Visualizing Uncertainty of Point Phenomena by Redesigned Error Ellipses

Christian E. Murphy[a]

[a] *Chair of Cartography, Technische Universität München, Arcisstraße 21, 80333 Munich, Germany; christian.murphy@tum.de*

Abstract: Visualizing uncertainty remains one of the great challenges in modern cartography. There is no overarching strategy to display the nature of uncertainty, as an effective and efficient visualization depends, besides on the spatial data feature type, heavily on the type of uncertainty. This work presents a design strategy to visualize uncertainty connected to point features. The error ellipse, well-known from mathematical statistics, is adapted to display the uncertainty of point information originating from spatial generalization. Modified designs of the error ellipse show the potential of quantitative and qualitative symbolization and simultaneous point based uncertainty symbolization. The user can intuitively depict the centers of gravity, the major orientation of the point arrays as well as estimate the ex-tents and possible spatial distributions of multiple point phenomena. The error ellipse represents uncertainty in an intuitive way, particularly suitable for laymen. Furthermore it is shown how applicable an adapted design of the er-ror ellipse is to display the uncertainty of point features originating from incomplete data. The suitability of the error ellipse to display the uncertainty of point information is demonstrated within two showcases: (1) the analysis of formations of association football players, and (2) uncertain positioning of events on maps for the media.

Keywords: Uncertainty Visualization, Statistical Graphics, Generalized Point Phenomena

1. Introduction

Geo-sensor networks generate more and more geo data, and a large proportion are collected as point features, i.e. as positional records. Spatial generalization of big geo data is the only way to visualize point features in a reasona-ble way. Information is decreased in order to enhance visual interpretability. In this process, cartographic generali-zation techniques will omit, simplify, displace or smooth spatial data. In many visualization tasks it is necessary not only to show averaged positions, but also to indicate the real distribution of the point phenomena. Representing the distribution of large point clouds in two-dimensional space via kernel density estimation (KDE) as an isarithmic map has become very popular. Though, KDE cannot fulfil all point cloud visualization tasks. When several varia-bles, in other words, multiple categories of point clouds are to be visualized, the continuous isarithmic depiction must be replaced by a discrete visualization tool. This can be viewed as a generalization step. Discrete point sym-bols may be able to show a characteristic value of the phenomena, like an average value, but point symbols cannot show the complete phenomena. A positional uncertainty of the true distribution is the consequence.

Error ellipses are a graphical, statistical tool to illustrate the correlation of two connected variables. They display maximum errors or confidence regions in a two-dimensional space. By viewing the sizes, the shapes as well as the orientations of error ellipses the user can quickly visually estimate the variances and covariance of two variables (Ghilani and Wolf, 2006). In surveying, error ellipses are used to show zones of uncertainty in order to visually com-pare the quality of surveying points within a geodetic network (Welsch et. al., 2000). Here, the ellipses show critical standard deviations in x and y directions based on confidence regions. The surveying engineer can instantly read from the error ellipses which surveying points have been measured with a high geometric accuracy and which sur-veying points are more inaccurate. The ellipses also show difference of qualities dependent on orientation. Using el-lipses to show uncertainties in cartography is surely not a new concept. In the 19th century Tissot (1881) described his distortion ellipses, known as Tissot's indicatrix. These ellipses are a visual tool to characterize local distortions of a map projection. The concept of ellipse use in geodetic networks and for map projection distortions can be easily transferred to the representation of discrete positional uncertainty. Buttenfield and Ganter (1990) made suggestions for visualizing cartographic metadata based on five types of data quality. Within their framework they state that the positional accuracy of discrete point data is appropriately visualized by size and shape, naming the error ellipse. Here, the author would like to emphasize his astonishment how so few applications in cartography are used to indi-cate positional uncertainty by the error ellipse, when cartographers could be more aware of this possibility.

2. 2. Principle of the Error Ellipses

The covariance of random variables is statistically specified within the covariance matrix. When the variance of a distribution in two-dimensional space is to be described, a 2 x 2 matrix is necessary. This covariance matrix fea-tures the variances as well as the covariances between the variables x and y.
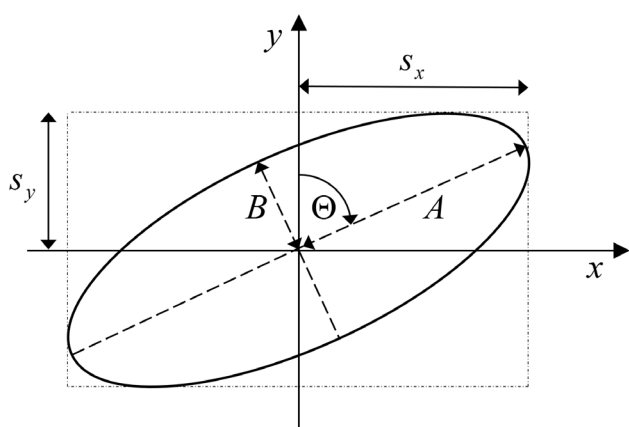
Fig. 1. Main values and geometric construction of the standard error ellipse

One can graphically show how the standard deviation and correlation between the two variables interact with an error ellipse. The principle of the error ellipse is drawn on Fig. 1. The large semiaxis A and the small semiaxis B set the lengths and size of the estimated error. The angle (formula 4) sets the orientation of the error ellipse. To define the error ellipse, the elements are calculated in the following four formulas (Pelzer, 1997), with the large semiaxis:

$$A^2 = \frac{1}{2}\left(s_x^2 + s_y^2 + w\right) \quad (1)$$

The small semi-axis as:

$$B^2 = \frac{1}{2}\left(s_x^2 + s_y^2 - w\right) \quad (2)$$

Using the following auxiliary quantity:

$$w^2 = \left(s_x^2 - s_y^2\right)^2 + 4s_{xy}^2 \quad (3)$$

And the angle of the error ellipse:

$$\Theta = \frac{1}{2}\arctan\left(\frac{2s_{xy}}{s_x^2 - s_y^2}\right) \quad (4)$$

It should be noted, that the lengths of the semiaxes and therefore the size of the error ellipse depends heavily on the chosen probability value.

3. Case study: the Analysis of Formations of Association Football Players

The suitability of the error ellipse to display the uncertainty of point information is demonstrated by using a positional sports dataset from association football. This association football dataset was collected at the Swiss

Federal Institute of Sport Magglingen (SFISM). Therefore, a local positioning system was installed around a football pitch to collect positions of all players during a football test match. The Inmotiotec local positioning system (Abatec group, 2012) features transponders for all players and collects data by gyroscope, compass and acceleration sensor. Base stations that are arranged around the pitch in order to receive the transponder signals for Wi-Fi positioning complete the surveying setup. The extracted values of the transponders as well as the Wi-Fi positioning measurements are the basis for deriving the local (x and y) coordinates. The spatial resolution in this data set is on cm-level, whereas the temporal frequency of positional data is one tenth of a second. The resulting dataset has ten time stamps for every second containing x and x coordinates of each player as well as the ball.
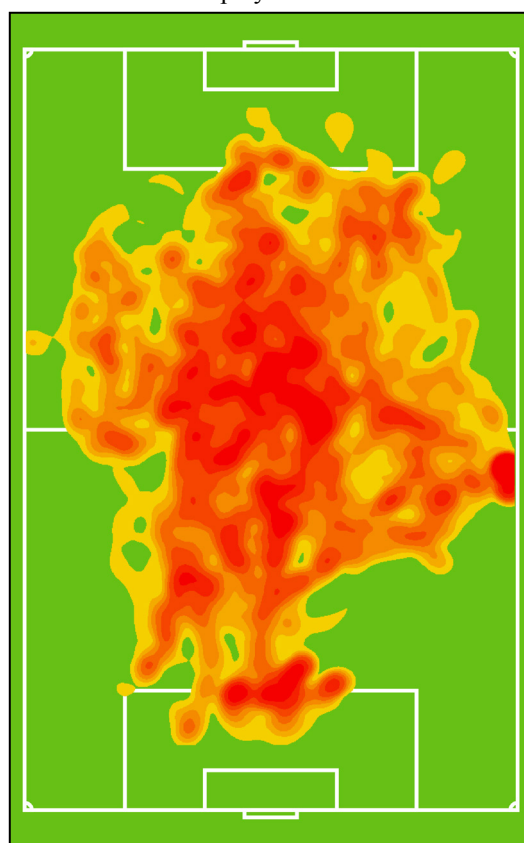

Fig. 2. Heat map symbolization for a single player in the 1st half of the recorded test association football match

A tactical insight can be given by movement analysis of the players. It is possible to create kernel density estimations over all time stamps and present these either as isarithmic maps or as gridded choropleth maps (in terms of sports analytics known as heatmaps) for single players. Fig. 2 shows an example heat map.

Although, these heatmaps give a nice overview to one player's movement, it misses context as the heatmap cannot represent multiple players at the same time. In other words, isarithmic maps have limitations when multiple distribution categories are to be presented. This makes comparisons of players as well as overall tactical analysis very difficult. For this reason, it is custom to

give an overview to the players' movement during a football match by presenting the average positions. A typical graphical representation used in sports analytics as well as in media shows the location of a players' average position on the field (Opta, 2017). Based on the recorded positions of individual players during the test match, Fig. 3 visualizes the average positions of one team.
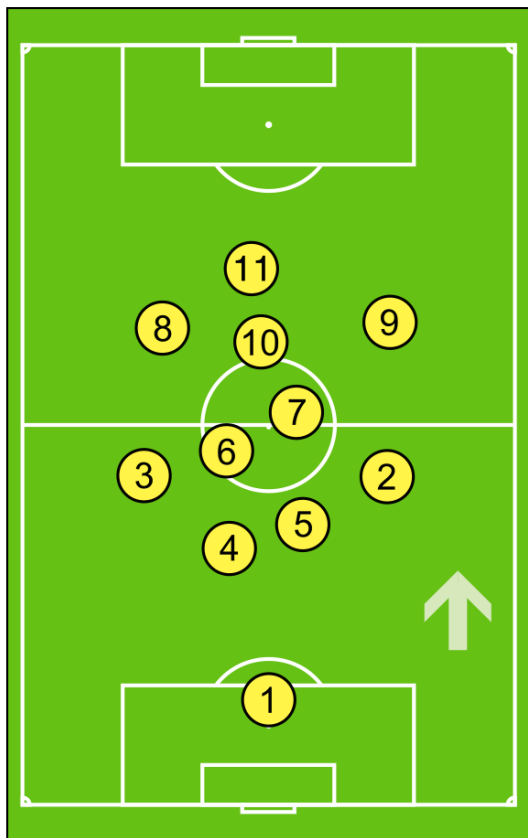

Fig. 3. Average position symbolization for the players of one team in the 1st half of the recorded test association football match

In Fig. 3 any user with a little insight into association football can read the players' roles as well as the teams' setup formation. It becomes understandable i.e. which players have a more defensive or advanced role, which players are the more wide players, as well as which players have a more isolated position. Though, much infor-mation concerning the movement of individual players remains concealed. The crucial information that is lost by solely showing average positions are:

- Major running directions
- Size of the operating area
- Overall running performance
- Hotspots of players' positioning

The error ellipse can effectively counteract the loss of this crucial information by maintaining the info graphics' average positioning style. By applying the error ellipse formulas (see section 2) individually to the time stamps of all players the info graphic in Fig. 4. can be produced. Hereby, the chi-squared distribution with two degrees of freedom was used to set confidence regions based on the variance of the x and y coordinates.
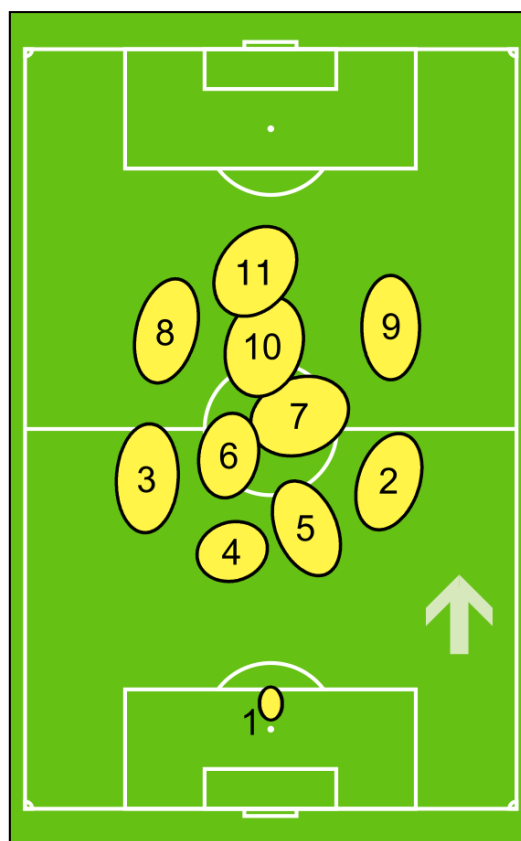

Fig. 4. Error ellipse symbolization for the players of one team in the 1st half of the recorded test association football match

Like in Fig. 3, in Fig. 4 it is still possible to visually estimate the centroids of the symbols and to derive information based on the teams' average positions. But here, the error ellipses reveal a lot of additional information. The error ellipses indicate the uncertainty of average positions by visually encoding in-depth statistics as a compact distributional summary. Even though, hotspots of individual players remain concealed, the degree of uncertainty shown reveals the players' operating range. I.e. in Fig. 3, the goalkeeper (nr. 1) has, expectably, the smallest running performance, whereas the operating areas of the midfield players nr. 7 and nr. 10 are very large comparing to other players of the team. Also, the major running directions become readable. As i.e. the players nr. 3 and nr. 9 run and operate in a more vertical routes, the central defender nr. 4 plays in more sideways areas. Many other findings based on major running directions, the size of the operating area, and the overall running performance can be intuitively made. The size of the error ellipses is dependent on the size of the confidence region. The higher the probability value the larger will the error ellipses be drawn. On the one hand the confidence regions need to have a well interpretable size, but on the other hand an error ellipse must not obscure important shapes of other ellipses. The probability value is adjusted by visual checking. In Fig. 3 the probability value is set to 0.3. That means, 30% of all time stamp coordinated of a player are located within the graphical ellipse. The absolute confidence region value is of less concern to the reader as the relative comparison is

much more important. Therefore, the probability value should be constant over a complete analysis series.

4. Case study: Uncertain Positioning of Events on Media Maps

Instead of producing an error ellipse symbol based on calculable geometric data, it is also possible to use it to vaguely indicate a positional uncertainty. Press reports are often accompanied by media maps as it is often crucial to know where an event has happened. There is a necessity to incorporate the communication of the localizing un-certainty in such media maps, which has been emphasized by Schiewe (2016). In many press report cases there is an uncertainty in the location of the event that is reported. The localizing of the event may have an error biased by direction. Then, it is required to communicate this uncertainty in an intuitive way. The example media map in Fig. 5 shows a landslide event in a typical manner for news agencies. At a certain place the road has been blocked by falling rocks. As the landslide blocking of the road is certain, the exact location of such instant press releases often have to be estimated.
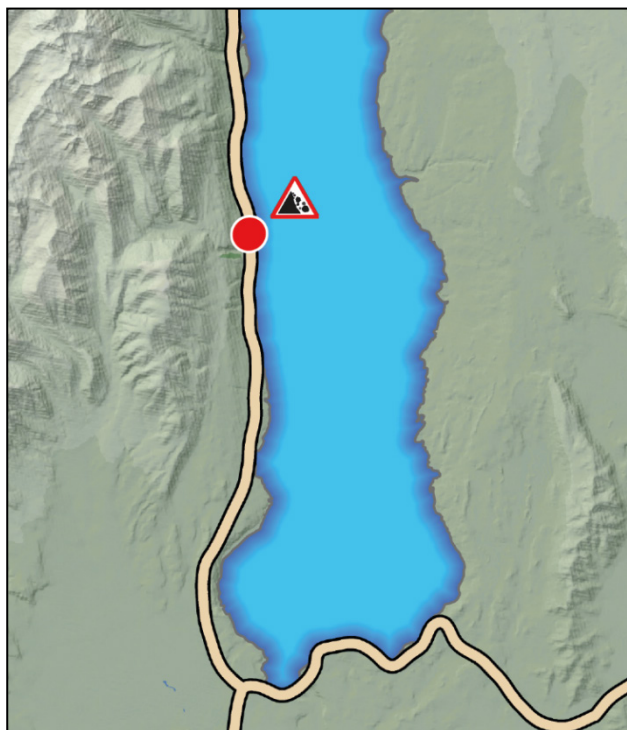

Fig. 5. Standard marking of an event on a media map

The location of the specific road will probably be well-known. This means, the positional accuracy of placing the event on the map will be depended on the angle. One can be sure that the event is on the street. Only the event's true position on the road is uncertain. Therefore, the uncertainty is larger in the directions of the further street course. In Fig. 6 it is shown how an error ellipse symbolization can communicate the greater positional uncertainty along the road. Here, the axes lengths of the error ellipse are not based on derived values. They are subjectively drawn to indicate the positional uncertainty. The user of this media map reads the, in comparison to Fig. 5, distorted symbol and can instantly assign the point

based event to a wider possible location area. In further user studies it has to be evaluated which majority amount of map readers can comprehend this visual metaphor of location uncertainty.
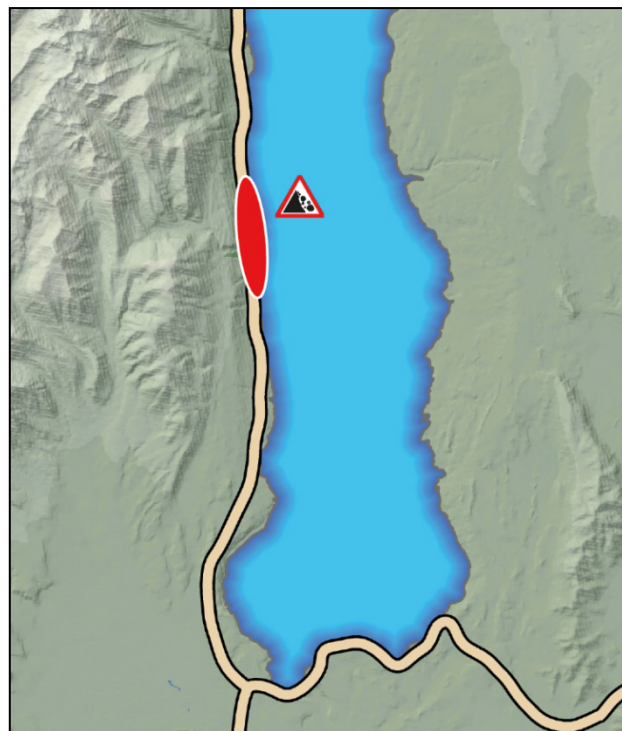

Fig. 6. Localization uncertainty indicated by using the error ellipse for an event on a media map

The graphical appearance of the error ellipses can be further developed by using visual variables appropriate for communicating uncertainty. The visual variables crispness, resolution and transparency have been assumed as being most useful for uncertainty representation (MacEachren, 1995). So it might seem obvious to graphically com-bine these variables with the error ellipse. A few years ago MacEachren et.al. (2012) presented now also empirical studies focused on uncertainty visualization. Among other valuable results, fuzziness was highlighted as a good visual variable to encode uncertainty, as higher fuzziness was intuitively perceived as more uncertain. Inspired by these findings is the adaption of the error ellipse from the previous Fig. 6. On the left image of Fig. 7 a radial fuzziness is applied to the error ellipse with the intention of showing possibilities of graphically modifying the error ellipse to enhance the understanding of localization uncertainty. This is complemented by another adaption, the radial transparency used on the right of Fig. 7.
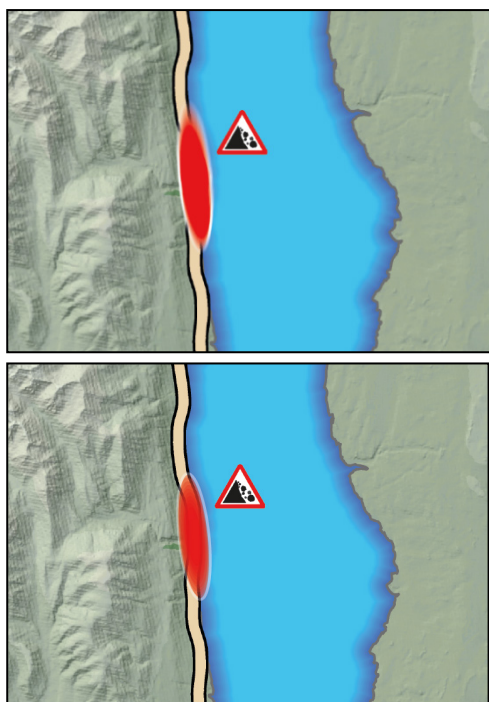
Fig. 7. Radial fuzziness (left) and radial transparency (right) applied to the error ellipse for intuitive understanding of uncertainty

5. Conclusions

The error ellipse can intuitively indicate positional uncertainties by visually encoding statistics as a compact distributional summary. It displays less detail than KDE, but it also takes up less graphical space and leaves room to show multiple distributions in a single visualization. The user can depict the centers of gravity, the major orientation of the point arrays as well as estimate the extents and possible spatial distributions of the point phenomena. The parameters of the error ellipse can either be mathematically derived from a distribution, or they are estimated by the map maker. This is dependent on the given task. When the ellipses are mathematically derived, the degree of probability has to be adjusted to the visualization in order to keep the semiaxes in a visually appropriate size that defies map symbol clutter. It has been shown that fusions of the ellipse with visual variables in context of uncertainty are quite possible. User studies are required to evaluate how intuitive and how effective the error ellipses communicate positional uncertainty are and if graphical modifications with the uncertainty variables enhance the effectiveness of the error ellipse concept. Although the author expects the error ellipse to be suitable for laymen it should also be defined for which user groups and which application fields the error ellipse symbolization is best.

6. References

Abatec group. (2012). inmotiotec Local Position Measurement. https://www.abatec-ag.com/en/inmotiotec. Accessed 17th of February 2017.

Buttenfield, B. P., & Ganter, J. H. (1990). Visualization and GIS: What should we see? What might we miss? Paper presented at the 4th International Symposium on Spatial Data Handling Zürich: University of Zürich: Department of Geography.

Ghilani, C. D., & Wolf, P. R. (2006). Adjustment Computations: Spatial Data Analysis (Fourth Edition ed.). Hoboken, New Jersey: John Wiley & Sons.

MacEachren, A. M. (1995). How Maps Work: Representation, Visualization, and Design. New York, USA: Guilford Publication.

MacEachren, A. M., O'Brien, R. E. R. a. J., Li, B., Swingley, D., & Gahegan, M. (2012). Visual Semiotics & Uncertainty Visualization: An Empirical Study. IEEE Transactions on Visualization and Computer Graphics, 18(12), 10.

Opta. (2017). Football Matchday Widget Showcase. http://www.optasports.com/football-matchday-widget-showcase.aspx. Accessed 27th of February, 2017.

Pelzer, H. (1997). Ausgleichsrechung. lecture notes. Universität Hannover.

Schiewe, J. (2016). Visualisierung von Unsicherheiten in Medienkarten. AGIT – Journal für Angewandte Geoinformatik, Wichmann-Verlag Berlin/Offenbach, 2-2016, 10.

Tissot, A. (1881). Mémoire sur la représentation des surfaces et les projections des cartes géographiques. Paris, Gauthier-Villars.

Welsch, W., Heunecke, O., & Kuhlmann, H. (2000). Auswertung geodätischer Überwachungsmessungen. Wichmann.