EnviroInfo 2010 (Cologne/Bonn)
Integration of Environmental Information in Europe

An urban health risk analysis for Berlin: exploration and integration of spatio-temporal information on the urban environment

Tobia Lakes[1], Ulf Leser[2], Cornelius Senf[1]
[1]Humboldt Universität zu Berlin, Geomatics Lab, Unter den Linden 6, 10099 Berlin
[2]Humboldt Universität zu Berlin, Informatics Department, Unter den Linden 6, 10099 Berlin
tobia.lakes@geo.hu-berlin.de
leser@informatik.hu-berlin.de
cornelius.senf@geo.hu-berlin.de

Abstract

Urban areas provide the living space for the majority of the world population. Well-being and health within urban areas is influenced by environmental and socioeconomic variables. To analyze the potential health risk in cities one has to integrate multi-dimensional and multi-scale information describing socioeconomic, environmental and health-related attributes. The aim of this paper is to study the urban health risk for different parts of the city of Berlin using a data-driven analysis approach. We focus on the detection and exploration of correlations between environmental, socioeconomic and health-related attributes in Berlin. We showcase the further study of selected correlations, including the biophysical and socioeconomic burden of certain diseases, and their visual exploration using advanced geovisualization.

1. Introduction

Urban health risk can only be analyzed by integrating socio-economic and environmental data, because the impact of urban livelihoods on health and well-being is strongly influenced by the characteristics of urban systems, such as densely built structures with little amount of green and open space, high population density, traffic, urban heat islands, air pollution etc. (Galea et al. 2005; Alberti, 2005; Frumkin 2002). In a number of studies negative impacts of bioclimatic stress, noise pollution, and a general decreased environmental living quality on the health status of an urban population have been identified (Patz et al. 2005). On the other hand, urban livelihoods benefit from the functions the urban ecosystem provides, such as regulation of temperature and air pollution by vegetation (Whitford et al. 2001). Moreover, the positive psychological impact of “nature in the city” in the living area of urban residents has been shown (Lafortezza et al. 2009). However, the analysis of the biophysical influences on urban livelihoods only reflects one part of the picture of the urban health risk. In addition, strong positive correlations between socio-economic factors and health status have been proven (Amarsinghe et al 2009).

Exploring local characteristics of urban livelihoods in their different health-relevant dimensions may be a promising approach to approximate the potential health risk. This is of particular importance since health data usually is available only on a coarse-grain scale, and not on the detailed level of individuals, households, or neighborhoods. Furthermore, up to now, only little is known about the spatial distribution of the urban health risk driven by environmental pollution and resources and the correlation of these factors with the socio-economic structure. Exploring and integrating all available data is particularly interesting following the idea of environmental justice studies (e.g. Kruize 2009) – is the environmental burden of disease equally distributed over the social status? This calls for a comprehensive health risk analysis, impos-

ing challenges on the integration of highly heterogeneous and often distributed data sets. Methods which allow the integration of environmental data in a spatio-temporal context by overcoming issues such as heterogeneity in data types, differences in spatial and temporal resolution, inaccuracies, multicolinearity etc. need to be developed and applied. Furthermore, qualitative, semi-quantitative, and quantitative information have to be analyzed in a joint approach.
In this paper, we study the urban health risk for different parts of the city of Berlin, including the biophysical and socioeconomic burden of diseases. We use data-driven analysis methods for the analysis of a multi-scale and multi-dimensional, large data set. We focus on (1) the exploration of correlations between environmental, socioeconomic and health attributes in Berlin, (2) the identification and further study of selected relationships and important attributes for further analysis, and (3) the geovisualization of multi-dimensional and multi-scale information of health risk .

2. Methods and Material

2.1 Study Area

We study the spatial patterns of health risk combining biophysical and socioeconomic vulnerability for the city of Berlin, the capital of Germany. Berlin has a population of about 3,400,000 inhabitants/800 km² and is divided in 12 districts of about 300,000 residents each (see Figure 1).

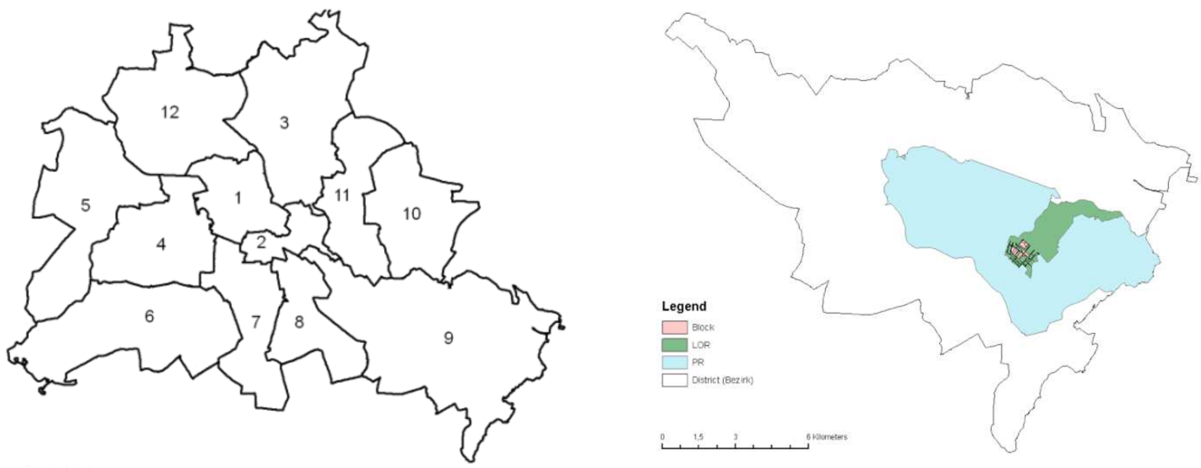

Figure 1: Study area of Berlin, Germany, and scale of analysis. Left: Berlin districts (1: Mitte, 2: Friedrichshain-Kreuzberg, 3: Pankow, 4: Charlottenburg-Wilmersdorf, 5: Spandau, 6: Steglitz-Zehlendorf, 7: Tempelhof-Schöneberg, 8: Neukölln, 9: Treptow-Köpenick, 10: Marzahn-Hellersdorf, 11: Lichtenberg, 12: Reinickendorf). Right: Multi-scale analysis: from district-level to block-level, for the example of Treptow-Köpenick

The city of Berlin is one of the greenest cities in Europe, although, at the same time, it contains plenty high-densely built areas. Major changes of urban land use have taken place in Berlin recently, following the reunification in 1989. In the inner city core, new and large-scale developments such as Potsdamer Platz, can be observed while at the same time a high number of brownfields remain. Thus, the city now shows a heterogeneous mixture of residential areas including block-built housing structures, social housing estates and prefabricated constructions, single-family housing, and new housing developments, such as townhouses.

Biophysical and in particular vegetation characteristics of the urban living environment strongly depend on the dominant urban land use and building types. These show a very heterogeneous and fragmented picture in Berlin. Parts of the inner core of Berlin consist, for example, of block-built structures which are characterized by high sealing degrees, only few green areas and high population numbers. High-rise precast concrete slabs in Marzahn-Hellersdorf, on the other hand, typically contain a large amount of green in the immediate neighborhood.

At the same time, the urban fabric of Berlin determines the development and proliferation of noise, namely the noise-creating traffic patterns and the urban soundscape in which noise is proliferated. Apart from the noise pollution it is particularly the climate and air quality that has a health-relevant impact. To improve air quality and noise pollution (with the ultimate goal to become an ecological and healthy city), an environmental zone has been introduced to Berlin in 2008 which restricts access of motor vehicles to parts of the inner city depending on the technical standard of the vehicle.

In Berlin, as in other European cities, social segregation appears as a small-scale heterogeneous pattern throughout the city. Pivotal transformations of the social structure since the 1990s derive from processes such as suburbanization and socially selective migrations as well as from gentrification (Häußermann and Kapphan 2000). The urban development index 2007, developed by the Berlin Senate Department, identifies the most deprived living environments in former west-Berlin inner-city areas, such as parts of Mitte and Neukölln, and in high-rise areas in the eastern and western outskirts of Berlin, such as Marzahn, Reinickendorf and Spandau. High social status groups, on the other side, concentrate in gentrified neighborhoods of the inner-city, such as parts of Pankow, as well as traditionally high-income neighborhoods of the outskirts, such as Steglitz-Zehlendorf or Köpenick.

2.2 Data

The available environmental, socio-economic and health data are heterogeneous in terms of their spatial resolution, up-to-dateness, and units of measurement (see Table 1). Data shows spatially inhomogeneous reference areas including grids and vector data on the levels of blocks, living environments ("Lebensweltlich orientierte Räume", LOR), planning environments ("Planungsräume", PR), and districts. Data was acquired from different sources.

Table 1: Data used for our analysis. Explanation: *Standardized Mortality Rate. 1 Derived from data on unemployment, received social benefits and number of foreign people. 2 Derived from social status and mobility data. 3 The Predicted Mean Vote is used as a bioclimatic indicator. 4 Sealing is assessed by using remote sensing classification. 5 Land use is defined by land use classes. 6 Land type is a pre-category of land use classes. 7 Building structure type is a specific definition of land use (different use- and builditypes). 8 Vegetation assessed from remote sensing data in proportion to total area of a LOR.

Nr.	Data	Entity	Year	Scale	Unit	Source
1	Social state[1]	LOR	2007	Ordinal	Classes	[1]
2	Social dynamics[2]	LOR	2007	Ordinal	Classes	[1]
3	Criminality (reported crimes)	District	2009	Ratio	Crimes per	[3]

					inhabitant	
4	Foreign people	District	2004	Ratio	%	[4]
5	Adolescents	Block	2008	Ratio	% (< 18 years)	[1]
6	Elderly	Block	2008	Ratio	% (> 65 years)	[1]
7	Population density	Block	2008	Ratio	People per km^2	[1]
8	Bioclimate[3]	Block	2009	Ordinal	Classified (z-scored PMV)	[1]
9	Sealing[4]	Block	2008	Ratio	%	[1]
10	Green area provision	Block	2009	Ordinal	Classified (m^2 per inhabitant)	[1]
11	Vegetation area factor[8]	LOR	2006	Ratio	% of area	Lab
12	Parks	LOR	2008	Ratio	% of area	[1]
13	Playgrounds	LOR	2009	Ratio	% of area	[1]
14	Land use[5]	Block	2008	Nominal	Text	[1]
15	Land type[6]	Block	2008	Nominal	Text	[1]
16	Building structure type[7]	Block	2008	Nominal	Text	[1]
17	Mortality	District	2007	Ratio	x/ 1000 inhab.	[4]
18	Traffic accidents	District	2007	Ratio	x/ 1000 inhab.	[1]
19	Traffic Accidents (with injured people)	District	2007	Ratio	x/ 1000 inhab.	[1]
20	Perceived childhood tests	PR	2007	Ratio	%	[2]
21	Child adiposity (at age of 6)	PR	2007	Ratio	%	[2]
22 m/w	Infection and parasitic disease	District	2000	Ratio	SMR*	[5]
23 m/w	Neoformations	District	2000	Ratio	SMR	[5]
24 m/w	Endocrine, nutritional- and metabolic dis.	District	2000	Ratio	SMR	[5]
25 m/w	Mental and behavior disorder	District	2000	Ratio	SMR	[5]
26 m/w	Dis. of nervous system	District	2000	Ratio	SMR	[5]
27 m/w	Dis. of circulatory system	District	2000	Ratio	SMR	[5]
28 m/w	Dis. of respiratory system	District	2000	Ratio	SMR	[5]
29 m/w	Dis. of digestive system	District	2000	Ratio	SMR	[5]
30 m/w	Dis. of cutaneous and subcutaneous	District	2000	Ratio	SMR	[5]
31 m/w	Dis. of geitourinary	District	2000	Ratio	SMR	[5]
32 m/w	Not classified findings of morbidity	District	2000	Ratio	SMR	[5]
33 m/w	Injuries and poisoning	District	2000	Ratio	SMR	[5]
34 m/w	External causes for morbidity	District	2000	Ratio	SMR	[5]

2.3 Method

To analyze potential health risk in Berlin, we acquired and pre-processed data in raster and vector format from different sources. We analyzed all data sets manually to derive defined reference units, precise semantics of fields, and metadata on the accuracy of the data. We stored all records in a tabular format and linked it to its corresponding vector file using surrogate keys. The vector files allow us to geocode and visualize the data in a geoinformation system (GIS). To perform spatial data mining across different scales (see below), we then intersected all data sets into a single table using topologic relations created in the GIS, in particular spatial containment relationships. For the analysis, we used two freely available data mining tools, namely RapidMiner [6] and Weka (Witten et. al, 1999), and the freely available statistical software R. First insights on thematic patterns and existing correlations, positive and negative ones, were achieved by calculating a correlation matrix. We then applied Naive Bayesian Classification on selected correlated phenomena (Hoffmann and Mahidadia, 2010). Finally we visualized derived patterns and I sights using ArcGIS (ESRI 2010).

3. Results and Discussion

We integrated the environmental, socio-economic and health data in one database and computed pair-wise attribute correlations to explore patterns and correlations in the dataset (aim 1). The correlation matrix shows strong correlations between a variety of different attributes.

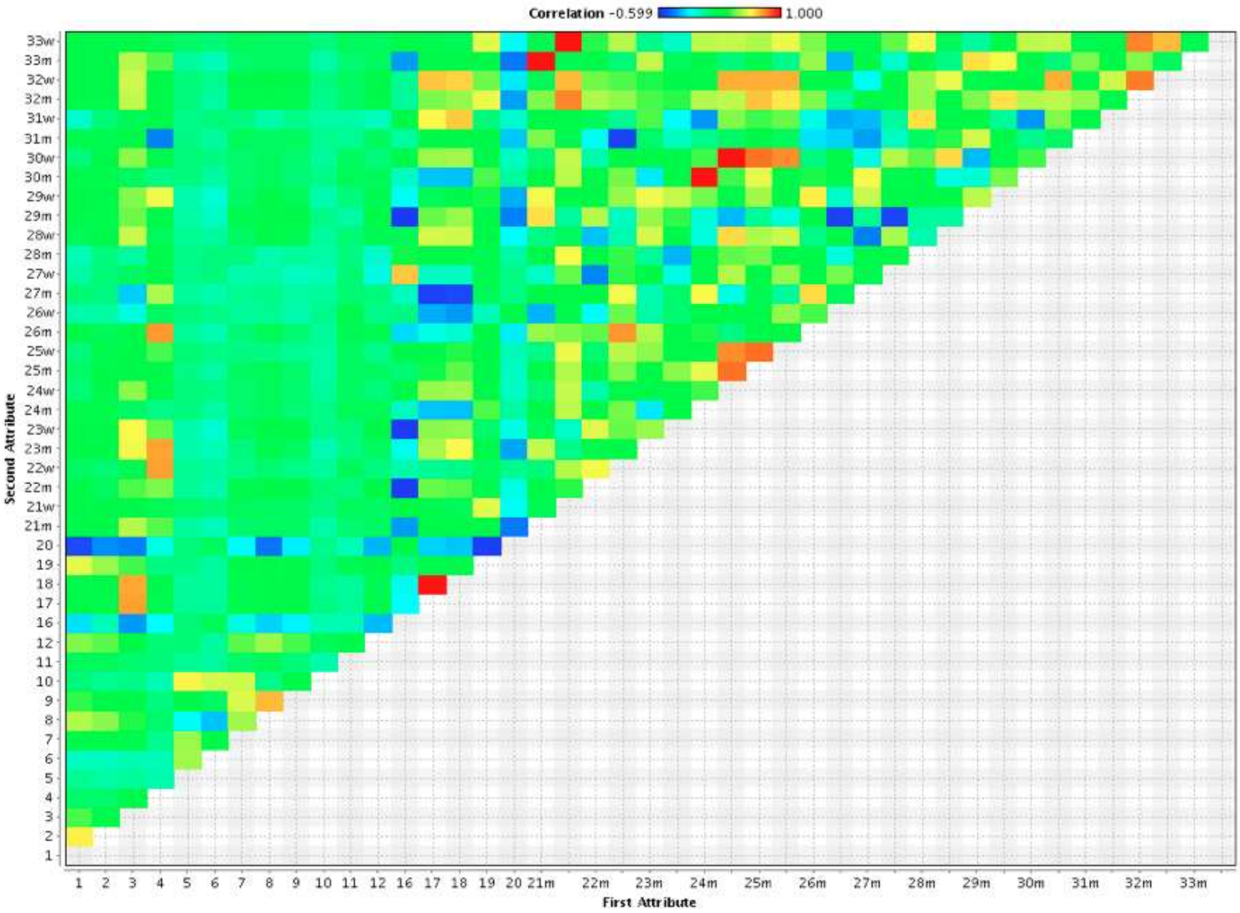

Figure 2: Correlation matrix between environmental, socio-economic and health data. See Table 1 for attribute information.

In particular, we are able to identify one cluster showing strong positive correlations between degree of sealing, bioclimatic conditions, and population density as indicated by the circle in figure 2. This underlines that the urban environment heavily impacts the urban climate which in turn exceeds a strong bioclimatic health impact. Furthermore, there are a number of moderate correlations between social status and environmental data on the one hand and health-related data on the other hand, pointing towards the importance of this social indicator in relationship to environmental pollution and health (indicated by the arrow in figure 2, Kruize 2007). Health datasets in general show strong correlations to other health-related data.
In the second step we identified and focused explicitly on selected relationships and important attributes for further analysis (aim 2). We show the investigated relationships between the different dimensions of datasets, namely socio-economic, environmental, and health-related in figures 3 and 4.

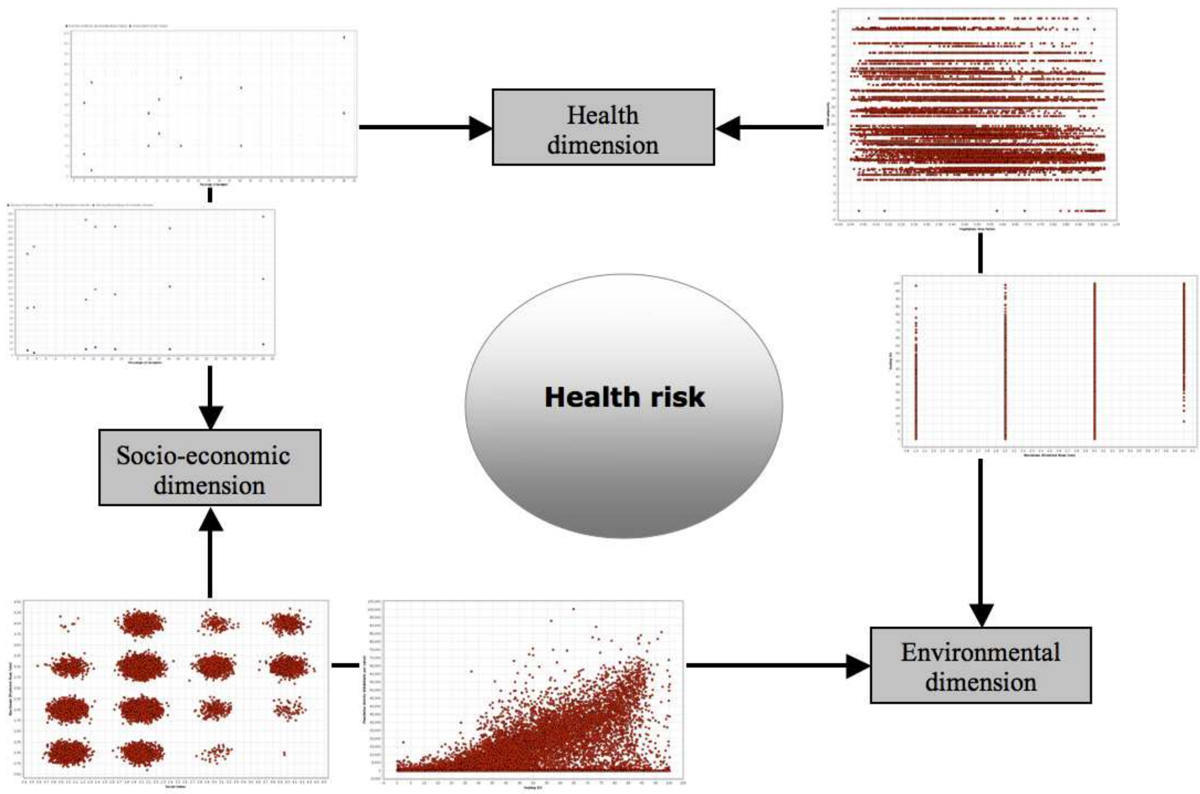

Figure 3: Integrated dimensions of health risk (clockwise x/y axis): vegetation area factor and child adiposity, sealing and bioclimate indicator, sealing and population density, social status and bioclimatic indicator, percentage of foreigners and mortality rates for diseases by gender.

Furthermore, we receive new insights from visualizing Naive Bayesian classifications for the target variable social status and calculate different density distributions of health and environmental dimensions.

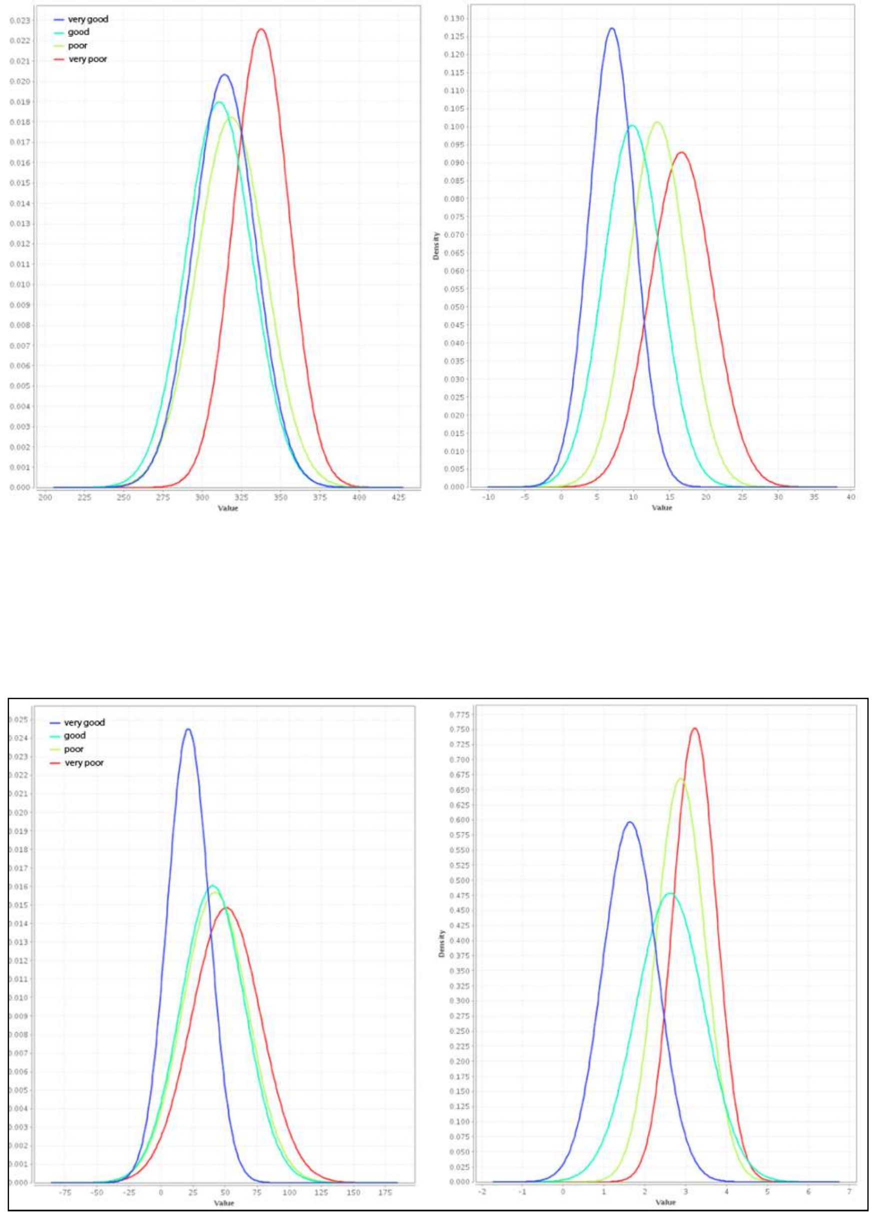

Figure 4: Social status in relation to: diseases of circulatory system (male) and child adiposity (upper figures); sealed surface and bioclimatic condition (lower figures)

For the very good, good and poor social status classes, circulatory system diseases show a similar density distribution. However, the status “very poor” is associated to a clearly higher average number of circulatory system diseases per 1000 inhabitants (~335). A similar observation holds for child adiposity, which has a lower average for the very good social status class while the three other social status classes show higher rates. These results underline the unequal burden of disease across social status which has been identified in other studies as well. In addition to these known phenomena, we are also able to identify additional environmental burden of disease in social status classes. While direct causal inferences are of course impossible, it is the double burden of disease, meaning the socioeconomic and environmental burden, which we are able to show with the correlation results between social status and environmental burden of disease. For the example of childhood adiposity this means that we find very high rates in the very poor social status class which comes along with a very high degree of sealing and low amounts of green space in those areas linked to a poor social status. On the other hand, sealing is minimal in the class of very good social status (~20 %). We can clearly identify a strong relationship of the environmental burden of diseases in different social status classes: very poor social classes are correlated with the worst bioclimatic conditions and, in contrast to this, a lower mean but higher variance is depicted for very good situated population. The latter might be due to the prevalent urban development processes of suburbanization and gentrification of higher income classes (see also Frumkin 2002). Very poor and poor social status classes are almost completely situated in unfavorable urban bioclimatic conditions such as in densely built structure with high sealing rates, low vegetation availability and with reduced air cooling.

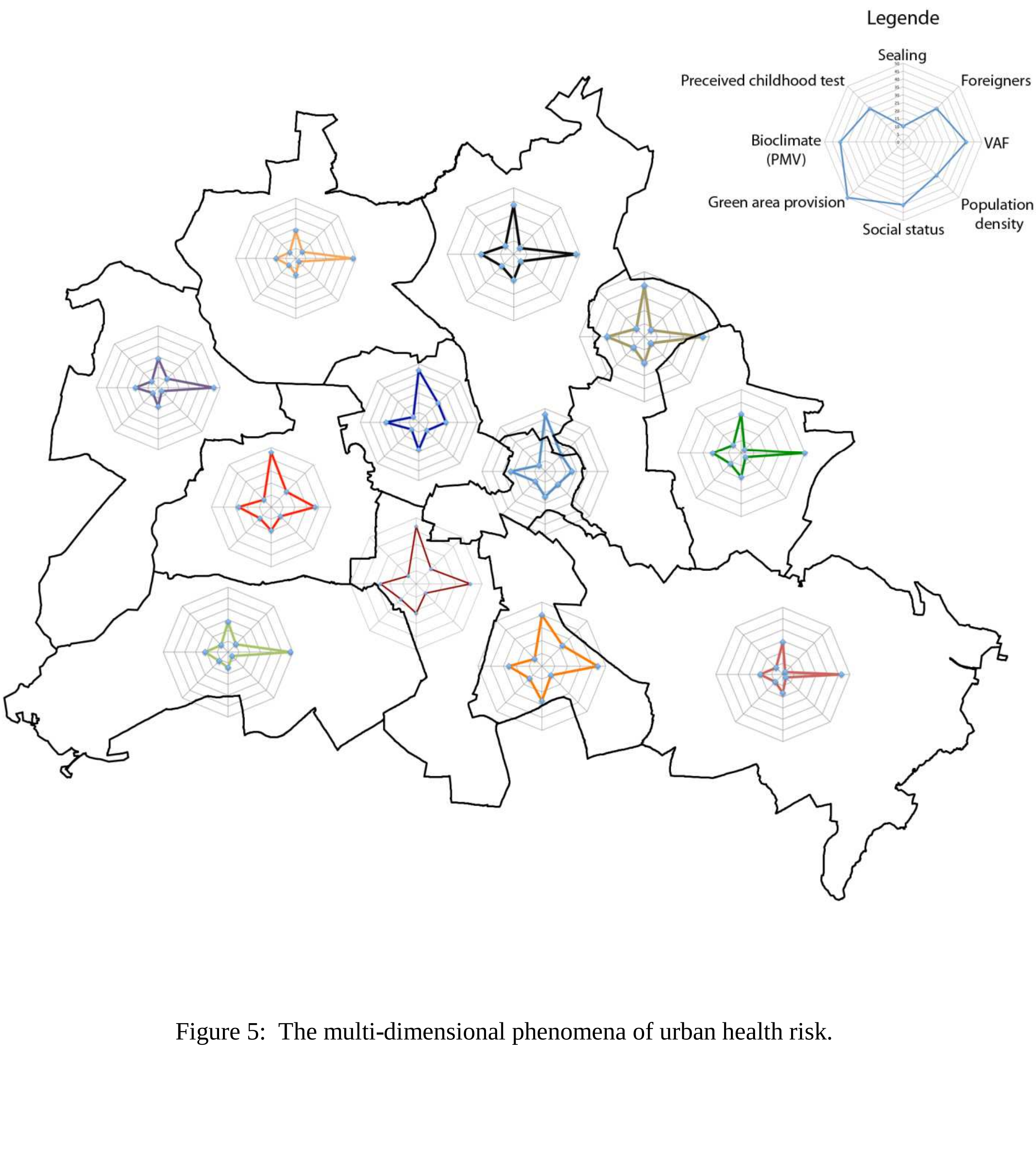

Figure 5: The multi-dimensional phenomena of urban health risk.

Visualisation of the results can help to show multi-dimensional phenomena of urban health risk for the whole city of Berlin (aim 3). Such a map allows comparisons between different districts and may provide new insights for decision-making, for example when addressing the increasing problem of childhood obesity; a health risk where the strong correlations to socioeconomic and environmental determinants have been described in other studies as well (Amarasinghe et al. 2009). We show that by integrating different vector and raster-derived data into a multi-scale analysis, new insights into aspects of environmental justice are possible, such as the higher risk of the environmental and social burden of disease (Kruize 2007). Particularly the freely available environmental information for risk analysis needs to be stressed which may be a good indicator for the potential health risk, when considering that in empirical studies precise and up-to-date health data is often lacking.

4. Conclusion and outlook

In this first stage of analysis we were able to show the multi-dimensionality of health risk in urban areas using data-driven techniques. On the basis of these insights further research is needed to exploit the benefits of spatial information by explicitly integrating topology and exploring multi-scale effects of environment and socioeconomic data on health. Above all, future work will also focus on a more detailed analysis in terms of spatial resolution, such as urban neighborhoods, to provide the necessary information for local urban health risk strategies.

5. Literature

Alberti, M., J. M. Marzluff, et al. (2003). "Integrating Humans into Ecology: Opportunities and Challenges for Studying Urban Ecosystems." BioScience 53: 1169-1179.

Amarasinghe, A., G. D'Souza, et al. (2009). "The Influence of Socioeconomic and Environmental Determinants on Health and Obesity: A West Virginia Case Study." International Journal of Environmental Research and Public Health 6: 2271-2287.

Frumkin, H. (2002). "Urban sprawl and public health." Public Health Rep. 117: 201-217.

Galea, S., N. Freudenberg, et al. (2005). "Cities and population health." Soc. Sci. Med. 60: 1017-1033.

Gruebner, O., Lakes, T., Krämer, A., Khan, M., M., H., Hostert, P. (2009). "Neighbourhood conditions, environmental and socio-economic factors explain health in informal settlements of Dhaka, Bangladesh". 7th International Scientific Conference on the Human Dimensions of Global Environmental Change. Bonn.

Häußermann, H. and A. Kapphan (2000). Berlin: von der geteilten zur gepalteten Stadt. Opladen.

Hoffmann, A., Mahidadia, A. (2010): "Machine Learning". In: Gaber, M. M. (Eds.): Scientific Data Mining and Knowledge Discovery (pp. 7-52). Springer.

Kruize, H. (2007). On environmental equity Exploring the distribution of environmental quality among socio-economic categories in the Netherlands. Utrecht.

Lafortezza, R., G. Carrus, et al. (2009). "Benefits and well-being perceived by people visiting green spaces in periods of heat stress." Urban For. Urban Green(8): 97-108.

Patz, J. A., Campbell-Lendrum, et al. (2005). "Impact of regional climate change on human health." Nature 438: 310–317.

Schneider, T., Goedecke, M., Lakes, T. (2007). "Berlin Urban and Environmental Information System: Application of remote sensing for planning and governance". In: Stefanov, W., Netzband, M., Redman, C. (Eds.) Applied Remote Sensing for Urban Planning, Governance and Sustainability (pp. 199-219). Springer.

Whitford, V., A. R. Ennos, et al. (2001). "“City form and natural process”—indicators for the ecological performance of urban areas and their application to Merseyside, UK." Landscape Urban Plan 57: 91-103.

Witten, I. H., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S. J. (1999). “Weka: Pracitcal Machine Learning Tools and Techniques with Java Implementations”, Tech. Rep. Department of Computer Science, University of Waikato, New Zealand.

[1] Umweltatlas Berlin: http://www.stadtentwicklung.berlin.de/umwelt/umweltatlas/
[2] Gesundheitsberichterstattung: http://www.berlin.de/sen/statistik/gessoz/index.html
[3] Polizeistatistik: http://www.berlin.de/polizei/kriminalitaet/pks.html
[4] Statistisches Landesamt: http://www.statistik-berlin-brandenburg.de/
[5] Mortalitätsatlas: http://www.mortalitaet.nicoleueberschaer.de/
[6] Rapid Miner: http://www.rapid-i.com