Space-Time Mapping of Mass Event Data

Christian E. Murphy

Abstract In conventional thematic cartography the visualisation techniques to symbolise spatio-temporal phenomena are limited. On a two dimensional map temporal changes can only be visualised adequately as time series or by animation. To simultaneously visualise thematic data in space and time a third dimension must be added. In this work conventional cartographic symbolization meets the space-time cube to create a holistic 3D spatio-temporal visualisation model. The two dimensional proportional symbol mapping technique is adopted and extruded into the third dimension to model the temporal factor. Kernel density estimation is performed on the time line to create a temporal continuous model from discrete points in time. The resulting visualisation model is implemented into an earth viewer to enable the user to freely navigate the phenomenon and visually detect anomalies without losing the overall view. This tool is evaluated by visualizing the events of a mobile phone location dataset over space and time in one single model.

Keywords Visual analytics · Geostatistics · Thematic cartography · 3D visualisation · Mobile phone location data

1 Introduction

There is a need for effective methods to exploit and use the hidden opportunities and knowledge resting in unexplored data resources (Keim et al. 2010). An effective approach to reveal the hidden information in large spatial databases is

C. E. Murphy (✉)
Department of Cartography, Technische Universität München, Arcisstraße 21,
80333 Munich, Germany
e-mail: Christian.Murphy@bv.tum.de

J. M. Krisp (ed.), *Progress in Location-Based Services*,
Lecture Notes in Geoinformation and Cartography, DOI: 10.1007/978-3-642-34203-5_3,

to combine computational tools with human understanding into visual data exploration and sense making. This approach is part of the visual analytics field, which consists of various research areas. Visual analytics has been introduced and specified as the science of analytical reasoning facilitated by interactive reasoning techniques (Thomas and Cook 2005).

Despite various existing methods for data mining and data visualisation, both visual data exploration and statistical analysis of temporal changes of a phenomenon remain difficult. Haggett (1990) defines four types of temporal change in spatial data: constants, trends, cycles and shifts. Constants are long periods of no change while trends refer to long-term linear changes, e.g. migration of objects. Cycles describe recurring patterns like for instance daily hotspots of rush hour traffic and shifts describe random changes that misfit overall trends and cycles. In the case of a mobile phone location dataset an overall daily cyclic pattern would be expected.

It is necessary to simplify large amounts of event data. These events cannot be analysed and visualised adequately and therefore a phenomena cannot be understood. The events, each with a distinct location, can be used as points. This is why kernel density estimation has become so popular in recent years for its ability to make large amounts of point data understandable. Derived from point distances and a kernel function it assigns density values throughout the study area of the examined phenomena. The resulting density values can be classified and colour coded which leads to an isarithmic mapping technique. This density surface displays a simplified and abstract presentation of the point cloud to provide a legible and comprehensive visualisation. To provide a spatiotemporal analysis of a spatiotemporal phenomenon a current practice is to investigate space and time separately. It seems reasonable to use kernel density analysis to investigate on the spatial behaviour. A typical spatial analysis by kernel density estimation of a mobile phone call location dataset is shown in Fig. 1.

To respect the temporal aspect of a spatiotemporal analysis the spatial investigation is enriched with a temporal analysis. In many cases this consists of a histogram simply showing quantities over time. The events are then classified into discrete intervals. An example using the same mobile phone call location dataset is shown in Fig. 2.

This implemented approach of separating the spatiotemporal analysis into a spatial analysis and a temporal analysis has, regardless of its usefulness for particular cases, major drawbacks. Time is not applied as continuous in opposition to its nature. The events in time are presented in discrete manner. Information is lost as the distribution of events within a class is neglected. But the crucial drawback is that the study area cannot be explored in detail over time. Spatiotemporal hotspots are therefore not identifiable as either the overall temporal hotspots or overall spatial hotspots only can be depicted. It is not possible to detect the density of a certain place at a certain time. The divisiveness of the analysis results into a visual

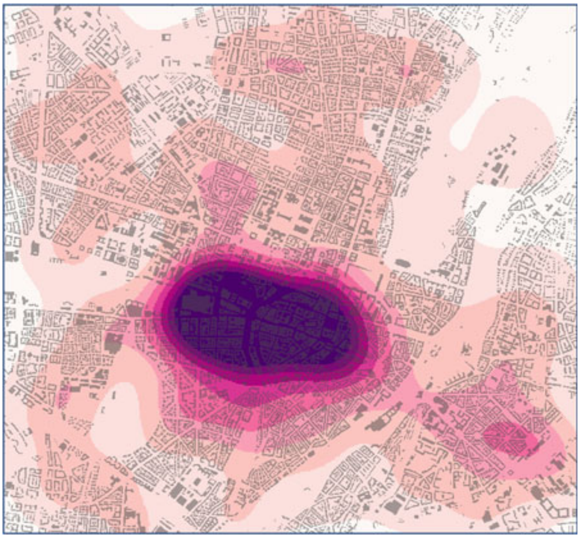

Fig. 1 Phone call densities in space

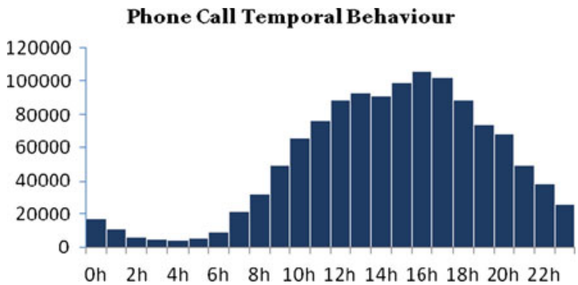

Fig. 2 A histogram showing phone call quantities over time

exploration of space *or* time rather than a visual exploration of space *and* time.

To enable the analyst to perceive for instance mobile phone call events as behaviour rather than multiple individual patterns, the analysis of all events must be presented in one display. This major principle for exploring data has been introduced for instance by Andrienko and Andrienko (2006) as "see the whole". For instance, a time-based animation for the exploration of spatiotemporal data does not fulfil this demand as different points in time can only be analysed in sequence. By the use of a temporal third dimension in a static view, time steps and positions can be explored according to the underlying scales without temporal limitations (Müller, Schumann 2003). The in a 3D presentation occurring handicaps such as occlusion and lost information on back faces can be tackled by advanced interaction techniques or additional cues (Aigner et al. 2007).

2 Input Data and Applied Method

2.1 Input Data

The in this work introduced visual analytics tool is applied on a mobile phone location dataset provided by Vodafone. For this dataset all outgoing Vodafone mobile phone calls were stored taken from an area approximately 7×7 km centred over Munich city centre during one week. Every mobile phone call was logged and assigned to the geographic location of a mobile phone cell. A resulting 1.5 million events occurred in 216 unique cells. For privacy issues all records were stored anonymously by removing the formal identifiers, such as phone number, phone ID, etc.

2.2 Applied Method

The coordinate pairs of the mobile phone base stations set the two dimensional distribution for all events. On the basis of the Space-Time-Cube concept, which was introduced by Hägerstrand (1970), the phone cells are presented in 2d space (along the x- and y-axis) and the height represents time (z-axis). Therefore, the time attribute of every event is used as the third dimension to enable a holistic space-time analysis. In this way the mobile phone call events can be defined as points in a 3D scatter plot.

Only if analytical methods are applied to compute expressive abstractions it is possible to analyse such large data sets efficiently (Aigner et al. 2007). In this case 1.5 million phoning events are generalised by a kernel density estimation, which is performed based on a formula stated by Scott (1992):

$$\widehat{f}_h(x) = \frac{1}{n \cdot h} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \quad (1)$$

With:

$\widehat{f}_h(x)$ = general Kernel Density
K = Kernel function
h = Kernel radius (bandwidth)
n = Number of points within kernel distance
$x_1, x_2 \ldots, x_n$ = points within the Kernel.

The kernel density function estimates a probability density function. This function is standardised by the number of points so that the integral of every probability density function is always equal to one. The kernel density function therefore represents relative point frequencies. As the density estimation is to be performed on the time axis over every base station, the density function has to be modified, to represent absolute (not relative) quantities. The in this work applied function (2) represents the absolute quantities of mobile phone call events, which allows visual comparison between different phone cells in sense of comparing absolute numbers.

$$q_h(z) = \sum_{i=1}^{N} K\left(\frac{z - z_i}{h}\right) \quad (2)$$

With:

$q_h(z)$ = Quantitative Kernel Density
K = Kernel function
h = Kernel radius (bandwidth)
$z_1, z_2 \ldots, z_n$ = points/events on the time axis within the kernel

A quantitative density function is estimated for all phone cells. It is necessary to use an identical kernel for all phone cells in order not to bias the comparison between different phone cells. A 1 h wide Gaussian kernel was used to calculate a density functions for every phone cell. These density functions are then used as rotating plane curves around z-axes perpendicular upon the two dimensional space. Each z-axis origins at a base station location so that a resulting 216 solid of revolutions represent the 3D symbols indicating the mobile phone call events in space and time. This statistical model indicates the amount of phone calls by its radius, which is the distance from time axis. The time value is given by the distance from the z-axis origin. A transverse section in x–y direction reveals the quantity of events at a certain time. This solid of revolution symbol is like an infinite number of two dimensional proportional symbol maps (one for every point in time) layered over each other.

The solid of revolutions are implemented into a Keyhole Markup Language (KML) file and rendered by the popular earth viewer Google Earth. A basic example of a solid of revolution is shown on Fig. 3. A Lambertian shading model is included to act as a diffuse reflecting texture over the solid figure and to enhance the 3D perception.

The solid of revolution persists of independently time coded sections. This empowers the user to visualise a certain time interval only. The earth viewer provides a comprehensive set of navigation tools to pan and zoom into every 3D location and to freely change the direction of sight. Therefore each solid figure of the model can be visually explored in detail. The through Google Earth supplied satellite imagery and aerial photography of the study area provides the users' orientation.

Google Earth features a time slider from Google Earth 5 and later. Whenever geodata containing time information is loaded into the viewer a time line function window called the time slider appears. The time slider enables a user to display spatio-temporal data of a specific point in time or to display a time interval. The user can also animate spatio-temporal data by playing the information in a visual sequence. In the case of the mobile phone events the user is given the power to

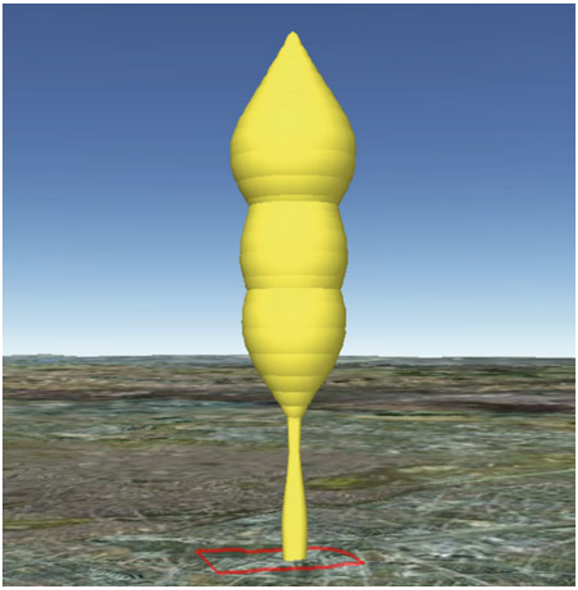

Fig. 3 A solid of revolution symbolising the mobile phone call events of a phone cell during one day, rendered in Google Earth

visualise the phone call densities at a specific day time (i.e. from 08.00 to 10.00). The user can then animate the model at his desired speed to highlight the phenomenon's density changes over the course of the day. The time interval settings stay hereby clearly in the user's visual field.

The assigning of day times to density values on the time axis is enhanced for the final visualisation by a bipolar colour scheme to identify night times in dark orange colours, early mornings in lighter orange over to a neutral light yellow colour for lunch time to a light red in the afternoons and dark red colours for late evenings (Fig. 4). This bipolar colour ramp is derived from the Lab colour space so that equal time differences are perceived as equal along the entire time axis.

3 Results

The resulting model for analysis of mobile phone events can be seen on Figs. 4 and 5. To symbolize the whole phenomena simultaneously in space and time integrated in one single model, space and time are treated as equal dimensions. The time axis is set as third dimension orthogonal to a plane two dimensional space. This Space-Time-Cube concept to represent multidimensional data has been applied by researchers prior to this approach. Tominski et al (2005) introduced the concept of 3D icons on a map display for representing spatio-temporal data. Better known as the pencil and helix icons this concept indeed has the capability to visualise higher dimensional data on distinct side surfaces of the icons based on the

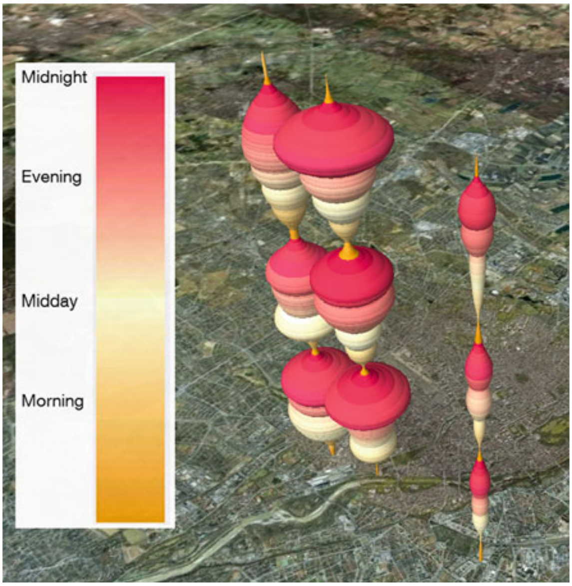

Fig. 4 Visualization of mobile phone call quantities during 3 days

visual variable colour, but has the handicap of information loss due to occlusion and hidden surfaces. Similar is the approach of Forlines and Wittenburg (2010) in which the multi-dimensional visualisation is based on the extrusion of a 2D radar chart. Here the visual variable size is used to indicate various quantitative variables in a specified horizontal angle. The in this work used solid of revolution symbols naturally show the same colour coding and convexity on every horizontal angle and therefore minimize the information loss due to occlusion and reduces the user′s cognitive workload. Also, it ensures that the general survey of the holistic spatiotemporal phenomenon remains allocated in every angle of view.

Within this approach the radii and accordingly the convexities of the solid of revolutions are derived by a kernel function from the temporal variable of occurring events. In this way it is ensured that time is treated as continuous. This is the major difference to Thakur and Ryne′s (2009) *data vases*. Although the appearance of the 3d data vases, in which polygonal disks are stacked for each time step, is alike to the here shown solids of revolution, the data vases are derived from independent values in time such as in a histogram. This has the benefit of simplicity in terms of the symbol setting but has a lack of continuity. The solids of revolution are derived from a rotating probability density function that estimates the distribution of a continuous valued random variable. In this case the variable is the amount of phone calls scattered on the time line. This avoids edge effects between different classes (for instance one hour classes) simply because no intervals are classified on the time line. The complete phenomenon is treated in respect to its continuous nature. The by the kernel inflicted data smoothing also guarantees individual events in space-time areas of sparse phone calls do not

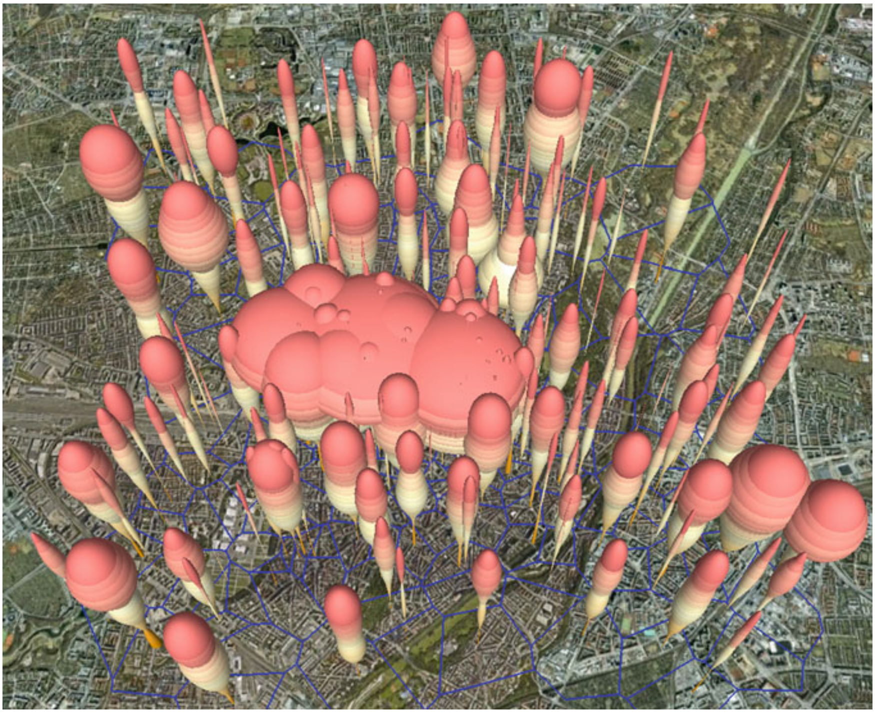

Fig. 5 Visualization of mobile phone call densities of one day until 6:00 p.m

interact strongly with the estimation. Outliers have a minimal influence on the solids of revolutions diameter. The kernel density estimation ensures that the major quantitative changes in time remain clearly visible.

One could argue that the kernel density estimation should not have only been executed on the time axis but also between phone cells in two dimensional space. But it has to be kept in mind that the events where collected through mobile phone base stations. Despite the clear boundary visualisation of cells by using Voronoi diagrams, the location of a phone call can only be determined by the coordinates of the connecting base station. The events are therefore only conceptual spatial points. O′Sullivan and Unwin (2003) state that one important criterion for using point pattern analysis (and therefore kernel density estimation) is to use true locations of event data. They further state that the points must be true incidents with real spatial coordinates. For this reason every phone cell is treated separately.

The overall analysis of the mobile phone call activities in Munich is shown on Fig. 5. A very high density can be identified in the middle of the study area which is the city centre. The perhaps not surprising phenomenon of a high mobile phone activity in city centres has been surveyed before, for instance by Ahas et al. (2010). But it is also possible to detect minor hotspots in other areas, most of them in the afternoon. At the same time regions with low mobile phone traffic can be seen.

The by common sense estimated daily cyclic behaviour of the phenomena can be better identified at a more tilted angle as depicted in Fig. 4. Peaks and valleys occur at nearly all base stations at similar times.

The mobile phone call events are symbolized by solids of revolution. The mobile phone call densities are indicated by the radius (or thickness), which is equivalent to the visual variable size. The temporal variable is indicated by a colour scheme driven by the visual variables brightness and hue. Size is classed as a quantitative visual variable by Bertin (1967/1984). Hence, the analyst is able to visually estimate the density from the solid of revolution′s radius. The visual variables brightness and hue plus saturation produce the colour. Colour empowers the observer to order (Green 1998). This insists the analyst can visually classify day times into a temporal order.

This 3D visualisation model can be used for various spatio-temporal phenomena. However, the space-time-cube based visualisation becomes limited when dealing with long period spatio-temporal phenomena. By extending the time line largely into the third dimension the resulting symbols have a great height. The model will appear oversized to the user and the overall view will be inhibited. A further problem would be the assignment of oversized symbols to their location. When dealing with long period spatio-temporal phenomena a solution to overcome this problem is to reduce the time/space ratio. As the time line has a spatial extent the vertical time scale in proportion to the horizontal space scale plays an important role on the model′s appearance. By inducing the model with a reduced time/space ratio the spatial extent of the symbols in the third dimension will be shortened to prevent the oversize symbols and to once again enable the overall view. The solids of revolution will be visualised in a more compact form. This could lead in some cases to surface artefacts simply for the reason that too much information is given on a spatially short timeline. The likeliness of overburdening the user′s visual estimation ability is high. The user would probably percept the diameter changes as noise. The decrease of the time/space ratio has therefore an impact on the choice of the kernel bandwidth. The choice of the bandwidth affects the probability density estimation strongly. The higher the time/space ratio is, the higher the kernel bandwidth must be. A larger bandwidth results into smoother changes in time. The visualisation model is more generalized with easier readable solids of revolution. Minor phenomenon changes in time are sacrificed to enable that the major phenomenon changes in time are well readable.

The Google Earth virtual globe provides the reference map to assign distinct places in Munich to every solid of revolution. Focusing on data subsets is enabled with the navigational functions onto any specific detail of the model. The analyst can freely move, view, zoom, tilt and rotate the current view to analyse the model from every angle and distance. With assistance of Google Earth′s time slider the course of the day can even be animated in this 3D environment. These tools improve the legibility immensely as the differences between neighbouring mobile phone cells can be compared more easily and the line-of-sight obstruction is minimized. The 3d visualisation exhibits its potency especially in approximately horizontal view in which identical time sections of different locations are all in one plane level and are therefore easily compared.

Every solid of revolution symbol is assigned to its location by the symbol root which is fixed to the earth surface. When using the time slider to focus on single time intervals this may not be the case. Then the symbols float at a certain height over the base map which constrains the user′s ability to identify its true location. Only tilting the model into a vertical view ensures that the user can assign a chosen time interval and symbol with its location. Here again it becomes definitive that this 3D dimensional spatio-temporal visualisation model is practicable for visual data exploration and the general overall view of a phenomenon.

Another benefit of the symbolisation method is that the solid of revolution′s smooth rotating curve respects the continuous nature of time. In addition, due to the natural geometry of a solid of revolution figure representing a mobile phone cell the horizontal angle of sight has no effect on the visual estimation. This also clears the back faces problem. Depth perception cues ensure the visual estimation of mobile phone call quantities. Especially the zooming, tilting and panning functions enable the user to obtain a 3D impression of the model by motion. Motion in combination with, the in this model offered, shading significantly reduces the errors of 3d structural information perception (Norman et al. 1995). As a consequence increases, decreases, stability and fluctuations of mobile phone calls can clearly be identified in spatiotemporal relation.

4 Conclusion

It has become easier in recent years to visualise and to disseminate geodata due to the emergence of virtual globes being able to run smoothly on various ubiquitous electronic devices. Either a standalone geographical information programme which provides the virtual globe or simply a browser plugin, which is for instance offered by Google, provides base map material for a thematic layer in a 3D virtual environment. Supported by XML notations such as KML it is possible to visualise spatiotemporal in a 3D view. The user of this visualisation can be either the professional analyst or an ordinary person. The earth viewers provide an easy to use time slider and a comprehensive set of navigation tools. To make spatio-temporal data understandable one can use all provided dimensions to enable a complete spatiotemporal analysis.

The in this work introduced visualization method ensures that the complete phenomenon can be visually examined in one view. All major and minor spatio-temporal hotspots (or coldspots) are clearly visible. The analyst can freely navigate within the phenomena to explore in detail any given location in space and time.

When visualising mass data an abstraction is essential to preserve the legibility. Kernel density estimation can generalise mass event data consisting of point clouds. This makes events displayable in a usable manner. Further research has to extend kernel density estimation applications for spatiotemporal and 3D data and should evaluate the usability of such 3D visualisations.

References

Ahas R, Silm S et al (2010) Using mobile positioning data to model locations meaningful to users of mobile phones. J urban technol 17:25

Aigner W, Miksch S et al (2007) Visualizing time-oriented data: a systematic view. Comput Graphics 31(3):17

Andrienko N, Andrienko G (2006) Exploratory analysis of spatial and temporal data: a systematic approach. Springer, Berlin

Bertin J (1967/1984) Semiology of graphics: diagrams, networks, maps. University of Wisconsin Press, Madison

Forlines C, Wittenburg K (2010) Wakame: sense making of multi-dimensional spatial-temporal data. International conference on advanced visual interfaces, ACM Press, Roma

Green M (1998) Toward a perceptual science of multidimensional data visualization: bertin and beyond. http://graphics.stanford.edu/courses/cs448b-06-winter/papers/Green_Towards.pdf, Retrieved Access 6 May 2011

Hägerstrand T (1970) What about People in Regional Science? Pap Reg Sci Assoc 24:15

Haggett P (1990) The Geographer‘s Art. Blackwell, Massachusetts

Keim D, Kohlhammer J et al (eds) (2010) Mastering the information age—solving problems with visual analytics. Druckhaus “Thomas Müntzer” GmbH, Bad Langensalza, Eurographics Association

Müller W, Schumann H (2003) Visualization methods for time-dependent data: an overview. In: Simulation Conference 2003. Proceedings of the 2003 Winter

Norman JF, Todd JT et al (1995) The perception of surface orientation from multiple sources of optical information. Percept Psychophys 57(5):8

O′Sullivan D, Unwin DJ (2003) Geographic Information Analysis. Wiley, Hoboken

Scott DW (1992) Multivariate density estimation: theory, practice, and visualization. Wiley, Canada

Thakur S, Rhyne TM (2009) Data vases: 2D and 3D plots for visualizing multiple time series. In: 5th international symposium on advances in visual computing: part II Las Vegas, Springer, Berlin

Thomas JJ, Cook KA eds (2005) Illuminating the path: the research and development agenda for visual analytics, IEEE

Tominski C, Schulze-Wollgast P et al (2005) 3D Information visualization for time dependent data on maps. In: 9th international conference on information visualisation, London, UK