USING SECONDARY DATA TO REDUCE MEASURING EFFORTS IN NATURALISTIC DRIVING OBSERVATION

FOR ENERGETIC SIMULATION AND EV-OPTIMIZATION

Jens-Olav Jerratsch, Thomas Herzlieb, Stefanie Marker
Technische Universität Berlin,
Chair of Naturalistic Driving Observation (TUB-FVB)
Berlin, Germany
Jerratsch@tu-berlin.de

***Abstract*— For advanced simulation of electric vehicle total energy consumption various parameters need to be considered. Mapping secondary data by time and location can help to reduce measurement efforts and enrich existing data. The methodology presented for ambient temperature mapping yields a mean difference of only 1.5 °C when compared to sensor acquired onboard data. Using TUB-FVB's total vehicle energy simulation for a small commercial vehicle this leads to a mean deviation in consumption of only 0.17 kWh/h or 2.6% respectively.**

***Keywords*— naturalistic driving observation, data logging economy, secondary weather data, energetic EV-simulation**

I. INTRODUCTION

Range anxiety and high (battery) costs are most important issues to be considered for mass market initiation of electric vehicles [1].TU Berlin has developed a total vehicle energy simulation that calculates optimal battery sizes for individual driver behavior [2] [3]. In order to obtain reliable results various input parameters (i.e. speed, slope, temperatures, humidity, solar conditions) are needed for energetic simulation. This is due to different traction and auxiliary components that in an electric vehicle (EV) draw their energy demand from the propulsion battery. Usually, these parameters are measured onboard during vehicle usage. If a vehicle lacks of built- in sensors or they cannot be read, for example because CAN-bus data is not available or cannot be decoded, additional sensors have to be installed for data acquisition purposes. From an economic perspective, especially in large scale observations, it is thus not always feasible to directly collect all needed data. Also large parts of existing data sets miss out on some of these parameters [4], [5], [6].

Temperature is a critical factor to estimate energy consumption in an electric car. Especially conditioning of battery and the passenger cabin on cold winter days can make up more than 50% of a vehicle's total energy demand [7]. For accurate energy demand estimation in individual cases, it is important to not only rely on statistical climatic conditions (i.e. average or min/max temperatures), but to make actual time and location specific temperatures available. Among other weather data, temperature can be obtained from (secondary) sources independent from vehicle operations. An example is the climate data center of Germany´s National Meteorological Service (DWD) [8].

In 2017 the project EN-WIN [9] was started with the goal to compare the usage and usability of light and heavy commercial vehicles in detail when driven by conventional and electric propulsion systems, respectively. To thoroughly understand energy demand and range-affecting factors a very diverse set of data from within and around the vehicle is to be assessed. In this context, TU Berlin's chair for naturalistic driving observation (TUB-FVB) has developed a methodology that uses secondary data to reduce measuring costs and enrich available data for advanced energetic analysis.

II. DATA BASE FOR TOTAL ENERGY SIMULATION FOR ELECTRIC VEHICLES

Compared to conventional vehicles the energetic demand of an EV can be simulated more accurately due to lower powertrain complexity. Most important parameters for energy demand are operational time, speed, slope, ambient and inner air temperature and humidity, solar conditions and actual vehicle mass. To obtain large data sets logging equipment needs to be simple, cost efficient and non-disturbing for drivers and passengers. Logging of vehicle speed, location and inner air conditions can be executed with standard GPS-loggers and simple interior climate sensors. If the vehicle's built in sensors cannot be used (i.e. because of CAN message encryption) measurement of data outside the vehicle tends to be more complex, pricy, subject to errors, and can cause deficits in passenger comfort due to wiring and outside sensors. Monitoring of outside temperature or humidity implies paying attention to solar conditions and shading, measurement of geographic altitude is often inaccurate or not available using standard GPS-devices. Generally, these outside data are not dependent on the vehicle but rather on the environmental conditions and can thus also be acquired through other measurement setups.

III. SECONDARY DATA ENRICHMENT METHODOLOGY

To minimize measuring cost for individual vehicles other data sources can be used to augment primary time- and location- (GPS-) based data. These data are referred to as secondary data in the study. This paper focuses on ambient temperature conditions to outline the underlying methodologies and present examples to the reader. Other relevant secondary data that can be used for enrichment are for instance additional weather parameters (precipitation, wind speed, solar radiation, humidity, etc.) or geographic topology.

Firstly, suitable data sources are to be identified. Weather stations throughout Germany can be analyzed to obtain climatic information. Most relevant is regional and temporal resolution as values are not available for every location and at any time. For weather data in Germany DWD's "climate data center" [8] provides hourly[1] temperature data (2m above ground) from more than 500 weather stations throughout Germany free of charge. Fig. 1 provides an overview on station distribution in the country.

The next question to be resolved is how to map the secondary data to time and location of the relevant vehicle and what accuracy can be expected. This problem is very common in meteorology [10] and this paper compares two basic methodologies to address the task in the vehicle data context: next neighbor (NN) and inverse distance weighting (IDW). More complex methods have not shown significantly better results [11] and are therefore not assessed in this context mainly due to higher costs and longer calculation times.

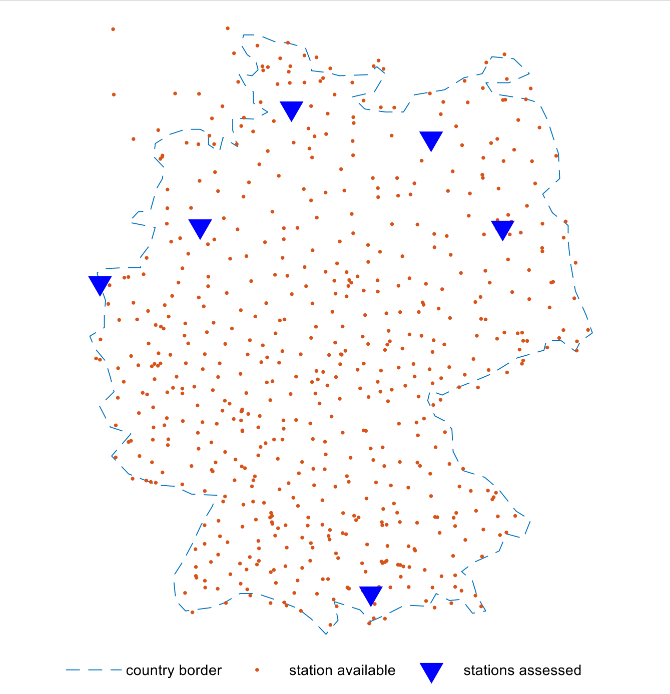

Fig. 1. DWD's station network for ambient temperature (own representation of [12])

[1]As of late the data resolution provided has been improved to intervals of 10 minutes by DWD which has not been considered so far in the example presented but is supposed to enhance the data quality in general.

The next neighbor method is the simplest method and finds the weather station with the lowest direct air distance for every route coordinate and uses the corresponding weather data to enrich the vehicle data. The main advantage is the very short computing time.

In order to efficiently apply the inverse distance weighting method a pre-selection of weather stations nearest to the vehicle is applied. These nearest stations are defined as those located in an area of right angular shape spanned by the route edge coordinates. This area is then iteratively enlarged so that it includes at least four weather stations. IDW then calculates a weighted average with respect to the distance between station and vehicle as displayed in equation (1). The term $z_{x,y,t,q}$ represents the estimated value at a (vehicle) location with coordinates x, y at time t. Its value is calculated of the weighted average over all selected n stations with $z_{i,t}$ being the measured data from the i^{th} station at time t, $d_{i,x,y}$ the distance between the location x, y and i^{th} station, and the weighting factor q that has to be defined individually to the respective case [13].

To evaluate the general accuracy of the data mapping, the average absolute difference $\bar{z}_q$ between estimated ($z_{x,y,t,q}$) and measured values at the location of the vehicle ($r_{x,y,t}$) for all m available data points of a test data sample is calculated (2). The optimal value for q is found by minimizing the average difference $\bar{z}_q$ over a range of q values (3). Hourly measuring values from DWD are linearly interpolated for relevant stations at the corresponding time intervals.

$$z_{x,y,t,q} = \frac{\sum_{i=1}^{n} z_{i,t} d_{i,x,y}^{-q}}{\sum_{i=1}^{n} d_{i,x,y}^{-q}} \quad (1)$$

$$\bar{z}_q = \frac{\sum_{j=1}^{m} \left| z_{x,y,t,q} - r_{x,y,t} \right|}{m} \quad (2)$$

$$q_{opt}: \min_{q \in [0.2 \ldots 5]} \bar{z}_q \quad (3)$$

Due to varying station density and parameter type also the accuracy of mapped secondary data values is varying. Assessing ambient temperature exemplarily, the following analysis is to show that sufficient data accuracy can be reached by mapping secondary data using the proposed method.

IV. VALIDATION

The presented method is validated in three steps: Firstly, next neighbor and inverse distance weighting methods are compared regarding general accuracy in order to decide what method is to be preferred further on. To avoid measuring issues regarding mobile measuring equipment and to make use of a wide data range, the DWD temperature data itself is used for this step. Data from one station at a time are erased from the database. Then the remaining stations are used to map weather data to the location of the erased station. The results are compared to the actual data from that particular station. Secondly the method that proved more reliable is used to map weather data to actual vehicle tracks. The mapped values are compared to data acquired onboard. Finally, TUB-FVB's total energy simulation is used to demonstrate the resulting difference in energy consumption using mapped data instead of real data.

For the stationary assessment the data from six weather stations have been selected in order to cover a range of different local characteristics like rural/urban, topography, station density, position in middle / at the edge of the station network. This is done in order to identify correlation between station characteristics and accuracy of mapped temperature values. In this way areas of potentially high accuracy can be separated from those with lower expected accuracy. TABLE I gives a summary of the stations assessed and their characteristics.

A. *Evaluation of proposed methods*

At first the optimal exponent q is identified to be 1.6 for the invers distance weighting method. Thus equation (1) can be simplified to equation (4)

$$z_{x,y,t} = \frac{\sum_{t=1}^{n} z_i d_{i,x,y}^{-1.6}}{\sum_{i=1}^{n} d_{i,x,y}^{-1.6}} \quad (4)$$

Fig. 2 displays a comparison of the results obtained by NN and IDW method respectively. The boxplots show the most relevant statistical parameters that describe the underlying distribution of differences. Both methods yield satisfying results. In average the deviation is 0°C, the absolute deviation is less than 1°C. Only very few outliers (about 1% of the regarded >50.000 data points) deviate more than ±5°C. In direct comparison shown in first two lines in TABLE II the inverse distance weighting method yields slightly better results, especially reducing the outliers. Thus, only the inverse distance weighting method is to be used further on in this study.

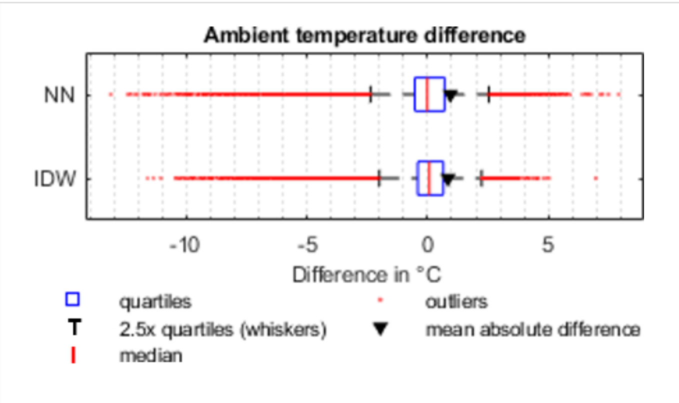

Fig. 2. Box plot of next neighbor (NN) and inverse distance weighting (IDW) methods' results in comparison.

TABLE I WEATHER STATIONS FOR TEMPERATURE ASSESSMENT

Station name	State	Characteristics
Tempelhof	Berlin	urban area, high station density
Goldberg	Mecklenburg-Vorpommern	low station density
Kleve	North Rhine-Westphalia	edge of the station network (border area)
Hohenpeissenberg	Bavaria	complex topography (alpine foothill region)
Itzehoe	Schleswig-Holstein	coastal region
Alfhausen	Lower Saxony	average station density

B. *Results using inverse distance weighting method*

Fig. 3 displays the results for the inverse distance weighting method for the six stations presented in TABLE I individually. The station "Hohenpeissenberg" (985m above sea level) shows the widest spread in deviations and proves to be responsible for practically all deviations below -5°C. The reason is the complex alpine foothill topography in the proximity and varying altitudes in this area. A possible explanation for high negative outliers (real temperature is higher than calculated temperature) is the appearance of inverse weather conditions.

The "Kleve" station is in the bordering area to the Netherlands and thus at the edge of the network. That explains the slightly larger deviations. The smallest deviations are observed in Berlin-Tempelhof. As there are multiple other stations in the direct proximity, this is not surprising.

In total this analysis shows that the method yields generally good approximations, only in complex (mountainous) topography sporadic deviations of more than ±5°C have to be expected. In all other regions deviations of more than ±2°C are very scarce.

C. *Improvement potentials*

In order to further improve reliability for stations in mountainous areas the impact of geographic altitude has been tested. Heyer [14] provides temperature gradients for mountainous areas that can be used to fit temperatures to geographic height of the targeted location. In average, the absolute mean deviation at Hohenpeissenberg station is improved from 2°C to 1.8°C. Regarding all six stations the absolute mean deviation is improved by only 0.04°C. This is due to very small altitude differences in the regions of all regarded stations except Hohenpeissenberg. As this study focuses on northern Germany, geographic height is neglected. In a different context it is strongly advised to take geographic altitude into account.

Another potential to improve reliability is the inclusion of other station networks to improve station density within Germany or to enlarge the area of application or to reduce edge effects respectively. As mentioned above the Kleve station yields slightly larger deviations than other northern German stations. The inclusion of additional station networks was rejected since it was (1) too cost-intensive, (2) too time-consuming due to additional data processing complexity and – most important – (3) the project EN-WIN focus is on tours within Germany, so that edge data would not need to be applied here. Again, in another context this improvement is to be considered.

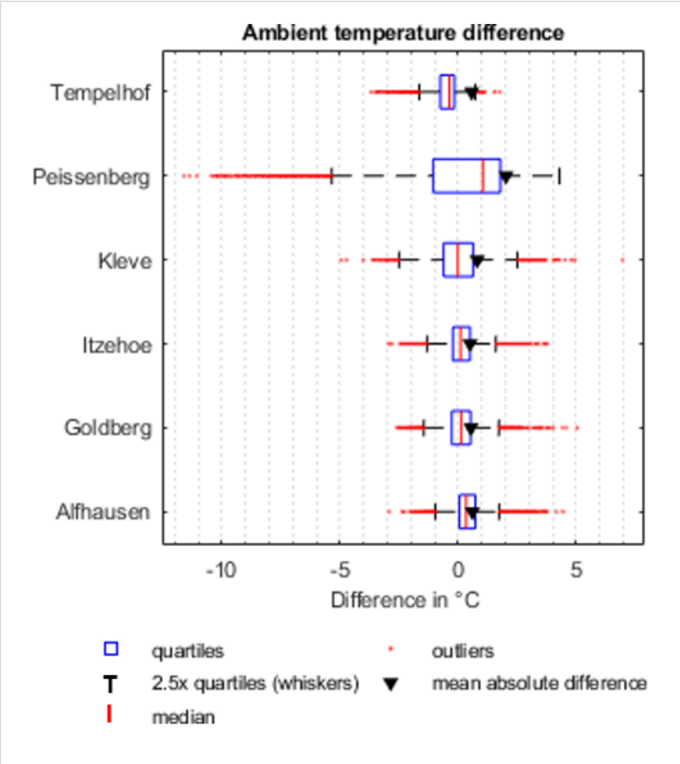

Fig. 3. Box plot of inverse distance weighting: results for six assessed stations individually.

TABLE II TEMPERATURE DEVIATION MEASURED VS SECONDARY DATA

Method station	Mean diff. $\bar{z}_q$ °C	Standard dev. °C	Mean abs. diff °C	25% quantile °C	75% quantile °C	Diff. ±2°C %	Diff. ±5°C %
NN all	-0.01	1.44	0.94	-0.53	0.70	89.2	98.8
IDW all	0,05	1.30	0.84	-0.41	0.65	91.2	98.9
IDW Alfhausen	0.44	0.59	0.60	0.05	0.73	96.1	100.0
IDW Tempelhof	-0.52	0.54	0.57	-0.76	-0.16	97.2	100.0
IDW Goldberg	0.14	0.49	0.54	-0.26	0.53	98.3	100.0
IDW Hohen-peissenberg	0.01	1.62	2.03	-1.07	1.79	63.5	93.45
IDW Itzehoe	0.17	0.49	0.50	-0.23	0.50	98.3	100.0
IDW Kleve	0.04	0.68	0.81	-0.62	0.64	93.8	99.9

D. *Mapping temperature to real vehicle trips*

This section gives an example how secondary data mapping is coupled to real driving profiles. For each route point the corresponding ambient temperature is calculated using the described IDW method and compared to the temperature acquired using the vehicle's intake air temperature sensor. Fig. 4 illustrates the temperature deviation over the course of a real driving profile in Berlin.

As electric heating is the main reason for increased energy consumption, ten winter days with high electric heating demand are selected to illustrate energetic impact of using secondary data mapping. TABLE III shows the results for all of the ten assessed trips in terms of daily mean values for actual and calculated temperature and the difference in energy demand using the two temperature values to feed the simulation. The energy demand is calculated using TUB-FVB's total vehicle energy model for a small commercial vehicle [3]. For the assessed cold days, the differences in energy demand vary by 0.17 kWh per hour or 2.6% of total vehicle energy demand on average. The last column illustrates that about 0.11 kWh additionally is needed per hour for every decrease of ambient temperature by 1K. The results show that small temperature deviations have minor impact on vehicle energy demand.

The generally higher differences between measured and mapped temperature values when compared to the stationary assessment allow for different interpretations. In all cases the measured temperatures are higher than the mapped. The largest differences appear in the first seconds of driving after a longer stop. This could be due to the vehicle sensor being influenced by local heat sources (i.e. waste heat from vehicle, warm road, buildings). More study would be necessary to assess whether any regularity can be derived to explain these differences.

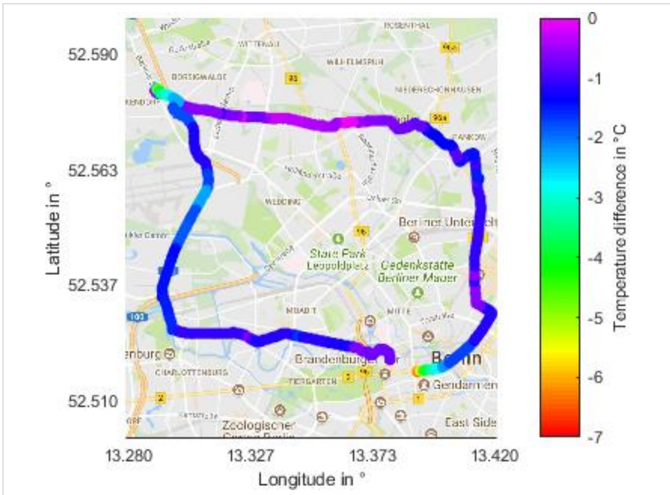

Fig. 4. Exemplary route profile in downtown Berlin illustrating difference between measured and calculated ambient temperature.

TABLE III COMPARISON OF ACTUAL AND SECONDARY TEMPERATURE DATA IN REAL DRIVING PROFILES ON SELECTED COLD DAYS (DAILY VALUES)

ID	Date mm/dd	Distance driven km	Temp. sensor $r_{x,y,t}$ °C (mean)	Temp. mapped $z_{x,y,t}$ °C (mean)	Vehicle energy demand difference (simulated) kWh/h	%	kWh/(hK)
1	11/26	51.1	2.9	-0.5	0.37	6.0	0.11
2	12/17	51.2	3.9	3.0	0.09	1.6	0.10
3	11/19	32.8	4.9	3.9	0.10	2.0	0.10
4	12/31	26.8	3.3	2.3	0.12	1.7	0.11
5	01/19	59.0	2.2	-0.4	0.31	4,6	0.12
6	12/12	113.6	3.2	2.5	0.07	0.8	0.10
7	02/04	62.0	0.8	-1.0	0.21	3.1	0.12
8	02/07	73.1	2.7	1.3	0.15	2.4	0.11
9	02/13	114.4	3.1	1.7	0.12	1.4	0.08
10	02/16	73.0	2.1	0.7	0.15	2.5	0.11
Avg	-	65.7	2.9	1.3	0.17	2.6	0.11

V. CONCLUSIONS

This study assesses the practicability of using secondary weather data in the context of driving profile data to reduce the need of onboard data acquisition or to enrich existing data that lack on data on ambient conditions, respectively. For the discussed purpose of vehicle energy demand simulation, the presented method yields good accuracy for ambient temperature data. In cases that do not allow reading built-in sensors or require additional sensors, secondary data might thus be used instead. Depending on intended application the potential of the presented mapping method for other secondary weather conditions like solar radiation, precipitation, or wind speed or topologic information like altitude has to be assessed.

Accurate weather data mapping is challenged by certain conditions: This is shown for ambient temperature in regions with complex (i.e. mountainous) topography. Further optimization as shown by taking into account altitude and temperature gradients can improve accuracy in these cases. Adding more weather stations, especially in areas with low station density or in neighboring countries can improve accuracy. Deployment of international or multiple national station networks could make the method applicable to border regions in Germany and also to other regions than Germany as well.

The mapped secondary temperature data is fed into TUB-FVB's electric vehicle energy simulation model for small commercial vehicles [3] to simulate total vehicle energy demand. The results are compared to the simulation of total energy demand using real onboard acquired data. The resulting mean difference in vehicle energy consumption is shown to be only 0.17 kWh/h or 2.6% respectively. Depending on vehicle type, battery size, and usage profile, this leads to a limited difference in terms of electric driving range. For the energetic simulation applied at TUB-FVB the use of secondary ambient temperatures has proven very useful, especially in cases where no ambient temperatures are available in data sets to be analyzed.

The authors would like to thank all contributors, especially the Volkswagen AG for supporting the development of the total vehicle energy model used at TUB-FVB and the EN-WIN project partners Fraunhofer IML, Hochschule Fulda, Florida Eis Manufaktur GmbH, Ludwig Meyer GmbH & Co KG, Meyer & Meyer Textillogistik GmbH & Co KG. EN-WIN is funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety in the program "Renewable Mobile".

REFERENCES

[1] U. Eberle und R. von Helmolt, „Sustainable transportation based on electric vehicle concepts - a brief overview," *Energy and Environmental Science,* pp. 689-699, 3 2010.

[2] S. Marker, „Battery Electric Vehicle (BEV) or Rage Extended Electric Vehicle? – Deciding Between Different Alternative Drives Based on Measured Individual Operational Profiles," *IFP Energies Nouvelle,* Bd. 68, Nr. 1, 2013.

[3] J.-O. Jerratsch, „Electrifying Commercial Vehicle Fleets – Energetic Simulation including Auxiliaries," in *Konferenzbericht 5.CoFAT,* Fürstenfeldbruck, 2015.

[4] UDIRVE, „European naturalistic Driving and Riding for Infrastructure & Vehicle safety and Environment," 06 2017. [Online]. Available: http://www.udrive.eu/. [Zugriff am 05 01 2018].

[5] J. Linssen, A. Schulz, S. Mischinger, H. Maas, C. Günther, O. Weinmann, E. Abbasi, S. Bickert, M. A. Danzer, W. Hennings, E. Lindwedel, S. Marker, V. Schindler, A. Schmidt, P. Schmitz, B. Schott, K. Strunz und P. Waldowski, „Netzintegration von Fahrzeugen mit elektrifizierten Antriebssystemen in bestehende und zukünftige Energieversorgungsstrukturen," Forschungszentrum Jülich, Jühlich, 2012.

[6] J. Grüner, C. Liebenow, M. Gobernatz und S. Marker, „Schlussbericht Projekt Ampere," Berlin, 2016.

[7] ElektromobilitätOnline, „ElektromobilitätOnline," 2017. [Online]. Available: https://www.emobilitaetonline.de/das-elektroauto/69-kapitel-6-die-nebenaggregate. [Zugriff am 08 01 2018].

[8] Deutscher Wetterdienst, „Climate Data Center," 2017. [Online]. Available: ftp://ftp-cdc.dwd.de/pub/CDC/. [Zugriff am 20 12 2017].

[9] Fraunhofer IML, „EN-WIN," 2017. [Online]. Available: www.EN-WIN.de. [Zugriff am 05 01 2018].

[10] K. Stahl, R. F. J. A. M. Moore und I. McKendry, „Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density," *Agricultural and Forest Meteorology,* pp. 224-236, 2006.

[11] P. E. Thornton, S. W. Running und M. A. White, „Generating surfaces of daily meteorological variables over large regions of complex terrain," *Journal of Hydrology,* pp. 214-251, 1997.

[12] Deutscher Wetterdienst, „Messnetzkarten," 2017. [Online]. Available: https://www.dwd.de/DE/leistungen/klimadatendeutschland/messnetzkarten.html?lsbId=343278. [Zugriff am 09 01 2018].

[13] P. Bartier und P. Keller, „Multivariate interpolation to incoporate thematic surface data using Inverse Distance Weighting (IDW)," *Computers & Geosciences Vol.22,* pp. 795-799, 1996.

[14] E. Heyer, Witterung und Klima, Stuttgart: Teubner Verlagsgesellschaft, 1993.