Fleet Weather Map
A Project to Integrate Floating Car Weather Data into the Field of Automated Driving

Meike Hellweg, M.Sc.
Institute for Measurement and Control Systems, Karlsruhe Institute of Technology (KIT)

Jens Nachtigall, B.Sc.
AUDI AG

Thomas Kratzsch, Dipl.-Met.
Deutscher Wetterdienst (DWD)

Roland Potthast, Prof. Dr.
Deutscher Wetterdienst (DWD)

Hella Riede, Dipl.-Chem.
Deutscher Wetterdienst (DWD)

Zoi Paschalidi, Dr. rer. nat.
Deutscher Wetterdienst (DWD)

Alexandros Bouras, M.Sc.
Deutscher Wetterdienst (DWD)

Walter Acevedo, Dr. rer. nat
Deutscher Wetterdienst (DWD)

Christoph Stiller, Prof. Dr.-Ing.
Institute for Measurement and Control Systems, Karlsruhe Institute of Technology (KIT)

Paper presented for presentation at the
“Preparing for Connected and Automated Vehicles,
Part 1: Technical Challenges and Developments” Session

of the 2019 TAC-ITS Canada Joint Conference, Halifax, NS

The project presented in this paper is co-funded by the Ministry of Transportation of the federal republic of Germany.

ABSTRACT

To lay the foundations for safe autonomous driving, detailed information about the current and upcoming weather conditions is needed. However, currently available wide-meshed weather data are insufficient to meet the necessary spatial and temporal resolution required. On the other hand economic constraints limit the extension of the observations network. The project presented in this paper explores to what extent vehicular sensor data may augment existing data sources to improve weather forecasting and nowcasting in general but for the road network in particular. This paper illustrates the limited applicability of raw sensor readings due to the impact of vehicle dynamics and a continuously changing environment. Subsequently, it investigates methods for filtering and adjustments to meet meteorological requirements. A preliminary, physically modelled algorithm focused on estimating a representative ambient temperature shows promising results in comparison to synoptic reference data. The paper closes with an analysis of further refinements and a continuous validation required as well as the need for further meteorological quantities that need to be integrated to complete a typical meteorological data collective.

1. Introduction

To ensure safe and predictive autonomous driving it is crucial to capture and map a vehicle‘s surroundings as precisely and reliably as possible. While the major tasks are arguably the detection of other road users, road infrastructure and obstacles, (road-) weather as an influencing factor is vastly understudied in this field. For various applications, current vehicles are equipped with a range of different sensors. However, their capabilities are limited, particularly regarding their use in meteorological applications and predictive functions. This is reflected in their limited range: Typically, detection of objects is only available for a distance of up to 250 m using a radar or laser scanning sensor. Assuming a speed of 130 km/h (= 80.7 mph) the vehicle can, at most, anticipate its surroundings for a window of 6.9 seconds. Meteorological sensors, on the contrary, are limited to the immediate surroundings of a vehicle. No vehicle sensors are able to detect slippery or icy road conditions ahead of the vehicle. Anticipating environmental conditions, therefore, can only be enabled through vehicle-to-vehicle communication or by building upon reasonable coverage of reliable and well maintained roadside weather stations. This is particularly substantial for winter conditions when such information are essential for sustainable and safe autonomous driving as well as to generally reduce the overall risks of weather-related accidents.

Currently, the German National Weather Service (DWD) receives road weather and condition data from the regional road authorities every 15 minutes. Clients who subscribe to the road weather information service see per-station data displayed on a map or in table format. Forecasts for road weather stations based on model output statistics (MOS) help gauge the future development of road conditions, especially to help the winter road service plan their operations. The coverage achieved by only using the static meteorological network with a temporal resolution of 10 up to 15 minutes, does not satisfy the high temporal and spatial resolution as required by autonomous driving functions. By relying only on static measurement data road conditions can be dangerously misinterpreted [Niskanen et al., 2018]. To enable pinpointing weather forecasts area-wide as required for autonomous driving functions, highly localized observations are needed, optionally aided by remote sensing data and specialized modeling.

The project “Flotten-Wetter-Karte” (Fleet Weather Map) aims to overcome these mutual limitations. For this project, the AUDI AG cooperates with the “Deutscher Wetterdienst” (national weather service of Germany) exploring the applicability of vehicles as mobile measurement systems to collect detailed data at neuralgic spots or hazard-prone locations.

The project seeks to prove that data from the sensor set installed in vehicles allow for a significantly higher resolution of weather products which, in turn, have the capabilities to enable safe autonomous driving functions. The project investigates the technical requirements for collecting floating car data and integrating it into road weather models and nowcasting. The main objective of the project is to increase the overall spatial and temporal resolution of the public weather observation network and the reliability and scope of nowcasting and weather models.

In the course of this project, the vehicle sensor data will be examined for their suitability in nowcasting products such as short-term heavy rain warnings and snow forecasts. The German weather forecast model COSMO-D2 is currently adapted to start a new data assimilation in the lowest model level near ground every 5 minutes. By effectively extending the observational network on basis of assimilating vehicle sensor data, weather products relevant in the boundary layer at much higher temporal resolution than currently implemented are generated. More detailed weather data would not only be essential for autonomous driving functions, but also a variety of applications beyond the automotive industry would benefit from that data, for example winter services [Karsisti and Sukuvaara, 2018]. The National Research Council of the United States stated in 2004 that "adverse weather is associated with over 1.5 million vehicular accidents each year". In America more than one fifth of the vehicle accidents are associated with inconvenient weather [Perchanok, 2016]. By having a reliable forecast for road conditions, more accurate estimations regarding necessity and extent of winter maintenance can be made individually for each road section. "The total annual global expenditure on the winter maintenance of roads is about €10 billion ($10 billion […]). The benefits of winter maintenance have been estimated to be about eight times the cost" [White et al., 2006]. These numbers highlight the need for reliable road weather forecasts with a respective temporal and spatial resolution.

Current projects like "Car sensors as mobile meteorological network" by the [KNMI Datalab, 2018], the work by [Habrovsky et al., 2018] and "roadweather online" by [Teconer, 2019] demonstrate the present necessity of weather data with a higher spatial and temporal resolution than currently available. The key feature of the project presented in this paper is the cooperation between a national weather service and a car manufacturer functioning as a potential swarm data supplier.

The paper is structured as follows. An overview of the overall setup and methodology used for the data acquisition in the light of the respective technical and legal constraints is provided in chapter 2. Chapter 3 lists the vehicle sensors employed for weather-related measurements and data sourced for reference purposes. Setups and methods on evaluating and calibrating vehicle sensor readings to render these applicable for meteorological data assimilation and inclusion into weather prediction models are presented in chapter 4. Preliminary experimental results from test vehicles are provided in chapter 5. The paper closes with a short summary and prospective next steps (chapter 6).

2. Technical Implementation

As outlined previously, the focus of this project is the establishment of the quality and reliability of weather data from vehicles for mutual application in weather modelling and vehicle safety applications. Particular challenges have been the establishment of a reference fully qualified for calibrating environmental sensors of a moving vehicle as well as the acquisition of a reasonable volume of measurement data under the desired real-world weather conditions.

Figure 1 illustrates the overall data flow which has been established between the prototypical vehicles, a project-specific cloud infrastructure and a project-specific data exchange interface of the national weather service. Due to the research nature of this project, the enrollment on any operational infrastructure was discarded as it requires a continuous adaption to a variety of interfaces, data models and signal sources while still meeting adequate security standards and fully complying with the General Data Protection Regulation (GDPR).

The GDPR has been in effect as of the first quarter of 2018 and requires explicit consent of the individual for processing of their personal data. Weather data from vehicles, which is particularly collected for studying their quality and reliability, requires unaltered geo- and time-references. Even though a connection to any individual is neither required nor intended, the possibility of tracing and distinguishing individual vehicles by their behavior and trajectories requires the implementation of the GDPR onto the entire data processing stream. Consequently, sensor data from customer vehicles cannot be used and the data collection in the course of this project is limited to a small number of test vehicles, company vehicles as well as vehicles of partners and project partnerships, totaling to less than one hundred vehicles. In addition to the GDPR, fleet data acquisition using non-standard, retrofitted logging devices of production vehicles is limited due to federal regulations in Germany which require explicit approval of the public authorities.

The German National Weather Service oversees a synoptical land observation network with almost 500 stations, which deliver state-of the-art accurate weather data every 10 minutes, but usually away from roads, which could disturb measurements. Responsibilities for the installation, maintenance and data are assigned to the municipal and local state authorities. This leads to heterogeneity in the overall observation network. Data from the around 1500 RWS are operationally transmitted only every 15 minutes. This is estimated to be too coarse in time for live calibration of a vehicle in motion.

Even though a widespread availability of the data would significantly leverage the potential of the experiments conducted in the course of this project, the overall objective is not ultimately impaired: Instead of training a machine learning algorithm that requires a high volume of labelled training and verification data, this project pursues a model-based approach.

Environment data as versatile and wide-spread as possible is gathered from a variety of sources (see chapter 3), each at specific locations and with an individual focus on their assets and drawbacks that need to be weighted and categorized. Consequently, a centralized cloud infrastructure for storage and processing has been established and is continuously enhanced according to the project's progress.

Due to the variety and geographic distribution of data sources and data types, processing is split into multiple stages yet sharing a common data structure and common interfaces:
Most importantly, the prototypical vehicles record full Controller Area Network (CAN) traces along with scaled but uncompressed video streams and (partially) raw sensor data on board the vehicles. As the relevant signal input set is continuously adapted over the course of the project, this allows to fully reconstruct and replay test drives as needed. While this is applicable for prototypical vehicles, only a limited subset of sensor signals is gathered from (production-grade) partner or company vehicles using retro-fitted logging hardware. Due to limitation of local storage and mobile bandwidth, a predefined, vehicle-specific data set is extracted and uploaded to the cloud infrastructure depending on the equipment of the vehicle. The same applies for auxiliary measurement setups as presented in the following chapter.

For development purposes of this project, geo- and time referenced data is collected frame-wise at 1Hz. This frequency more than satisfies typical model data assimilation and nowcasting input requirements of a weather service, especially if brought into a production vehicle fleet.

A scaling mechanism is therefore implemented that maps observations onto common tiles and a tile-layer per quantity. Meteorological quantities are processed from sensor readings or quantities estimated from indirect measurements - such as camera sensors - and explicitly do not allow conclusions about the originating vehicle or node. To allow subsequent selection, bias correction and weighting of quantities for integration into forecast models, each quantity is supplied with an individual confidence level. Establishing meteorologically profound requirements for the derivation thereof from a range of factors - the situational context in particular - is a dedicated ongoing field of action of this project.

As of writing, a pre-operational data push mechanism to the interfaces of the weather forecast provider has been established using batches of 1 minute windows. In return, for verification and development purposes, the project's cloud infrastructure periodically queries and stores synoptic and road weather observations as well as an operational nowcast output for distribution in vehicle safety functions. This output, however, has at this stage not yet been augmented with vehicular data but will be part of the functional analysis completing the overall research project.

3. Data basis for the project FloWKar

3.1 Data from vehicular environmental sensors

For economic as well as regulatory reasons the project FloWKar focuses on data obtained from a network of mass-produced vehicular sensors. Using measurements from these sensors for meteorological applications, however, has not yet been fully explored. Early experiments presented in this paper emphasize the need for dedicated methods and algorithms assuring applicability of the sensor readings in weather-related products.

The advantages of using measurements of mass-produced sensors dominate the drawbacks. First, the sensors are mounted in a wide variety of vehicles with the result that a large amount of data will be available. Secondly, difficulties with transferring sensors into series production are avoided while the business case for mass-produced sensors has already been established. For the reasons stated above, the project focuses on making use of production type sensors and determining the quality of the available data. Using supplementary sensors (not mounted as production grade sensor in the vehicle) is theoretically possible, but it does not help to increase the spatial coverage of weather-related data. Therefore, this approach does not contribute to achieving the objectives of the project FloWKar and is thus neglected.

The meteorological sensor data net evaluated in this project is composed of readings from mostly mass-produced vehicular sensors and is only supported by prototypical sensors. An overview of all available sensors is presented in Figure 2. For weather applications the most relevant signals recorded and delivered by a vehicle are the following (compare Table 1):

- atmospheric pressure
- air temperature
- relative humidity
- dewpoint temperature
- precipitation intensity
- friction estimate
- (road surface temperature)
- visible distance and
- global radiation.

These data are only useable with a timestamp and global navigation satellite system information (geographic reference).
Further quantities regarding the vehicle are taken into account for pre-processing, for example vehicle speed and engine temperature (compare Table 2). For development purposes all raw data are recorded with a frequency of 1 Hz and are directly stored in a matrix in the vehicle. This simplifies detecting dependencies between quantities as described in chapter 4.1.

A confidence measure is assigned to each measurement. Thereby, the reliability of every signal for each timestamp can be quantified and, for example, sensor failure can be indicated. For every timestamp all quantities are checked for physical plausibility and, if necessary, corrected by a vehicle agnostic algorithm, which is further described in chapter 4.1.

3.2 Data from reference stations

The outcome of this project ultimately depends on the reliability and availability of reference data that can be considered as ground truth. Manufacturer specifications for individual sensors are usually determined under static conditions. However, these do not take the influences of the vehicle dynamics in changing conditions into account. Therefore, meteorological reference data is required to qualify sensor performance under real-world conditions. Hence, laboratory experiments would lead to false assumptions.
This leads to the particular challenge of correlating individual measurement quantities of two independent highly dynamic systems (vehicle and weather) using intermediate static systems (weather stations).

Consequently, four different kind of reference stations are used in the course of this project (compare Figure 3). Two of which are stationary: synoptic stations and road weather stations. An auxiliary mobile weather station compensates for missing stationary reference data alongside test routes. While test drives can only cover short periods of time to collect measurements by mass-produced sensors, a hybrid weather station ("weather box") equipped with vehicular sensors enables continuous and independent data collection for long time studies.

Figure 3 depicts synoptic stations and road weather stations are summarized to one possible source of reference station, which is feasible due to their comparable measurement principles. Both types are stationary and can therefore only collect data from this individual location. The national weather service of Germany maintains the synoptic stations, whereas the road weather stations are supervised by the freeway directorate of the respective federal state. The data is collected by calibrated meteorological measurement equipment. The number and type of signals depends on the type of station. Road weather stations measure the standard quantities such as ambient temperature or atmospheric pressure, but also road surface temperature. For a synoptic station the detection of road surface temperature is impossible since synoptic stations are not located close to traffic ways like road weather stations. Instead they are commonly built offsite in rural areas, for example on an open meadow without roads, buildings or any other flow disturbing obstacles nearby. Through this setup the meteorological representativeness is guaranteed while the relevance to traffic applications is limited.
The project FloWKar mainly focuses on the comparability of data collected by vehicular sensors and by reference stations. In consideration of the comparability between the environmental conditions, it is decided on a case-by-case basis which available reference stations are suitable for the particular data set.
In terms of measurement reliability, both static stations (synoptic and road weather stations) are suitable to validate vehicular data since both fulfill the quality standard for their respective purposes. Major drawbacks are imposed by infrastructural design and different responsible authorities. These generally limit the temporal resolution to 10 to 15 minute intervals. Only selected, mostly experimental reference stations provide data with a temporal resolution of 1 minute.

The third possibility to collect reference data is using a mobile weather station. Similar to the previously described static reference systems, the mobile weather station also collects data with calibrated meteorological measurement equipment. Moreover, the mobile weather station comes with the advantage of mobility. Due to a mobile power supply, the system can be set up at any desired location, independently of synoptic stations, road weather stations or power connection. Using a car battery the operating time is estimated to last for 72 hours.
The temporal resolution of 1 second of the mobile weather station is suitable for a direct comparison to the vehicular sensor data, as these are also collected with a data rate of 1 Hz. Despite the higher temporal resolution and the spatial flexibility, the mobile weather station does neither replace the synoptic stations nor the road weather stations since the mobile weather station is not

operating at a fixed standardized height. Instead, the height is adjustable and thus directly adaptable to vehicle properties.

The remaining reference station is the so-called "weather box". It is the only source for reference data built with the same vehicular sensors as the prototypical vehicles used for the project and even includes a smaller version of the windshield complete with rain and light sensors. While the sensors are identical the surroundings differ. The vehicular sensors installed in the weather box are not subject to influences from vehicle integration or vehicle dynamics. Hereby, the quality and accuracy of the mass-produced vehicle sensors can be checked energy efficiently over a long period. The continuous measurement is independent of the installation in the vehicle and influencing factors on the data like vehicle speed as well as continuous offsets can be identified, quantified and revised.
The weather box is mobile with the restriction that a power outlet is needed. Theoretically, it can be set up to collect data at any location as long as power supply is ensured. The weather box is equipped with heating for the winter season. This prevents the windshield from freezing whereby measurements of the precipitation for temperatures below the freezing point are enabled.
Since the weather box uses the same sensors as the vehicle, it also provides data with a temporal resolution of 1 second. The weather box does not satisfy meteorological requirements for a reference station, thus it cannot not be considered as a reference station in a meteorological sense. It is used to determine the basic accuracy of the vehicular sensors without the influence of motion as well as the qualification thereof. Additionally, this setup allows for collecting data for extensive time periods.

Independent of the particular reference station it has to be decided on case-by-case basis, whether or which reference station is suitable to function as reference for validation. Thus, for example air temperatures measured by the vehicle above asphalt should not be compared to ambient temperatures collected by a synoptic station above lawn. In this use case the road weather station data might be of a bigger help instead, where the ambient temperature is also measured above asphalt.

4. Methodology

4.1 Physical correction of vehicular air temperature data in the vehicle

As stated in chapter 2, the raw data are collected on board of the vehicle and preprocessed before they are forwarded to the in-house backend. In the following, the paper focuses predominantly on the air temperature measurements' correction and validation. This is due to its fundamental role in weather modelling and the dependency of multiple other meteorological quantities thereon.

Ever since temperature is being measured on board of vehicles, it supports the control of the combustion engine and therefore undergoes basic filtering and attenuation (compare for instance patens [Brown, Hernandez and Kyrtsos, 2018], [Kintz, 2000] and [Poublon, 2000]) while meteo-rological requirements have not been in focus.

Every value of each signal undergoes a plausibility check, which verifies whether the value transferred from the sensor is within predefined bounds. These bounds represent its physical limitations. In case of the quantity air temperature these bounds range from -50 °C to 60 °C. Out-of-bound data is set to zero and its related confidence is reduced. For values within bounds no modification is performed during the plausibility check.

The actual correction for systematic distortions is carried out thereafter. The objectives are the detection and compensation of systematic effects distorting the measurements. For the success of the project it is crucial to accept locally induced fluctuations, for example changes due to variation in altitude or due to impacts of an urban area like an urban heat island. Known factors that possibly influence the air temperature measured by the vehicular sensors are listed in table 3, with highlighted local effects not to be considered during correction. For further weather-based signals, not described in more detail in the scope of this paper but evaluated in the cause of the project, other influencing factors need to be determined experimentally.

To quantify factors influencing air temperature, the exact position of the sensor in the vehicle is essential. The readings from multiple thermometers installed in each vehicle differ substantially due to their disparate mounting positions in the vehicle. Two exemplary chosen sensors are discussed in the following. Both are mounted in the front end of the vehicle. The first sensor is located in front of the radiator grill whereas the second sensor is attached under the windshield cowl, behind the engine (compare Figure 4). Both temperature values measured simultaneously during the same test drive are displayed in Figure 5. The difference between the two measured temperatures is significant and volatile. Around 08:45 UTC the temperatures differ by approximately 4 K whereas around 09:30 UTC their mismatch amounts to almost 30 K. The fluctuation of the temperature of the second sensor is more pronounced than the absolute temperature range of the first sensor. The experiments presented in this paper have incidentally been carried out on an AUDI S6 with a V8 bi-turbocharged engine. This engine configuration does not represent the average in the field and thus heat emissions are significantly higher than expected average heat emissions of passenger vehicles. Prospective experiments will take data from other configurations including electric vehicles into account. Meanwhile, this wide range of vehicle configurations helps to cover corner cases, and ensures the wide applicability of the algorithms developed for the analysis of car sensor data.

A first comparison of the displayed raw data from 25th January to reference data yields that all measurements by vehicular sensors deviate from the used reference data. Nonetheless, the peak temperature of the first sensor of 1.5 °C is more realistic and closer to the reference data than the temperature value of the second sensor with a maximum of 25.6 °C (Figure 5). In regard to a meteorologically representative measurement, the data from the first sensor in front of the radiator grill are chosen as basis for further evaluation. At this location the falsifying impacts of the surroundings are far less severe, but a physical correction is still required since the measurements at this location are still affected by the surrounding vehicle, although not as strongly as in the alternative locations.

An individual filter algorithm is implemented for all known falsifying factors on the considered meteorological signals. Combining the multiple filters to one is not feasible without distorting the correction due to interdependence of the different influencing factors. The order of the filters is fixed in the algorithm and was originally decided based on the extent of the individual influencing factor. The bigger the extent, the earlier the related filter is executed.

For the quality of the data it is crucial that locally induced real effects are not counteracted. The altitude is the environmental factor with the most significant influence on the air temperature. In Figure 6a the already known raw data of the air temperature (compare to Figure 5) are displayed. The comparison to the altitude of the vehicle shown in Figure 6b indicates the impact of this environmental factor. Generally it can be said that a rise in altitude leads to a decrease in temperature. This becomes particularly clear between 07:19 and 07:29 UTC. The opposite effect can be seen between 07:44 and 07:59 UTC where a decrease in altitude results in an increase in air temperature. These fluctuations induced by the altitude or other environmental circumstances must not be filtered in the following. Otherwise the vehicle data cannot resolve local effects and thereby the potential of vehicle data for all kinds of applications is narrowed.

As of writing, the environmental condition with the largest falsifying influence on the air temperature is the vehicle speed (compare Figure 6c) together with the engine temperature (compare Figure 6d). The lower the vehicle speed, the less air is taken through the radiator grill, which, in turn, leads to heat accumulation in the front end. This effect is particularly distinct for the prototypical vehicle used in the project due to the V8 turbo engine. For clarification, the scenario of a vehicle that stops or is being caught in slow, heavy traffic is described. When a vehicle with an engine at operating temperature stops, the air flow beneath the hood is cut off. Therefore, no fresh air with a temperature representative of the surroundings passes the thermometer. Instead, the air warmed-up by the engine is stuck beneath the hood and thus recorded by the thermometer. As a result, a rising ambient temperature is observable for example at 09:04 and 09:24 UTC (Figure 6a). This distorting effect is still visible during and after accelerating following a stop. In figure 6a this impact can be seen from 09:08 to 09:14 UTC, when the vehicle accelerated again after the previous stop and is driving with speeds varying in between 30 to 140 km/h (Figure 6c). Consequently, it is insufficient to only correct the temperature signal if vehicle speed falls below a certain threshold. Depending on the duration of the complete stop, the duration of vehicle speeds below a certain threshold, respectively, the air temperature needs a certain time to reach the ambient value. The algorithm developed compensates for the observed system's thermal inertia. For engine temperatures below the operating temperature of typically 90 °C, the described effect is observed limitedly. If the engine temperature is below 5 °C, this effect can be neglected completely. When the engine temperature increases according to engine run time (compare Figure 6d), the effect becomes visible and has to be taken into account.

Furthermore, the observed effect of allegedly rising air temperature depends on the rotational speed of the two radiator fans right behind the thermometer (compare Figure 6e). For a running engine both fans run at least at 10 % of their maximum capacity. A broad variety of conditions leads to changing rotational speeds of these fans. Especially reducing the rotational speed leads to an increase in the measured air temperature due to reduced ventilation at the thermometer. In Figure 5e a sudden increase in the fan performance from 10 to 33 % (fan 1) respectively 27 % (fan 2) can be observed at 09:04 UTC. In the following 5 minutes until 09:09 UTC the fan performance decreases (compare Figure 6e). Simultaneously, the vehicle stops which leads to an increase of the measured air temperature as described above. By the additionally overlaying decrease in fan performance the effect of the rising temperature is strengthened (compare Figure 6a).

On top of the described correction of the systematic deviations, the confidence for the corrected value of each individual quantity is estimated. The assigned confidence level declines while corrections by the algorithm are in effect. With a decline of the distorting effect the confidence is being elevated again. This ensures that measurements during ideal operating conditions maintain the highest confidence levels.

Observations have proven limitations of the correcting algorithm under marginal conditions. If not correctable with reasonable effort, these data points need to be identified and discarded by the algorithm before further processing on the backend.

To assure the quality of the processed sensor data and the reliability of the implemented algorithm, validation based on test drives including calibrated reference stations such as synoptic stations (see chapter 3.2) is conducted. Typically, reference data are only available for individual sections of the test drive due to the limited representativeness of each station depending on horizontal and vertical distance as well as differences in environmental conditions. Naturally, the location of the vehicle and therefore the surroundings change. Hence, reference data from reference stations in different surroundings with differing circumstances are of limited use for validation.

Ideally, data from multiple reference stations are available simultaneously to monitor possible deviations in the particular quantity. Although most of the measurement equipment of the reference stations is calibrated, a higher uncertainty of measurement compared to typical laboratory experiments has to be taken into account. Thus, data from two different reference stations may vary by a small percentage despite similar surroundings. The described deviations demonstrate the variance of the considered signal. As part of the validation process of the vehicular data only reference data with comparable environmental conditions are used.

An acceptable error margin is yet to be defined in accordance with the requirements of the weather service provider. For an eventual statistical assessment of the algorithms developed, reproducible driving scenarios along routes covered by reference stations as widely as possible are planned in cooperation with the weather service and the road authorities.

The algorithm developed in the project is designed vehicle-agnostic to support transfer to vehicle types different to the project vehicles by solely employing a dynamical correction of physically influencing factors, regardless of highly specific aspects such as the aero-dynamics of the vehicle body. The degree of correction is adjusted exclusively to the determined impact of the influencing effect. Thereby the algorithm automatically adapts to each vehicle individually, independent of vehicle properties, such as body type or engine size.

4. 2 Statistical Correction of vehicle data

Independent from the previously presented correction of systematic influences in the context of the individual vehicle, a statistical correction is performed on the backend based on collected fleet data. The statistical correction is used to define the bias of the individual sensors.

On a swarm data infrastructure, a diversity of sensors are expected. A large data volume exists, collected by different sensor designs with various measurement accuracies and step widths from a variety of vehicle types. Ideally, for each meteorological quantity the values are Gaussian distributed with a small variance between the values obtained from different vehicles.

Anticipating the future spatial aggregation of vehicle sensor data for anonymization purposes, the individual vehicle sensor data have to be treated in a statistical way to ensure best-possible data quality. Per aggregated area (polygon), only one value will be available for each quantity of vehicular measurements, instead of the whole collective of all $n_{vehicle}$ vehicles. A value representative for the polygon is determined by building the weighted average of all available data of the respective meteorological quantity with respect to the confidences and the bias assigned to the values, compare Equation (2). Data with a confidence below a certain threshold are excluded from the averaging, whereas data with confidence above the threshold are considered with the weight depending on their assigned confidence. Using a weighted average ensures that the average is substantially based on trustworthy measurement values and not distorted by faulty measurements.

$$j = 1, \ldots, n_{vehicles} \quad (1)$$

$$\mu_{Polygon} = \frac{1}{n_{vehicles}} \sum_{1}^{n_{vehicles}} \left\{ \left(x_j - \Delta x_j \right) * \frac{c_{x,j}}{100} \right\} \quad (2)$$

$$\Delta x_j = \alpha * \Delta x_{j,old} + (1 - \alpha) * (x_j - \mu_{Polygon}) \quad (3)$$

First, the extracted value from a vehicle x_j is adjusted by its bias Δx_j and subsequently multiplied with its confidence $c_{x,j}$ [%] divided by 100. This step is carried out for every vehicle. The sum over all vehicular values is calculated and divided by the amount of vehicles to obtain the weighted average $\mu_{Polygon}$ for a considered meteorological quantity x.
On basis of the average, the bias Δx_j for every meteorological quantity of each sensor is estimated (compare Equation (3)). The current bias is the deviation from the individual value x_j to

the weighted average $\mu_{Polygon}$. To monitor a possible drift of a sensor the bias from previous time steps are also considered as shown in equation 3. An α of for example 0.8 leads to a new bias Δx_j based to 80 % on previous deviations and to 20 % on the current bias for this sensor.

5. Preliminary Results

First results of the algorithm aiming to filter the bias induced by the vehicle dynamics as described in chapter 4.1 are presented in Figure 7a. The raw data of the air temperature (see figure 6a) are marked in grey, the corrected air temperature values based on the current state of the algorithm in blue. The uncertainty of measurement of the sensor (shaded in blue) is transferred to the corrected values without further adjustment. Thus, currently no uncertainty accounting for the correction is taken into account.

For the presented test drive three relevant sources of reference data are available as illustrated in figure 8. Data for the average ambient temperature measured by the synoptic station Gelbelsee are available every 10 minutes. One road weather station at the rest area along the freeway within a linear distance of about 500m (550 yards) from the synoptic station supplements the evaluation (compare figure 8b). 23 km (14 miles) up the freeway, a similar road weather station setup has been installed at the rest area at Offenbau (compare figure 8a). Both road weather stations at Gelbelsee and Offenbau are extended setups for which averaged data are available every minute instead of every 15 minutes. Reference data of all three stations are available for the whole extent of the test drive, but are only displayed when the linear distance to the vehicle was equal or less to 1 km (compare figure 7b) due to the limited representativeness in different environmental conditions.

During the evaluated test drive the synoptic station Gelbelsee is accessed (distance of vehicle to synoptic station smaller than 1 km) first from 07:28 to 07:44 UTC. Through these 16 minutes the vehicle was driving around the synoptic station with varying, radial distances (figure 7b). Due to the temporal resolution of 10 min of the synoptic station there are two reference values for this time span. In both cases the air temperature based on vehicle data is higher than the air temperature measured by the synoptic station. At 07:30 the total difference amounts to 0.9 K, at 07:40 UTC to 0.8 K.
Due to the linear distance of 500 m between the synoptic station Gelbelsee and the road weather station Gelbelsee it seems plausible that a comparison of the vehicular data to data from the road weather station for this time span is beneficial, but both measurement systems are exposed to different environmental conditions. The road weather station is set up next to the crash barrier of the freeway, whereas the vehicle measures on a dirt road surrounded by a meadow. Therefore, the vehicle data in the time span from 07:28 to 07:44 UTC should exclusively be compared to the data from the synoptic station as carried out above.
For the second period with the vehicle driving on paths around the synoptic station (09:13 - 09:34 UTC) the same effect of vehicular air temperatures being higher than ambient temperatures from the synoptic station is observed. For the time span of 21 minutes two ambient temperature values from the synoptic station differing by 0.1 K among each other are available. The vehicle data accounts for a temperature value in average 0.8 K higher than the reference data.
The accordance of the vehicular data and the road weather station Gelbelsee data for these two periods is not relevant due to the different environmental conditions as described above.

From 07:51 to 07:53 UTC and 09:01 to 09:08 UTC the vehicle stayed within a range of 200 m to the road weather station Gelbelsee. The vehicular data is measured closely to the freeway above asphalt for the selected time intervals. Hence, the surrounding conditions can be assumed identical. For both periods the vehicular air temperature is higher than the reference data from the road weather station Gelbelsee. In the first comparison interval of 2 minutes the difference accounts for 0.8 K. In the later period lasting 7 minutes one value of the reference data is within the bounds of measurement uncertainty of the vehicular values, but the majority differs by averagely 0.9 K compared to the corrected vehicular temperature values.

A similar picture emerges for both time spans with the vehicle staying next to the road weather station Offenbau. The vehicular air temperature is continuously higher than the reference data. In the period from 08:05 to 08:12 UTC they differ on average by 1.0 K, for the later period from 08:48 to 08:50 UTC by 0.6 K. For the second time interval in two of three cases the reference data from the road weather station do not exceed the error margins of the vehicular data.
It is noticeable that the temperature of the road weather station at 08:07 UTC lies 0.7 K above the previous and the following temperature value.

In addition, local differences in the vehicular data are noteworthy. Since the differences have been detected by the vehicle the according measurements took place with a temporal offset. Generally, comparing vehicle data from different locations automatically involves comparing measurements from different time intervals. Nevertheless, the comparison can be useful if the weather conditions do not change significantly at both locations during the meantime. During the evaluated test drive this condition is fulfilled with the sky being continuously cloudy without precipitation. Furthermore, no sunrise or sun set took place. The observed local differences are the following: First, the vehicular temperature values close to the road weather station Gelbelsee are consistently lower than those close to the road weather station Offenbau. There is a difference of 68 m in altitude between both stations. Secondly, the vehicular temperatures close to the road weather station Gelbelsee are above the temperatures measured close to the synoptic station Gelbelsee. The vehicle data indicate temperatures warmer by at least 0.48 K at the road weather station compared to the synoptic station which is underpinned by constant temperature values of both reference stations also differing by an almost constant offset.

6. Conclusions and Outlook

According to [Karisto and Nurmi 2016] using car observations in road weather forecasts reduces the forecast error considerably. Also [Bridge and Johns 2016] state the advantage of being able to measure road weather variables for the entire road network, e.g., by vehicle sensors, instead of relying on single point measurements from static road weather stations alone. The National Research Council of the United States [2010] highlighted that "Observational data with high temporal and spatial resolution are crucial to […] providing data for assimilation on models […]. This requires the synergistic combination of data from diverse sources.". This paper shows promising results in terms of the accuracy of vehicular air temperature data with high temporal and spatial resolution, encouraging the use of this data in addition to stationary observational data. Different mobile and stationary measurement stations have been introduced. These allow for close-to-production vehicle sensing or possess quality close to weather reference stations. As displayed in Figure 7 the current development of the algorithm used to correct the raw data measured by the vehicle successfully identifies and compensates structural deviations. The vehicle speed and the heat emission of the engine, respectively, are as of writing the major falsifying factors corrected by the algorithm. To conclude, it can be stated that the currently known erroneous fluctuations of the measured air temperature induced by the vehicle dynamics are detected and suppressed by the algorithm.

The evaluated test drive indicates a general offset of the vehicle data compared to the used synoptic station as well as the road weather stations. If the offset is attested by further test drives the algorithm will be expanded by a correction of the offset for all air temperature values measured by the vehicle.

Besides filtering the structural deviations the algorithm leaves fluctuations induced by highly local yet natural effects unchanged. This is a prerequisite for using vehicles as mobile measurement systems since local effects particularly matter for weather forecasts with high spatial resolution as they are needed for autonomous driving functions or winter maintenance services. In the evaluated test drives the primarily influencing local effect is the change of altitude. For more significant changes in altitude over a short period of time, the adjustment of the temperature spans over a longer period than the triggering factor due to the inertia of the sensor itself and its limited logical step-width of 0.5 K. Comparing the elevation profile to the measurement plots suggests a strong, local influence of the altitude. However, local effects such as this one are valuable local information and should thus not be corrected.

With the help of the elevation profile the noted differences within the reference data of the two road weather stations can be explained for the available data. The difference in altitude of 68 m may explain the deviation in temperature. Using the Boltzmann barometric equation assuming a linear decrease of temperature with height an air temperature of 0.44 K lower is expected for a change in altitude of 68 m [Berberan-Santos et al., 1996]. This value for a temperature difference induced by height fits to the temperature difference observed on average between the road weather stations Gelbelsee and Offenbau. Whether the occurring difference between these two stations can continuously be explained by this effect is to be investigated.

However, the deviations between the synoptic station Gelbelsee and the road weather station Gelbelsee cannot be explained by a different altitude, since their elevation only differs by 12 m. Neither an offset induced by differing calibrations of the measurement systems between the two stations seems to be the explanation, since the difference is also visible in the vehicular data. Instead the difference between the two stations is most probably induced by other local effects. In particular, this includes the different surrounding conditions. The synoptic station located on a meadow is not subject to effects induced by a surrounding with high traffic volume, for example radiation heat from the asphalt and turbulences caused by passing traffic. These local variations have to be considered for cross-checking vehicular data. The difference between the two stations located within 500 m linear distance demonstrates the limited spatial representativeness of measurements for meteorological applications with respect to road conditions and thereby highlights the need for fleet data to enable weather services to improve the spatial and temporal resolution of forecasts as well as their applicability to road weather.

Furthermore, the used reference data from the road weather station Offenbau illustrates that even reference data have to be examined critically in order to detect outliers and exempt these from further processing. As an example the substantial difference in ambient air temperature measured by the road weather station Offenbau has been mentioned in chapter 5. The significant difference of a single measurement value to adjacent values along the timeline indicates faulty data.

Overall, the paper provides first promising insight into using vehicular data for meteorological applications. Raw data from the vehicle have been proven not to match unconditionally with reference data. Nonetheless, a promising yet not fully verified way of correcting these sensor readings has been introduced. In the future the algorithm will be tested with data from further test drives and will be improved continuously.

In the course of the project FloWKar, further quantities including precipitation and road temperature will be evaluated, corrected if necessary and validated in upcoming experiments, aimed to cover most of the meteorological spectrum with vehicular sensors.

REFERENCES

Berberan-Santos, M. N.; Bodunov, E. N. and Pogliani L. 1996. "On the barometric formula". American Journal of Physics. Vol. 65 No 5, p. 404 – 412

Bridge, P.; Johns, D. 2016. "Road Weather Information Systems Reference Data" On: Standing International Road Weather Commission (SIRWEC) 2016

Brown, K.; Hernandez, E.; Kyrtsos, C. 2018. "Method of Determining Ambient Air Temperature" U.S. Patent No. 7,387,437, issued June 10, 2008

Habrovsky, R.; Bujnak, R.; Tarjani, V.; Vivoda, J. 2018. "The Overview Of Road Forecast Activities At Slovak Hydrometeorological Institute" In: Book of Abstracts of SIRWEC 2018, On: Standing International Road Weather Commission (SIRWEC) 2018

Karsisti, V.; Sukuvaara, T. 2018. "Verification Results For Road Surface Temperature Forecasts Utilizing Mobile Observations" In: Book of Abstracts of SIRWEC 2018, On: Standing International Road Weather Commission (SIRWEC) 2018

Karisto, V.; Nurmi, P. 2016. "Using car observations in road weather forecasting." On: Standing International Road Weather Commission (SIRWEC) 2016

Kintz, E.P. 2000. "Temperature Display for a Car Stereo" U.S. Patent No. 6,065,868, issued May 23, 2000

KNMI Datalab; Royal Netherlands Meteorological Institute 2018. "Car sensors as mobile meteorological network" [online]. Updated: October 2018 [viewed 31 March 2019] http://projects.knmi.nl/publications/fulltexts/product_datalab_carsensordata_v6.pdf

National Research Council 2010. "When Weather Matters: Science and Services to Meet Critical Societal Needs." Washington, DC: The National Academies Press. https://doi.org/10.17226/12888.

National Research Council 2004. "Where the Weather Meets the Road: A Research Agenda for Improving Road Weather Services." Washington, DC: The National Academies Press. https://doi.org/10.17226/10893.

Niskanen, A.; Tuononen, A. J.; Laine, J. 2018. "Fleet Based Road Weather Monitoring" In: Book of Abstracts of SIRWEC 2018, On: Standing International Road Weather Commission (SIRWEC) 2018

Perchanok, M. 2016. "The State of Road Weather Research" On: Standing International Road Weather Commission (SIRWEC) 2016

Poublon, M. J. 2000. "Ambient Temperature Learning Algorithm for Automotive Vehicles" U.S. Patent No. 6,088,661, issued July 11, 2000

Teconer "roadweather" [online], [viewed 01 April 2019] https://roadweather.online

White, S. P.; Thornes, J. E.; Chapmeann, L. 2016. "A Guide to Road Weather Information Systems" On: Standing International Road Weather Commission (SIRWEC) 2016

TABLES

Table 1: Meteorological quantities deduced or directly measured by vehicular sensors

Meteorological quantity	Series application	Unit	Resolution	Uncertainty / Accuracy
Atmospheric pressure	yes	hPa	7.9 hPa	±10 hPa
Air temperatue	yes	°C	0.5 K	±0,5 K (from 0 to -40 °C) ±0,3 K (for 25 °C)
Relative humidity	yes	%	0.7 %	±3 % (from 20 to 80 %RH)
Dewpoint temperature	yes	°C	0.1 K	
Precipitation	yes	$ml/m^2/s$		
Global radiation	yes	W/m^2		
Surface temperature	yes	°C	0.1 K	±0.8 K (for 0 °C)
Road friction	yes	-		
Visible distance	no	m		

Table 2: Quantities regarding vehicle properties measured by vehicle

Vehicle quantity	Unit
Latitude	°
Longitude	°
Altitude	m
Vehicle speed	km/h
Engine on/off	I/O
Engine temperature	°C

Table 3: Influencing factors on the air temperature; local effects not to be counteracted highlighted in grey

Meteorological quantity	Dependent on
Air temperatur	• Altitude • Local effects, for example: o Urban heat island o Areas where cold air is trapped o Inversion effects • Operating status engine (on/off) • Operating status ignition (on/off) • Vehicle speed (absolute value) • Engine temp • Duration of lower deviation of threshold for vehicle speed • Duration since when above threshold for vehicle speed • Fan performance • Global radiation

FIGURES

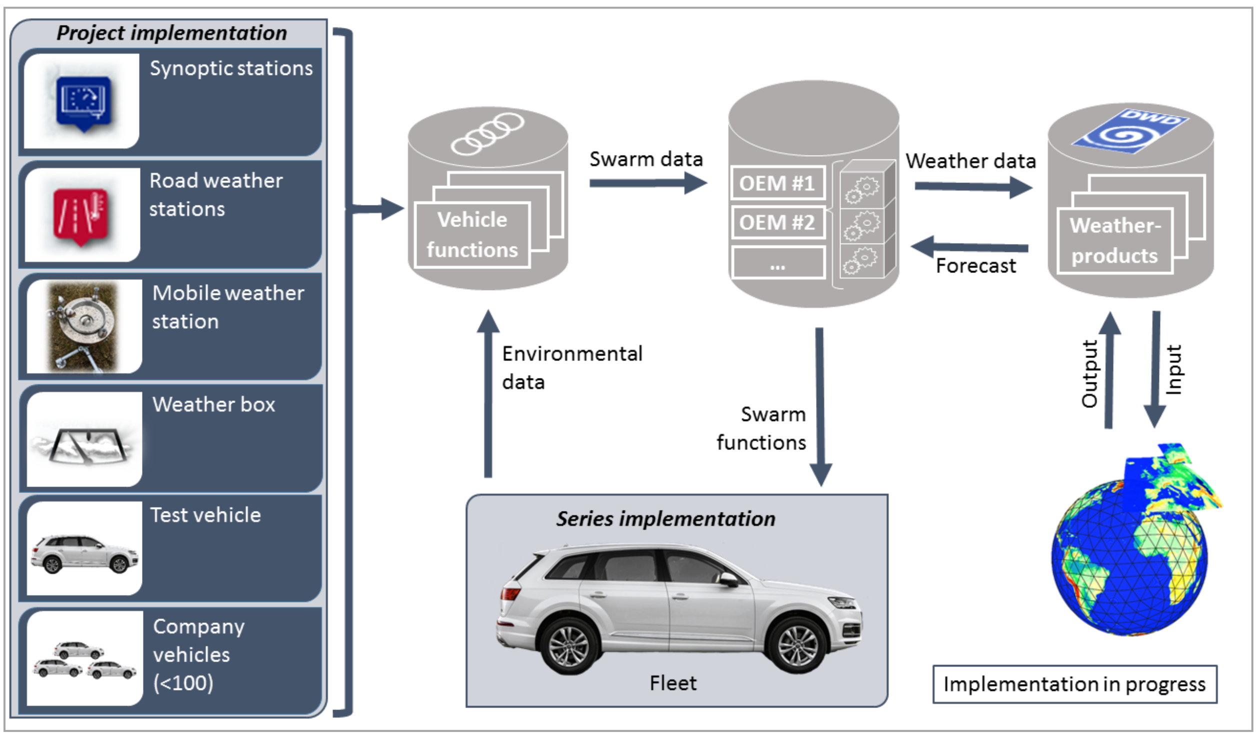

Figure 1: Technical implementation of the data transfer regarding meteorological application of diverse data sources for the project FloWKar

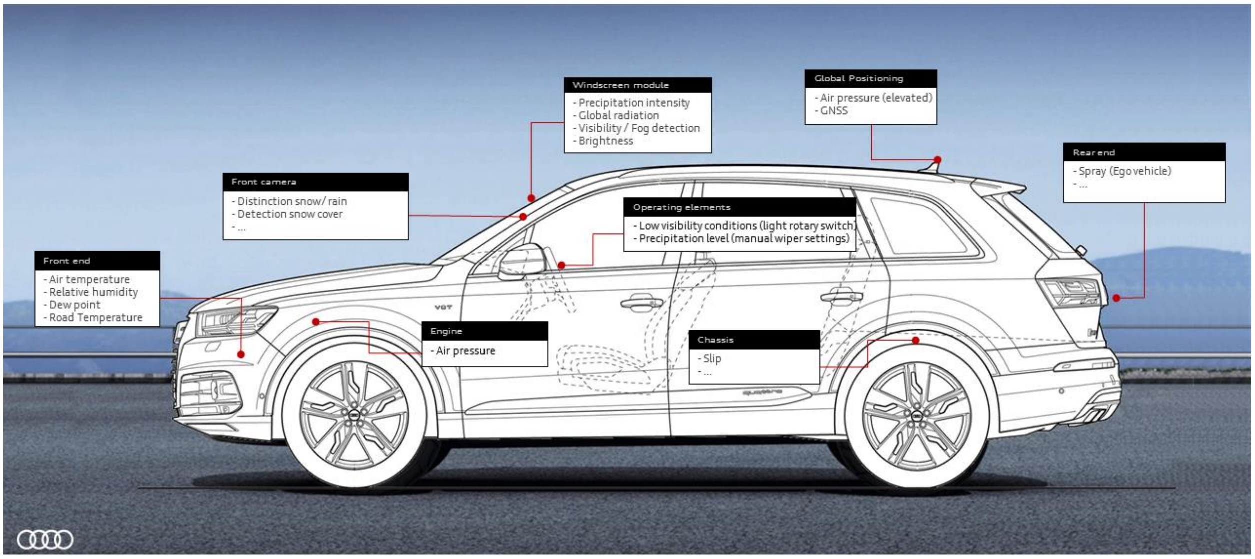

Figure 2: Overview of mass-produced sensors on vehicle

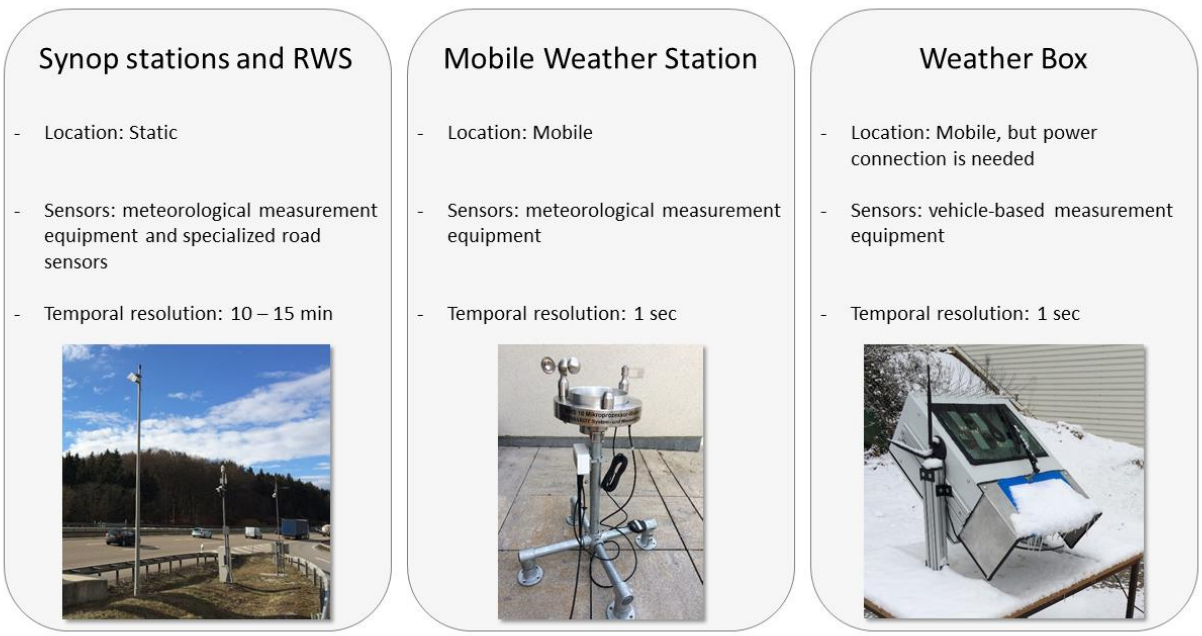

Figure 3: Available reference stations

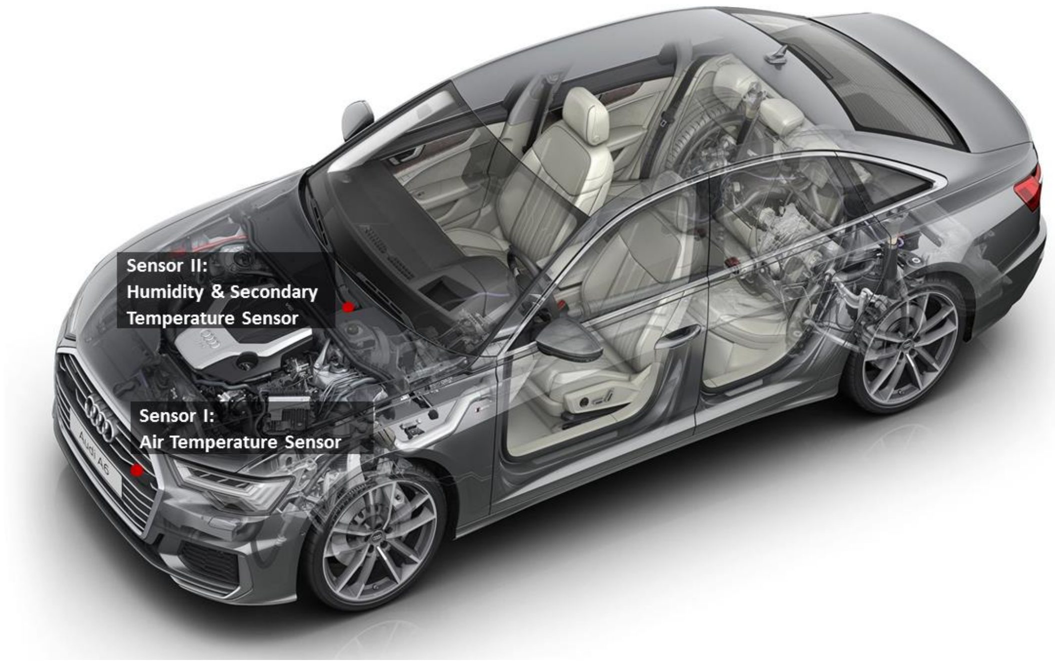

Figure 4: Mounting positions of the first and secondary air temperature sensor

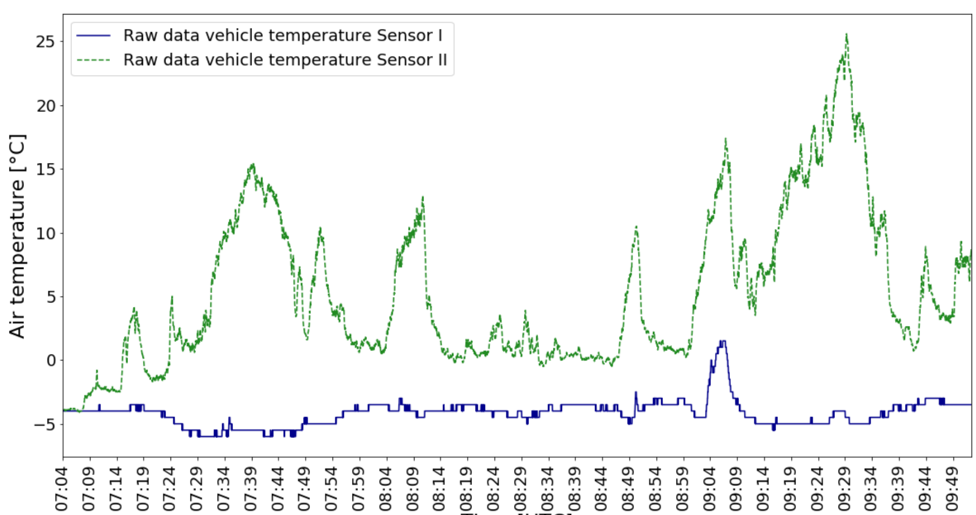

Figure 5: Air temperature measured by different thermometers at different mounting positions in the vehicle

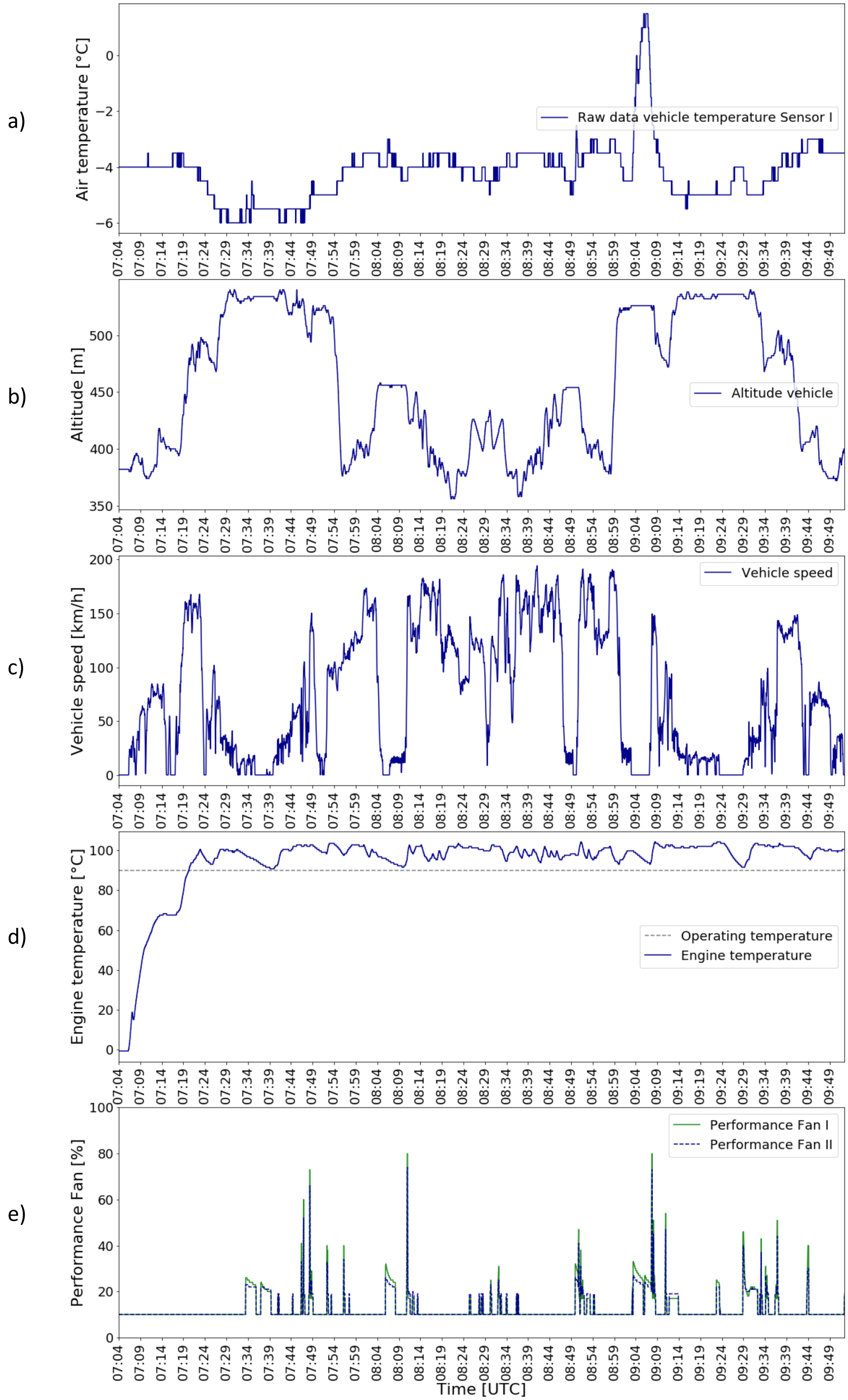

Figure 6: Extract from the raw data collected by the vehicle a) Air temperature Sensor I b) Altitude c) Vehicle speed d) Engine temperature e) Performance of fan I and II

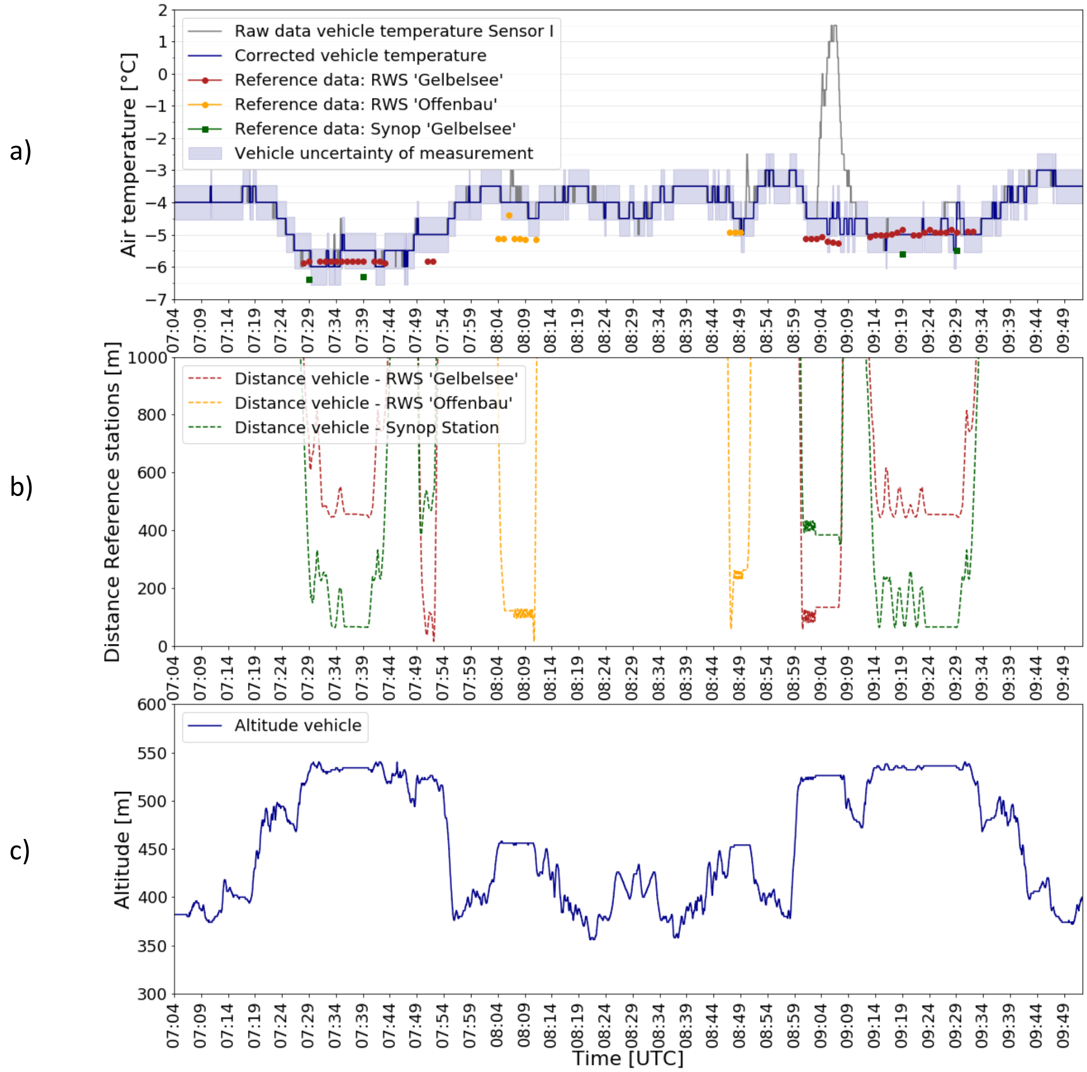

Figure 7: a) Corrected air temperature of Sensor I in comparison with reference data (RWS: road weather station) b) Distances to reference stations c) Altitude of vehicle

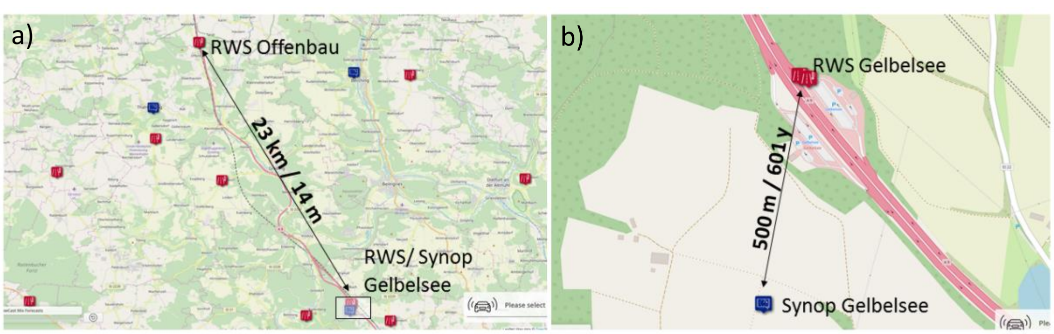

Figure 8: a) Area of experiment along the German freeway A9 between the road weather stations at Gelbelsee and Offenbau b) Zoom into the area between synoptic station and road weather station Gelbelsee