Eurographics Workshop on Urban Data Modelling and Visualisation (2015)
F. Biljecki and V. Tourre (Editors)

Representing Urban Phenomena in Their Context and at Different LoD: from Raw Data to Appropriate LoD

H. Pham[1], A. Ruas[1], T Libourel[2]

[1] French Institute of science and technology for transport, development and networks, France
[2] University of Montpellier, France

Abstract

More than half of the world population lives in cities today. This proportion rises to 80% in developed countries. The density of urban population causes environmental troubles such as noise, urban heat waves, and chemical pollutions or magnetic pollution. Sensors and models are used to improve knowledge related to these phenomena particularly in cities. The aim of our research is to propose methods to view these phenomena in contextualised ways and at different levels of details. In the context of data exploration, we wish to generate from the initial phenomena data other levels of detail to allow the visual perception of the information at different visual scale. We also propose symbols that resist as well as possible to scale change and without excessive covering the other information such as streets, buildings or names. The first solutions presented in this paper are implemented and illustrated through two examples: nocturne temperature in Paris with very sparse initial data and concentration of chlorine with very dense initial data.

Key-words: urban phenomena mapping, multi-scale, LoD, multiple representations, generalization, interpolation, semiotics

1. Introduction

Nowadays, more and more people live in cities. The concentration of urban population generates not only urban sprawl but also environmental problems such as noise, urban heat waves, and chemical or magnetic pollutions. As citizens we wonder if the air we aspire is pure, if the building we live in protects us from heat waves or how telephone signals impacts our health. If one of the objectives of smart cities is to improve the management of resources - such as water and electricity - another is to be able to view day by day what is going on in the city. Access to this information concerns the population as well as the technical services of the city. However today, at best, specific agencies are specialised in the computation and the communication of one type of phenomenon only. This is the case of AirParif, for example, which monitors a set of sensors (measuring NO_2, O_3, PM10, PM2.5) and proposes either the visualization of sensor values overlayed on Google Earth or a choropleth map including only the boundary of administrative unit around Paris (Figure 1).

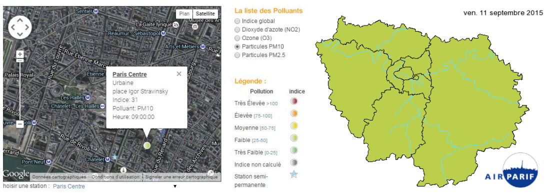

Figure 1: *Air Pollution visualization © AirParif*

Although these kinds of maps are useful, we wish to go one step further to propose processes and methods to map field values that describe one or several phenomena with other data that describe the space we live in. Based on web services and open source software such as PostGIS, we wish to propose ways to see phenomena either locally or globally and to see the interrelations between places we live in (the buildings), we move on (streets) and these phenomena. In order to illustrate our work we propose to map two phenomena: the temperature in the city and the water flow in a water pipe line. From initial data we wish to propose different levels of detail. For the water, we wish to show the outside – to see the connection between the water network and the buildings that use this water - and the inside, to see the water flow and the local concentrations of chlorine. For the temperature we wish to show the distribution of temperature in the city and also in each street. In this paper we do not propose to analyse the data but to map it. Even though scientific analytical and visual tools exist (e.g. Paraview), we use GIS suites in order to take advantage of geolocalisation and GIS functions. Moreover our aim is not to overlay graphical layers that are conceived independently from each other zoom on to adapt, at least partially, the symbolisation of phenomena to the zone we zoom to optimise the perception of the phenomena and of its geographical context. To achieve this, our first step is to generate different levels of detail (Figure 2).

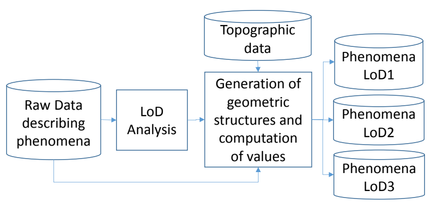

Figure 2: *Generation of LoDs from raw data*

The next section is a brief overview on current trends for mapping phenomena. We then summarize our proposal in section 3 and describe the implementation of our

DOI: 10.2312/udmv.20151346

solutions and present the graphical results through two examples: a study of temperature in Paris and a model for mapping drinking water flow in section 4 before concluding.

2. Elements for mapping phenomena

We use the term *phenomena* to describe physical events such as temperature, water flow and pollution. Phenomena are complex to measure and to model, and the representation is always an approximation of the reality. Phenomena belong to continuous field data and are represented by values on 2D or 3D grid nodes. Phenomena vary in time but we do not treat the dynamic aspect in this paper. In this paper *representation* refers to the graphic representation but it encompasses the *geometric structure* that holds the symbolisation. To represent a physical phenomenon, different kinds of software exist:

- Software that contains a model to calculate a phenomenon and a viewer to visualize it. For example, to simulate water flow in the pipe, physicists use Computational Fluid Dynamics (CFD) software such as FreeFem++ for the computation and Paraview for the visualization (see section 4.2) [WMC13].
- data analysis software, which has graphics capabilities such as R. R has functions for data processing: spatial analysis, calculating new data fields by interpolation and some capacity to map the results,
- GIS software. GIS has, a priori, all capacities to map geographic information. However, some are more devoted to continuous objects and others to discrete ones. In our cases we want to map discrete and continuous objects together, in a harmonious way.

As stated in the introduction, most phenomena maps are limited to choropleth or dot maps accessible through the internet (see http://aqicn.org/map/world for example). The representation of phenomena in itself has been studied by some authors but mainly to focus on the temporal aspect (e.g. [Yun01]) whereas we wish first to improve the visual aspect for each state. Very interesting platforms such as GeoVista Suite or CommonGIS, respectively developed at the universities of Pennsylvania and Fraunhofer, are providing solutions for multivariable information, data series and mobility but not for phenomena. [AAG03] gives an overview of methods and software to explore data but does not focusing on phenomena mapping. The representation of phenomena raises two issues: how to obtain an appropriate LoDs (which refers to multiple representation, generalisation and Interpolation) and how to find appropriate symbolisation in 2D or 3D (it refers to semiotics).

To represent phenomena on different scales, Frank and Timpf proposed a multi-scale data structure for cartographic objects: store objects at different LoDs in a database [FT94]. To generate coarse information, much literature exists related to Generalization techniques [MRS07]. To spread the information around initial data, several GIS interpolation methods exist and are compared in the work of Meng, Liu and Borders [MLB13]. For semiotic, and amongst others, Bertin [Ber67], demonstrated that the unsuitable symbolisation may disturb the understanding of maps and Christophe [Chr11] focused on the role of colours. To conclude much knowledge already exists on multiple representation, generalisation, interpolation and semiotics. We wish to use this knowledge to improve the representation of phenomena which has seldom been studied itself.

3. Representing a phenomenon

Our work focuses on the optimization of the presentation of phenomena. It encompasses:

- The visualisation of phenomena with other topographic data,
- The visualisation in 2D or 3D according to the needs of the use,
- The adaptation of the symbolisation according to the zoom and the angle of view. Whenever one of the parameters of views changes, the representation readjusts itself.

Most of the time, raw data that describe a phenomenon is modelled by a set of values on locations. Often, raw data is not suitable for visualization, because data is either too dense or too sparse to see all the particularities. By default, symbols overlap and hide the phenomenon or are too far away from on another. So, in order to visualize a phenomenon at different scales, we propose to create a new data set that encompasses a new geometric structure and new values attached to this structure. The new geometric structure might be another grid, a plan or a line network or even a node. Let's focus on grids first. Each grid corresponds to a level of detail. A new grid is created by raw data densification or generalization according to the required level of detail (fig2), a function is applied to compute the new values attached to the each of the node of the new grid.

For graphical representation purposes we add symbols to hold the values. Four geometrical structures may be used: a grid consists of nodes, a plan composed by polygons, a line composed of segments or a single node. A plan can be composed of regular square polygons or concentric areas resulting from classification. For larger scale, grids are efficient, for a smaller scale plan, a network or node can be more efficient (see section 3).

Rep(phenomena) = {(node, symbol (value)} | {(polygon, symbol (value)) } | {(segment, symbol (value)) } | (node, symbol (value)).

Before studying smaller scale we first concentrate on grids computed by generalisation or interpolation from the initial grid. Section 3.1 and 3.2 remind the classical generalisation and interpolation functions that we have implemented for grids and we develop the symbolisation.

3.1. Aggregation by generalization

Grid generalization builds a generalized grid which is less dense than the original grid and computes the value of the new nodes (nG) from the values of initial surrounding nodes (nR). To do so, we first create a box around a node of generalized grid: stepXY, stepZ are the dimensions of the box along the X-axis X, Y-axis and Z-axis. In a 2D mapping, stepZ is null.

All nodes of the original grid inside the box are taken for calculating the generalized value of the node. The assigned value to an unknown node is the mean, maximum or minimum of its neighbouring points according to the nature of the phenomenon. For example, for drinking water network, a lack of chlorine can cause sanitation problems; the generalized value thus takes the minimum of its neighbours. For most pollution, it takes the maximum value. In the absence of knowledge, the function is the average value of its neighbours:

$$v(nG_m) = max|average|min\left(v(nR_1), \dots, v(nR_N)\right)$$

3.2. Disaggregation by interpolation

When we disaggregate data we create more information. Either we use interpolation method or we have external knowledge to refine the data. Interpolation is a well known method to compute values of the grid denser than the original one.

The value of a new node (nI) may be computed by different formulas. By default, the closer the node, the greater its importance:

$$v(nI_k) = \frac{\sum_{i=1}^{N} \frac{v(nR_i)}{\mathrm{r}(nI_k, nR_i)}}{\sum_{i=1}^{N} \frac{1}{\mathrm{r}(nI_k, nR_i)}}$$

Where $\mathrm{r}(nI_k, nR_i)$ is the distance between nI_k and its neighbouring nodes (for r < l0) .

3.3. LoD, geometry and symbolization

In order to symbolize the phenomena at different LoDs, new geometries should be created, from the most detailed, to the simplest: a 3D grid, a 2D grid, a plan, a set of lines, a node. At the beginning of the process we start with a grid. If it is not dense enough the grid is interpolated, otherwise it is generalised. Then to reach small levels of details the geometry is simplified and the values are aggregated by mean, max of average functions. For a specific LoD we chose an appropriate symbol for which one of the characteristics depends on the value.

- For a grid, we add symbols on nodes. Symbols are characterised by a shape, a size and a colour
- For a plan, we can use large areas with a classification of values or a set of connected square like images (see Figure 10). Whatever the case, the symbolisation of each polygon will be mainly represented by color
- For a set of lines (see Figure 17), each segment will be represented by the line with a width and a colour.

The case of a grid is interesting because, if the symbols are not too big, a grid does not cover all space. Moreover, a grid can locally adjust its orientation according to the space the user zooms on in order to optimise the quantity of information perceived according to the available space.

In the same way a plan gives an overview and the idea of continuity. The plan should not cover all space but just the space between the buildings. We recall that the transparency solution, although it is often used, is not a good solution, as it decreases legibility and when two colours are mixed, it creates a third colour but does not allow to see the initial information from the two colours.

To sum up; to optimize the symbolization, first, we decide to represent the phenomenon on a grid or a plan(s) according to its context.

The grid allows to see the surrounding information of phenomena. If the phenomenon is based on a grid, the parameters of the grid must be chosen:
- step,
- number of levels Z,
- orientation around the axis Z,
- symbols on the nodes: type (punctual or area), size, colour family

The horizontal plan allows to see the continuity of phenomena. If the phenomenon is represented by plans, the parameters of one plan are:
- Z0
- size of one cell,
- symbols of polygons: classification, colour family

Sections 4 and 5 illustrate the symbolisation chosen according to the LoD.

3.4. Conceptual data schema

Figure 2 presents a simplified view on the data schema where raw data is distinguished from data for mapping. This is an extension of the model presented in [Rua15].

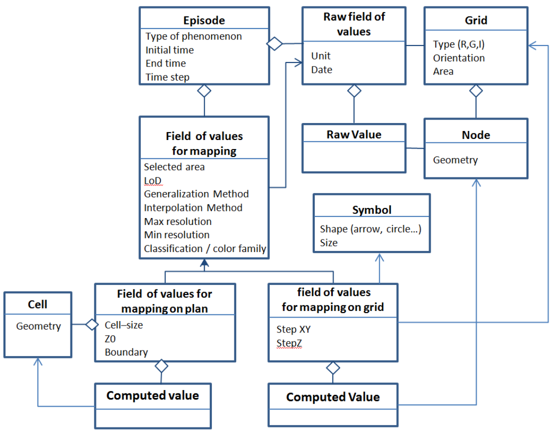

Figure 3: *The data schema of phenomena mapping*

In this model we distinguish raw data from data adapted to mapping. Data for mapping can be a grid, a plan or a set of lines according to needs and zooming. Each representation is characterized by a LoD, generated by generalization or interpolation and is mapped on screen in a resolution range. A phenomenon has thus different graphical representations that allow to improve the understanding.

4. Illustration of cases

In the following we present two cases to illustrate the concepts presented in section 3: the representation of temperature (4.1) which uses interpolation methods and the representation of chlorine in a water pipeline which uses generalization methods (4.2). Both phenomena are temporal, but in this paper we concentrate on the representation of one state only.

4.1. Meteorological data: from raw data to multilevel

Town Energy Balance, a specific model for cities

To better estimate the temperature in the cities in the future and anticipate the effects of climate change, researchers of Météo-France have developed a model, called TEB [HM08]. Validated in the 2000s, TEB model is based on physical equations of conservation and exchanges of energy and water. It takes into account the parameters of the surfaces that influence the atmosphere: vegetation, water surface, building, etc. It has already been tested in a dozen cities in the world, located at various latitudes: Paris, Montreal, Melbourne, etc.

TEB provides information on the flow of energy, water and CO_2 between the city and the atmosphere, urban microclimate of each ward, thermal comfort, etc. For each cell of the model, the city is represented by a street defined by its width, height and materials of its buildings, colors and thermal insulation of roofs, proportion of windows, etc. Spatial resolution can reach 100 meters. For our experiment we use the temperature computed by TEB in the streets in Paris and its suburbs at 2 AM, July 5^{th} 2010, with a 250m grid resolution. In the following we map this data at different LoDs.

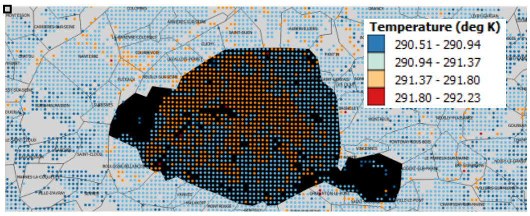

Figure 4: *The initial TEB output. A difference of temperature is observed between Paris and its suburbs*

Mapping the temperature data

To fully understand a phenomenon, we propose 4 LoDs:

- LoD1, global scale, allow a global view of urban heat island [1] in Paris

- LoD2, zoom on Paris, to detect heat wave area

- LoD3, zoom on a heat wave area, to study the variation of temperature in this zone

- LoD4, to see the temperature in a street.

With a 250 grid resolution, the density of raw data is low: this data grid satisfies the first level of detail but the points are too far apart at higher levels of detail. When we zoom in, the points are lost in space (Figure 5).

Figure 5: The initial grid is spaced too far apart

[1] *An urban heat island (UHI) is a city or metropolitan area that is significantly warmer than its surrounding rural areas due to materials and human activities. The phenomenon was first investigated and described by Luke Howard in the 1810s..*

Reversely, the quantity of nodes is high, so it is time consuming to load.

Propositions of mapping low density data

- **Densification of data by the interpolation method**

To accommodate the zooming process, we created 3 denser grids with cell size of 100m, 25m and 5m, for each LoD. The new grid values are computed by using the interpolation method presented above. To avoid errors of massive densification, we interpolate data progressively. Thus we calculate the 100m grid from the initial 250m grid, the 25m grid from the 100m grid and the 5m grid from the 25m one.

- **Temperature plan for overview**

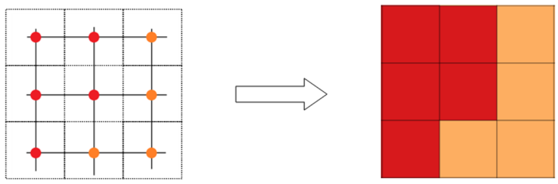

Figure 6: *The nodes are transformed into polygons to make a continuous plan of temperature*

As the human eye follows a continuous phenomenon more easily than a discontinuous one we propose to represent the temperature in a continuous plan. To do so each point is transformed into a square cell of the same size as the grid cell. To obtain the smoother plan, we choose a grid with the best resolution and a classification function.

- **Selected area (survey zone)**

In order to reduce loading time of data, we developed an extension of QGIS, which allows us to make all computations in a selected area only, and we do not waste time with other data that is outside of this area.

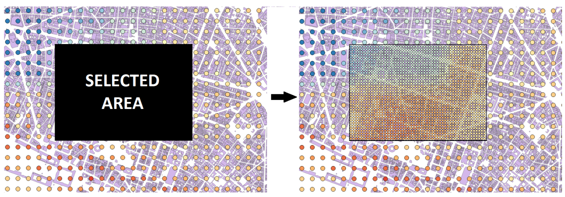

Figure 7: *Selection of an area and local computation of the new grid with more dense values*

Software concept

To represent the temperature, we utilized:

- PostGreSQL/PostGIS to store data
- QGIS and its plugin qgis2threejs for visualization

In addition, we developed three plugins of QGIS for data processing, which modify data in QGIS and in the database stored in PostGIS.

- (1) Select a survey area: create a new table of data that contains only data in the area selected by user,
- (2) Disaggregation by interpolation: create a new data grid denser than the initial one and compute its values
- (3) Transform a grid into plan: create a new plan from data with the same cell size

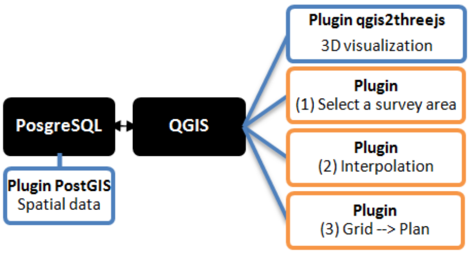

Figure 8: *Software concept*

Results

From raw data (250m grid), we created three denser grids (100m, 25m and 5m) and 1 plan (5m-cell size) to represent the temperature at different LoDs. Figure 4 illustrates the LoD1, with a 250m grid resolution. Figure 9 (LoD2) shows a full an overview of Paris.

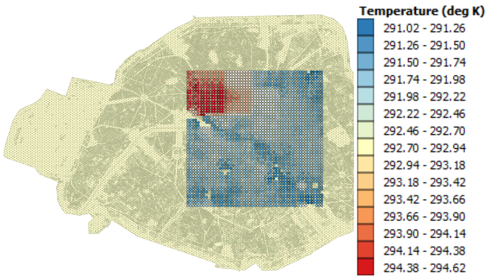

Figure 9: *100m grid (LoD2)*

Figure 10 illustrates the symbolisation with a plan. It allows seeing the continuity. Figure 11 and 12 are zooms, generated by progressive interpolation. Figure 12 is focused on a portion of a street (Boulevard Réaumur).

Figure 10: *Plan computed from 5m grid.*

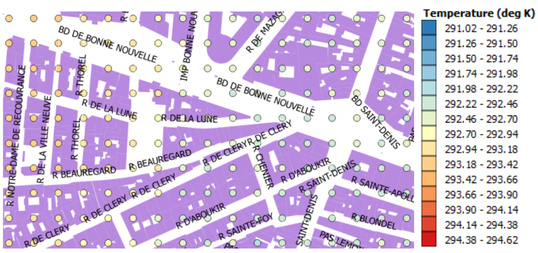

Figure 11: *25m grid satisfies the district scale (LoD3)*

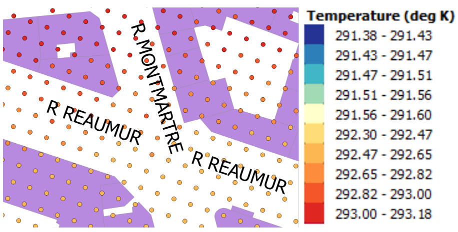

Figure 12: *5m grid adapts to a street view (LoD4).*

Figure 13 is a 3D representation where the grid is reoriented according to the orientation of the main street. Two levels are shown, one at 2m, the other at 5 m height.

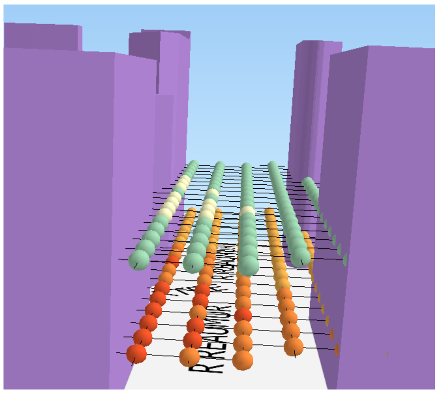

Figure 13: *5m grid adapts to a street view (LoD4) in 3D.*

4.2. Drinking water network

Data description

The initial grid describes the space inside the water pipe. It includes about 800.000 nodes for a 2km network. It is obtained from a model which computes pressure, flow velocity and chlorine concentration for each point. These data stimulate the propagation of chlorine in the network [WMC13]. In this paper, we focus only on the simulation of chlorine concentration in a steady state.

Mapping chlorine concentration

To survey water quality, we want to know what is happening in the pipe at different scales. We wish to map the data at three levels of details (LoD):

- LoD1, the district scale, gives an overview of whole network to quickly detect an anomalous zone

- LoD2, gives more details on the area,

- LoD3, is the pipe scale, to survey on the overdose or the lack of chlorine in the network

The first challenge is that the initial grid has too many points: about 800.000 nodes, so we cannot see them at LoD1 and LoD2 because points overlap. The second difficulty is the data density varies and the width of the pipe is not uniform, the largest section is 25cm with dozen points.

Propositions of mapping high density data

- **Use of the pipeline for low zoom levels**

Instead of displaying all the points of the calculation model, we represent the pipeline in small segments and small circles with a diameter equal to the largest diameter of the pipe. Each segment and circle takes the maximal value of the neighbour points. This presentation allows to detect anomalous areas and to have a global perspective of the entire network.

Figure 14: *LoD1, representing the data in segments*

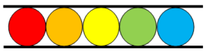

Figure 15: *LoD2, representation by a series of circles*

Values are computed by the method of generalization mentioned in section 3.

- **Replacing punctual symbols by area symbols**

By default, a symbol is a point with a size that does not vary according to map scale. When we zoom-out, symbols overlap (because the distance between the nodes decreases) and they may hide anomaly values. Reversely when we zoom-in the points become very small compared to the space between two nodes inside the pipe, they are lost in the void; the point density is too low to be visible. We propose to replace punctual symbols with area symbols (Figure 16). In such a case the area of the symbols varies according to scale. However when the zoom is too important, a new LoD is used.

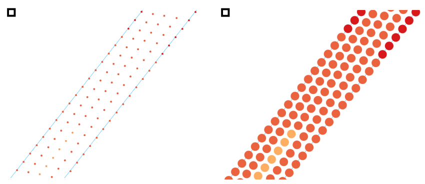

Figure 16: *Punctual symbols do not change when zoom (left) Surface symbols vary according to zoom (right)*

Implementation

To implement from initial data to web mapping, we use the following set of software:

- PostGreSQL/PostGIS to store the data,
- QGIS to create symbols,
- OpenLayers and GeoServe for web mapping,
- Web mapping with animation in JavaScript.

From raw data, we create two data fields by generalization to represent the concentration at different LoDs. They are illustrated in Figure 17 and 18.

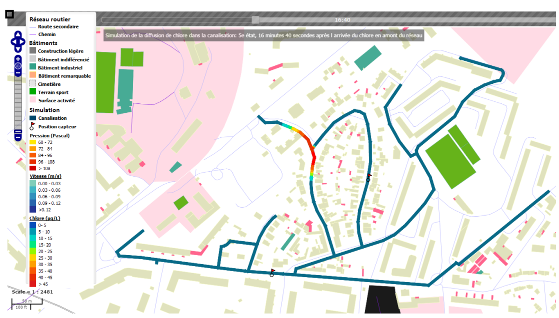

Figure 17: *LoD1, district scale, the most concentrated area of chlorine is detected quickly*

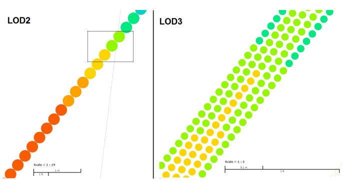

Figure 18: *LoD2 and LoD3, the area is shown in more detail*

5. Conclusion

We propose in this paper an approach for the multi-scale representation of data with any density: if data are too dense we generate other LoDs by generalisation. Reversely if raw data are too sparse, we generate a complementary grid or plan by interpolation. The appropriate LoD is automatically proposed according to scale. To do so, we propose a data schema (Figure 2) to represent a grid or plan and we propose to use area symbols that are more adapted to zooming. Our current research aims at optimizing the representation according to an observer by re-orientation of the grid/plan (Figure 13) and minimizing the information hidden by the buildings.

References

[AAG03] ANDRIENKO N., ANDRIENKO G., GATALSKY P.: Exploratory spatio-temporal visualization: an analytical view. *Journal of Visual Languages and Computing* (2003), vol. 14, iss. 6, pp. 503-541

[Ber67] BERTIN J.: Semiology of Graphics: Diagrams, Networks, Maps. *University of Wisconsin Press* (1983), (first published in French in 1967, translated to English by Berg W.J. in 1983)

[Chr11] CHRISTOPHE S.: Creative colours specifications based on knowledge (COLorLEGend system). The Cartographic Journal (2011), vol. 48, No. 2, pp. 138-145

[FT94] FRANK A., TIMPF S.: Multiple representations for cartographic objects in a multi-scale tree: an intelligent graphical zoom. *Computer Graphics* (1994), vol. 18, iss. 6, pp. 823-829

[HM08] HAMDI R., MASSON V.: Inclusion of a drag approach in the town energy balance (TEB) scheme: offline 1-d validation in a street canyon. *Journal of Applied Meteorology and Climatology* (2008), vol. 47, pp. 2627-2644

[MLB13] MENG Q., LIU Z., BORDERS B. E.: Assessment of regression kriging for spatial interpolation – comparisons of seven GIS interpolation methods. *Journal of Cartography and Geographic Information Science* (2013), vol. 40, No. 1, pp. 28-39

[MRS07] MACKANESS W. A., RUAS A., SARJAKOSKI L. T.: Generalization of Geographic Information: Cartographic Modeling and Applications, *International Cartographic Association* (2007), ISBN: 978-0-08-045374-3

[Rua15] RUAS A.: From a phenomenon to its perception: models and methods to represent and explore phenomena on GIS. *Modern Trends in Cartography Springer LNGC (2015),* ISBN 978-3-319- 07926-4, 259-268

[Yua01] YUAN M.: Representing complex geographic phenomena in GIS. *Cartography and Geographical Information Science,* Vol. 28, N° 2, 2001, pp. 83-96

[WMC13] WAEYTENS J., MERLIOT E., CHATELLIER P.: Fast inverse modeling technique to reconstruct laminar flow. *YIC 2013 ECCOMAS Young Investigators Conference Bordeaux, France* (2013)