Proseminar Proposal

Automatically Generated Temperature Map of the City of Bern by Using Machine Learning Approaches

University of Bern, Institute of Geography
Unit Geocomputation and Earth Observation

Submitted by
Nils Tinner
20-120-952
Patrick Bigler
20-100-178

supervised by
Prof. Dr. Benjamin Stocker

October 18, 2023

Abstract

The urban heat island (UHI) effect significantly impacts urban residents through thermal stress. This project aims to create a map of temperature distribution to address this issue in Bern. For this purpose, an algorithm from Tinner et al. (2023) were improved to calculate different models (multivariate linear regression, KNN, Random Forest). The models contains meteorologic seven variables from a automated weather station, sixty land use classes and temperature from the climate network. All models were classified according to bias (precision) and MAE (accuracy). The best model were used to create the map.

1 Background and Motivation

Anthropogenic climate change is expected to increase the amount, intensity, and duration of heat waves. The urban heat island (UHI) effect further amplifies this trend in urban environments (Burger et al., 2021; Gubler et al., 2021; Wicki et al., 2018). The UHI effect is expressed by higher air temperatures in urban areas compared to rural areas in the region (Oke, 2006). The effect is highest during night, as the emission of longwave radiation in urban environments is impaired and sensible heat fluxes are enhanced, whilst latent fluxes are reduced (Burger et al., 2021; Gubler et al., 2021). People living in urban areas are thus highly affected by the UHI effect via thermal stress (Burger et al., 2021; Wicki et al., 2018). Given that more than 75 % of the Central European population lives in urban areas, the increasing trend poses one of the major weather threats to people in urban environments (Wicki et al., 2018). Studying spatial temperature variabilities in urban areas is therefore crucial to implement adaptation measures to minimize effects on human health and the environment (Burger et al., 2021).

To capture small-scale temperature changes in these climatically complex areas, high spatial resolution measurement networks are needed. However, automated weather stations (AWS) are scarce due to their high costs. To tackle this problem, Gubler et al. (2021) developed a new type of low-cost measurement devices (LCDs). The LCD consists of a temperature logger and a custom-made radiation shield that is naturally ventilated. 79 LCDs were installed in the city of Bern, Switzerland (Gubler et al., 2021). Gubler et al. (2021) reported an overestimation of hourly mean temperature measurements by the LCDs (0.61 °C to 0.93 °C) compared to the reference stations (AWS) during daytime (06:00 – 22:00). During night-time (22:00 – 06:00), differences were much lower or even negative (-0.12 to 0.23 °C). But not only the LCD temperature and the anomaly between the LCDs and the AWS is interesting, but also the temperature distribution of the entire region of Bern, shown on a map.

Tinner et al. (2023) created a map of the distribution of temperatures in Bern for the current logger temperatures. This predicts the local temperature based on land use data as demonstrated by Burger et al. (2021). The approach used by Burger et al. (2021) is a multivariate linear regression Model. As the model only knows the current temperature distribution, its scope and statistical power are limited. This project aims to compare different machine learning techniques to a multivariate linear regression model for land use classes and meteorologic data, thus using collected data from the past four years of meteorological and logger measurements to train the models. Furthermore, this project aims to not only model the temperature as good as possible but maybe also use it to show (possible future) temperature distributions by given meteorologic data.

2 Objective

The goal of this project is to substantially improve the automatically generated temperature map based on an algorithm by Tinner et al., 2023. Furthermore, we want to demonstrate that machine learning is superior to multivariate linear regression. To quantify the quality of the models, we calculate the bias (for accuracy) and the MSE (for precision) for each model. We use the coefficient of determination R^2 to quantify what percentage of the total variance our models explain.

3 Implementation

3.1 Data

First, we want to introduce the source of the data we will use in this project. Because our project's basis is the model of Tinner et al. (2023), we use data for the years 2019-2022 from the network of the city of Bern. This is a numerical data set of the 2m temperature in about 80 locations, with a temporal resolution of 15 minutes for all LCDs. To improve the model of Tinner et al. (2023), we use additional data from the AWS at Zollikofen. For that we have ordered data from Meteo Schweiz. Now, we have access to seven meteorological variables (2m temperature, air pressure, relative humidity, global radiation, precipitation, wind speed, wind direction) with a temporal resolution of 10 minutes for the years 2019-2022. Further, we use the timestamp and the coordinates of Zollikofen. Because we want to capture small-scale temperature changes in a climatically complex area, we think knowledge about the land use is important. That is why we have access to 60 land use classes. The land use classes is a raster data set, where a numerical value is assigned in each pixel. Unfortunately, we do not know the original source. Since our secondary source is known and we trust him, we start our project anyway. But we will try to be able to specify the original source in our final workflow. Table 1 provides a brief overview of the data we will use.

Table 1: The table shows all the data which will be used in this project. Click on the source to get access to the raw data of this project

Source	Type	Kind of data	Resolution	Period
Network	Numeric	3m temperature, coordinates	15 min	2019-2022
Meteo Schweiz	Numeric	Meteorological Variable and coordinates	10 min	2019-2022
Burger et al.	Raster	60 different Land Use Classes	different	-

3.2 Methodology

The data of the years 2019 until 2022 of the urban climate measurement network of the city of Bern is read in and combined with metadata to locate the measurements. This data is combined with the appropriate land use values. It is not yet clear what exact land use values will be used for the model. These will be determined by comparing influence to runtime of the model. Then meteorologic values are added to each measurements. Now, the models are trained on the years 2020 to 2022 and the year 2019 will be used as a validation year. We will do this for a multivariate linear regression model, a KNN model and a random forest model. If we have time left, we will also try to calculate a neronal network model. Because both of us do not have any experience with that, we do not know whether we will able to do this.

As a visualization of the model, a map is then be rendered by combining the land use layers with meteorologic data. The model uses the locations of LCDs as kind of a reference values. Therefore, the temperature of a arbitrary point between two LDCs depends on the numerical values of the land use layers. If we run our code, then we obtain live data from the network as an input for our models. Based on this the model calculates the temperature distribution for the city of Bern. It will be also possible to give an input manually. This gives us the opportunity to get a glimpse into a possible future. Finally, we will determine the R^2, bias, and MAE for each model and use them to compare the models.

4 Responsibilities and Timeline

For our project, we split the responsibility for each task. But this does not mean that the responsible person works alone. All decisions are based on a collaborative exchange. Table 2 shows the tasks and who is responsible for them.

Table 2: The table shows the distribution of tasks. The listed person has the leadership for the respective task. However, the result is based on an equal professional exchange

Who	Task	Deadline
Both	Exchange so that both are on the same level of knowledge	Week 40
Both	Write a proposal	Week 41
Staff	Meeting with the staff to discuss the proposal	Week 41
Both	revise the proposal	Week 42
Both	Stepwise regression and decide which land use classes we use	Week 43
Nils	Random Forest	Week 45
Patrick	Regression and linear Regression model	Week 45
Patrick	KNN	Week 45
Both	Neuronal network (if possible)	Week 48
Both	Write and test the reproducible workflow	Week 48

5 Risks and Contingency

- Data loss and technical issues: To minimize the chance of data loss or technical issues all code and implementations as well as all data is backed up to github. If the data becomes to large there should either be an R-Script that generates the file or the file should be backed up onto external sources.
- Issues with coding: In case we face severe issues with the implementation, the staff panel will be consulted.
- It is possible that we are not able to create a neuronal network

6 Impact

Ideally, we can show that machine learning approaches are superior to classical linear regressions and therefore it is reasonable to implement them. Moreover, the generated model will be able to calculate the temperature of a given site based on land use and meteorologic data. This could be used for three things: First, one can calculate the temperature of the past, the present or the future. One could for example look at the heat of Bern during the last 100 years or use the predictions of MeteoSwiss to predict the urban heat island in the future if all variables are available. This could be of societal relevance since heat is a concern in cities especially with climate change. Second, a warning system could be implemented and warn people in affected areas. Third, one could also assess the excess deaths during the summer months of the past few decades by calculating the heat distribution of past summer nights and so better the understanding of heat deaths. These are just some of the possibilities a new model would offer.

7 Bibliography

Burger, M., Gubler, M., Heinimann, A. and Brönnimann, S. (2021), 'Modelling the spatial pattern of heatwaves in the city of bern using a land use regression approach', *Urban Climate* **38**. `https://doi.org/10.1016/j.uclim.2021.100885`.

Gubler, M., Christen, A., Remund, J. and Brönnimann, S. (2021), 'Evaluation and application of a low-cost measurement network to study intra-urban temperature differences during summer 2018 in bern, switzerland', *Urban Climate* **37**(100817). `https://doi.org/10.1016/j.uclim.2021.100817`.

Meyer, L., Gubler, M. R., Meier, F. and Brönnimann, S. (2021), 'Intercomparison and combination of low-cost urban air temperature measurement approaches', *Meteorologische Zeitschrift* **31**(2), 131–148. `https://doi.org/10.1127/metz/2021/1107`.

Oke, T. (2006), *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites: Instruments and Observing Methods. IOM Report No. 81.*, Canada: World Meteorological Organization. `https://library.wmo.int/index.php?lvl=notice_display&id=9262%20`.

Oke, T. R., Mills, G., Christen, A. and Voogt, J. A. (2017), *Concepts. In: Urban Climates*, Cambridge University Press, p. 14–43. `https://doi.org/10.1017/9781139016476.003`.

Stocker, B., Hufkens, K., Aran, P. and Schneider, P. (2023), *Applied Geodata Science (v1.0)*, Zenodo. `https://doi.org/10.5281/zenodo.7740560`.

Tinner, N., Gubler, M. R. and Brönnimann, S. (2023), 'Map of the temperatur distribution of the city of bern', *Scientific Data* .

Wicki, A., Parlow, E. and Feigenwinter, C. (2018), 'Evaluation and modeling of urban heat island intensity in basel, switzerland', *Climate* **6**(3). `https://doi.org/10.3390/cli6030055`.