Workshop on Visualisation in Environmental Sciences (EnvirVis) (2015)
A. Middel and K. Rink and G. H. Weber (Editors)

Spatial Aggregation of Mobile Transect Measurements for the Identification of Climatic Microenvironments

K. Häb †[1], A. Middel[2], B. L. Ruddell[3] and H. Hagen[1]

[1]Computer Graphics and HCI Group, University of Kaiserslautern, Germany
[2]Julie Ann Wrigley Global Institute of Sustainability, Arizona State University, Arizona, USA
[3]The Polytechnic School, Arizona State University, Arizona, USA

Abstract

Mobile transect measurements retrieve high-resolution observations revealing the spatial variation of atmospheric properties throughout an urban landscape. A sensor platform is moved through a study site with varying urban form, collecting a data set that can be used to investigate the multifaceted impacts of different building and landscape configurations on atmospheric properties. To generalize such findings, it is imperative to include transect runs representing different points in time and potentially different meteorological background conditions. However, the analysis of a set of mobile transect measurement runs is challenging because of the strict spatio-temporal dependence and multivariate nature of each recorded sample. In this study, we provide visual support for the identification of coherent climatic microenvironments within a study site using mobile transect measurements taken at different points in time and over diverse routes. A regular grid is used to spatially aggregate the data, and resulting summaries are classified according to similar multivariate relationships using clustering techniques. Finally, each grid cell is visualized using a radial glyph encoding cluster membership, predominant wind direction for each transect run, and the number of transect runs traversing this grid cell. The approach has been tested using a data set recorded in Gilbert, Arizona, USA, and it shows potential to identify spatially contiguous regions of similar microclimate.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.8]: Applications—; Computer Applications [J.2]: Earth and atmospheric sciences—

1. Introduction

Mobile transect measurements are frequently used in urban climatology to gain insight into the spatial variation of atmospheric properties. Based on the resulting observations, conclusions can be drawn about the impact of urban form on the surrounding climate [CPMB11, HSvH*14, SBC*09]. However, a single transect is not sufficient to generalize such findings because it represents a single meteorological background condition. Data has to be sampled repeatedly, resulting in a spatially dependent, multivariate, time-varying data set, which is difficult to analyze.

Aggregation techniques can be used to reduce the complexity of a mobile measurement data set and to facilitate reasoning about coherent local climate zones showing a similar value distribution over time. These zones share a coherent relationship between land use/land cover (LULC) and microclimate [SO12]. One approach to explore the multivariate and spatio-temporal information contained in a mobile transect data set could be to extract a multivariate time series for each sample location on a transect route, which could then be used to partition the underlying space into segments of similar temporal value distribution. This, however, is only possible if the observation routes are spatially identical.

This paper investigates an approach to comprehensively visualize data resulting from multiple transect runs over diverse routes. The overall goal of the visualization is to support the exploration of potential areas of similar multivariate value behavior under varying meteorological back-

† e-mail: kathrin.haeb@cs.uni-kl.de

DOI: 10.2312/envirvis.20151086

ground conditions. The approach is based on the spatial aggregation of multiple transect runs over a regular grid. Multivariate relationships are classified separately for each transect run using a combination of the self-organizing map (SOM) [Koh90] and the k-means clustering algorithm [WBS13, VA00]. Thus, per grid cell, we gain a summary about the multivariate behavior of all observations taken within this spatial compartment. To visually encode this information, we designed a glyph that represents the number of transect routes crossing the grid cell, their respective cluster membership at this location, and the wind direction that has been predominant during data collection in the field. The approach is implemented into TraVis, our framework for the visualization of mobile transect measurements [HMRH15].

2. Theoretical background and related work

Mobile transect measurements are a multivariate geospatial movement data set. Thus, each transect is a trajectory with a set of spatial locations $S = \{s_0, \dots, s_n\}$, a time stamp $t_i \in T$ associated with each s_i, and further attributes $A_{0,\dots,m}$, whose elements are either static over the entire trajectory or change dynamically with S and T [AA08, TSAA12].

Andrienko and Andrienko [AA10] describe a general framework for the aggregation of movement data sets. They distinguish two different views on movement: The *trajectory-oriented view*, which focuses on the movement of single entities, and the *situation-oriented view*, which focuses on the state of the entities at one or more points in time. For each of these two views, they give recommendations about applicable aggregation techniques, which are again based on two different views of space: The *space-centered view*, focusing on the space in which the movement occurs, and the *entity-centered view*, focusing on the movement itself. According to this theoretical framework, our analysis purpose corresponds to the category *space-centered* and *situation-oriented*, since we are interested in the relationship between trajectory attributes at certain points in time and space.

The traffic-oriented view on a car data set in Milano [AA08] is closely related to our visualization approach. The authors suggest to partition the space using a regular grid and to aggregate the data over each grid cell. The aggregation results are visualized using either small multiples showing the frequency of car traverses per grid cell and time step, or – similar to our solution – using radial glyphs that encode the traffic intensity per movement direction. Scheepens *et al.* [SWVDWvW11, SWvdW*11] combine multiple density maps to provide a comprehensive overview over multivariate movement data sets. In a later study [SvdWvW14], they use pie charts as glyphs on top of a map to summarize the number of certain objects, their heading, and the proportion of stationary objects within a spatial compartment. Bak *et al.* [BMJK09] aggregate their episodic movement data set using *growth-ring maps*. Also dealing with episodic movement data, Andrienko *et al.* [AAS*12] classify spatial situations by clustering feature vectors representing presence counts per location.

3. Methodology

3.1. Spatial aggregation of multivariate mobile transect measurements

A natural approach to spatially aggregate a set of trajectories is to partition the space into compartments and to summarize data collected for each of these compartments [AA08, SvdWvW14]. In our solution, a regular grid is spanned over the bounding box of all transect routes. Since spatial scale plays an important role during the analysis, the size of the grid cells can be adapted to the scale under investigation (Fig. 1). The observations are averaged separately for each transect run and each variable over each grid cell. Furthermore, to guarantee comparability of data sampled at different points in time and data represented in different units, all samples belonging to one transect run and one variable are scaled to the interval $[0, 1]$ based on their individual value range.

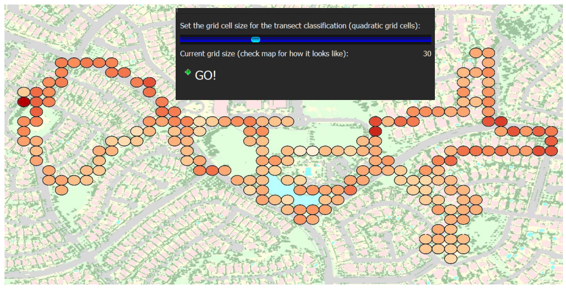

Figure 1: *Aggregating data over a regular grid.*

Since the transect routes vary, the number of mobile measurement runs traversing a grid cell also varies. Thus, an appropriate aggregation and visualization technique has to take this asymmetry into account. Since the overall goal of this study is to identify spatially coherent climatic microenvironments, data is clustered based on a user-defined set of variables. A combination of the SOM [Koh90] and the k-means clustering algorithm is applied to find a semantically meaningful structure within the data set [VA00, WBS13]. The SOM has been implemented roughly following [Koh90], using a rectangular grid. The initial cluster centroids for the k-means clustering are selected randomly.

In previous work [HMRH15], this technique was successfully used to partition a single transect measurement run into segments of similar multivariate relationships. In the current study, we make use of this finding and use the cluster membership of each aggregated sample as a summary measure for multivariate behavior at the corresponding spatial location. Since the SOM is computationally expensive, it is trained based on one, user-defined exemplary transect run, and then partitioned using k-means clustering [VA00, WBS13]. The quality of the partition can be explored using a parallel coordinates plot, which shows the clustering results as applied to the exemplary transect run. In this plot, classes are color-coded with a qualitative color-scheme [HB03] and can be brushed. If the user is not satisfied with the current partition, the number of cluster centroids for the k-means clustering can be interactively refined. Once the partition of the exemplary transect run is finished, the clustering results are applied successively to all other transects available in the data set.

The disadvantage of this approach lies in the potential sensitivity of the result on the selection of the first, "representative" transect run. Theoretically, it has to be chosen based on the number of distinct multivariate value combinations, which can then be classified appropriately and detected in subsequently added runs. A preliminary sensitivity analysis confirmed this hypothesis, and we are currently conducting further research on this issue.

3.2. Visualization approach and glyph design

To visualize this data on a map, a glyph is assigned to each grid cell that has been crossed at least once. The glyph was carefully designed to encode the...

- R1: ...number of transect runs traversing the grid cell.
- R2: ...cluster membership for each transect traversing it.
- R3: ...predominant wind direction during the time the transect has been conducted.
- R4: ...grid cell size over which the data has been aggregated.

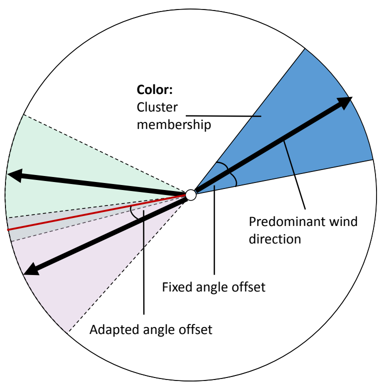

Figure 2: *Glyph design.*

The number of clusters (R1) has to be included into the glyph to enhance the perception of uncertainty associated with the visualization. If only one sample is responsible for the appearance of a grid cell, the reliability of conclusions drawn from this representation is reduced. Encoding the cluster membership for each transect traversing a grid cell is also mandatory for the visualization (R2) because it reveals information about multivariate relationships found at this location. The predominant wind direction for each transect traversing the grid cell (R3) can give further hints about the relationship between the values measured at this point and the LULC upwind. This holds especially true for atmospheric attributes, whose spatial distribution is dependent on mixing processes [Sch94]. Finally, the grid cell size (R4) hints at the spatial resolution of the resulting aggregation.

Figure 2 shows our glyph design in a schematic way. A circular layout is used, since this design enables us to easily encode the predominant wind direction at sampling time (R3) by the orientation of sectors. These sectors are created by applying a fixed offset angle left and right of the vector pointing into the wind direction, which is in our implementation given by the predominant wind direction measured at four weather stations surrounding the study site [Mes15, HMRH15]. The number of transect runs traversing the grid cell is encoded by the number of sectors arranged around an inner circle, fulfilling R1. Color coding these sectors according to the cluster membership fulfills R2. If two sectors would overlap due to similar wind directions at sampling time, the sector border between these two sectors is moved to the half-angle between the two respective wind directions. The grid cell size (R4) is proportional to the radius of the circle, which also prevents spatially adjacent glyphs from overlapping.

4. Use Case

The visualization was tested using a mobile transect measurement data set collected in a residential neighborhood in Gilbert, Arizona, USA. The data set was recorded on four different days in May 2014, September 2014, and February 2015. It consists of 21 transect runs with an average of 4333 sample points. For analysis, we considered five variables: Surface temperature, 1 and 2 m air temperature, and 1 and 2 m relative humidity.

As an exemplary transect run, we choose a run that has been conducted at September 15, 2014, at 0700 LST. It traverses a longer route, covering a potentially large number of different multivariate value configurations. The grid cell size is chosen to be 30 m. For the SOM, we use a field of $10 \, x \, 10$ neurons and let the training run over $N * 10$ iterations (N is the number of grid cells traversed by the exemplary transect run). Then, we apply a k-means clustering over 6 cluster centroids. Using the parallel coordinates plot, we find that the data was well-partitioned into distinct classes of multivariate relationships (Fig. 3a).

Then, we use the clustering results to classify the spatially aggregated data belonging to all other transect runs.

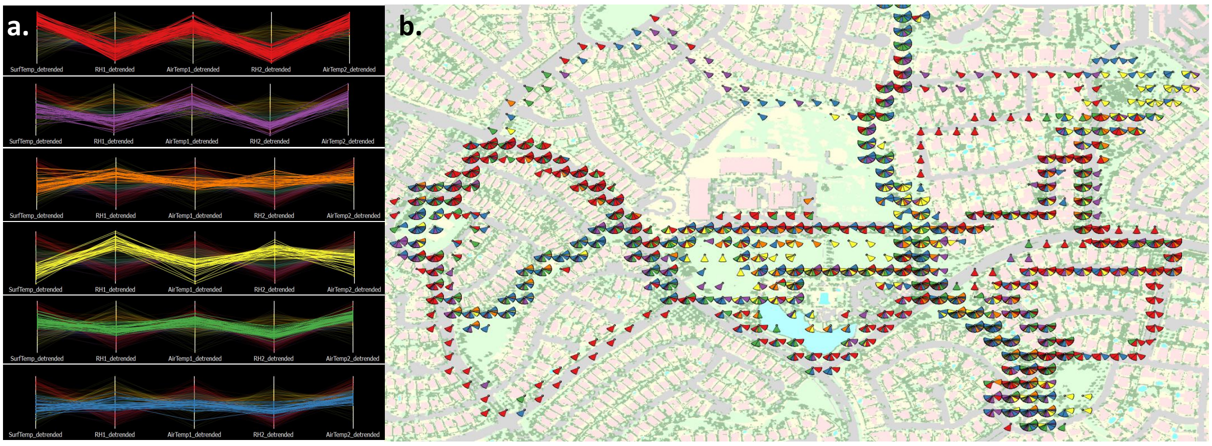

Figure 3: *a. Brushed parallel coordinates plots to visualize the meaning of the clusters. b. The glyphs combining the entire set of mobile transect measurements on a map (background map: [Env12]).*

In the resulting visualization (Fig. 3b), we can identify several patterns: First, it is obvious that the wind never comes from northerly directions. Second, red classes, associated with high surface temperatures, low humidities, and high air temperatures, appear frequently over asphalted areas and between arrays of houses, as would be expected for this kind of environment. Third: The yellow cluster, associated with low surface temperatures, high humidities and relatively low air temperatures, can predominantly be found in parks.

Coherent climatic microenvironments can qualitatively be identified by searching for patterns of predominant colors. However, this does not necessarily hold true for glyphs comprising a large number of different colors. In this case, it is not clear, whether the distinct clusters only correspond to the wind direction alone or also to other meteorological background conditions at sampling time, e.g., because the compartment joins data belonging to different times of a day or a year.

5. Conclusion and Future Work

In this paper, we described a visualization approach to visually identify climatic microenvironments within a study site based on a number of mobile transect measurements. We partition the space using a regular grid, before we aggregate the data associated with each grid cell by classifying it according to multivariate relationships and visualize it using radial glyphs. The glyph design enables the synchronous visualization of (a) the number of transect runs that contributed to the glyph, (b) the predominant wind direction at recording time, (c) multivariate relationships, and (d) the grid cell size. It supports forming hypotheses about the impact of urban design on microclimate, while also taking local data sparseness into account.

For our future work, we plan to explicitly incorporate the position of each transect in a temporal cycle, such as time of day or time of year, both of which are meteorologically relevant. Since our glyphs do not provide sufficient space to additionally encode this information, we plan to add a filtering capability that can be used to brush the glyphs on the map. We are also currently investigating the automatic computation of coherent microenvironments based on a metric that describes the similarity of the glyphs to each other. Based on this metric, we also aim at quantifying the sensitivity of the algorithm to different input configurations.

Acknowledgements

The authors wish to thank the reviewers for their valuable comments, as well as the Arizona State University 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, 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

[AA08] ANDRIENKO G., ANDRIENKO N.: Spatio-temporal aggregation for visual analysis of movements. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on* (Oct 2008), pp. 51–58. doi:10.1109/VAST.2008.4677356. 2

[AA10] ANDRIENKO G., ANDRIENKO N.: A general framework for using aggregation in visual exploration of movement data. *The Cartographic Journal 47*, 1 (2010), 22–40. URL: http://dx.doi.org/10.1179/000870409X12525737905042, doi:10.1179/000870409X12525737905042. 2

[AAS*12] ANDRIENKO N., ANDRIENKO G., STANGE H., LIEBIG T., HECKER D.: Visual analytics for understanding spatial situations from episodic movement data. *KI - Künstliche Intelligenz 26*, 3 (2012), 241–251. URL: http://dx.doi.org/10.1007/s13218-012-0177-4, doi:10.1007/s13218-012-0177-4. 2

[BMJK09] BAK P., MANSMANN F., JANETZKO H., KEIM D.: Spatiotemporal analysis of sensor logs using growth ring maps. *Visualization and Computer Graphics, IEEE Transactions on 15*, 6 (Nov 2009), 913–920. doi:10.1109/TVCG.2009.182. 2

[CPMB11] CHOW W. T., POPE R. L., MARTIN C. A., BRAZEL A. J.: Observing and modeling the nocturnal park cool island of an arid city: horizontal and vertical impacts. *Theoretical and Applied Climatology 103*, 1-2 (2011), 197–211. URL: http://dx.doi.org/10.1007/s00704-010-0293-8, doi:10.1007/s00704-010-0293-8. 1

[Env12] ENVIRONMENTAL REMOTE SENSING AND GEOINFORMATICS LAB AND CAP LTER: 4 band NAIP land classification of Central Arizona, 2012. Arizona State University. 4

[HB03] HARROWER M., BREWER C.: ColorBrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal 40*, 1 (Jun 2003), 27–37. doi:10.1179/000870403235002042. 3

[HMRH15] HÄB K., MIDDEL A., RUDDELL B. L., HAGEN H.: Travis – a visualization framework for mobile transect data sets in an urban microclimate context. In *Pacific Visualization Symposium (PacificVis), 2015 IEEE* (April 2015). 2, 3

[HSvH*14] HEUSINKVELD B. G., STEENEVELD G. J., VAN HOVE L. W. A., JACOBS C. M. J., HOLTSLAG A. A. M.: Spatial variability of the Rotterdam urban heat island as influenced by urban land use. *Journal of Geophysical Research: Atmospheres 119*, 2 (2014), 677–692. URL: http://dx.doi.org/10.1002/2012JD019399, doi:10.1002/2012JD019399. 1

[Koh90] KOHONEN T.: The self-organizing map. *Proceedings of the IEEE 78*, 9 (Sep 1990), 1464–1480. doi:10.1109/5.58325. 2

[Mes15] MESOWEST / UNIVERSITY OF UTAH: MesoWest data. Online: http://mesowest.utah.edu/, 2015. Last accessed: 02/10/2015. URL: http://mesowest.utah.edu/. 3

[SBC*09] SUN C.-Y., BRAZEL A. J., CHOW W. T., HEDQUIST B. C., PRASHAD L.: Desert heat island study in winter by mobile transect and remote sensing techniques. *Theoretical and Applied Climatology 98*, 3-4 (2009), 323–335. URL: http://dx.doi.org/10.1007/s00704-009-0120-2, doi:10.1007/s00704-009-0120-2. 1

[Sch94] SCHMID H.: Source areas for scalars and scalar fluxes. *Boundary-Layer Meteorology 67*, 3 (1994), 293–318. URL: http://dx.doi.org/10.1007/BF00713146, doi:10.1007/BF00713146. 3

[SO12] STEWART I. D., OKE T. R.: Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society 93*, 12 (2012), 1879–1900. URL: http://dx.doi.org/10.1175/BAMS-D-11-00019.1, doi:doi:10.1175/BAMS-D-11-00019.1. 1

[SvdWvW14] SCHEEPENS R., VAN DE WETERING H., VAN WIJK J.: Non-overlapping aggregated multivariate glyphs for moving objects. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE* (March 2014), pp. 17–24. doi:10.1109/PacificVis.2014.13. 2

[SWvdW*11] SCHEEPENS R., WILLEMS N., VAN DE WETERING H., ANDRIENKO G., ANDRIENKO N., VAN WIJK J.: Composite density maps for multivariate trajectories. *Visualization and Computer Graphics, IEEE Transactions on 17*, 12 (Dec 2011), 2518–2527. doi:10.1109/TVCG.2011.181. 2

[SWVDWvW11] SCHEEPENS R., WILLEMS N., VAN DE WETERING H., VAN WIJK J.: Interactive visualization of multivariate trajectory data with density maps. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE* (March 2011), pp. 147–154. doi:10.1109/PACIFICVIS.2011.5742384. 2

[TSAA12] TOMINSKI C., SCHUMANN H., ANDRIENKO G., ANDRIENKO N.: Stacking-based visualization of trajectory attribute data. *Visualization and Computer Graphics, IEEE Transactions on 18*, 12 (Dec 2012), 2565–2574. doi:10.1109/TVCG.2012.265. 2

[VA00] VESANTO J., ALHONIEMI E.: Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on 11*, 3 (May 2000), 586–600. doi:10.1109/72.846731. 2, 3

[WBS13] WANG N., BIGGS T. W., SKUPIN A.: 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 (2013), 107 – 120. URL: http://www.sciencedirect.com/science/article/pii/S0198971512000919, doi:http://dx.doi.org/10.1016/j.compenvurbsys.2012.10.005. 2, 3