PROSEMINAR PROPOSAL

A 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
DR. LAURA MARQUÉS
PEPA ARÀN

NOVEMBER 8, 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 by calculating different models (multivariate linear regression, KNN, Random Forest). The models contains three meteorologic variables (temperature, precipitation and wind) from a automated weather station in Zollikofen (WMO Standard), six land use classes, four geospatial layers to predict temperature of the climate network of the city of Bern. All models are classified according to bias (precision) and MAE (accuracy). The best model will be used used to create a 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 in 2018 (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 station (AWS) during daytime (06:00 – 22:00). During night-time (22:00 – 06:00), differences were much lower or even negative (−0.12 °C 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.

Nils Tinner created a map of the distribution of temperatures in Bern for the current logger temperatures. This predicts the local temperature based on geospatial 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.

2 Objective

This project aims to compare machine learning techniques (random forest and KNN) to a multivariate linear regression model with geospatial data (6 land use classes and 4 geospatial layers) and meteorologic data, thus using collected data from the past four years

(2019-2022) of meteorological and logger measurements to train the models. The model then predicts the local temperature as measured by the loggers based on meteorology and geospatial data. Furthermore, we want to explore if machine learning is superior to multivariate linear regression. The model will then also be used to show temperature distributions for a possible day as a map. Two independent work packages (WP1 and WP2) are defined in order to accomplish these project goals. Open science practices and reproducibility are included in WP1, while model implementation and evaluation are included in WP2 (details in chapter 4).

3 Implementation

3.1 Data

First, we want to introduce the source of the data that will be use in this project. We use temperature data for the years 2019-2022 from the network of the city of Bern. This is a numerical data set of the 3 m temperature in about 110 locations, with a temporal resolution of 10 minutes for all LCDs. This data is publicly available on BORIS. We use additional data from the AWS at Zollikofen, as it is the official meteorologic station of Bern. The three meteorological variables (2m temperature, precipitation and wind (direction and speed)) with a temporal resolution of 10 minutes for the years 2019-2022 will be used as well as the timestamp. Data of the previous one, three and five days will also be fed into the model. This data will presumably be uploaded to a repository to ensure availability since the IDAWEB-data service is open for scientific use but not completely open access. To show small scale urban patterns of temperature distributions, geospatial data as shown in Burger et al. (2021) is used. These are cantonal land use classes as well as federal geospatial data. All of the data is spatially averaged to obtain the layers most effective as shown in Burger et al. (2021). The land use data from the canton data will presumably also be uploaded to a repository since the data has to be downloaded manually from the webpage of the canton. The federal data is presumably directly downloaded from the web into R. 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
BORIS	Numeric	3m temperature, coordinates	15 min	2019-2022
IDAWEB	Numeric	Meteorological Variable and coordinates	10 min	2019-2022
Canton of Bern	Raster	8 different Land Use Classes	different	-

3.2 Methodology

The data of the LCDs of the years 2019 until 2022 is read in and combined with metadata to locate the measurements. This data is combined with the appropriate geospatial values according to the loggers location. Then meteorologic values are added to each measurement according to its timestamp. Now, the models are trained on the years 2019 to 2022. The models have as the predictor variable therefore the geospatial data (10 classes; table 2) and the meteorologic data of the current and the past one, three and five days to predict the temperature of a given logger which is the target variable. This will be done for a multivariate linear regression model, a KNN model and a random forest model. Separate LCD-locations will be used as validation data so that complete independence is guaranteed. Evaluation will be done by determining the R^2, bias, and MAE for each model and use them to compare the models.

Table 2: The table shows all the 10 geospatial data from (Burger et al., 2021) with their optimal zonal average distance which will be used.

Class	File Name	Resolution [m]
Land cover buildings	Bul_Raster_51	250
Open space sealed	SE_Raster_101	500
Open space forest	FO_Raster_51	250
Open space garden	GA_Raster_5	25
Open space water	WA_Raster_21	100
Open space agriculture	AC_Raster_101	500
Slope	SLO_21	100
Vegetation height	VH_WSL_21_21	100
Mean building height	BH_21	100
DEM	DEM	5

As a visualization of the model, a map is then rendered by combining the geospatial layers with meteorologic data. The model uses the locations of LCDs as reference values. Therefore, the temperature of a arbitrary point between two LDCs depends on the numerical values of the geospatial layers and meteorology in general. This meteorologic information can be chosen freely and may give insight into different possible heat distributions. Based on the geospatial layers and the meteorology the model calculates the temperature distribution for the city of Bern. This will be done for hypothetical days to explore the heat distribution on varying conditions.

All of the workflow will be done in R and be uploaded to github.com on a public repository with a GNU Affero General Public License. The workflow aims at full transparency and reproducibility.

4 Responsibilities and Timeline

For our project, we split the responsibility for each task. The tasks are completed independently of one another and form an individual work package. Nils is responsible for the first work package (WP1), which includes the open sciences workflow. Patrick is responsible for the models' implementation and evaluation in the second work package (WP2). In order to begin working on both work packages simultaneously, the models will be implemented and evaluated using temporary data from Burger et al. (2021), which will be substituted later with data of the WP1 (between week 46-47). Table 2 shows the tasks and who is responsible for them.

Table 3: The table shows the distribution of tasks. The listed person has the leadership for the respective task of a certain working package (WP)

Who	WP	Task	Deadline
Both	-	Exchange so that both are on the same level of knowledge	Week 42
Both	-	Write the proposal	Week 42
Staff	-	Meeting with the staff to discuss the proposal	Week 43
Both	-	Revise the proposal	Week 44
Nils	WP1	Data wrangling	Week 44
Nils	WP1	Making data available for open science	Week 45-46
Patrick	WP2	Linear Regression model	Week 45-47
Patrick	WP2	KNN	Week 45-47
Patrick	WP2	Random Forest	Week 45-47
Patrick	WP2	Evaluation of all models	Week 46-47
Both	-	Combine both work packages	Week 46-47
Patrick	WP2	Create a map	Week 46-47
Both	-	Buffer	Week 48
Both	-	Test the reproducible workflow	Week 49

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.

- In case we face severe issues with the implementation, the staff will be consulted.

- As we work with individual work packages that are to be combined at the end of the project, there is a risk that the parts will not fit together. Therefore, despite individual work, there must be good project coordination.

6 Impact

The generated model will be able to calculate the temperature of a given site based on land use classes, geospatial layers and meteorologic data. This could be used for three things: One can calculate the temperature of the past, the present or the future. Therefore, one could 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.

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.