TraVis – A Visualization Framework for Mobile Transect Data Sets in an Urban Microclimate Context

Kathrin Häb*
Computer Graphics and HCI Group
University of Kaiserslautern,
Kaiserslautern, Germany

Ariane Middel†
Julie Ann Wrigley Global Institute of Sustainability
Arizona State University
Tempe, AZ, USA

Benjamin L. Ruddell‡
The Polytechnic School
Arizona State University
Mesa, AZ, USA

Hans Hagen§
Computer Graphics and HCI Group
University of Kaiserslautern,
Kaiserslautern, Germany

ABSTRACT

In urban microclimate research, mobile transects are utilized to observe the relationship between atmospheric variables and the urban environment. However, the data sets resulting from mobile measurements are complex: They are spatially dependent, multivariate, and often-times multitemporal. At the same time, the spatial context of each observation – its field of view and area represented – is physically complex and dynamic. These properties make analysis and visualization challenging. We present a prototype visualization framework that assists researchers in the analysis of mobile transect measurements. The system enables users to visualize and explore observations as walls that delineate the transect route on a map. The observed attributes are stacked upon each other within these walls as ribbons to facilitate the qualitative analysis of spatial variability and multivariate correlations. The relationship between observations and spatial context can interactively be explored by moving a slider along the transect route. For each observation on the track, the spatially contextual source area is displayed and linked to a view of the fraction of land cover classes contained within the source area. These qualitative analysis capabilities are complemented by an interactive clustering interface, which allows for the classification of transect segments according to a coherent pattern of multivariate relationships between a user-defined set of observations. The framework was developed by a team comprising both visualization and urban microclimate researchers, and a case study shows its utility for this specialized application.

Index Terms: I.3.8 [Computer Graphics]: Applications—; J.2 [Computer Applications]: Earth and atmospheric sciences—

1 INTRODUCTION

Mobile transect measurements are an important tool for urban climatology. Several phenomena can be investigated using this technique. Research usually focuses on either the impact of urban form on microclimate [31, 6] or on the analysis of specific climatic phenomena, such as the urban heat island [15, 30, 23, 33] or park cool islands [7]. Studies also investigate health implications of the urban environment, e.g. through air quality measurements [9, 13] or through the analysis of thermal comfort [35].

Data sets from mobile measurements are multivariate and (oftentimes) time-varying trajectory data. A trajectory is defined as a time-ordered sequence of spatial locations visited by an entity [3, 4, 34, 25]. Data from mobile microclimate transect measurements build on these two elementary components, although sensors mounted on the moving platform add additional attributes. These attributes typically include air temperature, surface temperature, relative humidity, or air quality data, with varying values along the transect route. The time-varying component of the data set is introduced by repeating the transect measurements periodically. Furthermore, each sensor and therefore each trajectory attribute has a unique and dynamically varying spatial context and representative source area corresponding to the observation.

The multivariate, time-varying, and spatially contextual nature of mobile microclimate transect measurements makes the data set complex and its exploration time-consuming. The specific tasks during the analysis of mobile measurements depend on a central research question that is individual for each study. Nevertheless, two universal tasks could be identified by two domain scientists, who are the co-authors of this paper. Firstly, climatologists need to preprocess mobile observations to establish precise measurements of attributes and their spatial context. For example, this includes a time-detrending step to filter out potential impacts of changing reference weather during the course of a run [32]. Less trivially, the environmental attribute data need to be related to the spatial context by estimating the effective sensor field of view to precisely establish relationships between the observations and the properties of the (urban) landscape being traversed.

In this study, we introduce a prototypic framework for the analysis and visualization of mobile urban microclimate measurements that assists researchers conducting the described tasks. Similar to [34], our visualization relies on a hybrid 2D / 3D representation of the data in its geographical reference space using wall-layers stacked in z-direction on a two-dimensional land use / land cover (LULC) map. In contrast to [34], we do not use the third dimension to visualize the dynamics of one specific attribute on *different* trajectories. We rather use the individual wall-layers to display the values associated with different environmental attributes of the *same* trajectory, approximating the intuitive notion of measurement height above the ground that is inherent in the data.

The contribution of this study to the visualization domain is twofold. First, we develop an interactive visualization for *multivariate* trajectory attribute data, providing a detailed view of their dynamic behavior during a mobile transect measurement run. Second, the dynamic spatial context is not limited to visualizing a background-map, but its impact onto the trajectory attributes can interactively be explored employing the meteorological concept of a *source area*, which is a notion for the spatial field of view of this class of sensors [28]. An interactive clustering interface additionally enables the user to classify transect segments according to coherent patterns of multivariate relationships either between atmospheric attributes alone, or between atmospheric attributes and their associated land use fractions within the sensor-specific source area. Hence, the resulting framework comprises capabilities to qualitatively and quantitatively explore spatial and multivariate interdependencies within the data set under investigation.

The remainder of this paper is structured as follows: In Section 2, we describe similarities and differences between conventional trajectory data sets and mobile-platform transect data sets. Then, in Section 3, we review related work relevant to our study. In Section

*e-mail: kathrin.haeb@cs.uni-kl.de
†e-mail: ariane.middel@asu.edu
‡e-mail: bruddell@asu.edu
§e-mail: hagen@cs.uni-kl.de

IEEE Pacific Visualization Symposium 2015
14–17 April, Hangzhou, China
978-1-4673-6879-7/15/$31.00 ©2015 IEEE

4, we describe the features of our framework in detail, ranging from preprocessing to analysis capabilities. A use case is described in Section 5 before drawing a conclusion in Section 6.

2 Data set characteristics

Data sets from mobile transect measurements have several properties in common with conventional trajectory data. Using the notation described in [1, 34], the latter comprise a set of spatial locations per trajectory, $S = \{s_0, ..., s_n\}$, and a time stamp t_i associated with each element of S, yielding $T = \{t_0, ..., t_n\}$, with $n > 0$ and indicating the number of sampled positions. Using T and S, further sets of dynamically changing attributes $A_{0,...,m}$ can be computed per location and time stamp, including movement direction, velocity, and acceleration [4, 34]. Similar to the data sets visualized in [34], mobile-platform transect data sets involve additional attributes $A_{m,...,n}$ complementing S and T, such as air temperature, relative humidity, or surface temperature.

There are two major differences between conventional trajectory data and data from mobile transect measurements. Firstly, the attributes collected during mobile transect measurements are related to a certain height above ground level (AGL), which is either given by the mounting height of the sensors for in-situ measurements (e.g., air temperature) or by the location of the investigated object for remotely sensed data (e.g., surface temperature). This information is an important aspect during data analysis. It complements the 3D spatial locations of a moving entity's path in conventional trajectory data sets, which is usually described by a set of 3D geographic coordinates (latitude, longitude, and elevation above mean sea level (ASL)). Since the set of spatial locations is usually displayed on a two-dimensional map that represents the projection of the Earth's surface onto a plane, the measurement height AGL introduces a vertical offset dimension in addition to the resulting set of two-dimensional coordinates on that plane.

Secondly, the analysis of the spatial context of the data differs from usual trajectory data sets. In conventional trajectory data sets, the spatial context is analyzed to facilitate reasoning about certain movement patterns. For mobile transect measurements, the trajectory is exogenously fixed based on the aim to observe relationships between multiple environmental attributes and the spatial context, utilizing spatial context to explain attributes rather than the trajectory itself. The multiple sensors in a research transect each measure a different field of view, and as such each of multiple attribute observations has a distinct and separate relationship to the spatial context.

3 Related work

3.1 Mobile urban climate transect measurements

Mobile transect measurements are frequently applied in urban climatology to investigate the spatial and temporal dynamics of microclimate within the urban canopy layer, i.e. the space between ground and roof level.

Mobile transect observations usually include air temperature and relative humidity [7, 30, 15, 35]. Depending on the research goal of the study, additional variables are derived from these observations, e.g., the dew point temperature [31, 33], or measured. For example, Vanos et al. [35] observe total incoming shortwave radiation to investigate the moderating effect of parks onto the human energy budget. In urban areas, air quality observations are also an important source of information. Thus, flexible measurement systems have been developed for mobile air quality measurements. Notable examples are the AERO-TRAM project [13], or the Aeroflex bike [9].

Usually, measurements are conducted at a fixed level above ground. In some cases, the behavior of the observations with increasing height above the surface is an integral part of the research question. For example, Chow et al. [7] used mobile measurements at four heights to derive temperature profiles over a variety of surface types to investigate the characteristics of the nocturnal park cool island effect on the ASU campus in Tempe, Arizona.

LULC surrounding the sensors is frequently included into the analysis, using various methods. Sun et al. [33] relate their observations to the normalized differenced vegetation index (NDVI) within two constant radii around the sensor locations. Heusinkveld et al. [15] use the fraction of buildings, water, and vegetation to statistically analyze the relationship between observations and urban landscape context.

In urban climate literature, mobile transects are usually displayed as two-dimensional plots with the spatial location on the x-axis and the observed quantities on the y-axis [33, 35, 13]. Sometimes, the x-axis is complemented by LULC information located directly under the sensor platform to facilitate the relation of measurements to spatial context [6, 31]. Another frequently used visualization is a two-dimensional spatial plot of transect measurements on a map [15, 7, 9], displaying only one variable, one measurement height, or one point in time. In our framework, we combine the advantages of these visualization methods, displaying a user-defined number of variables as stacked ribbons on a background map, and transferring the generally two-dimensional diagrams into their native spatial context.

A common goal of the reviewed mobile transect literature is to understand the relationship between the spatial context and the observed environmental attributes. This relationship is complex, its analysis tedious, and its visual exploration limited using existing tools and methods. Therefore, our overarching research goal is the development of a tool that assists researchers during data preparation, transformation, exploration, and analysis of the relationship between multiple mobile platform observations and spatial context.

3.2 Visual analytics for geospatial trajectory data sets

Since mobile transect data sets share several characteristics with conventional trajectory data sets, the related work for this study mainly consists of visual analytics for movement data, which will be described in this subsection.

During the last decade, research has been conducted on the visualization of and interaction with large geospatial trajectory data sets. Standard trajectory data sets are multidimensional, and include space, time, and other attributes associated with movement (speed, direction, acceleration), moving entities (age, gender, vessel type), or – as in our case – contextual environmental attributes that change over the course of a trajectory. Depending on the research goal and the underlying application scenario, studies concentrate on one or more of these dimensions.

Studies focusing on the spatial dimension of movement data frequently cluster, filter, or aggregate trajectories to facilitate reasoning about explanatory patterns. Rinzivillo *et al.* [25] apply progressive clustering to a data set describing car movement in Milan, where simple distance functions based on start, destination, a set of waypoints, temporal distances between those points or a combination of these aspects can be used successively. Andrienko *et al.* [2] describe an iterative clustering approach on a trajectory data set, which relies on the initial definition of a "classifier" resulting from clustering a subset of the data set. This classifier is then applied to the remaining data, while it can also be interactively refined if necessary. Andrienko and Andrienko [3] use accumulations of significant points on trajectories (starting points, destinations, stops, turns) to partition the underlying space. Movement occurring between these areas is aggregated and visualized using a flow map. Krüger *et al.* [20] developed "TrajectoryLenses", an intuitive spatial filter interface for a large trajectory data set. Using three different kinds of TrajectoryLenses, a user can investigate the set of trajectories with common starting points, end points, intermediate route points, or a combination of these. A temporal filter can also

be applied to the data by selecting an interval on a hierarchical time slider.

In addition, studies include spatial semantics and a visualization of how trajectory patterns change over time. One example is described by Krüger *et al.* [19], where information about places of interest in direct adjacency to the trip destinations are streamed from a location-based application called Foursquare. A time-line view reveals which places were visited when and how often. Growth ring maps were introduced by Bak *et al.* [5]. Whenever an entity visits a specific location, the ring around that place grows. Spatial semantics and temporal information are included by color-coding the rings. In a study by Wang *et al.* [39], spatial semantics occur as an underlying road network. Trajectory data recorded by taxis are matched with this network to investigate the spatial distribution of traffic jams in Beijing, and the result is visualized as a 2D map. The average speed of the taxis is encoded as additional attribute through color-coding. A time-line view shows the temporal distribution of the traffic jams, while a graph projection view displays the interrelations between traffic jams over space and time.

The simultaneous analysis of space, time, and multiple attributes has also been addressed in recent studies. Hurter *et al.* [17] describe a visual query system to select and analyze data subsets in several dimensions. The query results can be spread out over several displays to efficiently analyze subsets of the data, while still having the remaining data at hand. The visualization relies on 2D-scatterplots, i.e., attributes are paired without spatial context.

Some approaches spatially plot multiple trajectory attributes. Lampe *et al.* [21] use quantitative differences between kernel density estimate plots to compare movement, taking into account different time intervals, or different attributes either associated with the moving entities or the environment. Interactive exploration of the data set is facilitated by multiple coordinated views that are connected to each other via brushing and linking. Scheepens *et al.* [27] develop a density map, in which several movement density fields can be combined by three different aggregation techniques. The final density is encoded using a height map, while the initial density fields are color-coded for differentiation in the final image. Later, this visualization was extended [26]; the kernel for the density estimation can now be dynamically adapted to attributes of interest, and the combination of single density fields was designed to be more flexible.

Attempts have been made to encode space, time, and additional attributes of *individual* trajectories in detail. Visualizing trajectories on this level is certainly not always advantageous. It can lead to enormous clutter, and the overview is lost, unless route variations are low or the number of trajectories is small. Both conditions may be met for microclimate mobile-platform transect data sets, which makes studies dealing with such detailed visualizations valuable for our work.

Ware *et al.* [40] developed "TrackPlot", a tool to visualize whale tracks. The authors encode a whale's trip through 3D-space as a ribbon, while the movement direction is shown by arrows on the ribbon's surface. The ribbon is wound around a central axis, highlighting the animal's roll behavior, and glyphs are attached to show certain swimming techniques.

The study most closely related to our work is the visualization framework introduced by Tominski *et al.* [34]. This approach combines time, space, and one attribute in a hybrid 2D/3D visualization display. The authors stack individual trajectories as color-coded ribbons along the z-axis, ordered by a user-defined attribute. The temporal dimension can be included into the visualization by *temporally* ordering the ribbons along the z-axis. Additionally, the temporal variation of attributes is investigated by using a "time lens", which is a circular display that can be moved to a position of interest. We rely on several design decisions presented in [34]. Although the authors refer to the clustering capabilities of their framework, a limitation is that only one trajectory attribute can be visualized in detail per display instance. We overcome this limitation by using the third dimension for the synchronous display of multiple attributes of interest that belong to the same trajectory. Furthermore, we enable the user to interactively and quantitatively explore the spatial context of the trajectories.

4 METHODS

Our framework combines important data transformation and contextualization steps, a visual interface, and interactive analysis capabilities to facilitate the exploration of mobile-platform transect data. Design decisions have been made in collaboration with urban climatologists using mobile transect measurements to observe the relationship of urban structure on microclimate in a desert city. These domain-experts co-authored this paper, i.e., the presented solution, TraVis, was implemented by a team comprised equally of visualization researchers and urban climate experimental scientists.

4.1 Data preprocessing

To enable queries and real-time interaction with the data, transect data files are loaded into a relational database (MySQL 5.6, Oracle Corporation). Each transect run is stored as a table in the data base. The results of the preprocessing steps are added as additional columns so that each preprocessing step has to be performed only once, while all intermediate and final results of the analysis are stored. Range, mean, and standard deviation of the attributes are provided through a database summary functionality.

Several data transformation steps are required to correct the spatial and temporal position of the transect observations for an accurate visualization. First, the time series recorded by the sensors during the transect is linearly time-detrended to estimate each observation's value at the transect's starting time. Reference conditions can change significantly during the time elapsed while the transect route is traversed, contaminating the observed relationship between spatial context and environmental attributes along the transect route [32].

Data transformation also includes Geographical Information System (GIS) georeferencing, projecting the attribute observations into the coordinate system of the spatial context maps. In our case, the GPS positions are given as geographical coordinates and projected into a UTM conformal projection to store all data in a Cartesian reference space. The GIS coordinate transformations are carried out using the open-source GDAL library [12].

After projection, we extract the LULC type directly underneath the sensor platform location from a high-resolution LULC map [10] and store them as additional attribute in the corresponding tables, i.e. each measurement tuple is complemented with the LULC information. This is the immediate spatial context of the mobile platform itself, and is important for our visualization strategy, as will be explained in the next Section.

Finally, a filtering step is conducted before transect data are rendered to avoid visual artifacts. We extract the tuples corresponding to temporary pauses of the mobile platform's movement. In this case, the filter is applied to tuples in which the platform moves slower than its average speed. Then, the extracted tuples are grouped according to their sequence on the trajectory. For each connected component, the central vertex of the resulting trajectory segment is used as a spatial representative, and is associated with attributes that have been averaged over all members of the particular trajectory segment. Note that the results of this preprocessing step are not stored in the data base, but calculated each time a transect run is queried. This way, potentially important information is not lost.

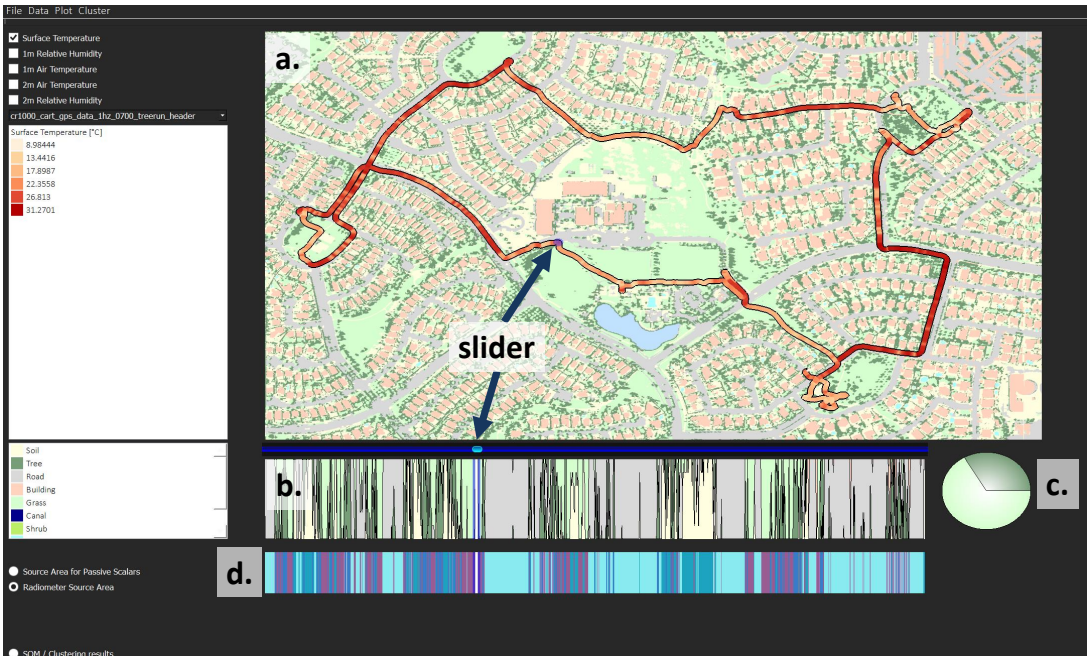

Figure 1: The graphical user interface. (a) The Map View. Plan-view overlay of a transect on the LULC layer, visualizing surface temperatures. Surface temperatures are higher (dark red) on asphalt, and cooler on grass LULC. (b) The Fraction Plot, showing the LULC fractions within the surface temperature infrared radiometer's field of view. (c) The pie-chart that complements the fraction plot. (d) Classification result based on surface temperature and land use fractions in the infrared radiometer's source area.

4.2 Data representation

The spatial context is represented by a high-resolution 2D LULC map, given by a georeferenced image file that is loaded into our map widget using the GDAL library [12]. Each pixel of this image is assigned a LULC class, and represents a spatial domain of 1 m^2 [10]. The displayed map is automatically scaled to the bounding box of the entire data set, while a visual buffering offset is added to each side. The LULC classes are by default semantically color-coded to facilitate quick perception of the spatial context, but the user may specify an individually preferred color scheme on-the-fly. The colors of the background map are rendered in low saturation to visually emphasize the stacked wall.

Similar to the visualization design described in [34], we use a hybrid 2D/3D representation of the trajectory data considered in our application scenario (Fig. 1a, Fig. 2). However, we do not use the third dimension to stack individual trajectories. In accordance with the semantics inherent in mobile transect data, we use the third dimension to display the multiple environmental attributes, stacked in order of the AGL offset height of each sensor. For our sample data set, this results in the following order: The surface temperature is rendered in the lowest layer, followed by 1 m relative humidity, 1 m air temperature, 2 m relative humidity, and finally 2 m air temperature as top layer.

The user selects a single transect run from a drop-down menu to the left side of the widget that provides all transects in the database as options. Then, the user can dynamically select a combination of environmental attributes and their sequence of display in the wall. It is also possible to only examine the value distribution of a single variable. Further layers can subsequently be added to or removed from the visualization display, enabling the user to choose another stacking order than the semantically meaningful default described above. User-defined adjacency of two specific layers in the wall is desirable to ease direct comparison of the two attributes they represent.

Figure 2 shows our wall-design in a close-up view, including all

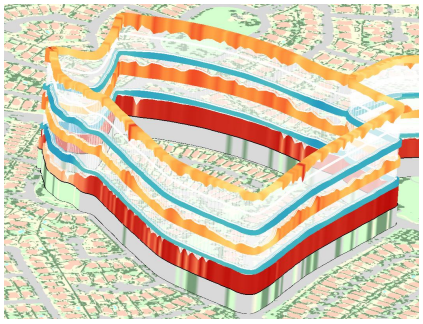

Figure 2: A wall in close up view, showing a segment of a transect run included in our sample data set: surface temperature, 1 m relative humidity, 1 m air temperature, 2 m relative humidity, and 2 m air temperature (from bottom to top). The lowest layer shows the LULC class of the particular route nodes; trees dark green, soil/gravel yellow, grass light green, hardscape light grey. Note the visual pattern of association between LULC, temperature, and humidity.

available environmental attributes in our sample data set. The attributes are visualized as ribbons stacked in the vertical wall that winds along the transect route. They are built around the underlying trajectories, which are created by connecting the points given by the GPS (the "nodes") with line segments. Using the concept of offset curves [16], we duplicate the trajectories on either side of the initial line to construct the bottom part of each wall layer. In contrast, the vertical thickness of an attribute's ribbon is not homogeneous over the course of the route, but is instead proportional to the measured values at each node. Between nodes the height is linearly interpolated to create a sense of spatial continuity.

A known drawback of 3D visualization is the difficulty to per-

ceive and compare magnitudes [8]. Relying on the height of the wall-layers as the only indicator for the value of an attribute at a certain location is thus insufficient. Therefore, the ribbons are additionally color-coded so that both the intensity of color and thickness of the ribbon are proportionate to the normalized value of the attribute. The color scale for the values depicted along the ribbons is predefined through a set of sequential colors in the ColorBrewer [14]. Lighter colors are associated with lower values and darker colors with higher values, following the cartographic convention of *"Dark equals more"* [14]. The user may specify an individually preferred color scheme for each ribbon. In this case we select red for temperature and blue for humidity attributes.

The semi-transparent surface between the ribbons serves three functions. First, it enhances the perception of the height and color of the individual ribbon. To ensure that this effect also applies to the uppermost ribbon, the top ribbon is inverted. Second, the surface visually connects individual wall-layers and enhances the perception of which points on the ribbons share a spatial location on the map. Third, the gaps that arise between the ribbons reduce occlusion because the semi-transparency provides a view behind the wall. However, a certain amount of occlusion is still possible, so the view can be rotated in three dimensions and translated or scaled to change the perspective.

An issue that cannot be solved using rotation, translation, and scaling is the occlusion of the immediate spatial context directly under the mobile platform transect by the wall itself. To avoid losing this important context information, a LULC class ribbon is added as the lowest layer of the wall, repeating the color-coding of the immediate spatial context.

4.3 Analyzing the impact of the spatial context on attribute values

4.3.1 Theoretical considerations

In micrometeorology, the field of view of an atmospheric sensor can be described using the *source area* [28] concept. Sensor height AGL, wind speed and direction, and atmospheric factors determine the shape and position of the *source weight function*, which describes the relative contribution of the different upwind locations to the sampled value [36, 28]. The source area is the projection of the isopleths of this source weight function to a 2D plane [28].

For the surface temperature, which is in this case measured using an infrared radiometer, the source area is a circular area that is the field of view of the instrument [29]. The radius r of this disc can easily be determined using the following equation [29]:

$$r = \frac{h}{\sqrt{\frac{1}{F} - 1}} \tag{1}$$

with h quantifying the mounting height of the radiometer in $[m]$. F is the dimensionless view factor that describes the ratio between a circular area on the ground and its surrounding annulus (limited by the horizon for a radiometer with a field of view of 180°) [29]. In our implementation, F is constantly set to 0.95, which means that the elements located within the resulting disc contribute to 95 % of the measured surface temperature signal [24].

The computation of a source area for passive scalars such as air temperature is more complex, because the heat fluxes determining this quantity underly turbulent transport mechanisms [29]. These, in turn, are dependent on several atmospheric properties that are prevalent at the time the quantity was measured. The land surfaces impacting the air temperature are located upwind of the sensor. Thus, the orientation of the source area's semimajor axis generally varies with the predominant wind direction at the spatio-temporal location of sampling. However, other factors also affect the source area; this has been a research area over the last 30 years, applying

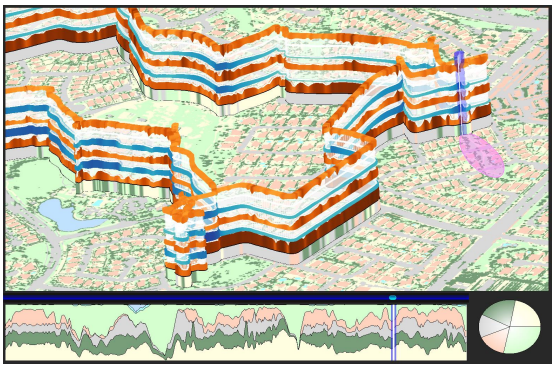

Figure 3: The slider can be used to interactively explore the spatial context of the data. The source area is attached to the bottom of the slider, while the fraction plot and the pie chart underneath the map display the composition of LULC fractions within the source area. The ellipsoid exemplifies the source area for air temperature measured at a height of $1\ m$.

methods such as Large Eddy Simulation or Lagrangian stochastic particle dispersion models [36].

Since the source area needs to be estimated for each node on each trajectory, the estimation technique must be computationally affordable. Therefore, we rely on Schmid's simple parametric model to determine a source area's shape and size [28]. Five parameters are required to compute the shape parameters for the resulting ellipsoid, with the orientation of the semimajor axis being parallel to the wind direction: measurement height, Obhukov length (a measure for the atmospheric stability), surface roughness length, friction velocity, and the standard deviation of lateral wind speeds. These parameters cannot be directly computed from our sample data set, but have to be determined using external information. For the prototype described in this paper, the orientation of the source area was estimated using averaged wind direction values retrieved from four weather stations surrounding the study area [22]. The other necessary parameters were selected such that the final ellipsoid approximates a realistic size.

4.3.2 Visualization

Depicting all source areas for each point on the route simultaneously would result in numerous overlapping areas, making it difficult to distinguish the source areas of individual points on the route. In our framework, the source areas for distinct sensors can be interactively explored by means of a slider, which can be moved along the transect route. The user can choose which source area is displayed by using the radiobuttons in our graphical user interface (Fig. 1). The slider consists of two parts: The visualization of the source area at the bottom, and a semi-transparent cylindrical bar. The latter allows for the unambiguous attribution of the slider elements to a particular spatial location traversed by the mobile platform. To ensure that the bar does not occlude the measurement values at the selected location, it is rendered transparently. A thin white plane is added to the bar, which indicates the exact slider position (Fig. 3).

Two different measurement field of views are computed within our framework as described in Section 4.3.1: The radiometer source area and the source area for passive scalar fluxes (e.g., air temperature). In our sample data set, the radiometer is mounted at a height of approximately 0.3 m, resulting in a disc radius of 1.28 m. Consequently, the circle that represents this source area is so small that it disappears underneath the wall, whose width on the map is broader than this radius. Furthermore, depending on the predominant wind direction, the source area for passive scalars might overlap with the

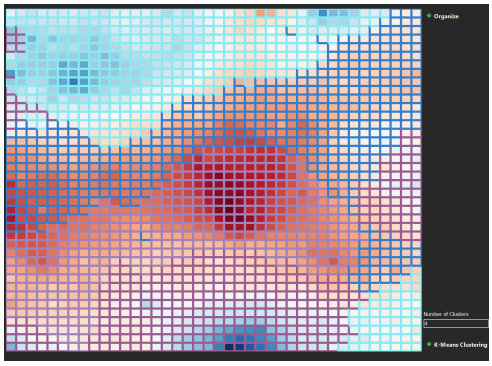

Figure 4: The cluster interface, which is linked to the main GUI depicted in Figure 1. In the background, the lengths of the weight vectors associated with the individual self-organizing map (SOM) nodes are visualized. The borders of the nodes represent each node's cluster membership after the k-means clustering has been applied.

transect route, so that parts of the resulting ellipsoid are occluded by the wall. Nevertheless, the fractions of the LULC classes within the source area can be detected in the LULC composition display located underneath the map, and the adjacent pie-chart that provides an alternative representation of the fractions. The map, the fraction plot, and the pie-chart are linked: While the slider indicating the position on the transect is repeated in the fraction plot, the pie-chart adapts dynamically to the values associated with the slider's position on the map and in the fraction plot. To ensure a smooth visual transition of the pie chart between different segments, the LULC classes are always allocated in the same order on the pie chart.

The source area for passive scalar fluxes is constructed using the shape parameters given by Schmid's parametric model [28], which include the upwind distance from the sensor to the near end of the ellipsoid, the distance to the far end, the distance to the point where the ellipsoid has the maximum width, and the length of the corresponding semi-minor axis. We use rational cubic Beziér curves to create conic sections that are attached to each other to construct the egg-shaped function [11]. The resulting shape is rendered semi-transparent, so that the LULC underneath this shape can still be recognized.

4.4 Classifying segments of the transect

The functionality of the framework is complemented by an interactive clustering interface, which can be used to classify the transect into segments of similar multi-attribute patterns. While the above described visualization and analysis capabilities facilitate the qualitative reasoning on the values and their interdependencies, the classification provides quantitative support.

The clustering is based on a combination of a self-organizing map (SOM) [18] and the k-means clustering algorithm (Fig. 4). This combination has been demonstrated to reliably structure the results of the SOM into a lower number of clusters, which is particularly useful if the number of nodes in the SOM is large [37]. Furthermore, it has also been successfully used in climate research, e.g. to visualize areas of similar temporal snowfall variation [38]. A drawback is the computation time, which does not allow for real-time interaction if a large number of SOM nodes and SOM training iterations is used. Another disadvantage of this approach lies in the dependency of the k-means clustering algorithm on the number and location of the initial cluster centroids. While we determine the initial positions of the cluster centroids randomly, a user can refine the number of cluster centers after visually inspecting the results of the SOM, or by considering hypotheses about the expected number of different microenvironments along the route.

We allow the user to interactively select the attributes to be clustered. Clustering some combinations of attributes might not lead to meaningful results. For example, there is no physical relationship between the infrared radiometer's source area and the sampled air temperature because the thermometer employed to observe the latter has a different field of view.

The results of the classification are visualized in a widget on our user interface (Fig. 1d), which is visually connected to the slider described above, so that the link between cluster membership, spatial context, and attribute values at that point on the transect can be interactively explored. Additionally, the clustering results are added to the wall as topmost layer.

5 Case study: Classifying segments of the transect route based on sampled atmospheric data

The visualization was tested using a mobile transect data set recorded on May 08, 2014, between 14:40 and 15:50 MST using a sensor platform mounted to a golf cart. Surface temperature, air temperature in 1 and 2 m and relative humidity in 1 and 2 m were sampled with a frequency of 1 Hz. The data were collected in a master-planned suburban community named Power Ranch, which is located in Gilbert, Arizona, USA. May is a hot and dry month in this semiarid climate region.

In order to investigate whether the observations along the transect route can be classified into distinct segments of similar variable behavior, the clustering capability of the framework is used. The clustering functionality utilizes the normalized values for air temperature at 1 m and 2 m, relative humidity at 1 m and 2 m, and surface temperature as input for the SOM, which consists of a network of 1600 nodes. After inspecting the results of the SOM training phase, we decide to partition our transect data set into 6 classes (Fig. 5a), which is also the number of the predominant LULC classes at our study site: soil, grass, tree, asphalt, building, and water (swimming pool and lake).

We inspect the classification of the points on the transect route by stacking the cluster membership as the topmost layer upon the wall. The cluster membership can now be examined in a standard 2D spatial plot (Fig. 5b). Comparing the SOM / k-means results with the land use patches surrounding the corresponding transect node, we find a visual correlation between the traversed patch and the cluster membership.

To investigate this further, we rotate the map to visually explore the meaning of the different clusters (Fig. 5c/d). Since the rotation reveals the value distribution of all selected attributes, we can relate the clusters to the multivariate value distribution shown on the stacked ribbons (also compare Table 1). Using this technique, we observe that the seemingly isolated segments belonging to cluster 6 near the lake in part B and in the south-eastern park (part D). This cluster corresponds to exceptionally high humidity (Fig. 5c, Table 1), presumably owing to the lake's open water surface evaporation and increased evapotranspiration in the park in part D.

Interestingly, park areas with grass LULC are associated with different microenvironmental clusters. Cluster 3 occurs near the lake in part B (Fig. 5b), in the north-eastern park (part C), in the park in the southern part of D, as well as during a small segment in the western park (western part of A). It is associated with a typical value distribution for vegetated areas, especially in terms of the low surface temperature. However, park areas are not consistently represented by cluster 3. Thus, the northern part of the park in part B largely belongs to cluster 2, which is defined by high surface temperatures (Fig. 5b, d, Table 1). Cluster 2 corresponds to the microenvironment of an asphalt-dominated LULC type that is adjacent to, and under the influence of, a grass covered park area. The same holds for the northern part of the park in the western part of A. The two park stretches in the southern part A and in the center of

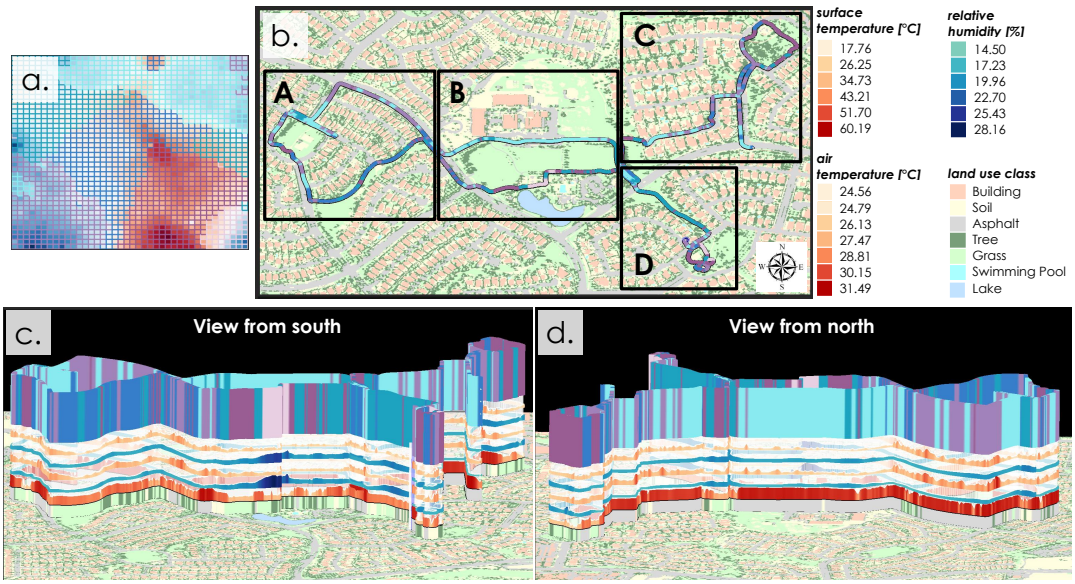

Figure 5: A use case for our framework. In this example, we analyze the classification of a transect route into similar microenvironmental segments, based on a combined SOM and k-means clustering algorithm [37, 38] over all measured quantities. This clustering is based purely on associations between (in order of stacking from surface) surface temperature, air temperature, and relative humidity at 1 m and 2 m AGL heights. (a) displays the visualization of the clustering output. (b) shows the distribution of the cluster members along the transect route from the birds-view perspective. Note the correlations between patches of land use and cluster membership. (c) and (d) visualize the meaning of the different clusters in terms of multivariate value distribution, with (c) showing the wall from the south, (d) from the north.

part D also belong to two different clusters. This can be explained by a difference in relative humidity, which is higher in the park stretch in part D (Fig. 5c).

Cluster	Average surface temperature [°C]	Average air temperature 1m [°C]	Average air temperature 2m [°C]	Average rel. humidity 1m [%]	Average rel. humidity 2m [%]
1	44.9	27.2	26.2	17.4	18.1
2	55.8	26.8	25.7	17.4	18.2
3	30.9	26.7	25.9	18.7	18.6
4	53.6	28.4	27.1	17.3	18.2
5	47.0	26.3	24.5	19.5	19.2
6	32.6	26.5	25.8	24.0	21.2

Table 1: The value distribution in the single clusters.

The area between two arrays of buildings is most consistently associated with cluster 4. This cluster is defined by high surface temperatures due to the asphalt cover, high air temperatures, and low relative humidity. There are, however, certain disruptions of this pattern, mostly by transect segments related to cluster 2. In these segments, the air temperature is lower, while surface temperature and relative humidity are similar to cluster 4.

6 Conclusion

A prototypic framework for the visualization and analysis of mobile environmental transect measurements is introduced based on contributions from the disciplines of visualization and urban climatology. Each measured attribute is visualized as a three-dimensional ribbon, whereas height and color emphasize the values measured at a particular location. Several ribbons can be vertically stacked in a wall to allow for the comparison of the value distribution between the attributes. A slider can be used to interactively investigate the composition of the *source area* at each measurement location, while the LULC composition within this field of view is encoded by a pie-chart and a fraction plot. Complemented by a clustering capability, the framework combines analysis capabilities for the exploration of the spatial context and the multivariate relationships between the measured attributes.

The framework has been tested using data from a mobile transect measurement campaign in Gilbert, Arizona, USA. Multivariate relationships between measured attributes could easily be detected, physically meaningful clusters determined, and connections between land use and observations drawn.

For our future work, we further extend the analysis capabilities of this framework by including the temporal dimension into the visualization and clustering capability. If more transect runs have been conducted on a similar route, the individual points on the trajectories have to be registered, so that the behavior of the measured variables at certain locations can be compared over time. We also plan to add the capability to evaluate multiple types of spatial context such as the height map of the urban environments features. Evaluation of the quantitative definitions and boundaries of each cluster will also be introduced. Furthermore, the estimation of the source areas for passive scalars will be improved and evaluated thoroughly. While we are currently only using wind direction data from surrounding weather stations to estimate the orientation of the source area, we are planning to integrate more data sources to retrieve information about the atmospheric stability and approximate turbulence characteristics at each measurement location. To refine the estimation of the passive scalar source area's shape, we are adding additional source area models that incorporate height maps and canopy structure.

Acknowledgements

The authors wish to thank the reviewers for their valuable comments, as well as Nils Feige and Diana Fernàndez Prieto for helpful discussions. Thanks also go to the ASU Environmental Remote Sensing and Geoinformatics Lab (ERSG) for providing the NAIP data set (additional support was furnished by the Gilbert F. White Environment and Society endowment. Source data: National Agriculture Imagery Program (NAIP), http://www.fsa.usda.gov). This

work was supported in part by the NSF Grant SES-0951366, Decision Center for a Desert City II: Urban Climate Adaptation, the NSF EaSM Program EF-1049251, the NSF LTER Program BCS-1026865, the Salt River Project grant to ASU, Alan and Sandra Ruffalo, and the Power Ranch Homeowners Association.

REFERENCES

[1] G. Andrienko and N. Andrienko. Spatio-temporal aggregation for visual analysis of movements. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pages 51–58, Oct 2008.

[2] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive visual clustering of large collections of trajectories. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 3–10, Oct 2009.

[3] N. Andrienko and G. Andrienko. Spatial generalization and aggregation of massive movement data. *Visualization and Computer Graphics, IEEE Transactions on*, 17(2):205–219, Feb 2011.

[4] N. Andrienko, G. Andrienko, N. Pelekis, and S. Spaccapietra. Basic concepts of movement data. In F. Giannotti and D. Pedreschi, editors, *Mobility, Data Mining and Privacy*, pages 15–38. Springer Berlin Heidelberg, 2008.

[5] P. Bak, F. Mansmann, H. Janetzko, and D. Keim. Spatiotemporal analysis of sensor logs using growth ring maps. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):913–920, Nov 2009.

[6] A. J. Brazel and D. M. Johnson. Land use effects on temperature and humidity in the Salt River Valley, Arizona. *Urban Forestry & Urban Greening*, 15(2):54–61, 1980.

[7] W. T. Chow, R. L. Pope, C. A. Martin, and A. J. Brazel. Observing and modeling the nocturnal park cool island of an arid city: horizontal and vertical impacts. *Theoretical and Applied Climatology*, 103(1-2):197–211, 2011.

[8] T. N. Dang, L. Wilkinson, and A. Anand. Stacking graphic elements to avoid over-plotting. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1044–1052, Nov 2010.

[9] B. Elen, J. Peters, M. V. Poppel, N. Bleux, J. Theunis, M. Reggente, and A. Standaert. The Aeroflex: A bicycle for mobile air quality measurements. *Sensors*, 13(1):221–240, 2012.

[10] Environmental Remote Sensing and Geoinformatics Lab and CAP LTER. 4 band NAIP land classification of Central Arizona, 2012. Arizona State University.

[11] G. Farin. *Curves and Surfaces for CAGD: A Practical Guide*. Academic Press, Inc., London, UK, 4th edition, 1997.

[12] GDAL Development Team. *GDAL - Geospatial Data Abstraction Library, Version 10.0.1*. Open Source Geospatial Foundation, 2013.

[13] R. Hagemann, U. Corsmeier, C. Kottmeier, R. Rinke, A. Wieser, and B. Vogel. Spatial variability of particle number concentrations and NOx in the Karlsruhe (Germany) area obtained with the mobile laboratory AERO-TRAM. *Atmospheric Environment*, 94(0):341 – 352, 2014.

[14] M. Harrower and C. Brewer. ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, Jun 2003.

[15] B. G. Heusinkveld, G. J. Steeneveld, L. W. A. van Hove, C. M. J. Jacobs, and A. A. M. Holtslag. Spatial variability of the Rotterdam urban heat island as influenced by urban land use. *Journal of Geophysical Research: Atmospheres*, 119(2):677–692, 2014.

[16] J. Hoschek. Spline approximation of offset curves. *Computer Aided Geometric Design*, 5(1):33 – 40, 1988.

[17] C. Hurter, B. Tissoires, and S. Conversy. FromDaDy: Spreading aircraft trajectories across views to support iterative queries. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1017–1024, Nov 2009.

[18] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, Sep 1990.

[19] R. Krüger, D. Thom, and T. Ertl. Visual analysis of movement behavior using web data for context enrichment. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE*, pages 193–200, March 2014.

[20] R. Krüger, D. Thom, M. Wörner, H. Bosch, and T. Ertl. TrajectoryLenses – a set-based filtering and exploration technique for long-term trajectory data. *Computer Graphics Forum*, 32(3):451–460, 2013.

[21] O. D. Lampe, J. Kehrer, and H. Hauser. Visual analysis of multivariate movement data using interactive difference views. In R. Koch, A. Kolb, and C. Rezk-Salama, editors, *VMV 2010: Vision, Modeling & Visualization*, pages 315–322, Siegen, Germany, 2010. Eurographics Association.

[22] MesoWest / University of Utah. MesoWest data. Online: http://mesowest.utah.edu/, 2015. Last accessed: 01/17/2015.

[23] D. J. Murphy, M. H. Hall, C. A. S. Hall, G. M. Heisler, S. V. Stehman, and C. Anselmi-Molina. The relationship between land cover and the urban heat island in northeastern Puerto Rico. *International Journal of Climatology*, 31(8):1222–1239, 2011.

[24] T. Oke. *Initial Guidance to Obtain Representative Meteorological Observations at Urban Sites*. Instruments and Observing Methods 81, WMO/TD 1250. World Meteorological Organization, 2006.

[25] S. Rinzivillo, D. Pedreschi, M. Nanni, F. Giannotti, N. Andrienko, and G. Andrienko. Visually driven analysis of movement data by progressive clustering. *Information Visualization*, 7(3-4):225–239, 2008.

[26] R. Scheepens, N. Willems, H. van de Wetering, G. Andrienko, N. Andrienko, and J. van Wijk. Composite density maps for multivariate trajectories. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2518–2527, Dec 2011.

[27] R. Scheepens, N. Willems, H. Van De Wetering, and J. van Wijk. Interactive visualization of multivariate trajectory data with density maps. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE*, pages 147–154, March 2011.

[28] H. Schmid. Source areas for scalars and scalar fluxes. *Boundary-Layer Meteorology*, 67(3):293–318, 1994.

[29] H. Schmid, H. Cleugh, C. Grimmond, and T. Oke. Spatial variability of energy fluxes in suburban terrain. *Boundary-Layer Meteorology*, 54(3):249–276, 1991.

[30] M. Sofer and O. Potchter. The urban heat island of a city in an arid zone: the case of Eilat, Israel. *Theoretical and Applied Climatology*, 85(1-2):81–88, 2006.

[31] L. B. Stabler, C. A. Martin, and A. J. Brazel. Microclimates in a desert city were related to land use and vegetation index. *Urban Forestry & Urban Greening*, 3(34):137 – 147, 2005.

[32] I. D. Stewart. A systematic review and scientific critique of methodology in modern urban heat island literature. *International Journal of Climatology*, 31(2):200–217, 2011.

[33] C.-Y. Sun, A. Brazel, W. Chow, B. Hedquist, and L. Prashad. Desert heat island study in winter by mobile transect and remote sensing techniques. *Theoretical and Applied Climatology*, 98(3-4):323–335, 2009.

[34] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko. Stacking-based visualization of trajectory attribute data. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2565–2574, Dec 2012.

[35] J. K. Vanos, J. S. Warland, T. J. Gillespie, G. A. Slater, R. D. Brown, and N. A. Kenny. Human energy budget modeling in urban parks in Toronto and applications to emergency heat stress preparedness. *Journal of Applied Meteorology and Climatology*, 51(9):1639–1653, 2012.

[36] T. Vesala, N. Kljun, U. Rannik, J. Rinne, A. Sogachev, T. Markkanen, K. Sabelfeld, T. Foken, and M. Leclerc. Flux and concentration footprint modelling: State of the art. *Environmental Pollution*, 152(3):653 – 666, 2008.

[37] J. Vesanto and E. Alhoniemi. Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on*, 11(3):586–600, May 2000.

[38] N. Wang, T. W. Biggs, and A. Skupin. Visualizing gridded time series data with self organizing maps: An application to multi-year snow dynamics in the northern hemisphere. *Computers, Environment and Urban Systems*, 39(0):107 – 120, 2013.

[39] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering. Visual traffic jam analysis based on trajectory data. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2159–2168, Dec 2013.

[40] C. Ware, R. Arsenault, M. Plumlee, and D. Wiley. Visualizing the underwater behavior of humpback whales. *IEEE Computer Graphics and Applications*, 26(4):14–18, 2006.