driving functions to further increase occupant safety, even in weather-related challenging situations. However, this is only one possible real-life application example and the high-resolution meteorological data shows cross-industry potential.

# A Appendix

## A.1 Sensors

| Measured quantity | Sensor | Measurement uncertainty | Measurement resolution | Operating range | Location of installation | Height |
|---|---|---|---|---|---|---|
| Air pressure | Bosch SMP 480 | ± 10 hPa | 7.91 hPa | [-40,130] °C | Engine control unit | N/A |
| Air temperature (BCM1) | Amphenol Advandes NKS8588 | ± 0.5 K | 0.5 K*1 | [-40,150] °C | Behind radiator grill | 0.32 m |
| Air temperature (MuFu) | Sensirion SHT20 | ± 0.3 K | 0.1 K | [-40,125] °C | Close to engine block | 0.80 m |
| Relative humidity | Sensirion SHT20 | ± 3 %rH | 1 %rH | [-40,125] °C | Close to engine block | 0.80 m |
| Dewpoint temperature | Sensirion SHT20 | No specifications in data sheet | | [-40,125] °C | Close to engine block | 0.80 m |
| Global radiation | Valeo | ± 7 %*2 | No public access to data sheet | | center of the windshield | 1.38 m |

*1: The sensor achieves higher measurement resolution, but is reduced by system-related limitations
*2: No data sheet of sensor publicly accessible. Measurement uncertainty from requirement specification of AUDI AG

Table 21: Tabular overview of the standard sensors which promise potential for meteorological applications on board of production vehicles.

| Measured quantity | Data source | Sensor | Measurement uncertainty | Measurement resolution | Operating range | Height |
|---|---|---|---|---|---|---|
| Air pressure | MWS | Bosch BMP180 | ± 0.8 hPa | 0,01 hPa | [-40,60] °C | 0.5 m |
| | WB | Bosch SMP 480 | ± 10 hPa | 7.91 hPa | [-40,130] °C | 0.5 m |
| | MMU | Transcat DPI 260 Series | ± 0.08 % | N/A | [0,50] °C | 2 m |
| Air temperature | MWS | Sensirion SHT25 | ± 0.3 K | 0,1 K | [-40,60] °C | 0.5 m |
| | WB | Amphenol Advandes NKS8588 | ± 0.5 K | 0.1 K | [-40,150] °C | 0.5 m |
| | MMU | Theodor Friedrichs Frankenberger Art | ± 0.1 K | N/A | [-35,45] °C | 2 m |
| Relative humidity | MWS | Sensirion SHT25 | ± 2%rH | 0.03 %rH | [-40,60] °C | 0.5 m |
| | WB | Sensirion SHT20 | ± 3 %rH | 1 %rH | [-40,125] °C | 0.5 m |
| | MMU | E+E EE33 | ± 1.3 %rH | N/A | [-40,180] °C | 2 m |
| Global radiation | Synop | Pyranometer CM11 (Kipp & Zonen) | ± 3 % | 1 W/m$^2$ | [-40,80] °C | 2 m |
| | MMU | Infrarot Strahlungs-pyrometer KT 15.82 EG & G Heimann | N/A | N/A | [-25,75] °C | 2 m |

Table 22: Tabular overview of the sensors of external data sources. Note, not all data sources mentioned in Chapter 3.2 are listed. The table is limited to data sources used as reference in Chapter 5.

## A.2 Physical model for air temperature

| | |
|---|---|
| F1 | [0.87,1.2] depending on EIT (larger EIT leads to smaller factor F1) |
| F2 | [0.2,1] depending on the change in altitude |
| ΔAlt | Change in Altitude since EIT is triggered |
| ΔDist | Distance travelled since EIT is triggered |
| ΔT | Change in temperature value since EIT is triggered |
| $\Delta T_{ave}$ | $= T_{corr,PHY} - T_{raw}$ (Averaging refers to 20 seconds for both values) |
| $\Delta T_{red}$ | $= (T^{i}_{raw} - T^{i-1}_{corr,PHY}) - 0.2 * \Delta T_{ave}$ |
| $T_c$ | $= T_{corr,PHY}$ being the corrected temperature value |
| $T_m$ | $= T_{meas}$ being the temperature value measured with vehicle-based sensor |
| $T_{m,dec}$ | $= T_{meas} - 0.2 * \Delta T_{ave}$ |
| $T_{m,inc}$ | $= T_{meas} + 0.05 * T_{meas}$ |
| $\tau_{Alt}$ | $= 300m$ (Threshold for Change in Altitude since EIT triggered) |
| $\tau_t$ | Threshold time, for cold engine: 300 seconds (5 min), for warm engine: 120 seconds (2 min) |
| $\tau_{Dist}$ | $= 2000m$ (Threshold for Distance travelled since EIT triggered) |
| $\tau_{EngOff}$ | $= 5\ min$ (Threshold for time engine is turned off) |
| $\tau_v$ | $= 5\ ^{km}/_{h}$ (Threshold for vehicle speed) |
| t | Time since EIT triggered in seconds |
| $t_{Engine\ off}$ | Time since engine is turned off |

Table 23: Tabular overview including explanations of all signals and quantities mentioned in Figure 11 in Chapter 4.2.3 for the physical model to correct the air temperature signal. Left column: Abbreviations, right column: Definitions and explanations.

## A.3 Neural net for air temperature

| Modell | Features | | | | | | | # Features |
|---|---|---|---|---|---|---|---|---|
| | T | EIT | ΔT | Alt | v | $D_{stop}$ | $Eng_{f/0}$ | |
| A | X | X | X | | | | | 3 |
| B | X | X | X | X | | | | 4 |
| C | X | X | X | X | X | | | 5 |
| D | X | X | X | | X | | | 4 |
| E | X | X | X | | X | X | | 5 |
| F | X | X | X | | X | X | X | 6 |

Table 24: Overview of features used for neural network models A-F and total number of features of.

| MSE | Modell | | | | | |
|---|---|---|---|---|---|---|
| | A | B | C | D | E | F |
| Train [K] | 0.68 | 0.62 | 0.57 | 0.64 | 0.50 | 0.48 |
| Validate [K] | 1.50 | 1.74 | 1.50 | 1.46 | 1.43 | 1.08 |
| Test [K] | 1.63 | 1.70 | 1.44 | 1.56 | 1.56 | 1.35 |

Table 25: Train, validate and test MSE for neural network models A-F. All models use same data base.

| Modell | Modified by feature X | MSE Test |
|---|---|---|
| F | reduced by "ΔT" | 7.77 K |
| F | reduced by "v" | 1.66 K |
| F | reduced by "$D_{stop}$" | 1.63 K |
| F | reduced by "$Eng_{f/0}$" | 1.56 K |
| F | reduced by "T" | 1.47 K |
| F | reduced by "EIT" | 1.41 K |
| F | - | 1.35 K |
| F | extended by "Alt" | 1.35 K |

Table 26: Test MSE for model F modified by one feature each time. Features listed according to their impact on test MSE with "Δ Temp Stop" having the biggest influence. Feature "Altitude", which is not considered in model F, does not change the Test MSE if it is considered. The reason for this is probably that there are too few situations in the training data in which a temperature change occurs during a change in altitude since it is a strongly local effect. Since there is no improvement when this feature is taken into account, it is not considered in model F.

| Hidden Layers | 6 Units MSE Train | 6 Units MSE Validate | 6 Units MSE Test | 8 Units MSE Train | 8 Units MSE Validate | 8 Units MSE Test | 10 Units MSE Train | 10 Units MSE Validate | 10 Units MSE Test |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.49 K | 1.15 K | 1.38 K | 0.49 K | 1.25 K | 1.49 K | 0.49 K | 1.33 K | 1.57 K |
| 2 | 0.50 K | 1.14 K | 1.39 K | 0.50 K | 1.22 K | 1.49 K | 0.49 K | 1.67 K | 1.43 K |
| 3 | 0.50 K | 1.29 K | 1.55 K | 0.46 K | 1.14 K | 1.52 K | 0.49 K | 1.21 K | 1.48 K |
| 4 | 0.49 K | 1.25 K | 1.53 K | 0.49 K | 1.16 K | 1.42 K | 0.48 K | 1.14 K | 1.40 K |

Table 27: Train, validate and test MSE for model F for different combinations of number of hidden layers and number of units. Increasing complexity of the network (more hidden layers and more units) does not improve the results. Smallest network with 1 hidden layer and 6 units produces best result.

| Learning rate | MSE Train | MSE Validate | MSE Test |
|---|---|---|---|
| 0.1 | 0.44 K | 1.22 K | 1.69 K |
| 0.01 | 0.44 K | 1.49 K | 1.98 K |
| 0.005 | 0.48 K | 1.18 K | 1.45 K |
| 0.002 | 0.49 K | 1.22 K | 1.45 K |
| 0.0015 | 0.48 K | 1.19 K | 1.45 K |
| 0.0012 | 0.49 K | 1.23 K | 1.46 K |
| 0.001 | 0.49 K | 1.18 K | 1.40 K |
| 0.0009 | 0.49 K | 1.17 K | 1.41 K |
| 0.0008 | 0.49 K | 1.18 K | 1.42 K |
| 0.0005 | 0.49 K | 1.19 K | 1.47 K |

Table 28: Train, validate and test MSE for model F for different learning rates. Learning rate of 0.001, which is set as default, results in lowest test MSE.

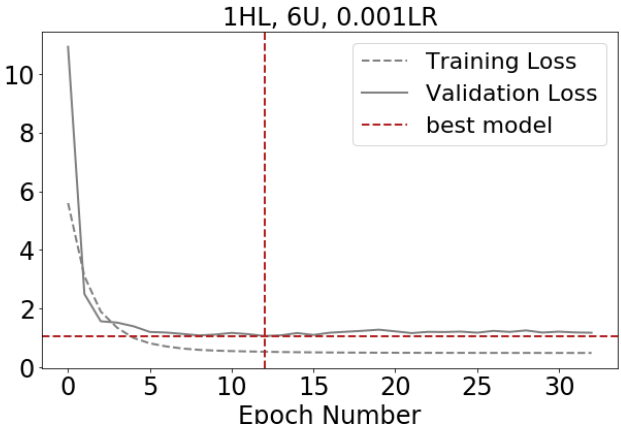


Figure 34: Training and validation loss (=MSE) of model F (best model) trained with 1 hidden layer (HL), 6 units (U) and a learning rate (LR) of 0.001. The best model is marked with the red lines at 12 epochs.

| | MSE | | |
|---|---|---|---|
| Run | Train | Validate | Test |
| 1 | 0.49 K | 1.18 K | 1.40 K |
| 2 | 0.49 K | 1.25 K | 1.48 K |
| 3 | 0.49 K | 1.22 K | 1.47 K |
| 4 | 0.49 K | 1.23 K | 1.45 K |
| 5 | 0.49 K | 1.22 K | 1.45 K |
| 6 | 0.49 K | 1.25 K | 1.48 K |
| 7 | 0.49 K | 1.20 K | 1.44 K |
| 8 | 0.49 K | 1.22 K | 1.45 K |
| 9 | 0.49 K | 1.18 K | 1.40 K |
| 10 | 0.49 K | 1.23 K | 1.46 K |
| Ø MSE | 0.49 K | 1.22 K | 1.45 K |
| Std dev | 0.0 K | 0.02 K | 0.03 K |
| Std dev in percent of Ø MSE | 0.00 % | 1.94 % | 1.87 % |

Table 29: Train, validate and test MSE for model F with 1 hidden layer, 6 units and a learning rate of 0.001 for ten different models (runs). All 10 models are trained with same boundary conditions to test reproducibility of results.

## A.4 Random forest for air temperature

| Modell | Features | | | | | | | | | MSE Test |
|---|---|---|---|---|---|---|---|---|---|---|
| | EIT | v | ΔT | $Eng_{\int 0}$ | $D_{stop}$ | $D_{EngOff}$ | T | $EIT_{max}$ | ΔAlt | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | |
| A | X | X | X | X | X | X | X | X | X | 1.01 |
| B = *A-(1)* | | X | X | X | X | X | X | X | X | 1.04 |
| C = *A-(2)* | X | | X | X | X | X | X | X | X | 1.89 |
| D = *A-(3)* | X | X | | X | X | X | X | X | X | 7.10 |
| E = *A-(4)* | X | X | X | | X | X | X | X | X | 1.06 |
| F = *A-(5)* | X | X | X | X | | X | X | X | X | 0.99 |
| G = *A-(6)* | X | X | X | X | X | | X | X | X | 1.06 |
| H = *A-(7)* | X | X | X | X | X | X | | X | X | 1.28 |
| I = *A-(8)* | X | X | X | X | X | X | X | | X | 1.03 |
| J = *A-(9)* | X | X | X | X | X | X | X | X | | 1.04 |
| *K* = *A-(5&8)* | X | X | X | X | | X | X | | X | 1.02 |

Table 30: Varying features for random forest for correcting temperature measurement. Start with model A with 9 features, model B-J represent variations of model A excluding a different feature for each model. Excluding feature "Duration Stop" (5) improves results, Feature "Max EIT Stop" (8) has lowest impact on results. Both features are therefore neglected for final model K, which is used for further evaluation. Tests conducted with 8 trees, each a maximum depth of 10 layers.

| MSE Test [K] | # Trees | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Max. depth | 8 | 10 | 12 | 15 | 20 | 30 | 50 | 100 |
| 4 | | | | 1.15 | | | | |
| 6 | | | | 0.98 | | | | |
| 8 | | | | 1.00 | | | | |
| 10 | 1.02 | 0.99 | 0.98 | 0.97 | 1.02 | 1.01 | 1.02 | 0.99 |
| 12 | | | | 1.00 | | | | |
| 14 | | | | 1.02 | | | | |

Table 31: MSE Test for different structures of random forest. Starting with 8 trees, each a maximum depth of 10 layers. First, only number of trees is increased (up to 100), while maximum depth is kept constant. Second, for forest with lowest MSE test, the maximum depth is varied. Lowest overall MSE test is achieved for a random forest with 15 trees, each a maximum depth of 10 layers.

| Model | MSE Test |
|---|---|
| K-1 | 0.97 K |
| K-2 | 1.00 K |
| K-3 | 1.03 K |
| K-4 | 0.99 K |
| K-5 | 1.00 K |
| K-6 | 0.96 K |
| K-7 | 1.00 K |
| K-8 | 1.00 K |
| K-9 | 1.00 K |
| K-10 | 0.99 K |
| Ø MSE | 0.99 K |
| Std dev | 0.017 K |
| Std dev in percent of Ø MSE | 1.71 % |

Table 32: Training 10 different RF models with same boundary conditions (Model K, 15 trees, each max. 10 layers deep, same data base) to test reproducibility of model.

## A.5 Neural net for relative humidity

| Ground Truth: Difference between MMU (reference) and | MSE | | |
|---|---|---|---|
| | Train | Validate | Test |
| A – vehicle-based raw signal of RH | 22.96 %rH | 23.20 %rH | 22.53 %rH |
| B – vehicle-based RH signal transferred from inside the sensor case to BCM1 sensor, without compensating for effects induced by vehicle dynamics | 24.46 %rH | 25.25 %rH | 24.14 %rH |
| C – vehicle-based RH signal corrected for effects induced by vehicle motion without transferring it to a more representative location | 23.10 %rH | 23.37 %rH | 22.69 %rH |

Table 33: Train, validate and test MSE for three models A, B and C varying in their used ground truth. All other parameters are identical for all three models.

| Modell | Features | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $EIT_{RH}$ | $\Delta RH$ | v | $D_{stop}$ | $RH_{raw}$ | $RH_{fran}$ | $RH_{vm}$ | $T_{Eng}$ | $T_{sc}$ |
| D | X | | X | X | X | X | X | X | X |
| E | X | X | | X | X | X | X | X | X |
| F | X | X | X | | X | X | X | X | X |
| G | | X | X | X | X | X | X | X | X |
| H | X | X | X | X | | X | X | X | X |
| I | X | X | X | X | X | | X | X | X |
| J | X | X | X | X | X | X | | X | X |
| K | X | X | X | X | X | X | X | | X |
| L | X | X | X | X | X | X | X | X | |
| M | X | X | | | X | X | X | X | X |

Table 34: Overview of different feature constellations tested for neural net for relative humidity.

| MSE | Modell | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | D | E | F | G | H | I | J | K | L | M |
| Train [%rH] | 37.32 | 22.97 | 23.02 | 23.18 | 26.75 | 73.98 | 24.31 | 24.10 | 26.41 | 23.01 |
| Validate [%rH] | 38.03 | 23.20 | 23.26 | 23.46 | 27.47 | 73.98 | 24.49 | 24.42 | 26.64 | 23.26 |
| Test [%rH] | 37.01 | 22.54 | 22.58 | 22.76 | 26.55 | 73.80 | 23.93 | 23.73 | 26.07 | 22.58 |

Table 35: Train, validate and test MSE for models D-M, all using different feature constellations.

| Hidden Layers | 6 Units | | | 8 Units | | | 10 Units | | |
|---|---|---|---|---|---|---|---|---|---|
| | MSE Train [%rH] | MSE Validate [%rH] | MSE Test [%rH] | MSE Train [%rH] | MSE Validate [%rH] | MSE Test [%rH] | MSE Train [%rH] | MSE Validate [%rH] | MSE Test [%rH] |
| 1 | 22.90 | 23.13 | 22.45 | 22.89 | 23.14 | 22.45 | 23.04 | 23.29 | 22.59 |
| 2 | 23.10 | 23.334 | 22.66 | 22.98 | 23.24 | 22.57 | 22.93 | 23.20 | 22.52 |
| 3 | 23.15 | 23.39 | 22.75 | 22.80 | 23.06 | 22.37 | 23.15 | 23.27 | 22.59 |
| 4 | 23.02 | 23.26 | 22.59 | 23.08 | 23.33 | 22.64 | 23.17 | 23.42 | 22.72 |

Table 36: Train, validate and test MSE for model M for different combinations of number of hidden layers and units.

| Learning rate | MSE | | |
|---|---|---|---|
| | Train [%rH] | Validate [%rH] | Test [%rH] |
| 0.02 | 25.73 | 26.02 | 25.37 |
| 0.017 | 24.02 | 24.27 | 26.63 |
| 0.015 | 12.38 | 12.38 | 12.04 |
| 0.014 | 17.59 | 17.79 | 17.26 |
| 0.012 | 17.56 | 17.79 | 17.27 |
| 0.01 | 18.05 | 18.27 | 17.70 |
| 0.008 | 23.15 | 23.39 | 22.70 |
| 0.005 | 23.00 | 23.25 | 22.55 |
| 0.001 | 22.80 | 23.06 | 22.37 |
| 0.0005 | 23.16 | 23.39 | 22.73 |
| 0.0001 | 37.98 | 38.70 | 37.72 |

Table 37: Train, validate and test MSE for model M with 3 hidden layers and 8 units for different learning rates.

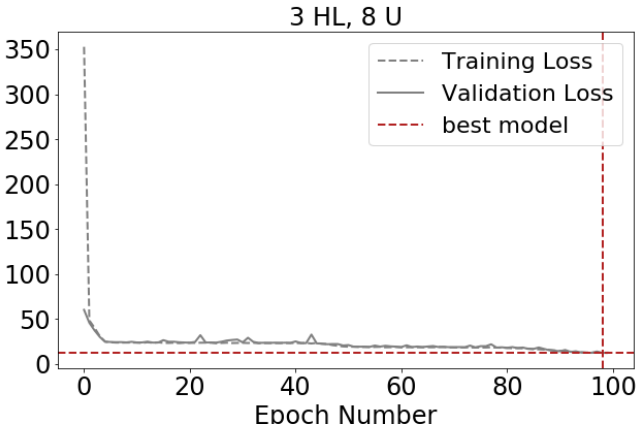


Figure 35: Training and validation loss (=MSE) of model M (best model) trained with 3 hidden layers (HL), 8 units (U) and a learning rate (LR) of 0.015. The best model is marked with the red lines.

a)

| Run | MSE | | |
|---|---|---|---|
| | Train | Validate | Test |
| M-1 | 12.38 | 12.38 | 12.04 |
| M-2 | 13.22 | 13.32 | 13.01 |
| M-3 | 13.76 | 13.51 | 13.37 |
| M-4 | 12.73 | 12.69 | 12.56 |
| M-5 | 16.44 | 16.54 | 16.31 |
| M-6 | 10.68 | 10.67 | 10.55 |
| Ø MSE | 13.20 | 13.19 | 12.97 |
| Std dev | 1.73 | 1.76 | 1.74 |
| Std dev in percent of Ø MSE | 13.11 % | 13.34 % | 13.42 % |

b)

| Run | MSE | | |
|---|---|---|---|
| | Train | Validate | Test |
| M-7 | 19.48 | 19.52 | 19.75 |
| M-8 | 13.52 | 13.32 | 13.44 |
| M-9 | 12.24 | 12.58 | 12.25 |
| M-10 | 14.54 | 14.25 | 14.29 |
| M-11 | 14.12 | 14.63 | 14.46 |
| M-12 | 14.08 | 14.04 | 14.12 |
| Ø MSE | 14.66 | 14.72 | 14.72 |
| Std dev | 2.27 | 2.25 | 2.37 |
| Std dev in percent of Ø MSE | 15.48 % | 15.29 % | 16.10 % |

Table 38: Train, validate and test MSE for six models with same boundary conditions (Model M, 3 hidden layers, 8 units) and a) identical data base for each of the 6 models M-1 to M-6 or b) newly mixed data base for all 6 models M-7 to M-12 to evaluate reproducibility of model.

# A.6 Random forest for relative humidity

| Modell | Features | | | | | | | MSE Test |
|---|---|---|---|---|---|---|---|---|
| | $EIT_{RH}$ | ΔRH | $RH_{raw}$ | $RH_{tran}$ | $RH_{vm}$ | $T_{Eng}$ | $T_{sc}$ | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | |
| A | X | X | X | X | X | X | X | 3.70 |
| B = *A-(7)* | X | X | X | X | X | X | | 5.02 |
| C = *A-(6)* | X | X | X | X | X | | X | 3.23 |
| D = *A-(5)* | X | X | X | X | | X | X | 3.76 |
| E = *A-(4)* | X | X | X | | X | X | X | 12.57 |
| F = *A-(3)* | X | X | | X | X | X | X | 5.22 |
| G = *A-(2)* | X | | X | X | X | X | X | 3.63 |
| H = *A-(1)* | | X | X | X | X | X | X | 3.91 |
| I = *A-(2&6)* | X | | X | X | X | | X | 3.60 |

Table 39: Varying features for random forest for correcting relative humidity measurement. Start with model A with 7 features (identical to final NN constellation), model B-H represent variations of model A excluding a different feature for each model. Excluding feature "Engine temperature" (6) improves results. Additionally excluding feature "Delta RH" (2) does not improve results further (model I). Therefore, only feature (6) is neglected for final model C, which is used for further evaluation. Tests conducted with 8 trees, each a maximum depth of 10 layers.

| MSE Test [%rH] | # Trees | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Max. depth | 8 | 10 | 12 | 15 | 20 | 30 | 50 | 100 |
| 4 | | | | 25.69 | | | | |
| 6 | | | | 10.43 | | | 10.52 | |
| 8 | | | | 5.77 | | | 5.70 | |
| 10 | 3.53 | 3.36 | 3.28 | 3.25 | 3.28 | 3.38 | 3.21 | 3.23 |
| 12 | | | | 1.99 | | | 1.90 | |
| 14 | | | | 1.21 | | | 1.18 | |
| 16 | | | | 0.91 | | | | |
| 18 | | | | 0.75 | | | | |
| 20 | | | | 0.68 | | | | |

Table 40: MSE Test for different structures of random forest. Starting with 8 trees, each a maximum depth of 10 layers. First, only number of trees is increased (up to 100), while maximum depth is kept constant. Second, for the forest with lowest MSE test (50 trees), the maximum depth is varied. The shallower each tree, the higher the MSE test. Additionally, the maximum depth is varied for a forest with 15 trees, since the MSE test is comparable and computational effort is significantly lower. Same Behavior as for 50 trees is visible. The deeper each tree can grow, the lower the MSE test. By increasing the maximum depth allowed for each tree by 2 in each run achieves a decrease in MSE test by 30 % for up to a maximum depth of 14. Further increasing in maximum layers allowed for each tree only reduces MSE test by less than 25 %. Therefore, a random forest with 15 trees and a maximum depth of 14 layes is chosen for further evaluation.

a)

| Model | MSE Test |
|---|---|
| C-1 | 1.21 %rH |
| C-2 | 1.25 %rH |
| C-3 | 1.19 %rH |
| C-4 | 1.18 %rH |
| C-5 | 1.21 %rH |
| C-6 | 1.22 %rH |
| C-7 | 1.23 %rH |
| C-8 | 1.26 %rH |
| C-9 | 1.19 %rH |
| C-10 | 1.27 %rH |
| Ø MSE | 1.22 %rH |
| Std dev | 0.029 %rH |
| Std dev in percent of Ø MSE | 2.41 % |

b)

| Model | MSE Test |
|---|---|
| C-11 | 1.21 %rH |
| C-12 | 1.17 %rH |
| C-13 | 1.16 %rH |
| C-14 | 1.18 %rH |
| C-15 | 1.19 %rH |
| C-16 | 1.14 %rH |
| C-17 | 1.33 %rH |
| C-18 | 1.31 %rH |
| C-19 | 1.20 %rH |
| C-20 | 1.19 %rH |
| Ø MSE | 1.21 %rH |
| Std dev | 0.059 %rH |
| Std dev in percent of Ø MSE | 4.91 % |

Table 41: Training 10 different RF models with same boundary conditions (Model C, 15 trees, each max. 14 layers deep) a) with identical data base for each of the 10 models C-1 to C-10 or b) with newly mixed data base for all 10 models C-11 to C-20 to evaluate reproducibility of model.

## A.7 Vertical sun incidence angle on vehicle

To calculate the position of the sun, the date, and the position of the vehicle are necessary. The date including a timestamp every second and the latitudinal ($l$) and longitudinal ($lo$) position of the vehicle are available in the data set and can be used for further calculation. [Gie20] uses the following formulas to calculate the sun height $S$:

$$S = \frac{asin(x) \cdot 180}{\pi} \tag{32}$$

with $x$ being defined by the following formula:

$$\begin{aligned} x = & sin(\frac{\pi}{180} \cdot l) \cdot sin(\frac{\pi}{180} \cdot d) \\ & + cos(\frac{\pi}{180} \cdot l) \cdot cos(\frac{\pi}{180} \cdot d) \cdot cos(\frac{\pi}{180} \cdot ha) \end{aligned} \tag{33}$$

Both parameters declination ($d$) and hour angle ($ha$) are calculated based on the time specification of the vehicle. The corresponding formulas are given in the next four equations.

The day count ($dc$) builds the basis for both parameters and is defined as:

$$dc = (month - 1) \cdot 30.3 + day \tag{34}$$

Based on Equation 34, the declination ($decl$) can be calculated:

$$d = -23.45 \cdot cos(\frac{\frac{\pi}{180} \cdot 360 \cdot (dc + 10)}{365}) \tag{35}$$

To calculate the hour angle ($ha$), the definition for the equation of time ($eot$) is needed, which is based on equation 34:

$$eot = 60 \cdot [-0.171 \cdot sin(0.0337 \cdot dc + 0.465) - 0.1299 \cdot sin(0.01787 \cdot dc - 0.168)] \tag{36}$$

Using the equation of time ($eot$), the hour angle ($ha$) can be calculated as follows:

$$ha = 15 \cdot (hour + \frac{minute}{60} - \frac{15 - lo}{15} - 12 + \frac{eot}{60}) \tag{37}$$

Based on the previous calculations, the slope of the windshield and the pitch angle of the vehicle, the vertical angle of incidence of the sun on the vehicle $\gamma$ is defined as follows:

$$\gamma = sl + S - P \tag{38}$$

With $sl$ being the slope of the windshield, $S$ the sun height and $P$ the pitch angle of the vehicle. Depending on the vehicle type the parameter S varies. The slope of the windshield of the project-internal test vehicles TC1 and TC2, is 42 °. The inclination or pitch angle P of the vehicle is output by the vehicle and is thus included in the collected data set. The pitch angle is given in °. Using equation 38 the elevation angle $\gamma$ of the sun on the vehicle can be calculated for each time point of the data set, as long as the time as well as position of the vehicle and the corresponding window slope is known.

## A.8 Additional results of quality limitations of used references

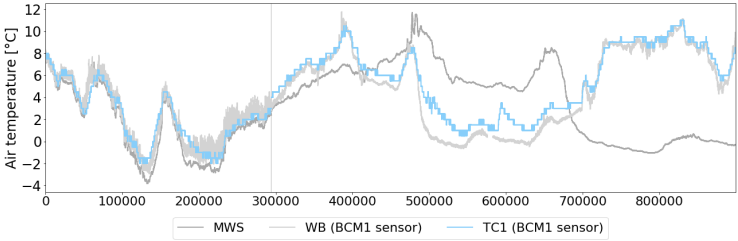


Figure 36: Time course of different temperature signals (darkgrey: MWS, lightgrey: WB, blue: TC1) for both stationary measurement campaigns in November 2020 and January/ February 2021. Temperature signal of MWS during second measurement campaign reveals the corruption of the signal and thus the impracticability as a reference signal.

## A.9 Additional results of the air pressure correction

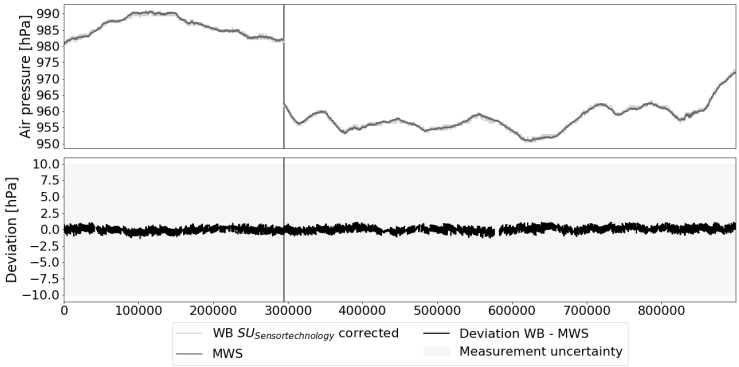


Figure 37: $SU_{Sensortechnology}$: Time course of air pressure signal for both static measurement campaigns separated by black vertical line. Upper plot: Air pressure measured by MWS used as reference to determine $SU_{Sensortechnology}$ for vehicle sensors mounted in WB, only corrected signal for WB shown. Lower plot: Deviation between corrected WB signal and MWS.

| TC2 | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 2 Feb 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|---|---|
| Data | Raw | Corr | Raw | Corr | Raw | Corr | Raw | Corr |
| Data points [sec] | 11108 | | 28783 | | 63383 | | 103274 | |
| Ave Dev [hPa] | -5.18 | 3.49 | -6.43 | 2.24 | 4.12 | 4.55 | -4.88 | 3.79 |
| Max Dev (abs) [hPa] | 12.08 | 10.41 | 15.86 | 10.50 | 12.24 | 13.34 | 15.86 | 13.34 |
| Std dev [hPa] | 2.54 | | 2.72 | | 2.78 | | 2.92 | |
| PD1MU [%] | 98.84 | 99.94 | 90.02 | 99.96 | 99.67 | 98.09 | 96.89 | 98.81 |
| PD2MU [%] | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

Table 42: Tabular overview of the results for raw data (Raw) and corrected data (Corr) of TC2 for the mobile measurement campaigns.

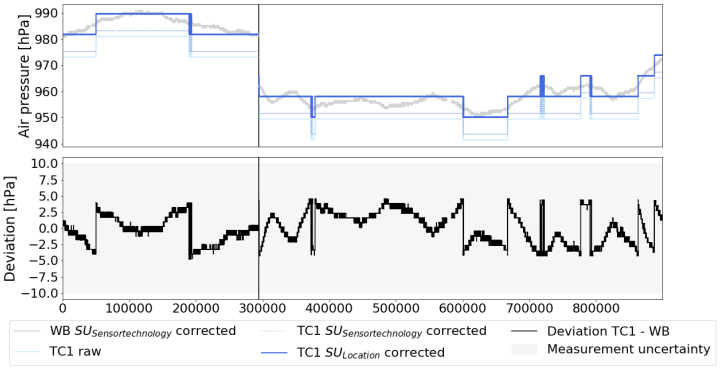


Figure 38: $SU_{Location}$: Time course of air pressure signal for both static measurement campaigns separated by black vertical line. Upper plot: Corrected air pressure of WB used as reference to determine $SU_{Location}$ for vehicle-based measurement of TC1, both raw and corrected signal for TC1 shown. Lower plot: Deviation between corrected TC1 signal and corrected WB signal.

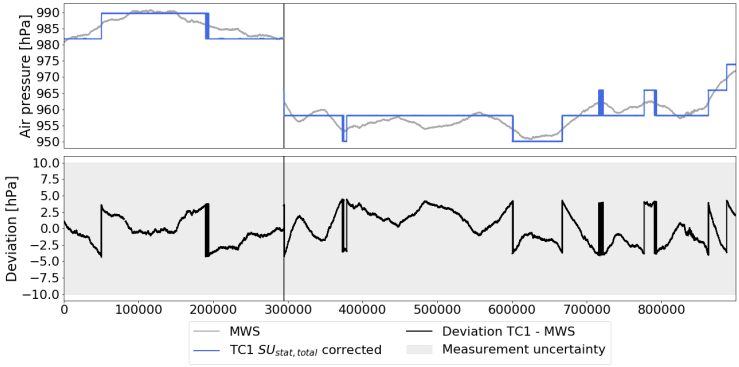


Figure 39: $SU_{stat,total}$: Time course of air pressure signal for both static measurement campaigns separated by black vertical line. Upper plot: Corrected vehicle-based air pressure signal of TC1 compared to MWS signal to control the total sensor uncertainty induced in static situations $SU_{stat,total}$ composed of $SU_{Sensortechnology}$ and $SU_{Location}$. Lower plot: Deviation between corrected TC1 signal and MWS reference signal.

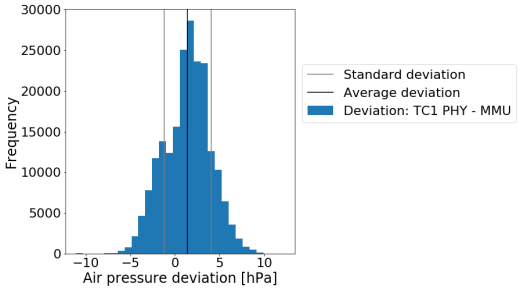


Figure 40: Histogramm of deviations between corrected vehcile-based air pressure of TC1 and MMU reference signal.

## A.10 Additional results of the air temperature correction

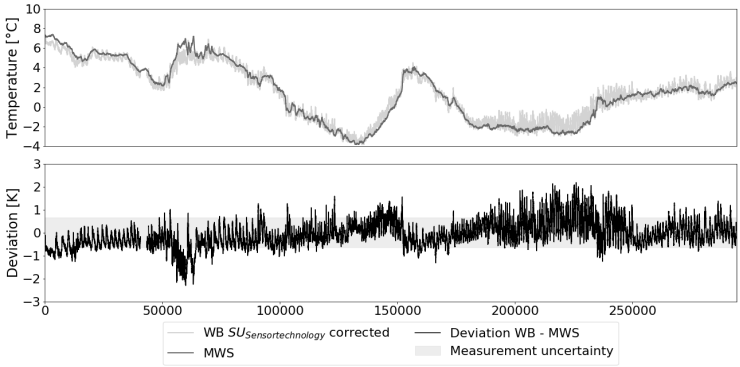


Figure 41: $SU_{Sensortechnology}$: Time course of air temperature signal for stationary measurement campaign in November 2020. Upper plot: Air temperature measured by MWS used as reference to determine $SU_{Sensortechnology}$ for vehicle sensors mounted in WB, only corrected signal for WB shown. Lower plot: Deviation between corrected WB signal and MWS.

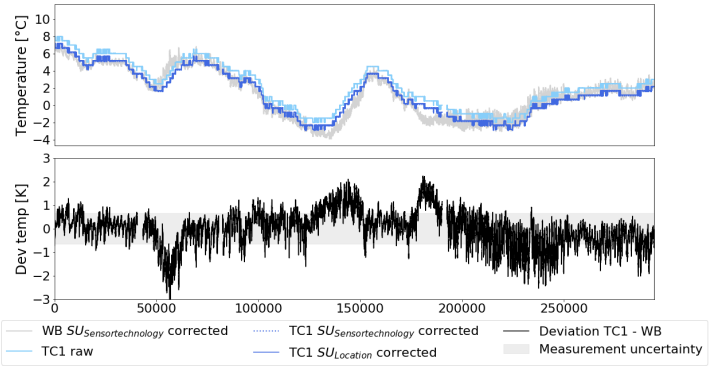


Figure 42: $SU_{Location}$: Time course of air temperature signal for stationary measurement campaign in November 2020. Upper plot: Corrected air pressure of WB used as reference to determine $SU_{Location}$ for vehicle-based measurement of TC1, both raw and corrected signal for TC1 shown. Lower plot: Deviation between corrected TC1 signal and corrected WB signal.

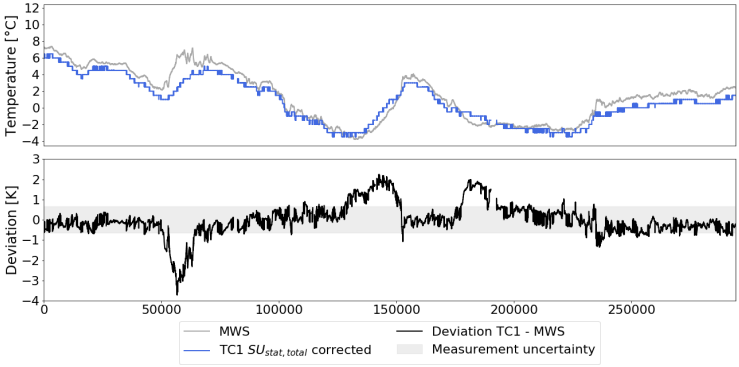


Figure 43: $SU_{stat,total}$: Time course of air temperature signal for stationary measurement campaign in November 2020. Upper plot: Corrected vehicle-based air temperature signal of TC1 compared to MWS signal to control the total sensor uncertainty induced in static situations $SU_{stat,total}$ composed of $SU_{Sensortechnology}$ and $SU_{Location}$. Lower plot: Deviation between corrected TC1 signal and MWS reference signal.

| TC1 | Stationary measurement campaign 1 Nov 2020 | |
|---|---|---|
| | Raw data | Corrected data |
| Data | 282050 seconds | |
| Ø Dev [K] | 0.81 | 0.00 |
| Max Dev [K] | 3.06 | 3.71 |
| Std dev [K] | 0.79 | 0.79 |
| PD1MU [%] | 40.17 | 76.71 |
| PD2MU [%] | 79.74 | 87.91 |

Table 43: Tabular overview of quality parameters determined for raw data and corrected data of stationary measurement campaign in November 2020.

| TC1 | Mobile measurement campaign 1 Jan 2020 | | | Mobile measurement campaign 2 Feb 2020 | | | Mobile measurement campaign 3 Sep 2020 | | | All three mobile measurement campaigns combined | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Offset** | **0** | **-0.81** | **-0.33** | **0** | **-0.81** | **-0.33** | **0** | **-0.81** | **-0.33** | **0** | **-0.81** | **-0.33** |
| Data | 89043 seconds | | | 53716 seconds | | | 64588 seconds | | | 207347 seconds | | |
| Ø Dev [K] | 1.23 | 0.42 | 0.90 | 2.26 | 1.45 | 1.93 | 1.92 | 1.11 | 1.59 | 1.71 | 0.90 | 1.38 |
| Max Dev [K] | 21.51 | 20.70 | 21.18 | 27.15 | 26.34 | 26.82 | 18.43 | 17.62 | 18.10 | 27.15 | 26.34 | 26.82 |
| Std dev [K] | 2.53 | 2.53 | 2.53 | 3.55 | 3.55 | 3.55 | 3.02 | 3.02 | 3.02 | 3.00 | 3.00 | 3.00 |
| PD1MU [%] | 66.32 | 50.02 | 73.99 | 36.25 | 41.32 | 43.80 | 48.92 | 31.19 | 49.25 | 53.11 | 41.90 | 58.46 |
| PD2MU [%] | 79.70 | 81.61 | 80.89 | 57.12 | 62.68 | 61.75 | 61.70 | 64.50 | 65.19 | 67.99 | 71.38 | 71.04 |

Table 44: Tabular overview to clarify the need for adjustment of the offset correction $SU_{stat,total}$. For all measurement campaigns individually, as well as for the entire data basis of the three mobile measurement campaigns, the results for the raw data (offset = 0 K), the data which were corrected using the previously determined $SU_{stat,total}$ (offset = -0.81 K), and the data which were corrected using the attenuated offset (offset = -0.33 K) are listed.

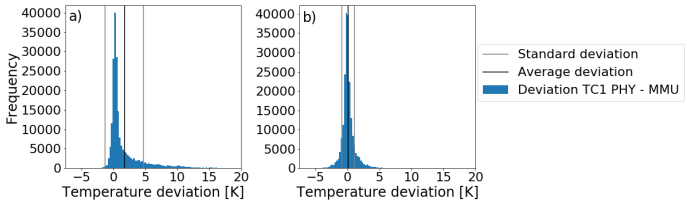


Figure 44: Histogram of deviation between reference (MMU) and vehicle-based a) raw data of TC1 and b) physically corrected data of TC1. Frequency for both histograms in absolute terms. In addition, the mean deviation (black) and the standard deviation (gray) are marked.

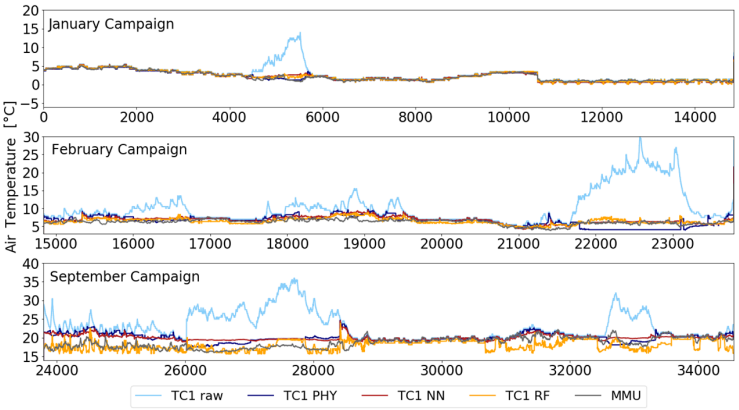


Figure 45: Time course of three five temperature signals (darkgrey: MMU, lightgrey: TC1 raw, blue: TC1 corrected based on physical model, red: TC1 corrected based on neural network, yellow: TC1 corrected based on random forest) for all three mobile measurement campaigns in January, February, and September 2020 spread over three plots. Each plot represents one measurement campaign. Note, the time axis (x-axis) is skewed differently for each plot due to the different amount of data available. For each plot, the y-axis maps air temperatures in °C over a span of 26 K. The respective section varies, however, as it is adapted to the respective air temperature values of the corresponding measurement campaign.

| TC1 | Mobile measurement campaign 1 Jan 2020 | | | | Mobile measurement campaign 1 Feb 2020 | | | | Mobile measurement campaign 1 Sep 2020 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Raw | PHY | NN | RF | Raw | PHY | NN | RF | Raw | PHY | NN | RF |
| Data | 14840 seconds | | | | 8952 seconds | | | | 10629 | | | |
| Ø Dev [K] | 0.70 | -0.08 | 0.05 | 0.04 | 4.10 | 0.38 | 0.63 | 0.24 | 4.29 | 1.27 | 1.29 | -0.78 |
| Max Dev [K] | 13.15 | 1.57 | 2.09 | 1.87 | 25.20 | 5.08 | 4.73 | 3.22 | 18.43 | 7.04 | 6.24 | 5.40 |
| Std dev [K] | 1.85 | 0.29 | 0.40 | 0.35 | 4.70 | 1.29 | 0.73 | 0.70 | 4.43 | 1.52 | 1.37 | 1.37 |
| PD1MU [%] | 86.52 | 96.16 | 91.69 | 91.32 | 16.66 | 42.71 | 56.75 | 57.60 | 29.38 | 38.38 | 36.17 | 37.54 |
| PD2MU [%] | 91.59 | 99.82 | 97.50 | 99.47 | 29.58 | 61.45 | 84.25 | 90.59 | 40.52 | 52.19 | 50.89 | 62.65 |

Table 45: Tabular overview of the quality parameters for the raw data and the results of all three models of the TC1 for all three mobile measurement campaigns.

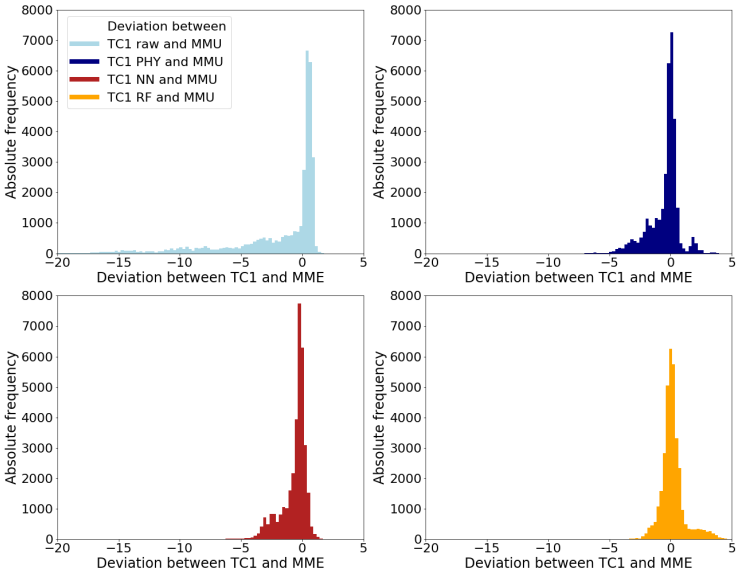


Figure 46: Histograms of anomalies of vehicle-based temperature signals: TC1 raw in lightblue, TC1 corrected based on physical model (PHY) in darkblue, TC1 corrected based on neural network (NN) in red, TC1 corrected based on random forest (RF) in yellow. Histograms show anomalies of all three mobile measurement campaigns in January, February, and September 2020 combined.

| TC1 RF | Mobile measurement campaign 1 Jan 2020 | | Mobile measurement campaign 2 Feb 2020 | | Mobile measurement campaign 3 Sep 2020 | | All three mobile measurement campaigns combined | |
|---|---|---|---|---|---|---|---|---|
| | Raw data | RF | Raw data | RF | Raw data | RF | Raw data | RF |
| Data | 14840 seconds | | 8818 seconds | | 10765 seconds | | 34423 seconds | |
| Ø Dev [K] | 0.70 | 0.04 | 4.03 | 0.24 | 4.33 | -0.77 | 2.69 | -0.16 |
| Max Dev [K] | 13.15 | 1.87 | 25.20 | 3.07 | 18.43 | 5.40 | 25.20 | 5.40 |
| Std dev [K] | 1.85 | 0.35 | 4.70 | 0.7 | 4.42 | 1.37 | 4.03 | 0.97 |
| PD1MU [%] | 86.52 | 91.32 | 16.91 | 57.58 | 29.01 | 37.82 | 50.70 | 65.94 |
| PD2MU [%] | 91.59 | 99.47 | 30.03 | 90.58 | 40.01 | 63.02 | 59.69 | 85.79 |

Table 46: Tabular overview of the correction results of TC1 for the random forest (RF) for the mobile measurement campaigns.

## A.11 Additional results of the relative humidity correction

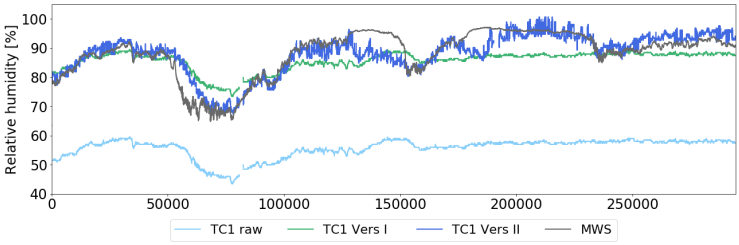


Figure 47: Time course of four different relative humidity (RH) signals. Reference signal of MWS in grey, vehicle-based raw signal of TC1 in light blue, pure offset correction (Version I) in green and transfer with additional offset correction (Version II) in blue. The x-axis contains all recorded data from the first stationary measurement campaign and thus corresponds to a temporal axis.

| TC1 | First static measurement campaign | | | |
|---|---|---|---|---|
| | Raw data | Version I | Version II (after Transfer) | Version II (after Offset) |
| Data | 288789 seconds | | | |
| Ø Dev [%rH] | -32.61 | -2.68 | 1.52 | 0.00 |
| Max Dev [%rH] | 45.13 | 15.44 | 17.01 | 15.49 |
| Std dev [%rH] | 5.47 | 5.47 | 3.83 | 3.83 |
| PD1MU [%] | 0.00 | 36.54 | 53.44 | 69.96 |
| PD2MU [%] | 0.00 | 58.52 | 87.45 | 86.75 |

Table 47: Tabular overview of the correction results of TC1 to compensate for effects induced in stationary situations. Version I executing a pure offset correction by 29.93 %rH. Version II is composed of a transfer or the raw signal and a subsequent offset correction by 1.52 %rH.

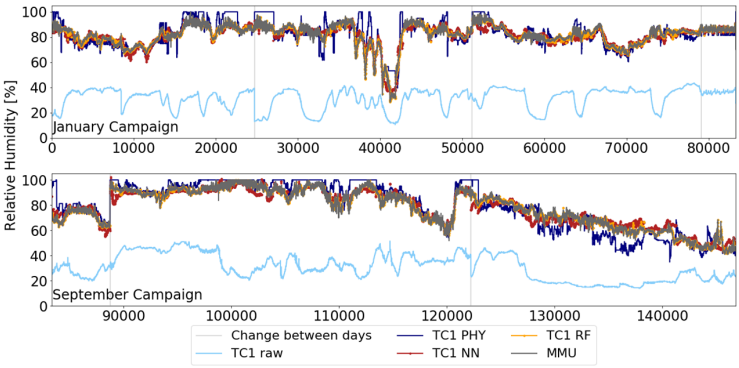


Figure 48: Time course of different relative humidity (RH) signals for data of January (upper plot) and September (lower plot) campaign. Reference signal of MMU in grey, vehicle-based raw signal of TC1 in light blue and vehicle-based signal corrected by random forest (RF) in yellow. Change between days during a campaign are marked as vertical lines. Due to the different amount of available data for each campaign the x-axis is skewed differently for the upper and lower plot.

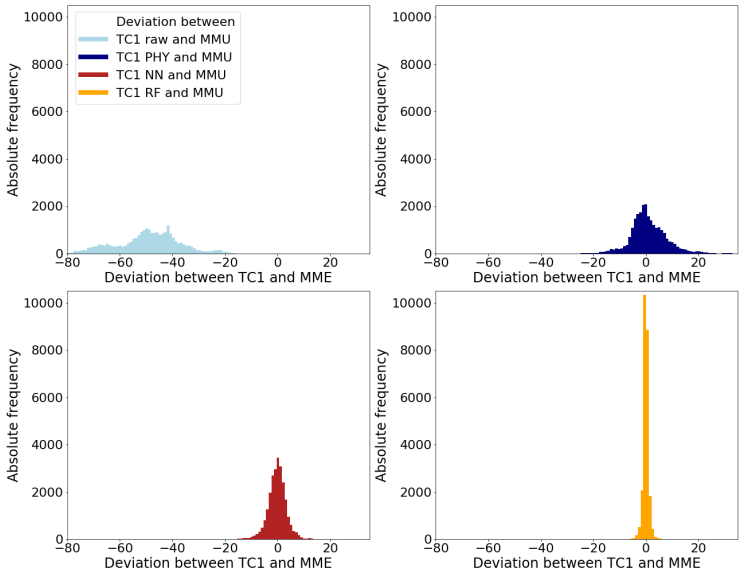


Figure 49: Histograms of anomalies of vehicle-based relative humidity signals: TC1 raw in light-blue, TC1 corrected based on physical model (PHY) in darkblue, TC1 corrected based on neural network (NN) in red, TC1 corrected based on random forest (RF) in yellow. Histograms show anomalies of all three mobile measurement campaigns in January, February, and September 2020 combined.

# A.12 Additional results of global radiation

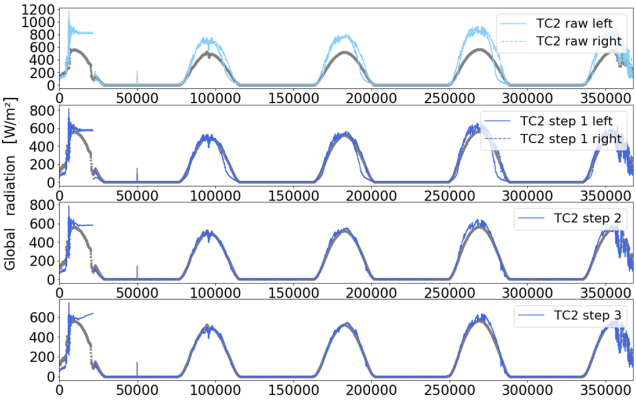


Figure 50: Overview of different stages of correction of global radiation values. Reference values of Synop station always shown in grey, vehicle-based values of TC2 in blue. a) Raw signal of TC2, b) TC2 signal corrected by a factor of 0.7 (step 1), c) combined TC2 signal based on weighting individual signals (step 2), d) final TC2 signal based on step 2 with correction by factor MA (step 3).

| TC2 | Second-by-second | 1 minute average | 1 minute Max | 1 minute Min |
|---|---|---|---|---|
| Selection criterion | TC2 at standstill and sun from the front | Sun from the front | Sun from the front | Sun from the front |
| Data | 4043 seconds | 366 minutes | 366 minutes | 366 minutes |
| Ø Dev [W/m²] | 27.45 | 70.12 | 87.45 | 20.96 |
| Max Dev [W/m²] | 702.37 | 406.95 | 636.30 | 508.85 |
| Std dev [W/m²] | 106.50 | 91.27 | 114.60 | 126.88 |
| PD1MU [%] | 27.23 | 14.48 | 15.03 | 8.74 |
| PD2MU [%] | 34.92 | 19.95 | 21.31 | 13.11 |

Table 48: Tabular overview of the correction results of TC2 for different data sections. The results of "second-by-second" data only consider data during standstill of TC2 and sun shining from the front onto the vehicle. For the other three categories either the average ("1 minute average"), the maximum ("1 minute max") or the minimum ("1 minute Min") of an interval of 60 seconds is considered for situations with the sun shining from the front onto the vehicle.

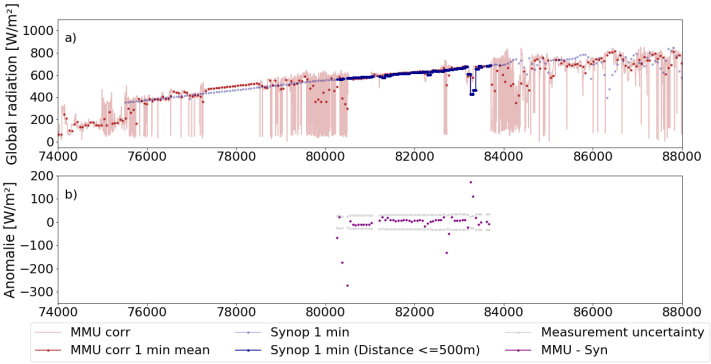


Figure 51: Comparison between global radiation measurement by Synop station "Weihenstephan" and MMU. a) shows minute data by Synop station in blue, which are weakened in color if the MMU is at a distance of more than 500 m to the Synop station. The measurements of the MMU are displayed in red; second-by-second data is weakened in color, data averaged over one minute is marked as red dots. The time interval with measurements of the Synop station drawn in dark blue, marks the range used for calibration of the MMU signal. b) shows the anomalies of minute data of the MMU to the Synop station in purple, the respective measurement uncertainty is marked in grey for the considered time interval.

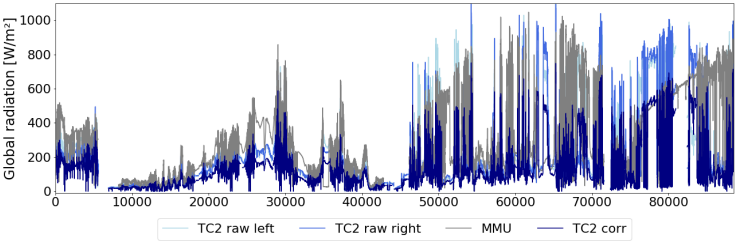


Figure 52: Different global radiation signals. Two vehicle-based raw signals in lighter blue, combined and corrected vehicle-based signal in dark blue and reference signal of MMU in grey.

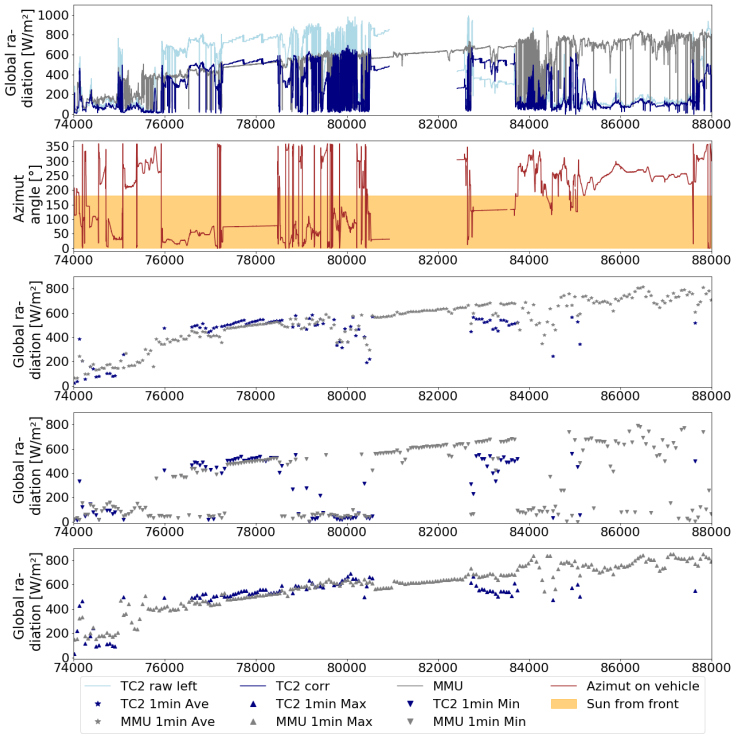


Figure 53: Different signals related to global radiation for a section of the September campaign. a) shows reference signal of MMU in grey, corrected vehicle-based signal in dark blue and one of the two vehicle-based raw signals in lightblue. b) shows the azimuth angle of incidence on the vehicle in brown. In addition, angles that mark an incidence of sunlight from the front (0-180°)on the vehicle are highlighted in yellow. c), d) and e) show different synthetic signals, where the time resolution of the reference signal as well as the vehicle-based signal is reduced to 1 min by averaging (c)), maximum value determination (d)) and minimum value determination (e)). Note, the vehicle-based signals in c)-e) are only displayed if sun shines from the front on the windshield.

## A.13 Additional results of the application to local danger warning

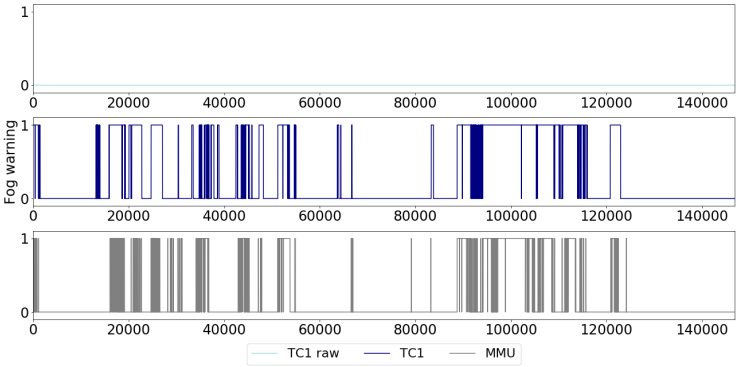


Figure 54: Status of fog warning: 1 represents triggered warning, 0 no warning for available data of mobile measurement campaigns. Upper plot: warnings based on vehicle-based raw signal, middle plot: warnings based on physically corrected values, lower plot: warnings based on MMU.

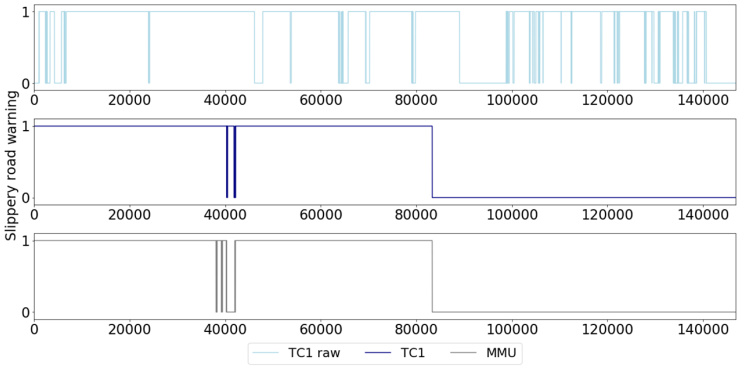


Figure 55: Status of slippery road warning: 1 represents triggered warning, 0 no warning for available data of mobile measurement campaigns. Upper plot: warnings based on vehicle-based raw signal, middle plot: warnings based on physically corrected values, lower plot: warnings based on MMU.

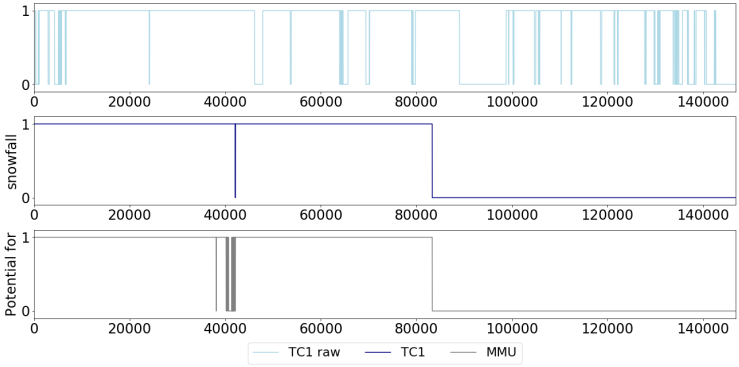


Figure 56: Status of warning of possible snowfall: 1 represents triggered warning, 0 no warning for available data of mobile measurement campaigns. Upper plot: warnings based on vehicle-based raw signal, middle plot: warnings based on physically corrected values, lower plot: warnings based on MMU.

# Bibliography


[AB09] Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[ACD+12] Amanda R. S. Anderson, Michael Chapman, Sheldon D. Drobot, Alemu Tadesse, Brice Lambi, Gerry Wiener, and Paul Pisano. Quality of Mobile Air Temperature and Atmospheric Pressure Observations from the 2010 Development Test Environment Experiment. In *Journal of Applied Meteorology and Climatology*, volume 51(4), pages 691–701, 2012. DOI: https://doi.org/10.1175/JAMC-D-11-0126.1.

[Amp15] Amphenol Advanced Sensors (data sheet). Provisional test specification NKS8588, 2015. Not published - Internal document.

[BHK08] Kenneth Brown, Esteban Hernandez, and Christos Kyrtsos. Method of Determining Ambient Air Temperature. US patent 7,387,437 B2. Also available via: https://patents.google.com/patent/US7387437B2/en?oq=US+patent+7%2c387%2c437+B2, last retrieved: 2021-08-22, 2008.

[BJ16] Paul Bridge and Danny Johns. Road Weather Information Systems Reference Data. In *Proceedings of the 18th SIRWEC conference*, Moving RWIS forward, Ft Collins, CO, USA, 2016.

[BKF18] Simtia Biswas, Teai J. Kwon, and Liping Fu. A Geostatistical Approach to Classification of Topography and Climate Zones for RWIS. In *Proceedings of the 19th SIRWEC conference*, Meteorological & climatological studies, page 79, Smolenice, SVK, 2018.

[Böh09] Klemens Böhm. Überwachtes Lernen II: Klassifikation und Regression - Neuronale Netze und Support-Vektor-Maschinen. Universität Karlsruhe (TH), 2009. https://dbis.ipd.kit

.edu/download/veranstaltung3-20090504.pdf, last retrieved: 2021-08-16.

[BPZ+19] Matthew Bartos, Hyongju Park, Tian Zhou, Branko Kerkez, and Ramanarayan Vasudevan. Windshield wipers on connected vehicles produce high-accuracy rainfall maps. In *Scientific reports*, volume 9, article 170, 2019. DOI: https://doi.org/10.1038/s41598-018-36282-7.

[Bra17] Mark Braun. Verbesserung der Kurzfristvorhersage von Niederschlagsereignissen mittels Fahrzeugsensoren - mobileVIEW, 2017. https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/verbesserung-de-kurzfristvorhersage-von-niederschlag-mittels-fahrzeugsensoren-mobileview.html, last retrieved: 2021-08-17.

[Bre01] Leo Breimann. Random Forests. In *Machine Learning*, volume 45, pages 5-32, 2001. DOI: https://doi.org/10.1023/A:1010933404324.

[Com04] Committee on Weather Research for Surface Transportation: The Roadway, editor. *Where the Weather Meets the Road*. National Academies Press, Washington, DC, USA, 2004. ISBN: 978-0-309-09136-7.

[Com10] Committee on Progress and Priorities of U.S. Weather Research and Research-to-Operations Activities, Board on Atmospheric Sciences, editor. *When Weather Matters: Science and Services to Meet Critical Societal Needs*. National Academies Press, Washington, DC, USA, 2010. ISBN: 978-0-309-15249-5.

[COW10] Joshua Conte, Akira Oga, and Robert Wunsche. External Temperature Display Control Algorithm. US patent 7,668,686 B2. Also available via: https://patents.google.com/patent/US7668686B2/en?oq=US+patent+7%2c668%2c686+B2, last retrieved: 2021-08-22, 2010.

[Dal21] Daltons Gesetz. In *Lexikon der Chemie*, 2021. https://www.spektrum.de/lexikon/chemie/daltonsches-gesetz/2204, last retrieved: 2021-08-17.

[DBW04] Michael J. DeRonne, Joshua J. Barhitte, and William Western. Ambient Air Temperature Prediction. US

patent 2004/0184509 A1. Also available via: https://patents.google.com/patent/US20040184509A1/en?oq=US+patent+2004%2f0184509+A1, last retrieved: 2021-08-22, 2004.

[Deu93] Deutsches Institut für Normung e.V. (Hrsg.). DIN 1345:1993-12, Thermodynamik Grundbegriffe, 1993. Beuth-Verlag, Berlin.

[Deu21] Deutscher Wetterdienst. Bodenwettermeldungen: Bodenwettermeldungen aus den verschiedenen Messnetzen des Deutschen Wetterdienstes und des Geoinformationsdienstes der Bundeswehr, 2021. https://www.dwd.de/DE/leistungen/bodenwettermeldung/bodenwettermeldung.html, last retrieved: 2021-07-17.

[Fro20] Jörg Frochte. *Maschinelles Lernen: Grundlagen und Algorithmen in Python*. Carl Hanser Verlag München, München, 3rd revised and expanded edition, 2020. ISBN: 978-3-446-46144-4.

[Gao08] Shixiang Gao. Method and System for Controlling a Climate Control System. US patent 7,392,662 B2. Also available via: https://patents.google.com/patent/US7392662B2/en?oq=US+patent+7%2c392%2c662+B2, last retrieved: 2021-08-22, 2008.

[Ger18] German law. General Data Protection Regulation, 2018. Entered into force on 25.05.2018. Also available via: https://gdpr-info.eu/, last retrieved: 2021-08-23.

[Gie20] Jürgen Giesen. Berechnung von Sonnenhöhe und Azimut mit Tabellenkalkulation, 2020. http://www.geoastro.de/SME/tk/index.htm, last retrieved: 2021-08-17.

[Hau10] Robert Hausen. Der Taupunkt - ein Multitalent in der Wettervorhersage. In *Thema des Tages des Deutschen Wetterdienstes*, 2010. https://www.wetterdienst.de/Deutschlandwetter/Thema_des_Tages/2483/der-taupunkt-ein-multitalent-in-der-wettervorhersage, last retrieved: 2021-08-23.

[HAVP+20] Meike Hellweg, John-W. Acevedo-Valencia, Zoi Paschalidi, Jens Nachtigall, Thomas Kratzsch, and Christoph Stiller. Using floating car data for more precise road weather forecasts. In *Pro-*

*ceedings of 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, Antwerp, BEL, 2020.

[HNK+19] Meike Hellweg, Jens Nachtigall, Thomas Kratzsch, Roland Potthast, Hella Riede, Zoi Paschalidi, Alexandros Bouras, John-W. Acevedo-Valencia, and Christoph Stiller. Fleet Weather Map - A Project to Integrate Floating Car Weather Data into the Field of Automated Driving. In *Proceedings of 2019 TAC-ITS Canada Joint Conference & Exhibition*, Halifax, CAN, 2019.

[HNN12] Taisto Haavasoja, Juhani Nylander, and Pauli Nylander. Experiences of Mobile Road Condition Monitoring. In *Proceedings of the 16th SIRWEC conference*, Winter Maintenance Methods, pages 1–7 (ID:14), Helsinki, FIN, 2012.

[HS20] Meike Hellweg and Christoph Stiller. Using artificial intelligence to correct weather related vehicle data. In *Eumetnet workshop on Artificial Intelligence for weather and climate*, Brussels, BEL, 2020.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, NY, USA, 2nd edition, 2009. ISBN: 978-0-387-84857-0.

[Hub16] Thomas Hubert. Outside Air Temperature Measurement Device and Method. US patent 9,435,694 B2. Also available via: `https://patents.google.com/patent/US9435694B2/en?oq=US+patent+9%2c435%2c694+B2`, last retrieved: 2021-08-22, 2016.

[HZ11] Jeffery S. Hawkins and Mark A. Zurawski. Method for Estimating Ambient Air Temperature prior to Combustion in an Internal Combustion Engine. US patent 20110106505 A1. Also available via: `https://patents.google.com/patent/US20110106505A1/en?oq=US20110106505A1`, last retrieved: 2021-08-22, 2011.

[KB09] Christos Kyrtsos and Kenneth Gerard Brown. Outside Ambient Temperature Initialization Technique. US patent 7,599,812 B2. Also available via: `https://patents.google.com/patent/US7599812B2/en?oq=US+patent+7%2c599%2c812+B2`, last retrieved: 2021-08-22, 2009.

[KB15] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *Proceedings of the 3rd International Conference for Learning Representations (ICLR)*, volume abs/1412.6980, San Diego, CA, USA, 2015.

[KBB13] Christos Kyrtsos, Michael Buckenmeyer, and Kenneth Gerard Brown. Ambient Temperature Estimation. US patent 8,577,514 B2. Also available via: https://patents.google.com/patent/US8577514B2/en?oq=US+patent+8%2c577%2c514+B2, last retrieved: 2021-08-22, 2013.

[KH09] Masatoshi Kubota and Masafumi Higashlyama. Outside Air Temperature Determination Appartus. US patent 7,572,055 B2. Also available via: https://patents.google.com/patent/US7572055B2/en?oq=US+patent+7%2c572%2c055+B2, last retrieved: 2021-08-22, 2009.

[KKS+19] Yoo-Jun Kim, Baek-Jo Kim, Yoon-Sook Shin, Hui-Won Kim, Geon-Tae Kim, and Seon-Jeong Kim. A case study of environmental characteristics on urban road-surface and air temperatures during heat-wave days in Seoul. In *Atmospheric and Oceanic Science Letters*, volume 12(4), pages 261–269, 2019. DOI: https://doi.org/10.1080/16742834.2019.1608791.

[KN16] Virve Karsisto and Pertti Nurmi. Using car observations in road weather forecasting. In *Proceedings of the 18th SIRWEC conference*, Poster session, pages 1–5 (ID:17), Ft Collins, CO, USA, 2016.

[KNSM17] Virve Karsisto, Pertti Nurmi, Timo Sukuvaara, and Kari Mäenpää. Towards Intelligent Real-Time Road Weather Services Utilizing Mobile Vehicular Data. In *Proceedings of 12th ITS European Congress*, 2017. ID: TP0915#.

[KR17] KNMI Datalab and Royal Netherlands Meteorological Institute. Car sensors as mobile meteorological network: The potential of car sensor data for meteorological observations, 2017. https://www.vetuda.com/wp-content/uploads/2017/10/Product_DataLab_carsensordata_v4_za_def.pdf, last retrieved: 2021-07-17.

[Kra21a] Thomas Kratzsch. Prognostizieren von Reifglätte. Personal communication, 2021.

[Kra21b] Thomas Kratzsch. Straßenglätte in Zusammenhang zur Feuchtkugeltemperatur. Personal communication, 2021.

[Kri07] David Kriesel. Ein kleiner Überblick über Neuronale Netze, 2007. https://www.dkriesel.com, last retrieved: 2021-08-15.

[LC79] Gösta H. Liljequist and Konrad Cehak. *Allgemeine Meteorologie*. Friedr. Vieweg & Sohn, Braunschweig/ Wiesbaden, GER, 2nd edition, 1979. ISBN: 978-3-3528-13555-3.

[LL20] Stefan Luber and Nico Litzel. Was ist ein Random Forest?. *BigDataInsider*, 2020. https://www.bigdata-insider.de/was-ist-random-forest-a-913937/, last retrieved: 2021-07-17.

[Mat14] Kenji David Matsuura. Vehicle Ambient Temperature Estimation System. US patent 2014/0044149 A1. Also available via: https://patents.google.com/patent/US20140044149A1/en?oq=US+patent+2014%2f0044149+A1, last retrieved: 2021-08-22, 2014.

[Mat21] MathWorks. Introducing deep learning with matlab, 2021. https://de.mathworks.com/content/dam/mathworks/ebook/gated/80879v00_Deep_Learning_ebook.pdf, last retrieved: 2021-08-17.

[MDPO10] Bill Mahoney, Sheldon Drobot, Paul Pisano, and Jim O'Sullivan. Vehicles as Mobile Observation Systems. In *Bulletin of the American Meteorological Society*, volume 91(9), pages 1179–1182, 2010. DOI: https://doi.org/10.1175/2010BAMS2954.1.

[Met16] Meteo France. Meteorological and climatic services for business. In *Meteo France Annual Report 2016*. Also available via: http://rapportannuel.meteofrance.fr/en/2016/services/1-4, last retrieved: 2021-08-16, 2016.

[MIO13] William P. Mahoney III and James M. O'Sullivan. Realizing the potential of vehicle-based observations. In *Bulletin of the*

*American Meteorological Society*, volume 94(7), pages 1007–1018, 2013. DOI: https://doi.org/10.1175/BAMS-D-12-00044.1.

[MS03] Raj Manakkalr and Hirohide Suda. Logic for Outside Temperature Display. US patent 6,665,629 B2. Also available via: https://patents.google.com/patent/US6665629B2/en?oq=US+patent+6%2c665%2c629+B2, last retrieved: 2021-08-22, 2003.

[MTS12] Jeffery D. Musiak, Brian J. Tillotson, and Charles B. Spinelli. Collection of Meteorological Data by vehicles. US patent 8,314,730 B1. Also available via: https://patents.google.com/patent/US8314730B1/en?oq=US+patent+8%2c314%2c730+B1, last retrieved: 2021-08-22, 2012.

[Mur12] Kevin P. Murphy. *Machine learning: A probabilistic perspective*. Adaptive computation and machine learning series. MIT Press, Cambridge, MA, USA, 1st edition, 2012. ISBN: 978-0-262-01802-9.

[Pou00] Mark J. Poublon. Ambient Temperature Learning Algorithm for Automotive Vehicles. US patent 6,088,661. Also available via: https://patents.google.com/patent/US6088661A/en?oq=US+patent+6%2c088%2c661, last retrieved: 2021-08-22, 2000.

[RHDJ95] Robert G. Rudzewicz, Robert P. Hennessee, Pascal Dutfoy, and Annette B. Juhasz. System and Method for Determining Ambient Temperature Outside of a Vehicle. US patent 5,416,728. Also available via: https://patents.google.com/patent/US5416728A/en?oq=US+patent+5%2c416%2c728, last retrieved: 2021-08-22, 1995.

[Rob12] Robert Bosch GmbH (data sheet). Barometric pressure sensor for engine management systems SMP480, 2012. https://www.datasheets360.com/pdf/-970403773916159829, last retrieved: 2021-08-17.

[Run10] Thomas A. Runkler. *Data Mining: Methoden und Algorithmen intelligenter Datenanalyse*. Computational intelligence. Vieweg

+ Teubner, Wiesbaden, 1st edition, 2010. ISBN: 978-3-8348-0858-5.

[RWLS18] Eike Rehder, Florian Wirth, Martin Lauer, and Christoph Stiller. Pedestrian prediction by planning using deep neural networks. In *Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5903–5908, 2018.

[SAL05] Leopold Super, Richard Avery, and John E. Longnecker. Method and System for Indirectly Estimating Ambient Air Temperature. US patent 2005/0071074 A1. Also available via: https://patents.google.com/patent/US20050071074A1/en?oq=US+patent+2005%2f0071074+A1, last retrieved: 2021-08-22, 2005.

[Sci20] Scikit-learn developers. sklearn.preprocessing.minmaxscaler, 2020. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html, last retrieved: 2021-08-18.

[Seg04] Mark R. Segal. Machine Learning Benchmarks and Random Forest Regression. *UCSF: Center for Bioinformatics and Molecular Biostatistics*, 2004. Also available via: https://escholarship.org/uc/item/35x3v9t4, last retrieved: 2021-07-17.

[Sen14] Sensirion (data sheet). SHT20 - Humidity and Temperature Sensor, 2014. https://www.sensirion.com/fileadmin/user_upload/customers/sensirion/Dokumente/2_Humidity_Sensors/Datasheets/Sensirion_Humidity_Sensors_SHT20_Datasheet.pdf, last retrieved: 2021-08-17.

[SKF18] Klaus Schürmanns, Sandra Krolewski, and Dietmar Fischer. Verfahren zur Bestimmung der Umgebungslufttemperatur eines Kraftfahrzeugs: Offenlegungsschrift. DE patent 102017206724 A1. Also available via: https://patents.google.com/patent/DE102017206724A1/de?oq=DE+patent+102017206724+A1, last retrieved: 2021-08-22, 2018.

[Son90] Dietrich Sonntag. Important new values of the physical constants of 1986, vapour pressure formulations based on the ITS-90, and psychrometer formula. In *Zeitschrift für Meteorologie*, volume 40(5), pages 340–344, 1990.

[Spe21] Lufttemperatur. In *Lexikon der Geographie*, 2021. https://www.spektrum.de/lexikon/geographie/lufttemperatur/4853, last retrieved: 2021-08-17.

[Ste83] Reinhold Steinacker. Diagnose und Prognose der Schneefallgrenze. In *Wetter und Leben*, volume 35, pages 81–90, 1983.

[Stu11] Roland Stull. Wet-Bulb Temperature from Relative Humidity and Air Temperature. In *Journal of Applied Meteorology and Climatology*, volume 50(11), pages 2267–2269, 2011. DOI: https://doi.org/10.1175/JAMC-D-11-0143.1.

[SW11] Claude Sammut and Geoffrey I. Webb, editors. *Encyclopedia of Machine Learning*, chapter Mean Squared Error, pages 653–653. Springer US, Boston, MA, USA, 2011. DOI: https://doi.org/10.1007/978-0-387-30164-8_528.

[THST00] Henk Taale, Aad De Hoog, Stef Smulders, and Onno Tool. The Results of a Dutch Experiment with Floating Car Data. In *Proceedings of the 9th IFAC Symposium Control in Transportation Systems 2000*, pages 192–197, Braunschweig, GER, 2000. IFAC.

[TNK21] Ari Tuononen, Arto Niskanen, and Janne Kovanen. Road Cloud, 2021. https://roadcloud.com/services/hyper-local-traffic-and-road-state-information/, last retrieved: 2021-08-17.

[Wal98] Werner Wallrafen. Method for Displaying of the Outside Temperature in a Motor Vehicle. US patent 5,737,243. Also available via: https://patents.google.com/patent/US5737243A/en?oq=US+patent+5%2c737%2c243, last retrieved: 2021-08-22, 1998.

[WBASA20] Curtis L. Walker, Brenda Boyce, Christopher P. Albrecht, and Amanda Siems-Anderson. Will Weather Dampen Self-Driving Vehicles? In *Bulletin of the American Meteorological Society*, pages E1914–E1923, 2020. DOI: https://doi.org/10.1175/BAMS-D-19-0035.1.

[Wer21] Martin Werner. *Digitale Bildverarbeitung: Grundkurs mit neuronalen Netzen und MATLAB-Praktikum*. Springer Vieweg, Wiesbaden, 1st edition, 2021. ISBN: 978-3-658-22184-3.

[Wet19a] Magnusformel. In *Wetterlexikon*, 2019. https://www.wetter.de/cms/wetterlexikon-magnus-formel-2399699.html, last retrieved: 2021-08-17.

[Wet19b] Spezifische Feuchte. In *Wetterlexikon*, 2019. https://www.wetter.de/cms/wetterlexikon-spezifische-feuchte-2400365.html, last retrieved: 2021-08-17.

[WFSF18] Sascha Wirges, Tom Fischer, Christoph Stiller, and Jesus B. Frias. Object detection and classification in occupancy grid maps using deep convolutional networks. In *Proceedings of 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3530–3535, 2018.

[WGMY17] Baowei Wang, Xiaodu Gu, Li Ma, and Shuangshuang Yan. Temperature error correction based on bp neural network in meteorological wireless sensor network. In *International Journal of Sensor Networks (IJSNET)*, volume 23(4), 2017. DOI: https://doi.org/10.1504/IJSNET.2017.083532.

[WH06] John M. Wallace and Peter V. Hobbs. *Atmospheric Science - An Introductory Survey*, volume 92 of *International Geophysica Series*. Academic Press, Burlington, MA, 2nd edition, 2006. ISBN: 978-0-12-732951-2.

[Wue99] Michael Wuertenberger. Process for Correctly Sensing an Outside Temperature. US patent 5,895,117. Also available via: https://patents.google.com/patent/US5895117A/en?oq=US+patent+5%2c895%2c117, last retrieved: 2021-08-22, 1999.

[YSYA19] Keisuke Yoneda, Naoki Suganuma, Ryo Yanase, and Mohammad Aldibaja. Automated driving recognition technologies for adverse weather conditions. In *IATSS Research*, volume 43(4), pages 253–262, 2019. DOI: https://doi.org/10.1016/j.iatssr.2019.11.005.

[ZN18] Oliver Zeigermann and Chi Nhan Nguyen. *Machine Learning - kurz & gut: Eine Einführung mit Python, Pandas und Scikit-Learn*. dpunkt.verlag GmbH, Heidelberg, GER, 2018. ISBN: 978-3-96009-052-6.