Workshop on Visualisation in Environmental Sciences (EnvirVis) (2016)
A. Middel, K. Rink, and D. Zeckzer (Editors)

*Short Paper*

# A Data-Driven Approach to Categorize Climatic Microenvironments

Kathrin Häb [†1], Ariane Middel[2], Benjamin L. Ruddell[3] and Hans Hagen[1]

[1]Computer Graphics and HCI Group, University of Kaiserslautern, Germany
[2]School of Geographical Sciences and Urban Planning, Arizona State University, Arizona, USA
[3]Fulton Schools of Engineering, Arizona State University, Arizona, USA

**Abstract**

*In urban climatology, identifying areas of similar microclimatic conditions helps to relate fine-scale urban morphology variations to their impact on atmospheric surroundings. Mobile transect measurements yield high-resolution microclimate data that allow for the delineation of these areas at a fine scale. However, the resulting spatio-temporal multivariate data is complicated and requires careful analysis and visualization to identify the emergent climatic microenvironments. Our previous work used a glyph-based visualization to comprehensively visualize spatially aggregated multivariate data from mobile measurements over diverse routes. This aggregation was conducted over a regular grid, and the utilized glyphs encoded multivariate relationships, average wind direction during data collection, number of transects traversing a grid cell, and grid cell size. In this paper, we reduce the visual complexity of the resulting map by coloring the background of the grid cells based on a comparison of the glyphs. The result is a gridded map that visually emphasizes spatial zones of similar multivariate relationships and that takes the information encoded by the glyphs into account. A preliminary evaluation shows that the described approach yields zones that line up with the physical structure of the study site.*

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

## 1. Introduction

Classifying a region into areas of similar climate characteristics has a long tradition in atmospheric sciences. On the global scale, the Köppen-Geiger classification structures the entire globe into zones of clearly distinguishable environmental conditions, based on vegetation groups, temperature regimes, and precipitation [KGB*06]. Similarly, Sayre et al. [SDF*14] delineate ecological land units on a global scale by aggregating bioclimate, landforms, lithology, and land cover data layers. At a smaller scale, the classification of urban areas into so-called climatopes is used to create urban climatic maps [RNK11, RSL*12]. Here, several information layers including urban morphology, land use, or atmospheric data, are combined to facilitate climate-sensitive and sustainable urban planning. Recently, the concept of local climate zones (LCZ) was developed by Stewart and Oke [SO12]. In this classification scheme, urban sites at the local scale are characterized using their morphological appearance, surface materials, and human activity – properties that are climate-relevant [SO12, BAB*15].

For the urban microscale, typically ranging between 1 *cm* and 1 *km* [EPW11], a generalizable climate classification was not developed so far due to the complex dynamics of microclimate. Mobile transect measurements can be used to retrieve fine-scale data to contribute to such classification efforts [HSvH*14, PLCC16]. To derive an appropriate and general set of distinctive properties for each class, a large number of observations is required.

The spatial aggregation of mobile measurements, or in more general terms, multivariate trajectories, is one step into this direction. Here, the links between trajectory attributes and the underlying spatial location is of predominant interest, so, following [AA10], each mobile measurement run can be treated as a sequence of independent measurement events [HMRH15, Häb15].

Spatial aggregation becomes easy if the mobile transect measurements are frequently repeated over the same set of routes, e.g., during larger campaigns using public transportation vehicles [HSW*14, BSK*11], or smaller campaigns using more flexible sensor transportation modes [EPP*12, LBCP15]. For a fixed route, a variable of interest can be aggregated over a set of waypoints [BSK*11, EPP*12], road segments [LBCP15], or a regular grid [HSW*14, AAS*12].

Not all mobile measurements are conducted over a fixed route. Data crowdsourcing is an example for vaying routes that depend on the individual movement of the participants. Examples for the latter cases can be found in [HSST12] and [WAK*10] for air quality measurements. Some mobile measurement campaigns use varying routes to research different climatic phenomena or to increase the

† E-mail: kathrin.haeb@cs.uni-kl.de



DOI: 10.2312/envirvis.20161105

spatial coverage of the collected data, so that a larger number of potential climatic microenvironments are represented in the samples. Both is the case for the sample data set used in this study.

In [HMRH15], we developed a glyph-based approach to visually delineate areas of similar urban microclimate based on mobile transect measurements traversing varying routes. Since the glyphs incorporate complex information and are thus also visually complex, interpreting them can be overwhelming. This study extends the work by [HMRH15], comparing the glyphs and coloring the background of the grid cells based on this comparison to facilitate the detection of potential climatic microenvironments [Häb15].

## 2. Mapping Approach

### 2.1. Spatial aggregation and glyph design

Similar to [AA08], [HMRH15] spatially aggregate the transect runs using a regular grid spanned over the bounding box of all transect routes. For each grid cell and individual transect run, measured attributes are averaged over the area covered by the grid cell and normalized according to their individual value range. A self-organizing map (SOM) [Koh90] is trained based on the aggregated and normalized values and partitioned using K-Means Clustering [VA00, WBS13]. In contrast to our previous work [HMRH15], we do not use the aggregated values of a single, "representative" transect run to train the SOM. Instead, a subset of the samples from all transect runs are used, while amount and location of these samples depend on the grid cell size. In a last step, the categories determined by the clustering procedure are applied to the entire set of aggregated and normalized values.

A glyph-based visualization shows the result of this aggregation approach; each traversed grid cell is assigned one glyph (Fig. 1). Following [HMRH15], the glyphs encode:

- the number of transects traversing that grid cell; each transect run is represented by one glyph segment
- the cluster membership of the run traversing that grid cell, encoded by the color of the corresponding glyph segment, whose meaning can be explored by means of a parallel coordinates plot [GCML06, VJN*15] (Fig. 1a)
- wind direction during data collection, given by the orientation of the corresponding glyph segment
- the grid cell size, which is proportional to a glyph's radius.

### 2.2. Glyph comparison and mapping technique

The research described in this paper is based on a mobile transect measurement data set with 21 runs. The data were recorded in a residential neighborhood in Gilbert, Arizona, USA and cover different times of day in three different seasons. Air temperature ($T_a$) and relative humidity ($RH$) were recorded in two different heights (1 $m$ and 2 $m$), along with surface temperature ($T_s$). The transects follow partially diverging routes [HMRH15].

To create a glyph map as described in Section 2.1 and as shown in Fig. 1, 2, and 3, we used the following setup:

- *Considered variables*: $T_s$; $RH_{1m}$; $T_{a,1m}$; $RH_{2m}$; $T_{a,2m}$
- *Grid cell size*: 30
- *SOM height and width*: 20

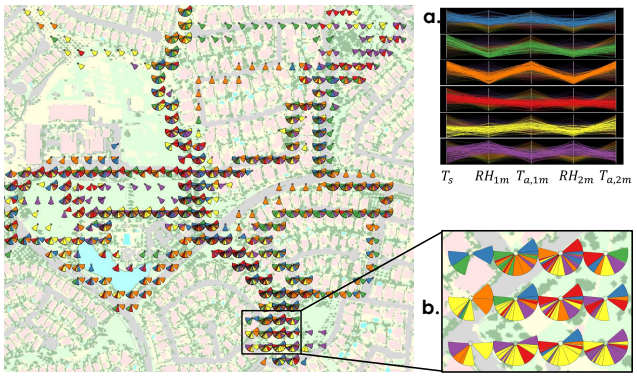



**Figure 1:** *Glyph-based visualization of climatic microenvironments in the Power Ranch community, Gilbert, AZ, USA (background map: [Env12]). (a) encodes the meaning of the colors used in the glyphs [GCML06, VJN*15] (colors based on [HB03]), while (b) shows a close-up of the glyphs.*

- *Number of SOM training iterations*: 10 x $N$ ($N$ = number of samples after aggregation)
- *Number of K-Means cluster centroids*: 6

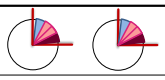
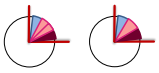
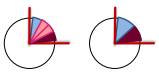
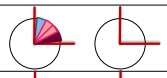
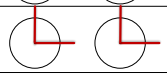
Fig. 1 reveals a first insight into potential microenvironments in a study area, and details about the multivariate value distribution can be explored by zooming into locations of interest. However, the display lacks a proper guidance on which glyphs share certain properties, and a structured comparison of glyphs that are located further apart from each other is difficult. Therefore, we color the traversed grid cells according to glyph similarity in addition to showing the glyphs at their location. This way, the distribution of similar grid cells is summarized, while the advantages of the previously described glyph-based visualization are still kept.

Grouping the glyphs according to similarity is complex because of the categorical nature of their appearance. The latter is given by the cluster memberships of the glyph's individual segments, each representing a certain category of multivariate relationships (e.g., a segment's cluster can represent "high surface temperatures and low humidities"). Another problem arises because the number of segments per glyph varies between grid cells. Finally, a segment's orientation and, hence, the distribution of cluster memberships around a 360° circle adds to the complexity.

Therefore, comparing the glyphs to each other requires a way to characterize their properties. In our solution, this description relies on the segmentation of the glyphs into eight sectors with a size of 45°. Per sector, the number of glyph segments is stored with each of their cluster memberships.

All glyphs in the display are categorized based on this information. Glyphs can only fall into the same category if they, for each sector, share the same amount of glyph segments, and the same distribution of cluster memberships inherent in these segments. Each glyph is assigned the resulting category, defined by the description explained above.

The actual comparison is conducted between these categories. To determine the amount of similarity, a "distance" between the glyph categories is defined. Using the cases shown in Table 1, the glyphs are compared, while corresponding similarity "points" are

| Case | Description | Graphical summary | Points (per sector) | Points (entire glyph) |
|---|---|---|---|---|
| 1 | - contain the same amount of segments<br>- distribution of cluster memberships is equal |  | 15 | w*15 |
| 2 | - do not contain the same amount of segments<br>- the same set of different cluster memberships is available in both glyphs. |  | 12 | w*12 |
| 3 | - do not contain the same amount of segments<br>- the set of cluster memberships glyph 1 is a subset of the cluster memberships in glyph 2. |  | 10 | w*10 |
| 4 | - may or may not contain the same amount of segments<br>- the set of cluster memberships in glyph 1 is different from that in glyph 2, but there is a percentage of agreement |  | $10p_a$ ** | w*$10p_a$ |
| 5 | - may or may not contain the same amount of segments<br>- there is no agreement between glyph 1 and glyph 2 in terms of the cluster memberships |  | 0.01 | 0.01 |
| 6 | One sector contains segments, while the other does not. |  | 0.01 | -- |
| 7 | Both sectors do not contain any segments. |  | 0.01 | -- |

** $p_a = \frac{2*\#segments\_agr}{\#segments_1+\#segments_2}$ w* = weight factor

$\#segments\_agr$ = number of segments with agreeing cluster membership (1 in the example given in the graphical summary of case 4)

**Table 1:** *Metric applied for glyph comparison [Häb15].*

summated. The distance between two glyphs, $\Delta_{glyph}$, is then given by

$$\Delta_{glyph} = \frac{1}{\sum Points} \quad (1)$$

with $\Delta_{glyph} = 0$ for glyphs of the same category.

The comparison is conducted in two steps. First, entire glyphs are compared to each other, independent from the described subdivision of glyphs into sectors. The points received for the cases depicted in Table 1 are multiplied by a weight factor $w$ that determines the importance of overall similarity ($w = 10$ in our solution). Then, compared glyphs receive additional points for agreement in corresponding sectors, which are added to the overall similarity score.

Both steps are necessary; comparing only sector-wise would cause large distances between glyphs with the same overall distribution of cluster memberships, but without agreement in corresponding sectors. On the other side, neglecting wind directions would render glyphs with a diverging circular setup similar. Adding points for corresponding segments in a sector decreases the distance between glyphs.

The grid cells are colored based on the pairwise distance between glyph categories. Sammon's mapping [Sam69] of the distances into the HSL color space is used to determine the color, i.e. grid cells containing similar glyphs are assigned similar colors [VJN*15, AAS*12, SR13].

Overall, the results look promising, as can be observed in Fig. 2. For example, grid cells containing glyphs with combinations of purple, yellow and red segments consistently appear in a light to mid purple color, while grid cells containing glyphs with predominantly orange, green and blue segments are consistently colored in dark green. If the latter colors are mixed with purple or red, the background color becomes dark purple. Grid cells containing numerous segments with a large variety of different clusters, especially those combining blue, red, and purple appear in blue (Fig. 2c). Fig. 2a shows that different combinations of cluster memberships lead to different colors, and similar glyphs receive similar background colors. Thus, the metric described here can be useful to highlight patterns of similar multivariate relationships, i.e. potential climatic microenvironments.

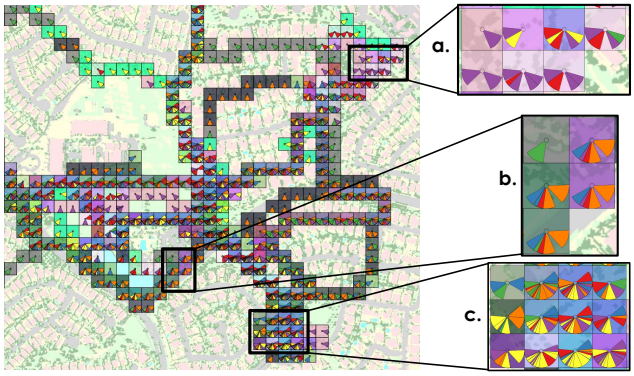



**Figure 2:** *Coloring the background of each grid cell according glyph similarity. In (a), the described procedure works well. In (b), background colors of similar glyphs are too distinct. In (c), a large variety of different clusters represented in one glyph is shown (background map: [Env12]; Fig. 1a shows the meaning of glyph segment colors)*

Fig. 2b reveals cases where the background color is more different as expected. While all glyphs in this detail show orange segments pointing east, different configurations of cluster memberships in the segments pointing west cause the background colors to diverge significantly. Similarly, grid cells with only yellow glyphs appear in turquoise, diverging from cells with glyphs combining, e.g., yellow and purple.

## 3. Results and Discussion

Fig. 3 illustrates the results of the approach, based on the data set and initialization described in Section 2.2. A qualitative analysis suggests that the physical structure of the study site (Fig. 3a) is represented by the colored grid cells (Fig. 3b). Parks stand out, showing a distinct pattern of turquoise and purple grid cells ("park cells"). However, different parks show different signals, owing to different humidity and temperature regimes. Thus, in the park on the left side of the study area only few turquoise and purple park cells appear. Using the corresponding glyphs for further analysis reveals that purple glyph segments are underrepresented if compared to, e.g., the park in the upper right of the map. Grid cells dominated by buildings and asphalt surfaces are colored dark green ("impervious cells"), showing high surface temperatures and low relative humidity (Fig. 1a).

Although the glyphs reveal the number of samples per grid cell, their background colors alone currently do not explicitly repre-


**Figure 3:** *Coloring a map according to glyph similarity to highlight zones of similar microclimate. (a) shows the physical structure of the study site (Google Earth, 2015), while (b) shows the colored grid cells and the corresponding glyph distribution (background map: [Env12], see Fig. 1a for the meaning of a glyph segment's color).*

sent the implied notion of uncertainty. The number of segments per glyph influences its position in the resulting distance field and therefore also its background color, but same visual cue is used to represent any other distance as well. This can lead to an ambiguous perception, if the glyphs are not shown. Therefore, for future work, we plan to include an appropriate visual cue to the visualization, while also quantitatively analyzing the described concept of glyph similarity. Furthermore, we plan to add the temporal dimension of visualized "zones" into the display.

## 4. Conclusion

We present an approach for the spatial aggregation and visualization of mobile transect measurements conducted over diverse routes. A glyph-based visualization summarizes the aggregation results for each grid cell in a regular grid that was spanned over the area covered by the entire set of routes, as described in our previous work [HMRH15].

To visually structure the display, glyph similarity is highlighted by coloring the background of the grid cells according to a metric that takes the glyph structure into account. This further facilitates the detection of climatic microenvironments, and represents a first step towards delineating climate "zones" at the urban microscale that are purely based on measured atmospheric variables.

## 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 DFG (IRTG 2057), the NSF Grant SES-0951366, 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. doi:10.1179/000870409X12525737905042. 1

[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. doi:10.1007/s13218-012-0177-4. 1, 3

[BAB*15] BECHTEL B., ALEXANDER P. J., BÖHNER J., CHING J., CONRAD O., FEDDEMA J., MILLS G., SEE L., STEWART I.: Mapping local climate zones for a worldwide database of the form and function of cities. *ISPRS International Journal of Geo-Information 4*, 1 (2015), 199–219. doi:10.3390/ijgi4010199. 1

[BSK*11] BUTTSTÄDT M., SACHSEN T., KETZLER G., MERBITZ H., SCHNEIDER C.: A new approach for highly resolved air temperature measurements in urban areas. *Atmospheric Measurement Techniques Discussions 4*, 1 (2011), 1001–1019. doi:10.5194/amtd-4-1001-2011. 1

[Env12] ENVIRONMENTAL REMOTE SENSING AND GEOINFORMATICS LAB AND CAP LTER: 4 band NAIP land classification of Central Arizona, 2012. Arizona State University. 2, 3, 4

[EPP*12] ELEN B., PETERS J., POPPEL M. V., BLEUX N., THEUNIS J., REGGENTE M., STANDAERT A.: The Aeroflex: A bicycle for mobile air quality measurements. *Sensors 13*, 1 (2012), 221–240. doi:10.3390/s130100221. 1

[EPW11] ERRELL E., PEARLMUTTER D., WILLIAMSON T.: *Urban Microclimate: Designing the spaces between buildings*. Earth Scan, London, Washington, 2011. 1

[GCML06] GUO D., CHEN J., MACEACHREN A. M., LIAO K.: A visualization system for space-time and multivariate patterns (VIS-STAMP).

*IEEE Transactions on Visualization and Computer Graphics 12*, 6 (Nov. 2006), 1461–1474. doi:10.1109/TVCG.2006.84. 2

[Häb15] HÄB K.: *Visualization and Analysis Techniques for Urban Microclimate Data Sets*. PhD thesis, University of Kaiserslautern, 2015. 1, 2, 3

[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. 2

[HMRH15] HÄB K., MIDDEL A., RUDDELL B. L., HAGEN H.: Spatial Aggregation of Mobile Transect Measurements for the Identification of Climatic Microenvironments. In *Workshop on Visualisation in Environmental Sciences (EnvirVis)* (2015), Middel A., Rink K., Weber G. H., (Eds.), The Eurographics Association, pp. 19–23. doi:10.2312/envirvis.20151086. 1, 2, 4

[HSST12] HASENFRATZ D., SAUKH O., STURZENEGGER S., THIELE L.: Participatory air pollution monitoring using smartphones. In *Mobile Sensing: From Smartphones and Wearables to Big Data* (Beijing, China, Apr 2012), ACM. 1

[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. doi:10.1002/2012JD019399. 1

[HSW*14] HASENFRATZ D., SAUKH O., WALSER C., HUEGLIN C., FIERZ M., THIELE L.: Pushing the spatio-temporal resolution limit of urban air pollution maps. In *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on* (March 2014), pp. 69–77. doi:10.1109/PerCom.2014.6813946. 1

[KGB*06] KOTTEK M., GRIESER J., BECK C., RUDOLF B., RUBEL F.: World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift 15*, 3 (2006), 259–263. doi:doi:10.1127/0941-2948/2006/0130. 1

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

[LBCP15] LECONTE F., BOUYER J., CLAVERIE R., PÉTRISSANS M.: Using local climate zone scheme for UHI assessment: Evaluation of the method using mobile measurements. *Building and Environment 83*, 0 (2015), 39 – 49. Special Issue: Climate adaptation in cities. doi:http://dx.doi.org/10.1016/j.buildenv.2014.05.005. 1

[PLCC16] PARECE T. E., LI J., CAMPBELL J. B., CARROLL D.: Assessing urban landscape variables' contributions to microclimates. *Advances in Meteorology 2016* (2016), 1–14. doi:10.1155/2016/8736263. 1

[RNK11] REN C., NG E. Y.-Y., KATZSCHNER L.: Urban climatic map studies: a review. *International Journal of Climatology 31*, 15 (2011), 2213–2233. doi:10.1002/joc.2237. 1

[RSL*12] REN C., SPIT T., LENZHOLZER S., YIM H. L. S., HEUSINKVELD B., VAN HOVE B., CHEN L., KUPSKI S., BURGHARDT R., KATZSCHNER L.: Urban climate map system for dutch spatial planning. *International Journal of Applied Earth Observation and Geoinformation 18* (2012), 207 – 221. doi:http://dx.doi.org/10.1016/j.jag.2012.01.026. 1

[Sam69] SAMMON J.: A nonlinear mapping for data structure analysis. *Computers, IEEE Transactions on C-18*, 5 (May 1969), 401–409. doi:10.1109/T-C.1969.222678. 3

[SDF*14] SAYRE R., DANGERMOND J., FRYE C., VAUGHAN R., ANIELLO P., BREYER S., CRIBBS D., HOPKINS D., NAUMAN R., DERRENBACHER W., WRIGHT D., BROWN C., CONVIS C., SMITH J., BENSON L., VANSISTINE D. P., WARNER H., CRESS J., DANIELSON J., HAMANN S., CECERE T., REDDY A., BURTON D., GROSSE A., TRUE D., METZGER M., HARTMANN J., MOOSDORF N., DÃIJRR H., PAGANINI M., DEFOURNY P., ARINO O., MAYNARD S., ANDERSON M., COMER P.: *A new map of global ecological land units – An ecophysiographic stratification approach*. Association of American Geographers, Wahington, DC, 2014. 1

[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. doi:10.1175/BAMS-D-11-00019.1. 1

[SR13] SARLIN P., RONNQVIST S.: Cluster coloring of the self-organizing map: An information visualization perspective. In *Information Visualisation (IV), 2013 17th International Conference* (July 2013), pp. 532–538. doi:10.1109/IV.2013.72. 3

[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

[VJN*15] VROTSOU K., JANETZKO H., NAVARRA C., FUCHS G., SPRETKE D., MANSMANN F., ANDRIENKO N., ANDRIENKO G.: SimpliFly: A methodology for simplification and thematic enhancement of trajectories. *Visualization and Computer Graphics, IEEE Transactions on 21*, 1 (Jan 2015), 107–121. doi:10.1109/TVCG.2014.2337333. 2, 3

[WAK*10] WILLETT W., AOKI P., KUMAR N., SUBRAMANIAN S., WOODRUFF A.: Common sense community: Scaffolding mobile sensing and analysis for novice users. In *Pervasive Computing*, Floréen P., Krüger A., Spasojevic M., (Eds.), vol. 6030 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010, pp. 301–318. doi:10.1007/978-3-642-12654-3_18. 1

[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. doi:http://dx.doi.org/10.1016/j.compenvurbsys.2012.10.005. 2