Visual Analytics of Volunteered Geographic Information: Detection and Investigation of Urban Heat Islands

Daniel Seebacher* University of Konstanz
Matthias Miller† University of Konstanz
Tom Polk ‡ University of Konstanz
Johannes Fuchs § University of Konstanz
Daniel Keim¶ University of Konstanz

Abstract

Urban heat islands are local areas where the temperature is much higher than in the vicinity and are a modern phenomenon that occurs mainly in highly developed areas, such as large cities. This effect has a negative impact on energy management in buildings and also has a direct impact on human health, especially for elderly people. With the advent of volunteered geographic information from private weather station networks, more high resolution data is now available within cities to better analyze this effect. However, such data sets are large and have heterogeneous characteristics requiring visual-interactive applications to support further analysis. We use machine learning methods to predict urban heat islands occurrences and utilize temporal and spatio-temporal visualizations to contextualize the emergence of urban heat islands to comprehend the influencing causes and their effects. Subsequently, we demonstrate the analysis capabilities of our application by presenting two use cases.

Index Terms: Human-centered computing—Visualization—Visualization systems and tools; Human-centered computing—Visualization—Visualization application domains—Geographic visualization

1 Introduction

The number of people moving into urban areas is constantly increasing. There are many potential reasons for this, such as better career opportunities or increased mobility. In 2014, about 54% of the population lived in urban areas; urbanization experts predict this number will rise to 66% by 2050 [34]. The urbanization process and the increase of the industrial sector resulted in numerous anthropogenic modifications to the environment such as buildings and streets, leading to a decrease of green spaces in urban areas [18].

This ongoing development has several negative consequences on human health, such as raised noise exposure, heightened air pollution, and increased heat stress [16]. Within cities in particular, sources such as industrial processes, building air conditioning, and transportation increase anthropogenic heat generation. Paving materials, such as black asphalt, generate more heat that is then trapped by tall buildings that disrupt the air flow [20]. These effects all contribute to an increase in temperature in urban areas and are known as the **urban heat island (UHI)** effect. The level of urbanization plays a crucial role in the severity and frequency of urban heat islands, as shown by the average temperature for different areas with different levels of urbanization in Figure 1.

The urban heat island effect has social, economic, and meteorological ramifications. First, it has direct negative consequences on human health. For example, urban heat islands can lead to dehydration, heat strokes and a generally increased mortality rate due to heat stress, especially for elderly people [16]. The likelihood of prolonged heat stress is further exacerbated by global warming. However, this development could be limited by reducing the quantity and intensity of urban heat islands.

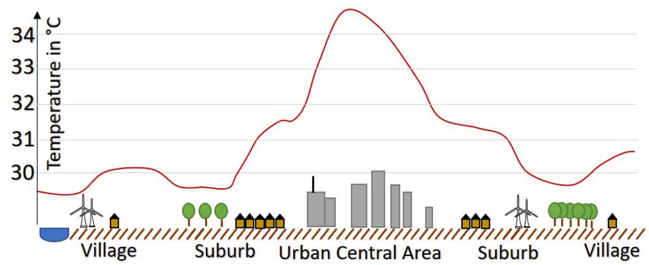


Figure 1: Average temperature at different urbanization area levels.

To reduce the number of urban heat islands in urban areas, their occurrence in cities must be made analyzable to enable an understanding of their underlying causes. This requires fine-grained resolution data that can provide urban planners or meteorologists with information on the effects of different building materials, specific building geometry, or green spaces. However, traditional data sources usually cannot meet these requirements. Satellite data, such as from MODIS, has a coarse spatial resolution of about 1km, as well as a coarse temporal resolution with only two measurements per day. National weather stations, such as those of the German Meteorological Service, are often located outside of cities and thus cannot provide information on the development of urban heat islands in urban areas. Volunteered geographic information (VGI) can help to close this gap in the data. One possible source is the private weather station network 'Weather Underground' [39] that provides access to over 25,000 weather stations in Germany alone, compared to approximately 300 stations provided by the German Meteorological Service. Thus, data with higher spatial resolution is available, providing a better foundation for spatio-temporal temperature models. The use of this data source also leads to new challenges, such as data uncertainty or the handling of hundreds or thousands of data stations, their analysis, and the visualization of relevant information.

The aim of this paper is to support domain experts, such as city planners or meteorologists, in the detection and investigation of urban heat islands and, by doing so, also improve their understanding of urban heat islands. We contribute an interactive visual analytics application that consolidates automatic predictive analysis and glyph-based overview visualizations. We tailor existing methods to this specific application to enable users to adjust input parameters and visually explore spatial and temporal information while considering the geographic context. The application is evaluated through two different use cases, highlighting the generalizability of our approach.

The remainder of the paper is structured as follows: In Section 2, we give an overview of related work and emphasize how we position our research. In Section 3, automatic methods for the detection of urban heat islands are presented and, in Section 4, we present suitable combinations of visualization and interaction methods. Afterwards, the capabilities of our tool are demonstrated

*e-mail: daniel.seebacher@uni.kn
†e-mail: matthias.miller@uni.kn
‡e-mail: polk@dbvis.inf.uni-konstanz.de
§e-mail: fuchs@dbvis.inf.uni-konstanz.de
¶e-mail: daniel.keim@uni.kn

IEEE Visualization in Data Science (VDS)
22 October 2018, Berlin, Germany

by two use cases in Section 5. The drawbacks and benefits of our approach and directions for future work are discussed in Section 6 with a subsequent conclusion in Section 7.

2 Related Work

The related work section is inspired by three main categories of research: (1) Automatic methods to detect and predict urban heat islands. (2) Spatial, temporal and spatio-temporal representations to facilitate the detection of interesting spatio-temporal patterns. (3) Visual Analytics methods to visually explore results of an automatic analysis for understanding causes of the emergence of UHI.

2.1 Prediction of Urban Heat Islands

There are two ways to predict urban heat islands. First, physical or numerical models can be used to formulate the behavior of complex systems by mathematical functions and to make them calculable. One example is the Urban Weather Generator from Bueno et al. [5] which uses different models such as the vertical diffusion model or the Urban Canopy and Building Energy Model to calculate hourly values of urban air temperature and humidity. This model achieved quite accurate results, with an expected error of around 1 Kelvin. One disadvantage of this approach is that it can only be used for the given task with the given parameters. Alternatively, machine learning methods can be used. Voelkel and Shandas [37] use Random Forest classification and Shao et al. [28] and Xi et al. [40] employ neural networks for the prediction of urban heat islands. These approaches were able to make reasonable predictions of urban heat islands, while having the advantage that the used methods are general and can be easily transferred to other areas. A drawback of the aforementioned methods is that for example neural networks are black-box methods, which make it hard to understand how the prediction results are achieved.

2.2 Spatio-Temporal Event Visualizations

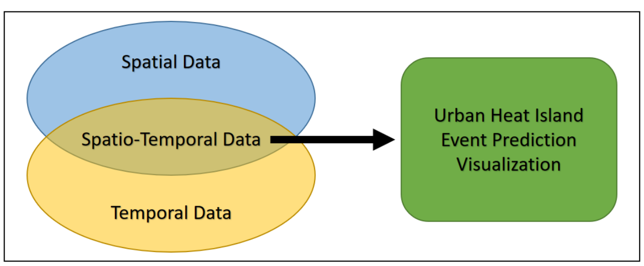


Figure 2: Spatial and temporal data serves as input for interactive visualizations of UHI event predictions.

Urban heat islands can be considered to be spatio-temporal event data, as illustrated in Figure 2. Heatmaps are often used for this type of data to convey spatial distribution and density [6, 9, 11, 23]. However, heatmaps are limited to the spatial domain, since they ignore additional information, such as directional information from trajectories or temporal information from spatio-temporal data [2]. Another common visualization for spatial data are choropleth maps [3, 15], which are similar to heatmaps, except that they use geographic boundaries, such as states or counties to aggregate the data. Similar to heatmaps, choropleth maps also struggle with the visualization of spatio-temporal data. Often animation is used to convey the temporal aspect of the data. However, studies show that larger animated choropleth maps cause issues such as change blindness and change blindness blindness [10, 17].

Visualizing spatio-temporal data, such as weather forecasts, there are two common visualization approaches. The first consists of multiple coordinated views which are connected through linking & brushing [7, 38]: the single visualizations represent only either spatial or temporal aspects. Users have to simultaneously read multiple views to understand connections within the data. Unfortunately, this approach leads to increased cognitive user load [12]. Thus, integrated approaches are more profitable. For example, the two-dimensional geographical space can be extended by the time dimension creating a three-dimensional space-time cube [14]. The second common visualization approach is to enrich maps with additional information, using, for instance, helix icons [32] or glyph-based visualizations [26, 27]. To only display temporal data, a glyph-based radial visualization can be more effective, where a user can pick particular time periods, as demonstrated by Fuchs et al. [13].

2.3 Spatio-Temporal Event Visual Analytics

The increasing amount of available data, partly due to volunteered geographic information, enables researchers to create prediction models for future trends or events or to identify causes and interrelations between outcomes and the input data. Incorporating the user into this automatic analysis is one of the core concepts of Visual Analytics [19]. Various domains benefit from incorporating domain experts in the analysis process, including sports [29], air traffic control [4] and firefighting [31]. The user-driven parameter steering and refinement of automatic predictive analysis results has led to high accuracy results [22].

2.4 Positioning

In this paper we build on the aforementioned related work and create a predictive visual analytics application by tailoring existing methods to the specific application of detection and the spatial and temporal study of urban heat islands. We create a geographic information system that uses a map-based spatial visualization combined with a glyph-based visualization to display the spatio-temporal distribution of urban heat island spots in cities. Additionally, we incorporate automatic analysis methods that can be interactively steered and refined by user interactions to enable the analysis of complex predictive visual analytics scenarios. We incorporate additional abstract and temporal data visualizations to steer the user to interesting periods of time and to examine the impact of meteorological and topographic features as land-use and land-cover on the prediction outcome.

3 Classification

The prediction of the urban heat island effect is a complex problem that depends on a multitude of factors that are responsible for the increased temperature phenomena occurring in urban areas. The ongoing urbanization process leads to anthropogenic changes in the land use, such as the replacement of vegetation and green areas by residential and industrial areas [33], which are essential for the mitigation of higher temperatures [30]. According to Lo and Quattrochi land-cover characteristics, such as the used building materials and their properties such as degree of absorption, radiation properties, albedo, and evaporation rates, are also influencing factors [21].

In our work, we follow an approach that is similar to Voelkel and Shandas [37] or Shao et al. [28]: we combine land use and land cover features with meteorological features and employ machine learning algorithms for the prediction of urban heat islands. We describe the data set that we used in detail in Section 3.1 and present an evaluation of different machine learning algorithms and their performance in Section 3.2.

3.1 Data Foundation

For the meteorological data, such as temperature and precipitation, we collect global and local weather data for individual cities in Germany. The global meteorological data for a city is provided by the German Meteorological Service (DWD). DWD data sets have a good quality with high accuracy and low uncertainty. However, most DWD stations are situated outside of the city area, thereby

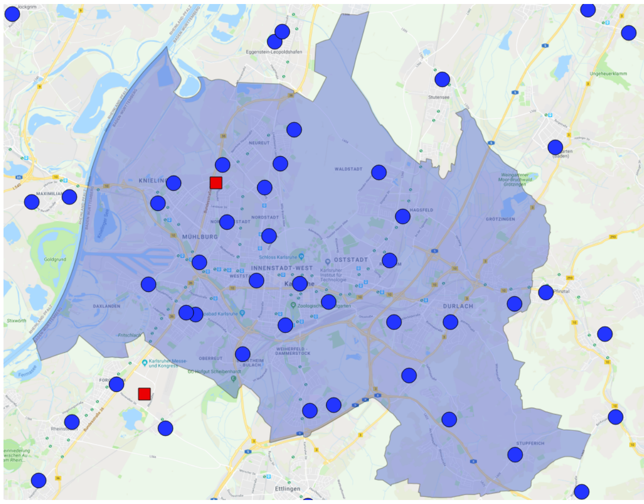


Figure 3: Distribution of private weather stations (blue circles) and stations from the DWD (red squares) in and around the city of Karlsruhe. The volunteered data set provides a more fine-grained spatial distribution: 28 stations in the city of Karlsruhe.

providing no measuring points within cities where the UHI effect occurs. To fill this gap in the data, we rely on more localized data recorded by private weather stations. Well-known providers of such volunteered data sets include Weather Underground [39], openSenseMap [25], and EUWeather [8]. However, one must be careful in using these data sets, as the quality of the data is not examined thoroughly. We faced an issue with stations that are located inside of buildings, which are characterized by providing constant temperature values at room temperature, and stations with strongly deviating or opposing temperature patterns. Nevertheless, even after applying these filtering steps, cities such as Karlsruhe (about 300,000 inhabitants) contain 28 stations distributed within the city, allowing a closer look at the emergence of urban heat islands. The distribution of these stations is shown in Figure 3. The distribution of the weather stations corresponds approximately to the distribution of the built-up areas and consequently serves as an acceptable data foundation. For example, a large area in the north of Karlsruhe is visible in which no weather stations are present due to the forest area in this region. Since forests are not suffering from the urban heat island effect the prediction of the UHI effect is not interesting. In contrast, there are several stations in the city center which are only a few hundred meters apart. Throughout Germany, we have over 1800 stations located within city boundaries. These city stations can be used to identify places and times of the emergence of urban heat islands. A heat island is defined as a location that has a higher temperature than its surrounding area. We calculate the average temperature of the surrounding stations within a radius of 1 km using a bilinear interpolation strategy. Thus, closer stations have a higher influence and more distant stations a lower influence. Afterwards, we introduce a threshold value that determines the amount of temperature difference (with respect to the average temperature values of neighboring stations) needed for a specific location to be classified as an urban heat island. Through initial experiments, we have determined this threshold to be 1° Kelvin, which is the minimum annual mean air temperature difference, that a heat island is warmer than its surrounding, according to US Environmental Protection Agency [1]. We perform this thresholding hourly for each station over the course of the year 2016, resulting in a binary classification dataset including $\approx$ 16 million data points.

We enrich this large data set by meteorological information provided by the DWD, namely: *Air Temperature, Soil Temperature, Relative Humidity, Cloudiness, Precipitation, Air Pressure, Sunshine Duration, Wind Direction,* and *Wind Speed*. Additionally, we add *land-use* and *land-cover* information from OpenStreetMap. We considered the surrounding area within the radius of 1 km around each station to be relevant and aggregated the surface area characteristics into five different categories: *Water Area, Green Space, Sand/Stone Area, Residential or Institutional Buildings,* and *Industrial Territory*. These five categories were chosen in consideration of heat absorption and radiation characteristics of different surfaces.

3.2 Classification of Urban Heat Island Events

To identify locations where the UHI effect occurs and to enable an analysis of the influencing causes, we train prediction models with well-known machine learning algorithms. To compare our results to those of Shao et al. and Xi et al. we compare Neural Networks, Random Forests, Decision Trees and Naive Bayes against each other. To decide which machine learning method is the most appropriate for our application, we examine their performance in a quantitative evaluation. We perform a 5-fold cross validation, i.e., we divide our dataset into five folds and use four folds for the training and one fold for the testing of the classifier. The results of this evaluation are reported in Table 1. In our case, the Random Forest classification has the best performance, with an accuracy of 82.3% and a Cohen's κ score of .647. Thus, we decided to use Random Forest classification for the prediction of the emergence of the UHI effect. However, this does not mean that this evaluation contradicts the results of Shao et al. and Xi et al. as we have used standard parameters in our case and have not optimized the classifiers, since this is not the main contribution of this work.

Classification Model	**Accuracy**	**Cohen's K**
Random Forest (RF)	82.3 %	.647
Decision Tree (J48)	79.2 %	.584
Multilayer Perceptron (MLP)	75.9 %	.519
Naive Bayes (NB)	52.7 %	.055

Table 1: Classification model evaluation results sorted by accuracy.

4 System

During our research on the subject area and in preliminary discussions with domain experts, we have have identified three requirements that our system must meet to support experts in analyzing the circumstances of the emergence of urban heat islands:

V1) *When do urban heat islands occur?* We need suitable visualizations that enable the experts to analyze when urban heat islands are emerging and if there are temporal pattern, such as correlations in the occurrence of urban heat islands.

V2) *Why do urban heat islands occur?* To understand why urban heat islands emerge, we connect temporal and geo-spatial information about the occurrence of urban heat island events. This enables us to bring the occurrences into its context.

V3) *How can we mitigate the urban heat island effect?* In order to mitigate the urban heat island effect in cities, we need to understand what effects planned urban measures and climate changes have on the emergence of urban heat islands. Therefore, we have to enable the experts to play through what-if scenarios to check whether, for example, planned structural changes promote or reduce the emergence of heat islands.

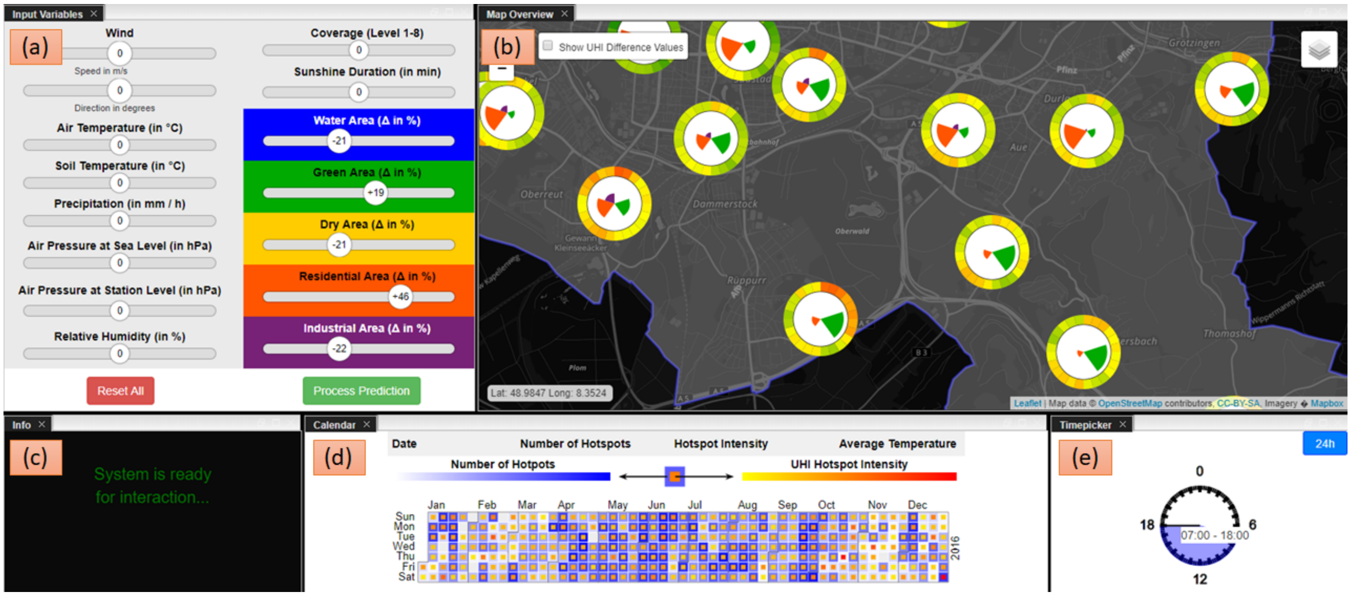


Figure 4: Our system consists of five components that can be flexibly resized by the user: the *Feature Input Panel* **(a)** allows to predict the influence of 15 environmental parameters on the UHI event outcome. The colors correspond to the Nightingale glyph in **(b)**. The *Nightingale glyph map* **(b)** visualizes the temporal occurrence of UHI events and the surrounding spatial context of stations in Karlsruhe. The *system status panel* **(c)** displays information feedback about the system's state to the user. The interactive *calendar view* **(d)** depicts the number of hotspot and their intensity for each day over an entire year. Users can select time ranges to filter the Nightingale glyphs in (b). The *time selector* **(e)** enables users to apply an hourly-based filter to both the Nightingale glyph map (b) and the calendar view (d) using the clock metaphor.

To support the analysis of our predictions of the occurrence of the urban heat islands and to tackle our three stated requirements, we have developed a visual-interactive application, which is shown in Figure 4. In particular, we are analyzing the urban heat islands in 2016 for the city of Karlsruhe in Germany. As shown in Figure 3, we investigate 28 stations in detail. The application consists of several components, which can be resized and rearranged by the user, facilitating the focus on the most relevant components for a given task. We use a calendar-view visualization to display the distribution and intensity of urban heat islands for 2016. These visualizations can also be used for the temporal filtering of date and time ranges. In addition, we offer a map visualization in which we use a combination of a radial glyph and a nightingale chart to encode the temporal distribution of urban heat island events and additional context information such as surface characteristics in the surrounding of the stations on the map. Finally, we offer an intuitive input panel that allows the user to investigate arbitrary what-if scenarios. To do this, the user can adjust one or more variables of interest, which alter the input for the prediction with the trained Random Forest classifier. In the following, we describe all used components in detail.

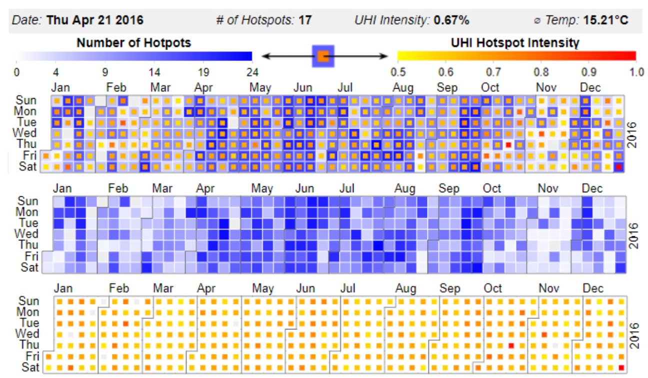


Figure 5: Calendar overview: provides a temporal overview using a weekly structure. Each day is represented by square that simultaneously provides both information about the UHI intensity and the number of hotpots. Users can alter the visualization to focus on only the number of UHI (center) or the UHI intensity of each day.

4.1 Temporal Components

To fulfill requirement **V1**, the visualization of the temporal distribution of the occurrences of the urban heat islands over the year and for the temporal filtering of the data we have implemented a calendar view, as depicted in Figure 5 and a clock-inspired visualization, as shown in Figure 7. These components were selected to use known metaphors such as calendars and clocks to visualize the temporal distribution of the occurrence of the urban heat islands and to enable the application of hourly and date range filters on the data.

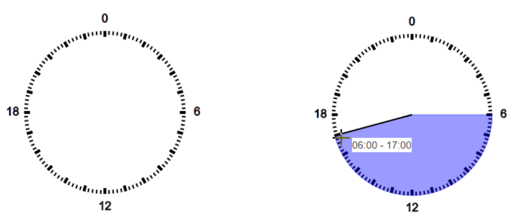
On the one hand, we visualize linear time series data to enable temporal correlation exploration: e.g., the frequency of UHI events over a year. On the other hand, we have to be able to reveal cyclical time series patterns providing valuable information to domain experts, for example an increase or a decrease of the occurrences at the weekends. In order to address these two requirements, we decided to use a calendar view. Each day is represented by a date cell. This facilitates the users to determine whether the occurrences are due to the season, the day of the week, or some other correlation [35]. We use color to encode the number of UHI as well as their intensity, because color allows us to better visualize the progression. This also facilitates to quickly recognize outliers, for example, a day without UHI in a week with many UHI. However, color as a visual channel is less effective for encoding of actual values than, for example, length. Therefore, we offer tooltips to display the exact values on demand.

In the outer area of each date cell, we encoded the number of identified hotspots of the respective date by using a linear mapping of the number of hotspots to a continuous white-blue color scale

Figure 6: Calendar view legends. The **number of identified hotspots** is encoded by a continuous white-blue color scale. The inner part of the square displays the **average UHI hotspot intensity** ($> 50\%$) of all identified hotspot stations on a respective day using a continuous yellow-red color scale.

as described in Figure 6. Similarly, we represent the average urban heat island intensity by exploiting a continuous yellow-red color scale. Details-on-demand are interactively available on hover to get precise information such as number of hotspots and hotspot intensity. Additionally, the calendar view is linked to the other visualizations and allows to filter the data by selecting an individual day or a user-defined period of time, which then updates the information visualized in the other components. Likewise, if filtering is performed in another visualization, the calender view will be updated as well.

In addition to the calendar view, we have implemented a clock-based time selector. This component enables the user to filter the represented information by the time of day, which allows for an in-depth analysis of the occurrences of urban heat islands within different time periods of the days. This clock-based time selector and an example showing the selection of a time period between 6 a.m. and 5 p.m. is depicted in Figure 7.



(a) Default presentation of the *Time Range Selector*: no interval selected.

(b) Hourly interval filter selection example: 6 a.m. to 5 p.m.

Figure 7: Clock-based visualization for the time range selection.

4.2 Spatio-Temporal: Nightingale Glyph

To meet requirement **V2**, we need to bring the temporal occurrences of the urban heat islands into their spatial context. Hence, we have decided to apply a radial glyph-based strategy. This strategy enables us to visualize both the temporal distribution and further information directly on the map, which, according to Fuchs et al. [12], requires less cognitive user effort than linking multiple views. Additionally, a radial glyph-visualization of the time-oriented data has proven to be more effective for specific tasks, such as picking particular temporal locations, as highlighted in a study by Fuchs et al. [13].

In Figure 8, the design of our glyph-based visualization is introduced. Its inner segments can be used to encode different information. In our case, we encode the distribution of land use and land cover features in the vicinity of a station with the help of a Nightingale Rose Chart [24], where we map the fraction of the total area to the radius. Although this introduces a bias for areas with a larger fraction of the total area, since they are assigned a disproportionately large area in the chart, this facilitates users to recognize the predominant areas more easily. Alternatively, the inner area segments could be altered to convey meteorological or statistical features. The outer ring, we visualize the temporal intensity of urban heat island events of a station in each time segment using a continuous color scale from green to red. Green color indicates low UHI event occurrences whereas red color indicates more UHI events in a specific time slot. The time segments are arranged radially clockwise around the inner

(a) Nightingale glyph: The inner group represents the geographic section of the glyph representation. Every segment indicates the amount of a certain area, that is located within a radius of 1km around a WU station. To the right a colored legend describes the five different area types.

(b) The outer group visualizes the temporal progress of the UHI index using a clockwise ordering. Every segment indicates the UHI intensity of the respective temporal unit (hour, day, month, date range). Here, 24 hours are displayed: one segment per hour.

Figure 8: An explanation of the Nightingale glyph visualization.

visualization area and the first time segment of the selected period starts at 12 a.m. The time segments can be chosen arbitrarily so that users can retrieve details about the intensity of the urban heat islands for every day hour or the months of the year.

The nightingale glyphs are positioned on the map at the geographic position of stations available in the data, allowing us to encode additional context information on the map, as well as the temporal distribution of the occurrence of UHI events for the respective station. Also, the color scale of the time segments can also be changed to visualize different information. We use this feature to enable the user to inspect the outcome of a user-steered what-if scenario. Instead of merely showing the distribution of the occurrence of urban heat islands after a change in the input parameters, the user can switch to a blue-white-red color scale, which visualizes the difference between the prediction and the original values. A reduction in the number of urban heat island events within a period is indicated by a blue color, an increase in red. White color indicates no change for a specific time segment. The color tone is mapped to the strength of the difference, with a strong increase or decrease being mapped to a strong red or blue color. This approach was chosen to emphasize the temporal aggregated changes of UHI events which are illustrated in the example in Figure 10.

4.3 Feature Input Panel

To support exploration and to investigate the effects of climate or land use changes, as stated in requirement **V3** , we provide a Feature Input Panel (FIP) that is presented in Figure 9. Users can use the FIP to interactively steer various parameters. With the help of this graphical interface, users can generate both simple and complex scenarios fitting their questions. For example, how could a change in the cityscape such as an increase of the industrial sector or a climate change resulting from global warming affect the UHI situation?

In the input panel, sliders are provided that allow the user to manipulate values for any of the 15 parameters used for the prediction of urban heat island events. In the standard configuration, the sliders are set to 0, which corresponds to no change in the values from the default state. Moving the slider to the left decreases a value, while moving it to the right increases it. The user can adjust these sliders to interactively generate delta values that are then applied to the measured values of the regarded time period to alter the input values for the prediction. This can be used, for example, to increase the temperature to simulate scenarios with warmer climate or reduced wind speeds like large buildings in the cityscape are acting as wind breaks. If the users change one of the land use characteristics, such as the residential areas, the sliders of the other land use characteristics are automatically adapted. For example, if the residential area is increased by 40%, the water, green, dry and industrial selectors are reduced by 10% each. This ensures that the area used for the prediction always remains at 100% to simulate more realistic situations. Moreover, users can arbitrarily adjust all parameters of interest simultaneously to generate more complex scenarios.

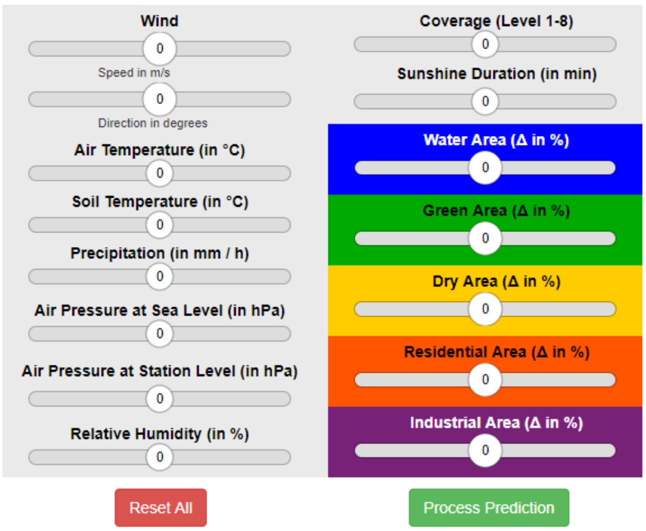


Figure 9: The FIP consists of two main sections: Meteorological (in light gray: from *wind* to *sunshine duration*) variables and spatial (multicolored; *water area* to *industrial area*) features. Users can steer multiple variables to process flexible prediction scenarios.

5 Use Cases

In this section, we want to highlight how our system can help domain experts in the investigation of the emergence of urban heat islands. We show the usefulness of our trained classifier in combination with the selected components presented in the previous section in two use cases. The first use case is the analysis of the impact of industrialization in urban areas and in a second use case we examine the impact of global warming. Both use cases will be performed for stations in the city of Karlsruhe in 2016.

5.1 Increase of Industry Area Level

To investigate the impact of increased industrialization in urban areas on the occurrence of urban heat islands we simulate this scenario by using the introduced Feature Input Panel, as described in Section 4.3 by increasing the delta of the industrial area to 60%. As described earlier, this decreases the delta of the water, green, dry and residential area by 15% each, to keep the area at 100%. After the delta values are chosen, the prediction can be repeatedly executed to simulate a cityscape with 60% increased industrial area. The results of both the initial and the modified scenario are shown in Figure 10.

The cityscape in Karlsruhe, which can be explored in Figure 10a through the nightingale chart glyphs, is dominated by much green space. This leads to a small number of urban heat islands. However, this can be greatly changed by the change of the spatial surface characteristics. By increasing the industrial area by 60% and reducing the other areas by 15% each, we create the scenario depicted in Figure 10b. A direct comparison with Figure 10a shows that this increase of industrial areas would increase the number of urban heat islands. The difference view that is displayed in Figure 10c supports this result and emphasizes the UHI prediction differences for each station. The glyphs in Figure 10c are almost always consistently red, which indicates that, compared to the standard situation, the number of urban heat islands has risen sharply and thus shows the negative effects of increased industrialization without compensating areas.

(a) Default occurrences of UHI in Karlsruhe when applying hourly aggregation.

(b) UHI occurrences, after a 60% increase of the industrial sector in the stations' surrounding.

(c) Difference view showing the changes between the initial situation depicted in view (a) and the adjusted situation displayed in view (b).

Figure 10: Effects of increased industrial area level on the occurrence of UHI events. With the help of the difference view, users can be observe how growth of the industrial sector leads to a considerable increase of urban heat island events.

5.2 Warmer Climate

With the help of our Feature Input Panel, even more complex scenarios can be simulated and analyzed. For example, we can explore the effects of global warming on the severity of urban heat islands in cities. To create such a scenario, we need to adjust meteorological parameters in our Feature Input Panel as shown in Figure 11. The parameter selection simulates a warmer and drier weather situation which could be a result of global warming processes. The prediction is then recalculated to enable analysis of the selected scenario.

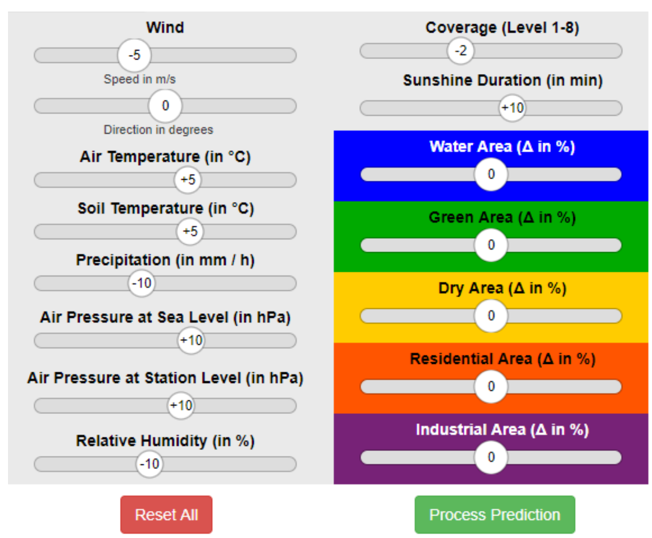
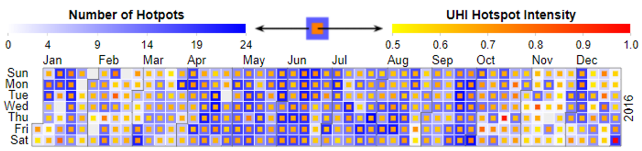
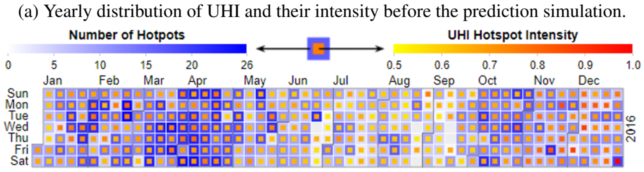


Figure 11: Adjusted feature parameters to simulate a warmer and drier weather situation.

The effects of this simulation can be analyzed using the calendar view in combination with our clock-based time selector to focus on interesting day hours. This enables us to investigate how the annual distribution of urban heat island events develops in the context of global warming and what impact it has at various times of the day. For example, Figure 12 shows the annual distribution of urban heat islands before and after simulation. The calender view shows that the number of urban heat islands is declining in summer. This result seems counter-intuitive at first, since one expects that warmer weather, especially in summer, will create more heat islands in urban areas. However, the conditions are changing in such a way that there are no more isolated urban heat islands, since it seems that the situation is turning into a rather global heat problem. In addition, the prediction shows a shift of the occurrences of heat island events from summer to spring and even winter. This can be an adverse development, as prolonged exposure to heat can lead to health problems, especially for the elderly [36].



(a) Yearly distribution of UHI and their intensity before the prediction simulation.



(b) Yearly distribution of UHI and their intensity after the prediction simulation.

Figure 12: Overview of the emergence of urban heat islands and their average intensity per day for a year before and after executing a warm weather simulation.

Figure 12a shows that the UHI hotspot intensity distribution seems to be equally distributed in the summer months. Generally, higher number of heat islands per day have a slight correlation with a higher intensity. Some outliers exist in February and October with a low number of hotspots with high intensity values. The distribution after the prediction, depicted in Figure 12b, shows that the intensity of the UHI is increasing for the spring months. The intensity of UHI events from October to December has increased drastically indicating a unfavorable future trend if the temperature increases over the years as a result from global warming.

We can further investigate this scenario by using our clock-based time selector. We examine the first half of the night, from 8 p.m. to 1 a.m. and the second half of the night, from 2 a.m. to 7 a.m as shown in Figure 13 to identify UHI event differences between day and night hours. We can identify that the number of urban heat islands is increasing in the second half of the night compared to the first half. This development can be a result of lower humidity and lower wind speeds that reduce heat dissipation in cities or that the surrounding of hotspots is cooling down more quickly than the UHI hotspots. This is why for example industrial areas remain warmer overnight and act as urban heat islands. Increased temperatures at night prevent the affected population to recover from the warm temperatures during the day and should be considered by city planners for decision making in urban planning.

6 Discussion and Future Work

In this section, we discuss the key topics presented in this work, limitations of our approach, and ideas for future work.

Using the example of the city of Karlsruhe in 2016, we have demonstrated the possibilities offered by our application in two specific use cases. On the one hand, we can support city planners in their urban planning by showing them the impact of changes in the cityscape. On the other hand, meteorologists can use our application to investigate the effects of different meteorological effects on the occurrence of urban heat islands. We had an informal discussion with a domain expert from the DWD to identify potential fields of application of our system. For instance, the expert emphasized that local communities would be interested in our visual application but

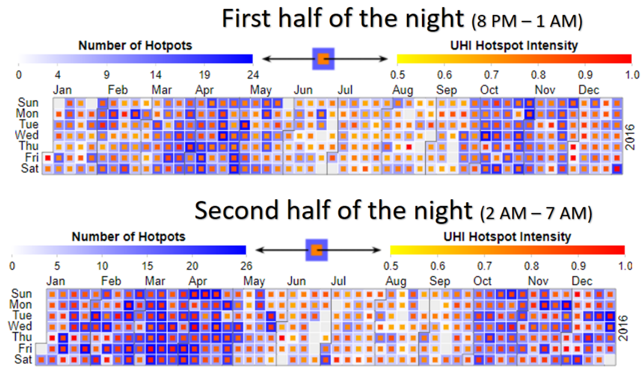


Figure 13: Comparison of UHI hotspots characteristics for the first and second half of the night.

it is essential to revise methodical and climatological concepts to improve the implemented model. The expert highlighted that our application targets a relevant problem domain that will become more severe with the progressing climate change in the future. Another possible application would be the investigation of hotspots in the vicinity of sensible infrastructures, such as hospitals, retirement homes, or schools. City planners are particularly in need of climatological evaluation to understand the emergence of a specific hotspot in a city.

Additionally, we got valuable feedback on additional influencing variables that should be included in the prediction model of urban heat islands, such as the morphology and physical properties of urban surfaces and buildings, thermal properties like heat absorption, and surface radiation.

In our application, we consider only local hotspots. In some use cases, it is required to understand the temperature intensity of the whole city compared to the rural surrounding of the city. Since Karlsruhe is mainly surrounded by further smaller cities, we considered to analyze rather local UHI to be more promising. Nevertheless, our system could benefit from adding functionality that enables meteorologists to investigate a city's climate with the immediate rural surrounding to identify temporal patterns of the upscaled approach.

A current limitation of our application is that it requires extensive domain knowledge and that it cannot guide users' attention to relevant sections of the visualization to emphasize outliers or extreme values. Additionally, we currently do not support an exploratory analysis of an ensemble of generated scenarios, but rely on a rather confirmatory analysis of user-generated scenarios. In addition, the difference view should be implemented in all components. So, strong deviations from the initial situation can be detected not only in the spatial distribution of UHI, as is currently possible, but also in time, if the calendar view would offer such a difference view. Further extensions we are currently working on include the spatial and temporal extension of the application by supporting the analysis of several cities over several years. Another future step is the evaluation of our current system, as well as ideas for future work, in a formal expert study. Finally, we plan to extend this work to enable the analysts in this domain to better analyze and understand the context of urban heat islands in cities, to facilitate the finding of the relevant causes, and to identify suitable mitigation strategies.

7 Conclusion

The main contribution of this paper was to provide an application that supports domain experts in analyzing the urban heat island effect in city areas and to investigate the impact of changes in the cityscape or the climate. To do so, we used volunteered geographic information along with other data sources to train and evaluate a classifier to predict the emergence of urban heat islands. We enabled the visual analysis of these predictions using an adaptive workspace combined with various visualizations. For the temporal analysis of the distribution of the urban heat islands, we developed a calendar-based view and a clock-based time selector, which is used for the filtering. Additionally, we presented a nightingale chart glyph, which provides insights into the temporal distribution of occurrences of urban heat island events, as well as providing contextual geographic information. By placing them at their respective positions on the map, we integrate the temporal and geographic characteristics. We have presented by two examples how our application can help to visually analyze simulations of urban heat island event predictions, how different what-if scenarios can be investigated, and what conclusions can users draw from the presented information.

Acknowledgments

The project BigGIS (reference number: 01IS14012) is funded by the Federal Ministry of Education and Research (BMBF) within the frame of the research programme “Management and Analysis of Big Data” in “ICT 2020 – Research for Innovations”.

References

[1] United States Environmental Protection Agency (EPA). Heat Island Effect. `https://www.epa.gov/heat-islands`, 2014. [Online; accessed 10-July-2018].

[2] C. Bauckhage, R. Sifa, A. Drachen, C. Thurau, and F. Hadiji. Beyond heatmaps: Spatio-temporal clustering using behavior-based partitioning of game levels. In *Computational Intelligence and Games (CIG), 2014 IEEE conference on*, pp. 1–8. IEEE, 2014.

[3] C. A. Brewer, A. M. MacEachren, L. W. Pickle, and D. Herrmann. Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers*, 87(3):411–438, 1997.

[4] J. Buchmüller, H. Janetzko, G. Andrienko, N. Andrienko, G. Fuchs, and D. A. Keim. Visual analytics for exploring local impact of air traffic. In *Computer Graphics Forum*, vol. 34, pp. 181–190. Wiley Online Library, 2015.

[5] B. Bueno, L. Norford, J. Hidalgo, and G. Pigeon. Prediction of the urban heat island effect to be used in building energy analyses. *Proceedings of SimBuild*, 5(1):236–245, 2012.

[6] N. Cao, C. Lin, Q. Zhu, Y.-R. Lin, X. Teng, and X. Wen. Voila: Visual anomaly detection and monitoring with streaming spatiotemporal data. *IEEE transactions on visualization and computer graphics*, 24(1):23–33, 2018.

[7] A. Diehl, L. Pelorosso, C. Delrieux, C. Saulo, J. Ruiz, M. E. Gröller, and S. Bruckner. Visual analysis of spatio-temporal data: Applications in weather forecasting. In *Computer Graphics Forum*, vol. 34, pp. 381–390. Wiley Online Library, 2015.

[8] European Weathernetwork (EWN). EUweather. `https://euweather.eu/`, 2014. [Online; accessed 10-July-2018].

[9] N. Ferreira, L. Lins, D. Fink, S. Kelling, C. Wood, J. Freire, and C. Silva. Birdvis: Visualizing and understanding bird populations. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2374–2383, 2011.

[10] C. Fish, K. P. Goldsberry, and S. Battersby. Change blindness in animated choropleth maps: an empirical study. *Cartography and Geographic Information Science*, 38(4):350–362, 2011.

[11] D. Fisher. Hotmap: Looking at geographic attention. *IEEE Transactions on Visualization & Computer Graphics*, (6):1184–1191, 2007.

[12] G. Fuchs and H. Schumann. Visualizing abstract data on maps. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pp. 139–144. IEEE, 2004.

[13] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 3237–3246. ACM, 2013.

[14] P. Gatalsky, N. Andrienko, and G. Andrienko. Interactive analysis of event data using space-time cube. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pp. 145–152. IEEE, 2004.

[15] J. G. Gimpel, F. E. Lee, and J. Kaminski. The political geography of campaign contributions in american politics. *The Journal of Politics*, 68(3):626–639, 2006.

[16] R. Godfrey and M. Julien. Urbanisation and health. *Clinical Medicine*, 5(2):137–141, 2005.

[17] K. Goldsberry and S. Battersby. Issues of change detection in animated choropleth maps. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 44(3):201–215, 2009.

[18] E. Kalnay and M. Cai. Impact of urbanization and land-use change on climate. *Nature*, 423(6939):528, 2003.

[19] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon. Visual analytics: Definition, process, and challenges. In *Information visualization*, pp. 154–175. Springer, 2008.

[20] H. H. Kim. Urban heat island. *International Journal of Remote Sensing*, 13(12):2319–2336, 1992.

[21] C. Lo and D. A. Quattrochi. Land-use and land-cover change, urban heat island phenomenon, and health implications. *Photogrammetric Engineering & Remote Sensing*, 69(9):1053–1063, 2003.

[22] Y. Lu, R. Garcia, B. Hansen, M. Gleicher, and R. Maciejewski. The state-of-the-art in predictive visual analytics. In *Computer Graphics Forum*, vol. 36, pp. 539–562. Wiley Online Library, 2017.

[23] R. Maciejewski, S. Rudolph, R. Hafen, A. Abusalah, M. Yakout, M. Ouzzani, W. S. Cleveland, S. J. Grannis, and D. S. Ebert. A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics*, 16(2):205–220, 2010.

[24] F. Nightingale. Diagram of the causes of mortality in the army in the east. *Public domain. Retrieved February*, 21:2015, 1858.

[25] openSenseMap. openSenseMap. `https://opensensemap.org/`, 2014. [Online; accessed 10-July-2018].

[26] C. Rohrdantz, M. Krstajic, M. El-Assady, and D. A. Keim. What's Going On? How Twitter and Online News Can Work in Synergy to Increase Situational Awareness. In *Published at the 2nd IEEE Workshop on Interactive Visual Text Analytics "Task-Driven Analysis of Social Media" as part of the IEEE VisWeek 2012, October 15th, 2012, Seattle, Washington, USA*, 2012.

[27] D. Seebacher, J. Häußler, M. Hundt, M. Stein, H. Müller, U. Engelke, and D. Keim. Visual analysis of spatio-temporal event predictions: Investigating the spread dynamics of invasive species. 2017.

[28] B. Shao, M. Zhang, Q. Mi, and N. Xiang. Prediction and visualization for urban heat island simulation. In *Transactions on edutainment VI*, pp. 1–11. Springer, 2011.

[29] M. Stein, J. Häußler, D. Jäckle, H. Janetzko, T. Schreck, and D. A. Keim. Visual Soccer Analytics: Understanding the Characteristics of Collective Team Movement Based on Feature-Driven Analysis and Abstraction. *ISPRS International Journal of Geo-Information, Special Issue Advances in Spatio-Temporal Data Analysis and Mining*, 2015. doi: 10.3390/ijgi4042159

[30] T. Susca, S. R. Gaffin, and G. Dell'Osso. Positive effects of vegetation: Urban heat island and green roofs. *Environmental pollution*, 159(8-9):2119–2126, 2011.

[31] A. D. Syphard, V. C. Radeloff, N. S. Keuler, R. S. Taylor, T. J. Hawbaker, S. I. Stewart, and M. K. Clayton. Predicting spatial patterns of fire on a southern california landscape. *International Journal of Wildland Fire*, 17(5):602–613, 2008.

[32] C. Tominski, P. Schulze-Wollgast, and H. Schumann. 3d information visualization for time dependent data on maps. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pp. 175–181. IEEE, 2005.

[33] A. R. Trihamdani, T. Kubota, H. S. Lee, K. Sumida, and T. T. T. Phuong. Impacts of land use changes on urban heat islands in hanoi, vietnam: Scenario analysis. *Procedia Engineering*, 198:525–529, 2017.

[34] United Nations. World's population increasingly urban with more than half living in urban areas. `http://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html`, 2014. [Online; accessed 28-June-2018].

[35] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, pp. 4–9. IEEE, 1999.

[36] S. Vandentorren, P. Bretin, A. Zeghnoun, L. Mandereau-Bruno, A. Croisier, C. Cochet, J. Ribéron, I. Siberan, B. Declercq, and M. Ledrans. August 2003 Heat Wave in France: Risk Factors for Death of Elderly People Living at Home. *The European Journal of Public Health*, 16(6):583–591, 2006.

[37] J. Voelkel and V. Shandas. Towards systematic prediction of urban heat islands: Grounding measurements, assessing modeling techniques. *Climate*, 5(2):41, 2017.

[38] J. A. Walsh, J. Zucco, R. T. Smith, and B. H. Thomas. Temporal-geospatial cooperative visual analysis. In *Big Data Visual Analytics (BDVA), 2016*, pp. 1–8. IEEE, 2016.

[39] Weather Underground Organization. Weather Underground. `https://www.wunderground.com/`, 2014. [Online; accessed 28-June-2018].

[40] W. Xi and P. He. Prediction of urban heat island intensity in chuxiong city with backpropagation neural network. In *Advances in Neural Network Research and Applications*, pp. 29–36. Springer, 2010.