

Spatio-temporal Visualization of Interpolated Particulate Matter (PM_{2.5}) in Beijing

Andreas Keler and Jukka M. Krisp

Applied Geoinformatics, University of Augsburg, Germany · andreas.keler@geo.uni-augsburg.de

Full paper double blind review

Abstract

People in growing urban areas are more and more influenced by emissions coming from numerous vehicles and factories. In this paper we inspect the concentration of particulate matter (PM_{2.5}) visually over time. This information stems from a data set of air quality measurements from 36 static sensors in Beijing over one year (from 8.02.2013 till 8.02.2014). One possibility for creating an overview for 36 positions with varying PM_{2.5} measurements in time is the use of interpolation techniques. In our approach, we generate surfaces of PM_{2.5} concentration using inverse distance weighting (IDW). The resulting surfaces represent interpolated PM_{2.5} values, based on averaged PM_{2.5} information (e.g. average of one day). We create simple interactive visualizations using points as surface representations. Each surface point within the 3D visual analysis display exhibits its PM_{2.5} value by differing coloration and z-value (height component). The interactivity consists of using selection circles for stacked 3D displays of interpolated PM_{2.5} surfaces for different times (time series). The aim of this visual information analysis is the possible detection of periodical hotspots of high PM_{2.5} concentrations, which might be useful for people with respiratory diseases. For the detection of dynamic PM_{2.5} hotspot variations, we introduce thresholds for querying only the highest PM_{2.5} values of the surfaces. Afterwards, these points are aggregated into convex hulls (polygons), with the idea of comparing the size and shape of the PM_{2.5} hotspots in each created surface. The change of position and size of these polygons over time may be an indicator for air quality changes within an urban environment. Considering the above, this may be a starting point for the conception of a personalized routing solution for pedestrians or vehicle drivers with respiratory diseases, who want to avoid these hotspots of high PM_{2.5} concentrations.

1 Introduction

In today's environment, especially in urban areas, we can observe changes in air quality due to an increasing number of people, vehicles, and factories. For ZHENG et al. (2013) the air quality in urban spaces is characterized by non-linear variations. Additionally there are multiple factors influencing air quality like weather, traffic, or land use type (ZHENG et al. 2013). Especially higher PM_{2.5} concentration is an indicator of adverse health effects (CAO et al. 2013). Therefore, the analysis of air quality is required, especially its influence on human health. This can include, for example, spatial analysis of air pollution and mortality (JERRETT et al. 2005). Since sensors for air quality measurements are often static (CAO et

al. 2013), the resulting information is limited to point positions. One possibility for the visual analysis of georeferenced point data sets can be the creation of surfaces using point interpolation techniques.

1.1 Point Interpolation Techniques

There are numerous approaches in environmental and health research using interpolation techniques for air quality data, especially for PM concentration (TRAN et al. 2012, WONG et al. 2004, JERRETT et al. 2005). Spatiotemporal interpolation of fine PM2.5 has already been performed using inverse distance weighting (LI et al. 2014), radial basis functions (LOSSER et al. 2014), and kriging (MWENDA & SHI 2012). Within this thematic field, especially when creating continuous fields from point data, the term uncertainty becomes really important, because it can be represented differently (VČKOVSKI 1995). The reason is that after interpolating point information, there are additional variances stemming from the influence of the used interpolation technique on a map. Therefore SENARATNE et al. (2012) describe how to compare interpolated PM10 concentration and its uncertainty with different types of maps.

ZHENG Y et al. (2013) compare linear point interpolation with an individual interpolation method by the name of U-Air for PM10 measurements for 22 stations in Beijing. The results of the comparison show that U-Air can consist of more detailed PM10 distribution than linear interpolation. This shows that the used method for point interpolation has a decisive influence on the resulting surfaces.

There are numerous already implemented point interpolation methods nowadays used in GIS software. Well-known point interpolation techniques include Kriging, first described by KRIGE (1951), the mentioned linear point interpolation, spline interpolation, and inverse distance weighting (IDW), which was introduced by SHEPARD (1968). The general idea of the last mentioned interpolation method is that each measured value has a weight that is inversely proportional to the distance to the estimated point values. For WATSON & PHILIP (1985) one restriction of IDW is that estimated values have to be within the range of the extrema of the sample values. For our approach, we use IDW, and do not consider this as a disadvantage, due to the sparse amount of points and the simplicity of this method.

1.2 Defining a Continuous Field

In general, there is a differentiation between discrete and continuous geodata. Discrete data refers to discrete objects in space with defined boundaries, like, for example, waterbodies, roads, or buildings. On the other hand, there is continuous data describing surfaces or fields, with a certain value for each location (in 2D). For GOODCHILD & GLENNON (2009) many geographic phenomena are continuous. Besides air quality measures, there are numerous other phenomena that appear as continuous fields. This includes an additional time component of continuous data, which extends continuous fields to progressively varying continuous fields. Examples for the last-mentioned term are atmospheric movement, ocean circulations, temperature, humidity, or vehicle and pedestrian movements. Some of these progressively varying surfaces can be modelled by deriving information from multispectral imagery data. The other case is to produce interpolated continuous fields based on discrete data, due to unavailability of continuous geodata. In case of points with time components, this step consists of point interpolation techniques for different moments or time windows.

Based on the knowledge of Tobler's First Law of Geography (TOBLER 1970), which states that nearby objects or things are more similar to each other than more distant things, point interpolations can be proved by intentions. Therefore, GOODCHILD (2009) states that this assumption allows the interpolation of continuous surfaces based on point observations. Nevertheless, some difficulties in standard statistical methods can be carried jointly, as interpolated surface properties are assumed to be uniform (GOODCHILD 2009).

1.3 Ideas for Visualizing Progressively Varying Continuous Fields

There are numerous possibilities for progressively visualizing varying continuous fields. One possibility may be the use of animation, where each static state of the investigation areas is pictured successively. In most cases this option does not allow interactive visual data analysis, and serves only for a rough overview of the changes. Our approach consists of averaging measured PM2.5 concentrations for one day, and performing inverse distance weighting on this averaged information for 36 points in Beijing. The raster data sets are subsequently converted into sets of points. Afterwards, we iterate this step for a rash of days, and define the z-values as averaged PM2.5 concentration. We visualize each surface with darkening coloration and increasing z-coordinates for higher values in a three-dimensional coordinate system, and optionally perform queries on selecting points that match road elements. The final step consists in generating a stacked view of these 3D views with ascending order by day from the bottom to the top. By using selection circles, we detect suitable thresholds for the extraction of polygons representing high PM2.5 concentration areas for each surface.

2 Description of the Test Data

Our test data set is freely available and was provided by Microsoft Research Asia (ZHENG et al. 2013, 2014). The data format and its properties will be described in the following. There are 36 monitoring stations, which are described in a separate CSV file with their respective station name, identification, and geographic coordinates. Table 1 provides a structural overview of the list of monitoring stations.

Table 1: Structure of the used CSV file "Station"

Columns	Station ID	Station name	Latitude	Longitude
Unit	None	None	Degree (°)	
Meaning	Each station with unique id	Chinese name	A station's position	

The measurements of each of the 36 inspected stations are provided in another CSV file, which is simply called "crawled data". An overview of this data set is provided in table 2. In this table, the data structure is described, showing only the selected air quality parameters with their respective units and meanings.

Table 2: Structure of the used CSV file “CralwedData”

Columns	Unit	Meaning
Station ID	None	Each station with unique id
Time	None	Data updated time
PM2.5 AQI value	None	Current AQI of PM2.5

Besides Time and PM2.5 AQI value, the attribute Station ID is used in our approach (see table 2). The Station ID is used for connecting each measured PM2.5 value with the exact station position. Since we want to investigate the values of PM2.5, the other 7 attributes are neglected for our study: PM10 AQI, NO2, Temperature, Pressure, Humidity, Wind, and Weather. The question is now how to interpolate 36 positions with differing PM2.5 values. It should be noted here that the inspected data set includes some missing PM2.5 measurements, which means that the time of acquisition is not always constant. Therefore, we average the measured PM2.5 values for certain time windows to guarantee its appearance in all 36 positions. As a result, we can say that the temporal accuracy of the data will be reduced. In our examples, we aggregate the PM2.5 values to daily average PM2.5 values, which are the basis for the further creation of interpolated surfaces. The positions of these stations are pictured in figure 1.

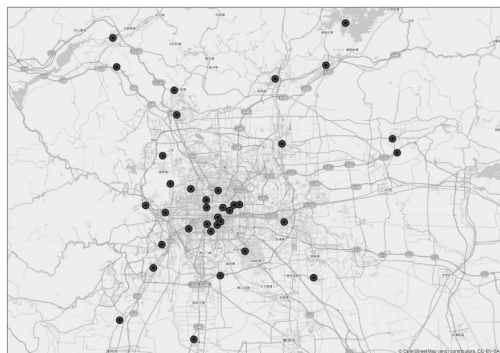


Fig. 1: Positions of the 36 stations in Beijing, where air quality was measured, projected on an OSM basemap

In figure 1, the positions of the measuring stations in Beijing are shown on an OpenStreet-Map (OSM) basemap. In the central area, a dense distribution of the points is detectable. This fact might be useful for the later interpolation step, because it provides insight into the reliability of the interpolations results, which are in our case surfaces of PM2.5.

3 Visualization of Interpolated PM2.5 Surfaces

Our idea for visualising the presented test data set is based on creating a time series of interpolated surfaces for the description of a progressively varying continuous field. In the following we describe the surface generation workflow, and show how to visualize interpolated PM2.5 information in 3D. One extension of this visualization is a stacked view of

multiple surfaces from different times, with interactive selection possibilities. Based on this interactivity, we propose a technique for extracting PM2.5 hotspots, by introducing thresholds, point clustering, and convex hull calculation.

3.1 Description of the Used Method for Spatiotemporal Visualization

As already mentioned, the first steps consist of connecting the 2 CSV files by joining them with the attribute station ID, and of averaging the measured PM2.5 values for certain time windows. Afterwards, the points are projected onto a map. Depending on their spatial distribution, we select a suitable point interpolation technique, which is in our case inverse distance weighting (IDW), for creating raster surface representations with a cell size of 2 meters. We provide this technique using a variable search radius, which is dependent on the number of nearest input points. In our case we use 12 input points to interpolate the PM2.5 value for each cell in the resulting raster surface.

Afterwards, we derive center points of the raster cells for representing the surfaces with points. This workflow for creating interpolated PM2.5 surfaces is pictured in figure 2.

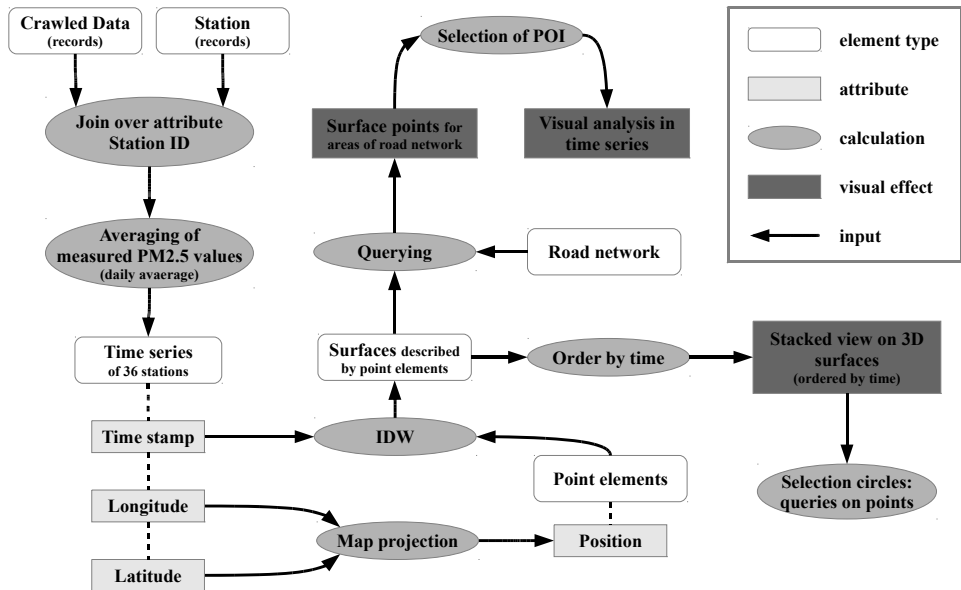


Fig. 2: Workflow for creating interpolated PM2.5 value surfaces and three-dimensional surface visualization

The end product of the presented workflow in figure 2 is a series of interpolated PM2.5 surfaces, which are represented by regularly ordered points.

3.2 3D surface Representations of Interpolated PM2.5 Values

Our interpolated PM surfaces are now visualized in 3D by using the third coordinate axis to depict PM2.5 values for certain two-dimensional positions. As already mentioned in the

workflow in figure 2, we have the option of querying the surface points that match the road network. The reason why we only use interpolated PM2.5 values that match the road network of Beijing is to provide orientation for the visual analysis of a possible end user. An example of an interpolated PM2.5 surface for one day is pictured in figure 3, where the z-value of the surface starts at the fixed registration point of PM2.5 is equal to 0. For reasons of orientation, this point starts in figure 3, where the road network is shown (bottom).

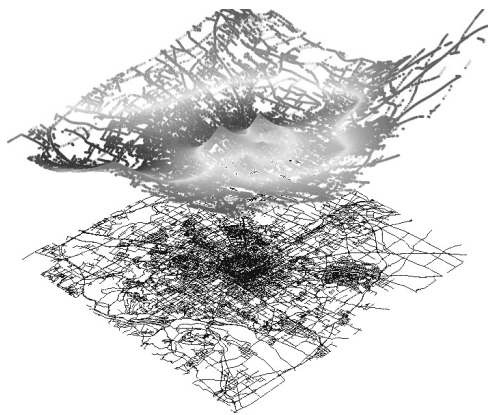


Fig. 3:

Three-dimensional representation of an interpolated PM2.5 surface in Beijing reduced to road network areas (top), and its corresponding road network based on OSM data (bottom)

In figure 3, a surface is presented with colors, showing the classified interpolated PM2.5 values and varying height depending on the relative distribution of these values. By interactively changing the 3D viewing angle, we can detect two bigger PM2.5 concentration peaks in the center of Beijing.

3.3 Comparison of Multiple Interpolated PM2.5 Surfaces within a Stacked 3D View

After the generation of six surfaces representing PM2.5 values for certain time windows, an extended visualization technique is used for the spatiotemporal analysis. The displays for comparing different PM2.5 surfaces consist of stacked 3D surfaces similar to the view in figure 3. The z-axis represents two attributes in this display: the time component in decimals (number of 3D surfaces), and PM2.5 concentration (z-axis within each 3D surface). With the use of a selection circle we further inspect the interpolated values. This inspection is deduced from the need of gaining insight into certain surface partitions with respect to the time component. Selecting a certain amount of surface points enables the derivation of histograms for interpolated PM values. With this method it is possible to visually compare surfaces of different time windows by their differences. Further analysis of the surfaces is possible with calculating differences between the interpolated PM2.5 values in time.

3.4 Creating Hotspot Polygons by Introducing Constant Thresholds and Aggregation

After inspecting the PM2.5 maxima on the surfaces, we introduce a constant threshold for applying a query for the higher PM2.5 values. Based on this selection, certain areas of the surfaces represent multiple hotspots within the investigation area. For detecting areas with

higher PM2.5 concentration we first provide a distance-based clustering on the selected surface points (former cell size: 2 meters), with a search distance of around 50 meters, and aggregate the resulting point clusters by creating convex hulls (polygons). For the step of creating convex hulls we use the approach based on JARVIS (1973). After this aggregation step is used for four example time windows, we derive different polygons for different time windows. Afterwards, these different polygons will be compared with each other by visual inspection. The workflow for creating PM 2.5 hotspot polygons is pictured in the diagram in figure 4.

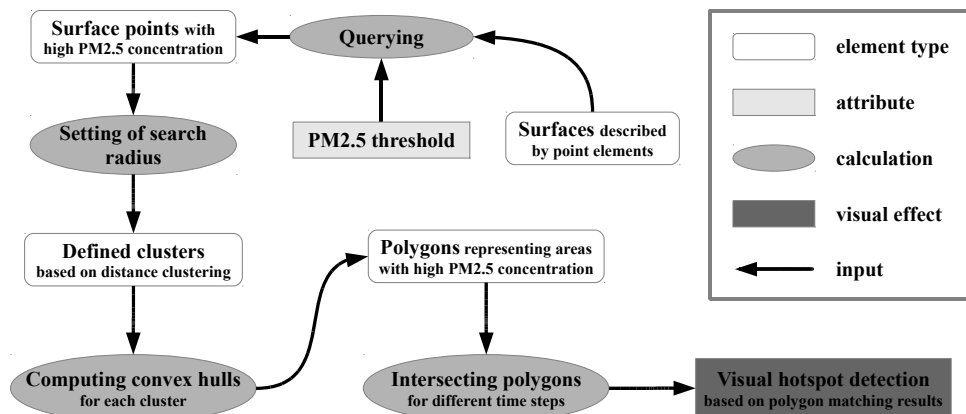


Fig. 4: Workflow for creating PM2.5 hotspot polygons

The resulting polygons are computed for four day averages, with the aim of testing possible polygon intersections. Intersected areas may show certain periodically appearing PM2.5 hotspots.

In case hotspot polygons are created based on a consistent PM2.5 threshold, empty surface point selections are generally possible: PM2.5 maximum is below the selected threshold. This may result in missing hotspot polygons within a series of inspected PM2.5 surfaces.

4 Results

In the following we will show example results for the visualization of PM2.5 surfaces.

4.1 Visual Analysis Using a Stacked View for Three-dimensional Surfaces

The visual analysis of multiple PM2.5 surfaces (day averages) is now presented with an example. In this stacked 3D surface view we have a number of interpolated PM2.5 surfaces from different times displaced on the z-axis. This series of surfaces is ordered by time, with a surface arrangement of ascending days from the bottom to the top. Each fixed registration point of each PM2.5 surface is equal to 0. The idea behind this is to avoid intersections of different surfaces. Further analysis on the interpolated surfaces consists of detecting points of interest (e.g. PM2.5 extrema) within the investigation area, and their selection and in-

spection over a period of time. One example of such linked or connected view is pictured in figure 5, where a certain area in Beijing is selected.

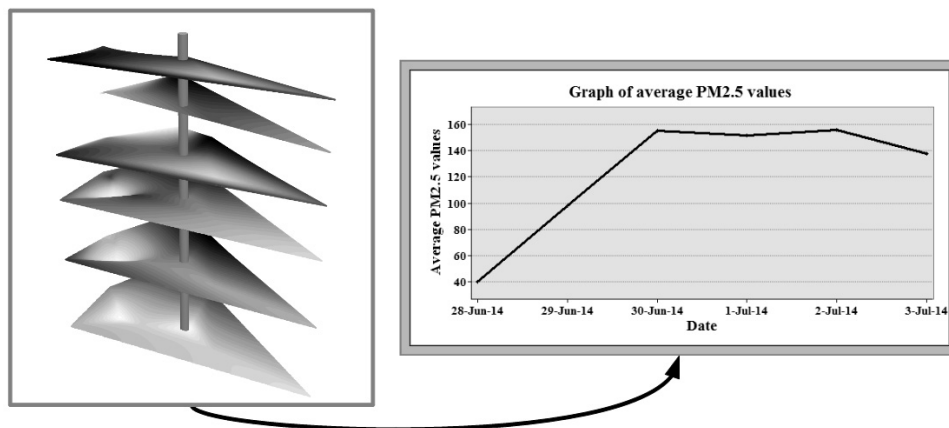


Fig. 5: Example of a stacked view of a series of interpolated PM2.5 surfaces (left), and corresponding graph of average PM2.5 values for a selection circle query (right)

Figure 5 shows on the left hand side the stacked surface series visualization with six surface layers. Darker colorations come along with higher z-values, and conclude higher PM2.5 concentrations. Additionally, there is a column on respectively the same geographic location, which is based on an extruded selection circle, and represents an interactive selection possibility. We use this spatial query for averaging values for the same selected area. One example of such PM2.5 average results is shown on the right hand side of figure 5, with a Time-PM2.5-diagram. The variation of interpolated PM2.5 values within the same area during six days can be detected in this chart.

4.2 Visual detection of PM2.5 hotspots by using polygons

After the generation of interpolated PM2.5 surfaces, we introduce polygons that are derived by certain selected thresholds, and represent higher PM2.5 concentrations. This method is described by a workflow diagram in figure 4. In figure 6, one example of four daily average PM2.5 hotspot polygons is pictured in a map view.

The setting of the PM2.5 threshold in this example is $230 \mu\text{g}/\text{m}^3$ and consistent for all derived polygons of four days. Besides some intersecting polygons, there are large variations of the polygons in shape and size, which shows high concentration variations between the days. Overlapping patterns can be described formally by type of interaction.



Fig. 6: Comparison of PM2.5 hotspot polygons for four different time windows

5 Conclusion

Our method for the visualization of PM2.5 by three-dimensional interpolated surfaces delivers good results for getting an overview about the relative changes of PM2.5 concentration over time. The stacked view is useful for comparing concentrations of different times, and hotspot polygons help to immediately detect higher values. Nevertheless, it is difficult to detect detailed information on air quality using our visual analysis method. Additionally, we need expert knowledge for defining critical concentrations, especially when creating polygons. Further knowledge on other air quality measures is needed for detecting correlations, which may include linked visual analysis displays or the use of visual geodata mining techniques.

6 Outlook

Besides inspecting PM2.5 surfaces derived using IDW, we can test other interpolation techniques on this data. Additionally, time-dependent PM2.5 hotspot polygons can be tested for their use for personalized routing solutions. Another question is how to guarantee reasonable routing results based on costs of continuous fields, which are connected to transportation networks. This conception was presented by KARRAIS et al. (2014), with some differing examples of how to connect interpolated point information represented by raster data and urban road networks represented by vector data. Depending on the used point interpolation technique, and on the connection with road elements, different paths with differing accumulated costs were computed for the same start and end points. Nevertheless, no clear answer was given on the best solution, which opens initiatives to conduct further tests on this complex of problems. Additionally, it must be mentioned that the presented polygons in this paper can be used for personalized routing solutions as obstacle polygons. They represent high PM2.5 values, and could be avoided within a personalized routing solution for people with respiratory problems.

References

- CAO, J., CHOW, J. C., LEE, F. S. C. & WATSON, J. G. (2013), Evolution of PM2.5 Measurements and Standards in the U.S. and Future Perspectives for China. *Aerosol and Air Quality Research*, 13, 1197-1211.
- GOODCHILD, M. F. (2009), Geographic information systems and science: today and tomorrow. *Annals of GIS*, 15 (1), 3-9.
- GOODCHILD, M. F. & GLENNON, J. A. (2008), Representation and computation of geographic dynamics. In: HORNSBY, K. S. & YUAN, M. (Eds.), *Understanding Dynamics of Geographic Domains*. CRC Press, Boca Raton, 13-30.
- JARVIS, R. A. (1973), On the identification of the convex hull of a finite set of points in the plane. *Information Processing Letters*, 2, 18-21.
- JERRETT, M., BURNETT, R. T., MA, R. J., POPE, C. A., KREWSKI, D., NEWBOLD, K. B., THURSTON, G., SHI, Y. L., FINKELSTEIN, N., CALLE, E. E. & THUN, M. J. (2005), Spatial analysis of air pollution and mortality in Los Angeles. *Epidemiology*, 16, 727-736.
- KARRAIS, N., KELER, A. & TIMPF, S. (2014), Routing through a continuous field constrained by a network. *Extended Abstract Proceedings of the GIScience 2014, Movement and Context*, 57-61.
- KRIGE, D. G. (1951), A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52 (6), 119-139.
- LI, L., LOSSER, T., YORKE, C. & PILTNER, R. (2014), Fast Inverse Distance Weighting-Based Spatiotemporal Interpolation: A Web-Based Application of Interpolating Daily Fine Particulate Matter PM2.5 in the Contiguous U. S. Using Parallel Programming and k-d Tree. *International Journal of Environmental Research and Public Health*, 11, 9101-9141.
- LOSSER, T., LI, T. & PILTNER, R. (2014), A Spatiotemporal Interpolation Method Using Radial Basis Functions for Geospatiotemporal Big Data. *Proceedings of the 5th International Conference on Computing for Geospatial Research and Application (COM.Geo 2014)*, 4-6 August 2014 (Washington, DC), 17-24.
- MWENDA, K. M. & SHI, X. (2012), Spatial Interpolation of Fine Particulate Matter in New Hampshire using Landuse-based Kriging. *Extended Abstract Proceedings of the GIScience 2012, Extended abstracts (posters)*, Columbus, Ohio, 09/2012.
- SENARATNE, H., GERHARZ, L., PEBESMA, E. & SCHWERING, A. (2012), Usability of Spatio-Temporal Uncertainty Visualisation Methods. *Proceedings of the 15th AGILE International Conference on Geographic Information Science – Bridging the Geographic Information Sciences*. Springer, 3-23.
- SHEPARD, D. (1968), A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 23rd ACM national conference*, 517-524.
- TRAN, H. N., LEELASAKULTUM, K. & MÖLDERS, N. (2012), A Tool for Public PM2.5-Concentration Advisory Based on Mobile Measurements. *Journal of Environmental Protection*, 3, 1671-1688.
- VČKOVSKI, A. (1995), Representation of Continuous Fields. In: *Proceedings of the AUTOCARTO 12 Conference*, 127-136.
- WATSON, D. F. & PHILIP, G. M. (1985), A Refinement of Inverse Distance Weighted Interpolation. *Geoprocessing*, 2, 315-327.

- WONG, D. W., YUAN, L. & PERLIN, S. A. (2004), Comparison of spatial interpolation methods for the estimation of air quality data. *Journal of Exposure Analysis and Environmental Epidemiology*, 14, 404-415.
- ZHENG, Y., LIU, F. & HSIEH, H-P. (2013), U-Air: When Urban Air Quality Inference Meets Big Data. 19th SIGKDD conference on Knowledge Discovery and Data Mining (KDD 2013).
- ZHENG, Y., CHEN, X., JIN, Q., CHEN, Y., QU, X., LIU, X., CHANG, E., MA, W-Y., RUI, Y. & SUN, W. (2014), A Cloud-Based Knowledge Discovery System for Monitoring Fine-Grained Air Quality. no. MSR-TR-2014-40, March 2014.