Assessing the Impact of Design Decisions on the Usability of Uncertainty Visualization: Noise Annotation Lines for the Visual Representation of Attribute Uncertainty

Christoph Kinkeldey*, Jennifer Mason, Alexander Klippel**, Jochen Schiewe***

* Lab for Geoinformatics and Geovisualization (g²lab), HafenCity University Hamburg, Germany
** GeoVISTA Center, Department of Geography, The Pennsylvania State University, USA

Abstract. We report findings from a web-based experiment on noise annotation lines, a method to represent attribute uncertainty. We tested and compared three design aspects of noise annotation lines and evaluated how different design variations influence user performance. We systematically varied the number of uncertainty categories, noise width, and noise grain. Our results show that the number of uncertainty categories significantly influences user performance but that certain design characteristics can counterbalance the negative effect of an increased number of categories. Additionally, we were able to show that performance decreases if uncertainty changes continuously.

Keywords: uncertainty; attribute uncertainty; visualization; noise annotation lines; evaluation; thematic map

1. Introduction

Visualization is a common means to represent uncertainty in spatio-temporal data. A variety of methods for this purpose exists, especially in the area of GIScience and scientific visualization (Brodlie et al. 2012, Slocum 2003). Different types of uncertainty are often identified (geometric, attribute, temporal, or combinations) and displayed in different ways (e.g., static/dynamic, integrated view/adjacent view, interactive/non-interactive). Several typologies were created to support the choice of a suitable method for different purposes (MacEachren et al. 2005, Senaratne et al. 2012).

With regards to integrated views a basic distinction between intrinsic and extrinsic approaches can be made (Gershon 1998). Intrinsic approaches utilize visual variables from existing objects in the visualization to represent uncertainty, mostly including visual variables derived in cartography. Besides the seven visual variables described by Bertin (1983), other variables including color saturation, symbol focus, and clarity were added (MacEachren 1992, MacGranaghan 1993). Extrinsic approaches, on the other hand, incorporate additional graphical objects to represent uncertainty, e.g. glyphs (Pang 2001) or other objects such as bars or dials that are added to the display. Unlike most intrinsic variables, they can be visually separated from the content.

Although a variety of methods for representing uncertainty have been implemented, most of them have not or have only partially been evaluated with regards to aspects of usability. Moreover, most of the studies focus on intrinsic methods.

In this paper we evaluate an extrinsic method we call *noise annotation lines*, a method first described as "procedural annotations" by Cedilnik and Rheingans (2000). It is a promising way to display attribute uncertainty in maps with heterogeneous geometries, e.g. land cover maps, and like most uncertainty visualization methods it has barely been evaluated. An exception is a qualitative evaluation by Zuk and Carpendale (2006), who evaluated procedural annotations from a theoretical standpoint using heuristics from the theories by Bertin, Tufte, and Ware. They point out that the data-ink ratio of a noise grid is relatively small and they hypothesize that with other annotation types more uncertainty categories would be possible to represent. They do not provide evidence for this but they remark that a more formal testing of procedural annotations, especially concerning perceptual aspects, would be interesting.

This work contributes to the evaluation of extrinsic uncertainty visualization methods by testing usability aspects of noise annotation lines, focusing on the impact of design factors on the usability of the method.

2. Uncertainty Visualization Using Noise Annotation Lines

Many thematic maps (e.g., land cover maps) contain objects of high geometric variability, that is, objects that differ considerably in size and shape. Representing uncertainty integrated into such maps can lead to cluttered displays. Most commonly used uncertainty visualization methods such as color saturation or transparency work well with uniform areas but become harder to perceive for areas that are geometrically diverse. Extrinsic methods, especially those based on uniform grids, seem promising because they are independent of the underlying geometry.

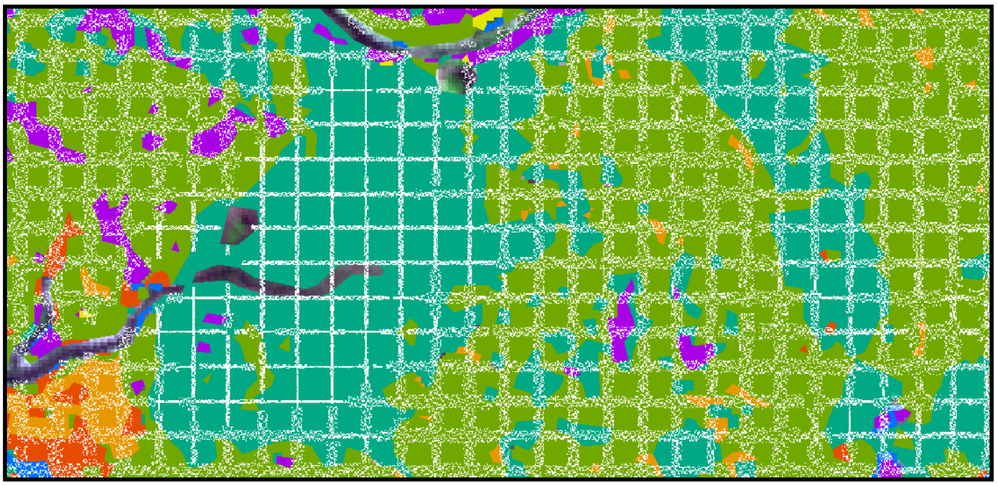

Figure 1. Noise annotation lines representing classification uncertainty on a vegetation land cover map. The local width of the noise grid indicates the degree of uncertainty (larger noise width = higher degree of uncertainty).

Using noise annotation lines, a regular grid is placed onto the map and is distorted locally to represent the degree of uncertainty. While the authors of the original paper proposed four different versions of annotations (variation of width, sharpness, noise and amplitude) we decided to implement and evaluate the noise grid because we expected noise to be a suitable metaphor for uncertainty. In a qualitative pre-test people found the noise display particularly intuitive.

Noise annotation lines consist of a noise grid that is varied in size locally to represent the level of uncertainty (the more uncertain, the greater the width) whereas the number of noise particles and their size (“grain”) remains constant. The grid representation generalizes the original uncertainty distribution as only the values covered by the lines are represented. However, since the size of the grid cells can vary according to the scale of the map, a compromise can be made between maximum coverage of uncertainty information and minimum occlusion of the underlying content. This characteristic makes this method promising for use in maps.

3. Methods

An experiment was conducted to learn more about the usability of noise annotation lines as uncertainty visualization technique. The central aspect was the impact of different designs of the noise grid on the effectiveness and efficiency of the display.

3.1. Research questions

Our goal was to answer the following research questions:

1. How do different design parameters impact the effectiveness and efficiency of noise annotation lines as a representation of attribute uncertainty in a thematic map?
2. How does the maximum number of categories affect user performance?
3. Can users accurately compare the overall degree of uncertainty between two defined areas when the values vary within the areas ("continuous" uncertainty)?

3.2. Independent variables

The appearance of noise annotation lines can be changed by altering different design parameters. Our hypothesis was that these changes may have an impact on the effectiveness and efficiency of the uncertainty display. The following three major design parameters were chosen as factors: number of categories of uncertainty, the width, and the grain of the noise grid (Table 1).

The width of the noise grid is defined for the display of maximum uncertainty (100%) with respect to the size of the grid cells (Figure 2). A lower noise width covers a smaller area with noise, however there is also less space to represent varying uncertainty values. Consequently, the choice of this parameter is a compromise between interference with the underlying content and the number of categories that are discernible. If the grid width is too large, the gaps between the noise lines become so small that within highly uncertain areas you cannot see the grid structure anymore. This already occurs with a noise width of 60% of the grid cell size. On the other hand, a width of less than 40% does not seem to be suitable to represent more than three categories of uncertainty (as revealed by a pre-test). Thus, we chose 40% and 50% of the grid cell size as levels of this factor.

The grain of the noise particles is the second design parameter we are manipulating (Figure 3). A finer grid consists of more and smaller particles, while a coarser grid contains less particles that are larger. Since we kept a constant resolution across all maps in this study (see section 3.4) we implemented 1x1 pixels and 2x2 pixels for the factor “grain”.

The number of uncertainty categories is varied as the third factor. With more categories the difference between two adjacent values decreases from 33% to 20% uncertainty (Table 2). We chose a minimum of four categories because a pre-test revealed that three categories (i.e., 0%, 50%, and 100% uncertainty) are straightforward to discriminate in contrast to four categories which already lead to errors. Five and six categories seemed to be more challenging, so we hypothesized that user performance would decrease in comparison to four categories.

Factor	Number of levels	Levels
Noise width	2	Small (40%), Large (50%)
Noise grain	2	Fine (1x1), coarse (2x2)
Uncertainty categories	3	4, 5, 6 categories

Table 1. Factors used in the study.

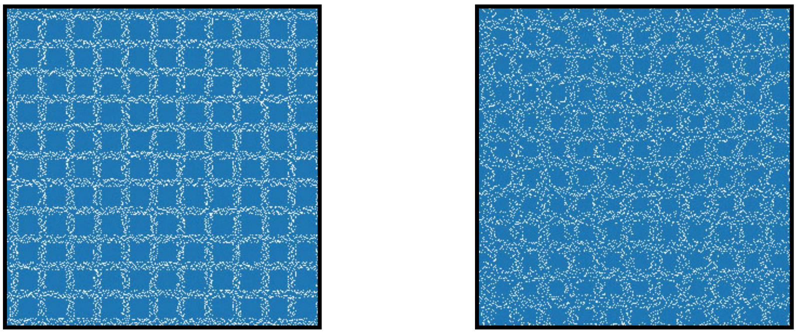

Figure 2. Design parameter “Noise width”. Both grids represent the same degree of uncertainty (100%), but with different widths: 40% (left) and 50% (right) of the grid cell size.

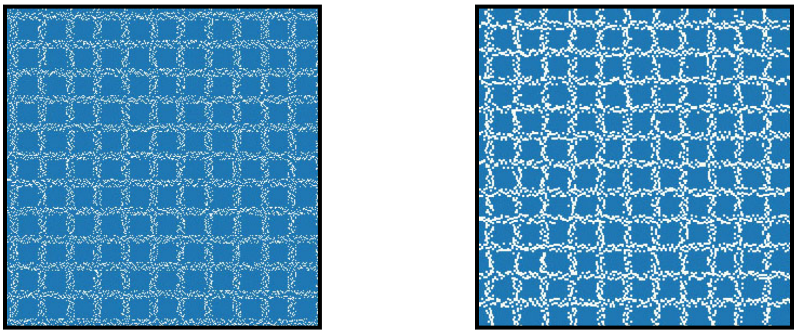

Figure 3. Design parameter “Noise grain”. Both grids represent the same degree of uncertainty (100%), but with different grain sizes: “fine” (left) and “coarse” (right).

	Step	Levels
4 categories	33%	0%, 33%, 66%, 100%
5 categories	25%	0% 25%, 50%, 75%, 100%
6 categories	20%	0%, 20%, 40%, 60%, 80%, 100%

Table 2. Categories of uncertainty used in the study.

3.3. Tasks

For the main part of the study we chose the following task: A comparison of uncertainty categories between two defined areas in the map. This approach allowed us to compare a total of 150 maps (see below). Still, this is a realistic task, e.g., when analyzing a land cover map: A qualitative, pair-wise comparison of different objects from the same class regarding their uncertainty. The survey question remained the same for all maps: “Which area is more uncertain?” Potential answers included “A is more uncertain”, “B is more uncertain”, “A and B are equal” and “I can’t tell”. As the questions were mandatory the latter answer was included to minimize nonsense answers when participants could not read the map.

3.4. Maps

We created 10 maps per factor combination to establish 10 repetitions. All maps were taken from the same vegetation land cover map representing equally sized areas (100 m x 100 m). Furthermore, the size of the noise grid cells in all maps was the same. The maps showed two kinds of uncertainty: Discrete and continuous. “Discrete” maps show a constant uncertainty value for each map object whereas “continuous” maps varied in value within each area (Figure 4).

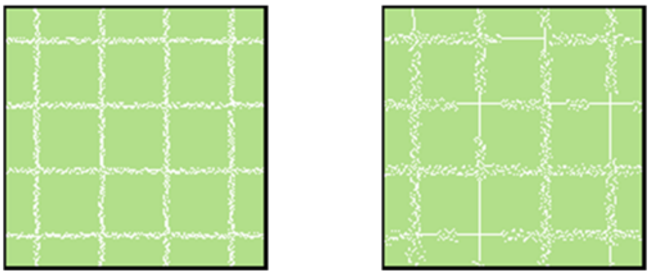

Figure 4. Discrete (left) and continuous (right) uncertainty distribution.

In addition, we varied the background colors according to a qualitative color scheme recommended by ColorBrewer (http://colorbrewer.org/). The utilized color scheme (“Paired”) is indicated to be colorblind-safe and laptop-/LCD-friendly. In each map, two square areas in the size of 3 x 3 noise grid cells were drawn on areas of the same color and labeled ‘A’ and ‘B’ (Figure 5). We placed the squares A and B on areas with the same background color, either light blue or light green. Those two colors have a very similar contrast distance from the white color of the grid so we vary the color but not the contrast between grid and background.

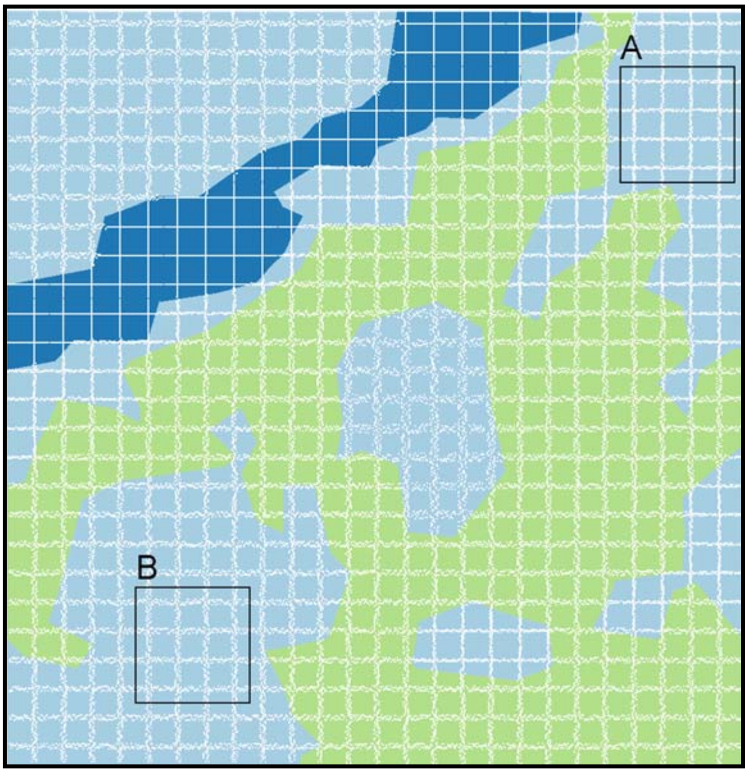

Figure 5. Example map from the study with 40% noise width, coarse grain and four uncertainty categories. Participants were asked to compare the degree of uncertainty in areas A and B. Area B is more uncertain than A.

3.5. Survey

We used a 2x2x3 factorial design and 10 repetitions per factor combination for discrete uncertainty visualizations. In order to test the third research question (discrete versus continuous), we added maps showing a continuous uncertainty distribution. For this, we used constant values for noise width and grain and just varied the number of uncertainty categories. The 3 levels shown with 10 repetitions resulted in 30 questions for this section. In the survey, the maps with object uncertainty and those with continuous uncertainty were randomized. In total, each participant answered 150 map questions.

We conducted the experiment as a web-based survey which made it possible to use Amazon Mechanical Turk (see sectionParticipants3.6). While we were not able to control the display type, color calibration, or distractions that potentially influence the participant, we did not expect relevant differences when viewing our maps on different displays. Several studies show that lab experiments are comparable to online experiments (Mason & Suri 2012).

The study had four parts:

1) Introduction: The participants were provided with an introductory explanation of uncertainty and noise annotation lines. Three figures of noise annotation lines were shown (no, medium and high uncertainty) to clarify the method. We also included a note to not use a smartphone or similar device and when using a tablet, to not zoom in and out to make sure that all subjects see each map as a whole when answering the questions.
2) Personal information: We asked for gender, age and a self-assessment in terms of experience with uncertainty visualization in maps.
3) Map questions: This main section of the study showed the maps including uncertainty. In each map, the two areas (A and B) were compared. In order to avoid bias and learning effects we randomized the order of the questions.
4) Comments: An opportunity to comment on the survey.

All questions except the comments at the end were mandatory. LimeSurvey (http://www.limesurvey.org) was used to deploy the survey, which is a user-friendly software freely available under an open source license. We used version 1.92+as problems with the randomization occurred in the latest version.

3.6. Participants

We recruited participants using the online crowdsourcing service Amazon Mechanical Turk (AMT, http://mturk.com). The reasons for utilizing this service are threefold: It was easy to recruit the subjects, we aimed to obtain participants with different backgrounds and expertise (not only from our domain) and third, paid participants were likely to be motivated to finish the survey even though it took 20 to 30 minutes. Participants were reimbursed with 50 cents for their participation.

The majority declared themselves as female (13 out of 22) and only 9 as male. Regarding age most of the participants classified themselves to be between 20 and 29 years old (9 out of 22), followed by 50 to 59 years (7/22). Very young and very old people as well as the group 40 to 49 years were barely represented.

4. Results

Concerning the subjects' experience with uncertainty maps, we asked three questions: If they had known about the concept of uncertainty before, how often they use maps, and if they had seen a map including uncertainty information before. From these answers we determined a level of experience per participant (little, average, extensive experience). Half of the participants (11 out of 22) had little experience while one-quarter had average experience (5/22) or extensive experience(6/22) with uncertainty maps. This is not surprising as one can expect that participants acquired via Amazon Mechanical Turk will be primarily lay people.

None of the participants produced data that should be considered an outlier (Tabachnick & Fidell 2007). We analyzed both data accuracy and time it took participants to answer questions. Given the difficulty of controlling time (compared to lab experiments) it does not come as a surprise that time did not show any significance and will not be considered further. Accuracy is recorded as the percentage of correct answers for a set of maps that belong to a factor combination. We had 10 maps for each combination in our 3 (levels of uncertainty) x 2 (noise width) x 2 (noise grain) factorial design. In case a participant responded that he or she was not able to provide an answer, we treated this situation as a missing value. There were only 68 missing values out of 3300 (~2%) such that sufficient responses were collected for each participants and each map.

Repeated measures ANOVA with the three factors revealed the following: Maulchy's test of sphericity shows that for some factors (levels of uncertainty) and factor combinations the assumption of sphericity is violated. Hence, we are using the Greenhouse-Geisser correction as suggested by Tabachnick and Fidell (2007). Of the three main effects, only the levels of uncertainty are statistically significant ($F(33.4,1.6)=6.96$, $p=.005$), while noise and particles did not significantly change the perception of uncertainty. However, both non-significant main effects show a statistically significant interaction effect with the levels of uncertainty. Levels of uncertainty * noise width: $F(42,2)=3.69$, $p=.038$); levels of uncertainty * grain: $F(42,2)=3.94$, $p=.042$). Figure 6 shows a graph that supports these findings visually.

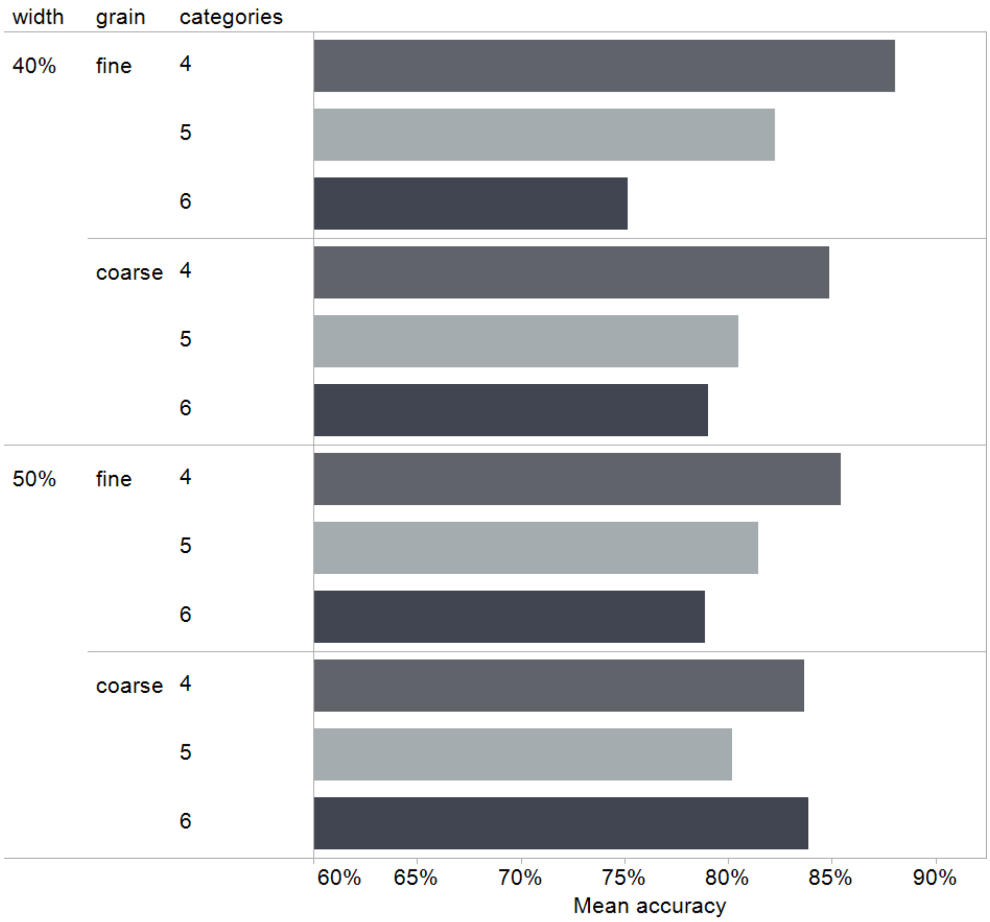

Figure 6. Mean accuracy for different factor levels: Accuracy values decrease when more categories of uncertainty are displayed, esp. with a smaller grid width. The combination of larger grid size and coarse grain does not show this behavior: The accuracy with 6 categories is roughly as high as with 4 categories (last row). Please note that the chart starts at 60% accuracy.

A second repeated measure ANOVA compared the "discrete versus continuous" presentation of uncertainty (see section 3.4), that is, whether uncertainty within an area of interest is changing or not. The second main factor is the levels of uncertainty. Maulchy's test of sphericity was not statistically significant, however, we adopted the Greenhouse-Geisser correction. The main factor of the discrete-vs.-continuous comparison showed a statistically significant trend ($F(21,1)=4.07$, $p=.057$). The main factor levels of uncertainty was statistically significant ($F(37.14, 1.77)=13.34$, $p<.001$). There was no significant interaction between these two factors.

5. Discussion

The results offer insights into various aspects of uncertainty visualization. The first thing to note is that, as expected, the number of categories has an influence on people's abilities to make judgments about uncertainty. Simply put, the fewer categories participants had to distinguish the better their performance was in terms of correct answer. This result can be explained by numerous studies in both perception and cognition literature that indicate that the more information that has to be distinguished and kept in working memory, the more difficult it is to reason with this information (Lloyd & Bunch 2005).

However, and this is an important finding for research on visualizing uncertainty, the downward trend of performance can be stopped using appropriate visualizations. In the case reported here, the visually more salient visualization means that complement uncertainty visualizations (the two salient levels of *noise width* and *noise grain*) were able to leverage the negative effect of an increased number of uncertainty categories (as revealed by the significant interaction effects and the graph in Figure 6). This is especially true for the combined effect of high values of noise width and noise grain.

The comparison of discrete and continuous visualizations of uncertainty paint the general picture, that is, the more complex the information is that is offered to participants, the more errors they make. We only compare the lowest level of grain and width for both discrete and continuous uncertainty visualization. In this combination it is clear that the increase in the number of categories leads to worse performance and that on all levels the performance is worse if more than one category of uncertainty is present in the areas which participants compared.

6. Conclusion

We have presented a study to evaluate the noise annotation lines for visualizing attribute uncertainty. In a web-based survey, users had to compare the uncertainty values for two equally-sized areas. This was done for discrete and continuous uncertainty representations (see section 3.4).

From the experiment, the following results could be concluded:

- the number of uncertainty categories has a significant influence on the participants' judgment (with more categories, user performance decreases)
- the variation of the design parameters *noise width* and *noise grain* did not significantly change user performance

- there is a decrease of user performance when more uncertainty categories have to be distinguished, but this can be counterbalanced with changes in the design of the noise grid
- more complex uncertainty information (“continuous” uncertainty) leads to a significant decrease in user performance
- there were no significant effects with respect to response time

In this experiment, we did not consider the aspect of intuitiveness. As we assume that noise is a good metaphor for uncertainty, it would be interesting to evaluate if people can intuitively understand and utilize noise annotation lines in comparison to other extrinsic (e.g., glyphs, sine amplitude) or intrinsic approaches (e.g., saturation, opacity or whiteness). Besides, we did not evaluate the influence of background colors with different contrast levels. An interesting aspect would be if and how different colors affect the usability of the method.

References

Bertin, J (1983) Semiology of graphics. Madison, Wis: University of Wisconsin Press.

Brodlie, K, Allendes Osorio, R& Lopes, A (2012) A Review of Uncertainty in Data Visualization, in Dill, John, et al. (Hg.): Expanding the Frontiers of Visual Analytics and Visualization: Springer London, 81–109.

Cedilnik, A, Rheingans, P (2000). Procedural Annotation of Uncertain Information: Proceedings of IEEE Visualization IEEE 2000, 77–84. URL: http://www.cs.umbc.edu/~rheingan/pubs/grids00.pdf

Gershon, N (1998) Visualization of an Imperfect World. IEEE Computer Graphics and Applications 18(4), 43-45.

Lloyd, R, Bunch, R (2005) Individual differences in map reading spatial abilities using perceptual and memory processes. Cartography and Geographic Information Science 32(1), 33–46.

MacEachren, A M (1992) Visualizing Uncertain Information. Cartographic Perspectives 13, 10–19.

MacEachren, A M, Robinson, A, Gardner, S, Murray, R, Gahegan, M, Hetzler, E (2005) Visualizing geospatial information uncertainty: What we know and what we need to know. Cartography and Geographic Information Science 32, 139–160.

MacGranaghan, M (1993) A Cartographic View of Spatial Data Quality. Cartographica: The International Journal for Geographic Information and Geovisualization 30(2), 8–19. URL: http://dx.doi.org/10.3138/310V-0067-7570-6566.

Mason, W, Suri, S (2012). Conducting behavioral research on Amazon's Mechanical Turk. Behavioral Research Methods, 2012 Mar, 44(1), 1-23. URL: http://dx.doi.org/10.3758/s13428-011-0124-6.

Pang, A (2001) Visualizing Uncertainty in Geo-spatial Data: Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology.

Senaratne, H, Gerharz, L, Pebesma, E, Schwering, A (2012) Usability of Spatio-Temporal Uncertainty Visualisation Methods, in Gensel, J, Josselin, D, Vandenbroucke, D (Eds.): Bridging the Geographic Information Sciences: Springer Berlin Heidelberg. (Lecture Notes in Geoinformation and Cartography), 3-23. URL: http://dx.doi.org/10.1007/978-3-642-29063-3_1.

Slocum, T A (2003) Thematic cartography and geographic visualization. 2nd Ed. Upper Saddle River, N.J., London: Prentice Hall.

Tabachnik, B G, Fidell, L S (2007) Using multivariate statistics. Boston: Pearson Education

Wittenbrink, C M, Pang, A T, Lodha, S K (1996) Glyphs for visualizing uncertainty in vector fields: Visualization and Computer Graphics, IEEE Transactions on Visualization and Computer Graphics, 2(3), 266–279.

Zuk, T, Carpendale, S (2006). Theoretical Analysis of Uncertainty Visualizations. In: Erbacher, R F, Roberts, J C, Gröhn, M T, Börner, K (Eds.): Visualization and Data Analysis 2006, Proceedings SPIE & IS&T Conf. Electronic Imaging