Enhancing Uncertainty Communication in Time Series Predictions: Insights and Recommendations

A. Karagappa[1], P. Kaur Betz[1], J. Gilg[1], M. Zeumer[1], A. Gerndt[1] and B. Preim[2]

[1]Institute for Software Technology, DLR, Braunschweig, Germany
[2]Faculty of Computer Science, University of Magdeburg, Magdeburg, Germany

Abstract

As the world increasingly relies on mathematical models for forecasts in different areas, effective communication of uncertainty in time series predictions is important for informed decision making. This study explores how users estimate probabilistic uncertainty in time series predictions under different variants of line charts depicting uncertainty. It examines the role of individual characteristics and the influence of user-reported metrics on uncertainty estimations. By addressing these aspects, this paper aims to enhance the understanding of uncertainty visualization and for improving communication in time series forecast visualizations and the design of prediction data dashboards.

CCS Concepts
• *Human-centered computing → Visualization theory, concepts and paradigms; Empirical studies in visualization; Visualization design and evaluation methods;*

1. Introduction

In time series forecasts, a model predicts the future values of a variable based on its past data as well as other contributing factors, using statistical or machine learning techniques. It finds application in several areas including weather forecasting, predicting growth indicators (e.g. GDP) and predicting daily cases of infectious diseases (e.g. COVID-19). These forecasts are always accompanied by uncertainty. The uncertainty can arise in many different stages of the visualization pipeline, see Figure 2 [SSK*16]. In the figure, there are three types of uncertainty depicted:

(i) Data and model uncertainty (**U1**): This uncertainty arises from factors like data collection, preprocessing, the prediction model, and parameterization.

(ii) Visualization uncertainty (**U2**): This uncertainty stems from how uncertainty is visually encoded and how our brains naturally perceive it.

(iii) Perception and uncertainty awareness (**U3**): This uncertainty relates to differences between users due to individual characteristics between them.

Often the visualization authors do not have control over the data and model uncertainty, therefore, the goal is to clearly communicate this while minimizing the visualization uncertainty and aiding the perception and uncertainty awareness in the user.

The uncertainty is often expressed in terms of a credible or a confidence interval. In a literature review on the reliability of predic-
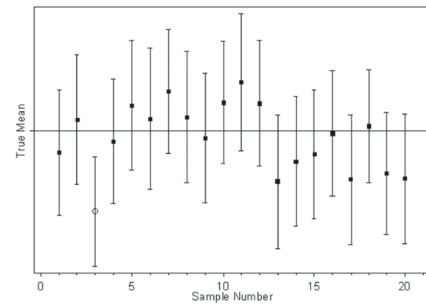

Figure 1: For 20 independent samples from the population, nineteen of their 95% confidence intervals (marked by black squares) contain the true mean, while one (marked by a hollow circle) does not [TT10].

tions on COVID-19, the reliability of cumulative cases was found to be reported using either confidence intervals (CI) or credible intervals (CrI) [GSKG21]. While both the CI and CrI are measures of uncertainty, they differ in their nature. A 95% CrI suggests there is a 95% probability that the true value lies within the predicted interval given the distribution of the observed data. This interpretation is often applied to a confidence interval, but is incorrect [HMRW14]. A 95% CI implies that when a study is conducted multiple times, each time with different samples from the same population, each sample will produce its own 95% CI. We expect 95% of these CIs to contain the true population mean, and this expectation holds with 100% certainty, see Fig 1 [TT10].
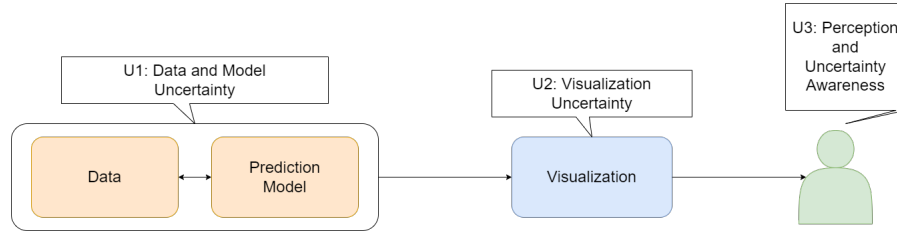
Figure 2: A simplified version of the knowledge generation model [SSK*16]

Given the wide range of applications, naturally, the users are diverse. It is essential that the uncertainty accompanying time series visualizations is understood by the diverse users and considered for decision making. To achieve this, we must ensure that: 1. the graphical representation of uncertainty clearly communicates its nature and value, and 2. the complete information required by the user for their decision making tasks is made available alongside this graphical representation.

Graphically, the uncertainty in time-series forecasts is often represented with a confidence band in a line chart (see Figure 3), or some variation of it. Previous research has shown that with some visualizations, users are more likely to adopt an interpretation of uncertainty that closely resembles a normal distribution. It is important to understand which individual properties of a user contributes to their interpretation of uncertainty and their accuracy in estimating it from a visualization. This should allow us to design the uncertainty visualization so that disparities between user perceptions of uncertainty may be minimized. Additionally, it is unclear whether the nature and complexity of the uncertainty are communicated well enough, contextually, for a user to include the uncertainty estimation in their decision-making process.

In order to answer these questions, we conducted two user studies, with observed variables on individual characteristics, task performance, and user-reported metrics. The results of these studies have found evidence in support of five major considerations in developing visualizations for prediction charts:

- Developing and utilizing standard uncertainty terminology and visualization techniques
- Meeting the different information needs of users
- Neutralizing the effects of numeracy
- Emphasis on clutter reduction and aesthetics in design
- The inclusion of interactive techniques for increased comprehensibility

2. Motivation

In 2020, we saw the rise of many Public Health Dashboards that displayed the predicted daily cases, hospitalizations and even deaths due to COVID-19. The COVID-19 Projections dashboard by the Institute for Health Metrics and Evaluation (IHME) [Ins], was likely the first dashboard available for providing forecasts related to the COVID-19 pandemic. WHO/Europe stated that since the start of the COVID-19 pandemic, it increasingly referenced IHME forecasts in providing guidance and recommendations to countries

in the WHO European region with respect to containing the pandemic [Wor].

Undeniably, these visualizations played a significant role in public safety and well-being. Simultaneously, there were concerns over the validity of the predictions and their usefulness to policymakers. One particular concern was whether the graphical representation of uncertainty in line charts is effectively conducive to understanding uncertainty in predicted peak daily values, with instances of the uncertainty bounds, and consequently the projections, being interpreted incorrectly in both formal and social media [JLJ20].

This concern is heightened by reports from other national-level COVID-19 dashboard development teams, indicating challenges in presenting data clearly and understandably [BID*22]. The challenges faced by dashboard development teams were likely exacerbated by the rapidly evolving nature of the COVID-19 virus and time constraints in the design and development process. Additionally, authors generally omit uncertainty from their visualizations due to lack of canonical forms of uncertainty visualizations and expected difficulty for users in interpreting them [Hul20].

In their examination of Epidemiological Scenarios for Infectious Diseases (ESID), a visual analytics tool aimed at projecting simulated developments of infectious disease spread (see Figure 3), Betz et al. discovered issues of misinterpretation and misuse arising from inadequate treatment and communication of uncertainty [BSG*23]. As a result, we find the need to establish guidelines that help visualization authors effectively communicate uncertainty based on empirical findings.

3. Related Work

For representing uncertainty, Olston and Mackinlay highlighted the necessity for distinct visual representations of bounded and statistical uncertainty, addressing the potential misinterpretation that led to the development of the confidence interval band [OM02]. Subsequent studies revealed variations in the interpretation of confidence bands and its variants, with some aligning more closely with a credible interval [TTvE14]. Although the results provide valuable insights, the recent misinterpretation of credible intervals underscores the need for a study where uncertainty is precisely defined, and suitable representations are identified to ensure comprehensive information for user trust. An individual's numeracy influences their uncertainty estimation [TTvE14, TvEBvB19], thus there is a desire to explore whether numeracy is impacted by other factors. This exploration aims to inform the design of visualizations that facilitate training or support interpretation for a broad user base.
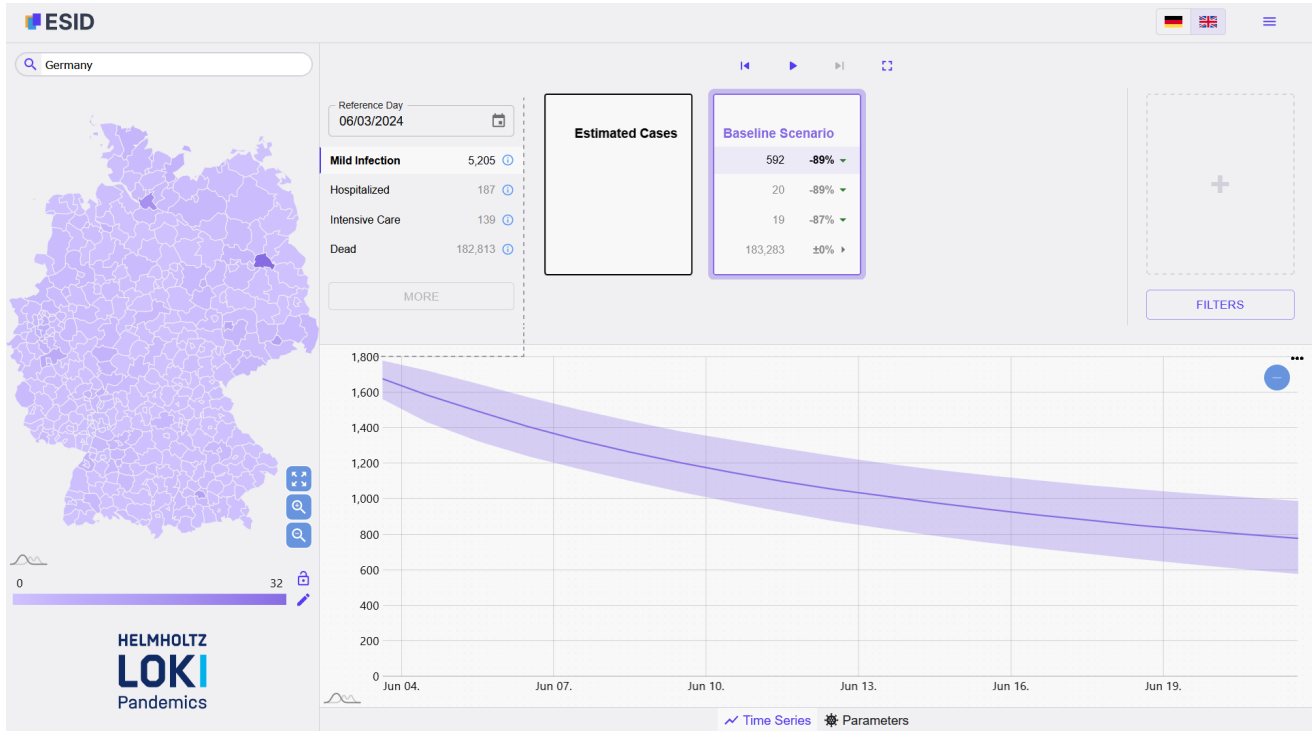
Figure 3: A snapshot of the Epidemiological Scenarios for Infectious Diseases (ESID) application, a visual analytics application developed to show the results of epidemiological simulations. On the left hand side is a map of the German counties with a customizable heat legend. At the top section is a purple scenario card representing one simulation scenario. Left of the scenario cards is a list of infection states or aggregations of infection states, like infected, hospitalized, and dead. The bottom portion contains a line chart comparing the scenarios over a predefined timeline. [BSG*23].

Trust in uncertainty visualizations and visual processing has been a subject of study [EPX22]. Yet another crucial aspect is whether users feel assured in their understanding of the given information [YS12]. Consequently, we investigate to determine if such assurance can be influenced by specific characteristics of the visualization. While visualization authors may face limitations in the available information [Hul20], the goal is to assist them in designing dashboards that consider factors within their control.

Lastly, we have realised a gap in the literature concerning evaluations of whether prediction dashboards adequately present the information needed to make decisions based on uncertainty. Our studies explored these gaps and sought to identify their potential sources.

4. Method

Our work focuses on two sets of factors that inform the design process of uncertainty visualizations: those contributing to visual uncertainty (U2 in Figure 2) and those affecting uncertainty perception of individuals (U3 in Figure 2). The factors considered include:

- Visual uncertainty factors:
  - Task performance under different techniques
  - User-reported metrics of visualization qualities, such as clutter and aesthetic.

- Uncertainty perception factors:
  - Individual characteristics, including numeracy, field of work, and familiarity with line charts.
  - Information needs
  - User-reported metrics of user performance, such as success and difficulty.

To summarize, the questions we want to answer are:

- **Task Performance:** What is the impact of different visualization techniques on the participants' uncertainty estimation?
- **Information Needs:** Does the information provided by the uncertainty predictions visualizations meet the needs of the users?
- **User-Reported Metrics:** How do the effects of varying visualization techniques (like clutter and aesthetics) influence users' evaluations of task difficulty, and their perceived level of success in task performance?
- **Individual Characteristics:** Do individual characteristics shared among target user groups, such as area of study, frequency of visualization use, the highest degree correlate with numeracy, and subsequently task performance?

To answer our research questions, we conducted one user study with a diverse user base and a second user study with the target user group. The target user group consists of participants from local health authorities, medicine and neuroscience. These individuals typically work with medical images, epidemiological data, and un-
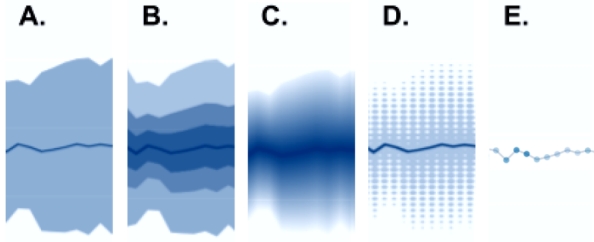
Figure 4: Uncertainty represented as A. Confidence Band, B. Overlapping Bands, C. Blur, D. Circular Glyphs and E. Colored Markers

certainties within their respective fields. They also carry some degree of job responsibility in being able to understand and interpret epidemiological predictions. In the design of public health dashboards, they are usually among the main stakeholders [BSG*23], and thus results from them are deemed important in the visualization design.

4.1. Stimuli

In the initial user study, we inspected five uncertainty visualizations, as shown in Figure 4. The techniques chosen are A. Confidence Band [OM02], B. Overlapping Bands of varying credible intervals, C. Band represented by decreasing colour saturation, D. Band represented by decreasing area of circular glyph and E. Colored Markers of varying colour saturation [SZB*09].

The user studies consisted of three parts; collection of participants' data, a numeracy task, and visualization tasks.

4.2. Participants Data

The collection of individual information included participants' highest degree, field of study and the frequency with which they interact with visualizations (see Q.P1 - Q.P3, Supplementary Material).

4.3. Numeracy Task

To measure the users' numeracy, we adopted three questions from the J. Weller et al. numeracy scale whose difficulties span over easy, intermediate and difficult (see Q.N1 - Q.N3, Supplementary Material) [WDT*13]. For the final study, we included two additional questions to assess their comprehension of a normal distribution (see Q.N4 a and b, Supplementary Material).

4.4. Visualization Task

The visualization task was divided into two subtasks. The first subtask featured three questions concerning probability estimation (see Q.T1, Q.T4 and Q.T7, Supplementary Material). The participants were tasked with assessing the likelihood with which the predicted value falls within a certain range. This is done in a

decision-making context, although no decision is asked to be made [HQC*19,CQHP22]. The uncertainty estimations are elicited in the form of a range slider with the minimum and maximum marked at 0 and 100, respectively. Following each probability estimation, users were asked to declare whether the information provided was enough to solve the task (see Q.T2, Q.T5 and Q.T8, Supplementary Material). Only in the final study, if they required more information, they were prompted to describe what particular information they needed (see Q.T3, Q.T6 and Q.T9, Supplementary Material).

The second subtask focused on gathering user-reported metrics. Two questions pertained to the task (see Q.T10a and b, Supplementary Material) and two questions pertained to the characteristics of the visualizations (see Q.T10 c and d, Supplementary Material). Questions Q.T10 a and b seek to evaluate the users' perceived difficulty of the task and their perceived success in the task, taken from the NASA-TLX questionnaire [NAS88]. Questions Q.T10 c and d seek to evaluate the users' perceived clutter in and perceived aesthetic of the visualization technique. This is essential to understanding what aspects of a visualization an author must pay attention to in order to improve the communication of uncertainty [SSK*16]. The four user-reported metrics were measured on a five point Likert scale.

5. Results

The evaluation of the results was performed using the R programming language and associated statistical tools and libraries.

5.1. Measures

In our studies, all participants were exposed to all visualization techniques, and the resulting data was not normally distributed. Therefore, to determine if there were significant differences in uncertainty estimations across these techniques, we conducted a Friedman test using the `stats::friedman.test` function. Following this, we performed a post hoc Nemenyi test using the `PMCMRplus::frdAllPairsNemenyiTest` function to identify significant pairwise differences between the various uncertainty visualization methods.

The performance metric used to validate the uncertainty estimation was the normalized sum of all three uncertainty estimations under a visualization technique. The normalization was performed by calculating the range of the sum of the uncertainty estimations across all participants, then applying the following formula:

$$\text{Performance} = \frac{\max(U_1 + U_2 + U_3) - (U_1 + U_2 + U_3)}{\max(U_1 + U_2 + U_3) - \min(U_1 + U_2 + U_3)}$$

where U_1, U_2, and U_3 represent the three uncertainty estimations under each visualization for each participant. U_N (where $N = 1, 2, 3$) is calculated as the absolute difference between the user estimate and the true value. This was chosen as the performance measure as it considers the overall uncertainty by aggregating estimations from different ranges of uncertainty variation across techniques. Furthermore, normalization enhances interpretability by scaling the metric to a common range, facilitating clearer insights into relative performance.

Lastly, Spearman Rho is used to find correlation between ordinal-ordinal and continuous-ordinal values, utilizing the `stats::cor` function. Ordinal encoding is employed where necessary. This correlation can be seen in Figure 6, where only the correlations with a *p* value below 0.05 are visualized. The *p* values are calculated using the Spearman method from the `rstatix::cor_pmat` function. The correlation values are interpreted according to Table 1 [DR04].

Spearman's Rho	Interpretation
0.01 - 0.19	None or very weak relationship
0.20 - 0.29	Weak relationship
0.30 - 0.39	Moderate relationship
0.40 - 0.69	Strong relationship
≥ 0.70	Very strong relationship

Table 1: Interpretation of Spearman's Rho correlation [DR04]

Points are awarded in the numeracy task based on question difficulty: 1 point for the easiest question, 2 points for the next, and so on. The lowest possible score is 0. In the initial study with three questions, participants could score up to 6 points, with a mean possible score of 3. In the second study with five questions, the maximum score was 15 points, with a mean possible score of 7.5.

5.2. Initial User Study With a Diverse Audience

The initial user study involved 94 participants from diverse fields, ranging from high school to individuals holding a PhD or advanced degrees. The Friedman test shows significant ($p < 0.001$) differences between the performance of users with different visualization techniques. A post hoc Nemenyi test finds significant ($p \leq 0.001$) differences between *Colored Marker* and all other techniques. As shown in Figure 5A, the *Colored Marker* visualization technique performed the worst in the initial user study. This could be attributed to the lack of area representation in *Colored Marker* that might have allowed users to build heuristics to estimate uncertainty in the other techniques. Therefore, in the second user study, we do not consider the *Colored Markers* technique owing to its bad performance. We also omitted *Confidence Band* as it did not show significantly better performance than other techniques and also does not qualitatively depict the nature of a credible interval.

Additionally, in the initial study, we also observe that perceived aesthetics and perceived success exhibit a strong negative correlation with perceived clutter (see Figure 6A). Visual information processing in novices have shown to be negatively impacted by clutter in financial visualizations, in comparison to experts [OTMH19]. Additionally, aesthetic visualizations have previously shown to display a higher level of user patience, resulting in lower task abandonment and erroneous response [CM07].

5.3. Second User Study With the Target User Group

To know the information needs of the specific group, the next study was conducted with 31 participants. In this study, we compared *Blur*, *Overlapping Bands* and *Circular Glyphs* (C, B and D in reference to Figure 4). In the following, we present the results according to the objectives defined earlier.
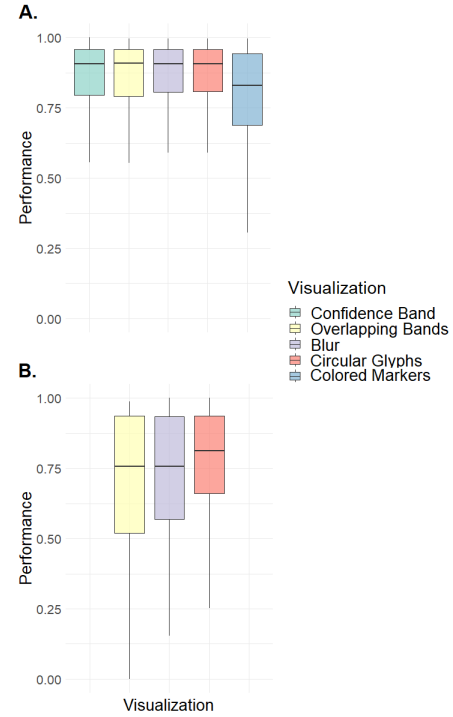
Figure 5: Boxplots showing the performance of visualization techniques. **A:** The five visualization techniques from the initial study ($p < 0.001$). The performance under *Colored Marker* is significantly worse than all the other techniques. **B:** The three visualization techniques from the second study ($p \leq 0.05$). *Colored Marker* is not considered for the second study due to its poor performance in the initial study. *Confidence Band* is also excluded as it did not show a significant difference compared to other techniques and does not qualitatively depict the nature of a credible interval. In the second study, *Circular Glyphs* shows a significant difference with *Blur* ($p = 0.058$) but no other significant differences were observed. *Circular Glyphs* demonstrates higher median performance and smaller variability.

Task performance: The Friedman test shows significant ($p < 0.05$) differences between the performance of users with different visualization technique. A post hoc Nemenyi test reveals a trend towards significant differences (p = 0.058) between *Circular Glyphs* and *Blur*. *Circular Glyphs* demonstrate a higher median performance and lower variability compared to *Blur* and *Overlapping Bands* (see Figure 5B). It is challenging to understand why, because there was no observed correlation between the visualization technique and the factors that influence usability. Earlier empirical findings demonstrate that, in quantitative perceptual tasks, the use of area outperforms colour hue [CM84]. This observation may offer insights into why *Circular Glyphs* is more effective than *Blur*. Furthermore, minor inaccuracies in estimating the size of an area result in only minor misperceptions of the associated quantitative value being encoded [Mac86].

Information needs: In approximately a third of the uncertainty estimation task instances, participants reported that they did not have the necessary information to make the required estimate. The in-
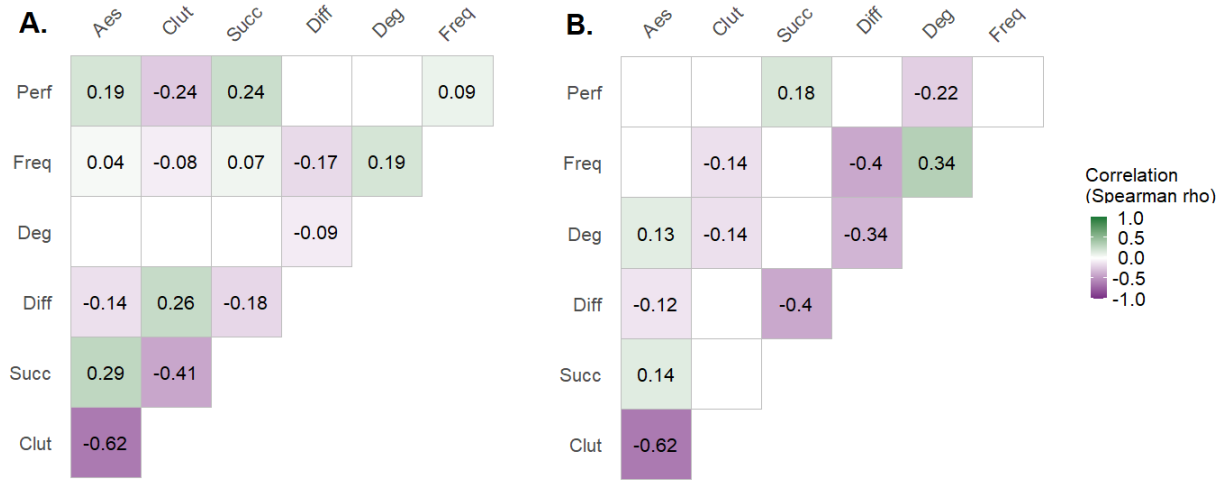
Figure 6: Correlation matrix of performance, individual differences, and user-reported metrics in **A**. the initial study and **B**. the second study. Abbreviations: Perf (Performance of user in uncertainty estimation tasks, with range from Worst to Best), Freq (Frequency with which the user interacts with visualizations, with range from Rarely to Once a week), Deg (Highest Degree held by the user, with range from High School to PhD or Higher), Diff (Perceived difficulty of task given a visualization technique, with range from Not hard at all to Very hard), Succ (Perceived success of task given a visualization technique, with range from Poor to Good), Clut (Clutter of a visualization technique, with range from Uncluttered to Cluttered), Aes (Aesthetic of a visualization technique, with range from Bad to Good).

formation needed by participants can be divided largely into two groups: Statistical Information and Model Information. Statistical Information refers to the information needed to make the correct association between a credible interval and likelihood, while Model Information refers to the information regarding parameters chosen and model used in the prediction.

User-reported metrics: Similar to the findings of the first study, we observe a strong correlation between perceived clutter and aesthetic. However, perceived success does not correlate with clutter; instead, it correlates with perceived difficulty. This perceived difficulty, in turn, correlates with individual characteristics, such as a user's degree and frequency of interaction with line charts.

Individual characteristics: Users' frequency of interaction with line charts has a weak correlation with their numeracy (Rho = 0.27). Higher numerate users (numeracy > 7.5) exhibit lower errors across all visualizations combined, with a narrow error range. In contrast, those with low numeracy show higher and more variable errors.

6. Guidelines

Given the results of the user study and the relationship between the measured metrics, we have formulated four guidelines for uncertainty visualization designers.

6.1. Standardizing Uncertainty Terminology and Visualization Techniques

There exists a discrepancy in the terminology used to describe uncertainty of predictions in the present literature, including but not limited to: Credible Interval, Prediction Interval, Simulation Percentile, and Uncertainty Interval. While the choice of terminology
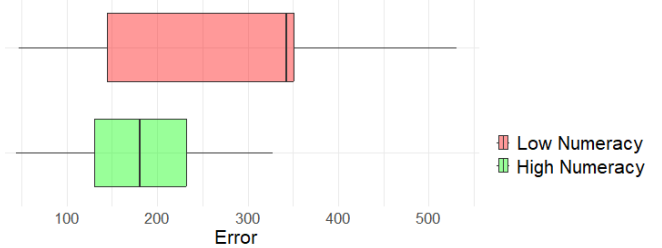

Figure 7: Boxplots comparing total user errors across all visualization techniques combined between higher numerate (numeracy > 7.5) and lower numerate users.

used for uncertainty ranges depends on the specific modeling approach and uncertainty quantification, both researchers and users of visualizations can benefit from standardizing the terminology. When illustrating uncertainty ranges visually, it is important to balance clarity and completeness. Whether it is a precise interval or a probabilistic distribution, the graphical representation and terminology should accurately reflect this.

6.2. Meeting the Informational Needs of a Diverse Audience

For users to make an informed decision, it is advisable to include the following information:

- **Statistical Information**
  In the second user study, participants either explicitly stated or hinted towards needing more statistical information that would help them link the terms *likelihood* and *uncertainty interval*. A participant suggested a calculation tool to help make this connec-

tion. The uncertainty showed was credible intervals, but some participants expressed confusion regarding the distinction between confidence intervals and credible intervals.

- **Model Information**
  Participants in the second user study also stated that the information provided about the parameters in the prediction model and its *historic accuracy* was necessary to adequately make an estimation. Providing access to previous forecasts with uncertainty in conjunction with truth data could be used to show historic accuracy.

Tool-tips or hover-over explanations, annotations, and narrative approaches should be considered to fill these gaps in knowledge.

6.3. Equalizing Effects of Numeracy in Uncertainty Estimation

Minimizing the cognitive effort needed for making many inferences is beneficial to users with lower numeracy [PHSD07]. The role of uncertainty and the risk of excluding it in the decision-making process should be explicitly displayed as people tend to underweight probable outcomes compared to certain ones [KT79]. Consider hiding visual objects that do not contain relevant information when presenting uncertainty. Removing non-quality information will improve comprehension for those with lower numeracy skills. Additionally, use consistent and comprehensible formats. For instance, simplify uncertainty ranges to ensure they are easily understandable and manageable.

6.4. Clutter Reduction and Aesthetic Design Influence on Uncertainty Estimation

Incorporating the range of uncertainty intervals and avoiding the overlap of uncertainties should be integral to the design process. If not carefully managed, improper handling of these elements can lead to axis realignment and overwhelming information, thereby compromising the effectiveness of the visualization. Decluttering graphs involves removing backgrounds, chart borders, gridlines, and unnecessary axis lines, adding white space between major elements and replacing legends with direct labels [ALX*22]. In a study conducted, grayscale visualizations of ensemble predictions were more trusted than colour-coded visualizations [PFCB23]. The authors initially predicted that adding colour may add complexity and reduce clarity, whose effect would increase with the number of forecasts but found it to be consistently true irrespective of the number of forecasts.

6.5. Increase Comprehensibility Through Interaction

Provide facilities for better comprehension of uncertainty visualizations. For instance, a visualization may incorporate two movable horizontal lines that allow users to render a calculated probability of the predicted value of each displayed model falling within the selected range defined by these lines. Rather than policy makers having to examine the given information and make educated guesses about the uncertainties associated with various intervention strategies to identify the most suitable one, they can adopt a reverse approach based on their existing data. They can start by stating,

for example, that they currently have X hospital beds available and the goal is to limit the number of infections to Y to accommodate this capacity. Then, they can evaluate various predictions resulting from varying intervention plans and select the one that best aligns with these predefined requirements. In essence, it simplifies decision making by working backward from their existing constraints and goals.

7. Conclusion and Future Work

In this paper, we presented the results of our two user studies which we conducted to understand the factors influencing visual uncertainty and individual uncertainty perception. The initial study involved 94 participants from diverse backgrounds, assessing five different uncertainty visualization techniques. The results indicated that *Circular Glyphs* and *Blur* showed the best performance, while the Colored Markers technique performed the worst. We found that higher numeracy and frequent interaction with visualizations positively correlated with better task performance and perceived success. Additionally, increased perceived clutter and decreased aesthetics negatively impacted perceived difficulty and actual performance. The second user study targeted 31 participants from local health authorities, medicine, and neuroscience, focusing on the three best-performing techniques from the initial study: *Blur*, *Overlapping Bands*, and *Circular Glyphs*. In this study, *Circular Glyphs* slightly outperformed the other techniques. Participants frequently reported needing more statistical and model information to make informed uncertainty estimations. Similar to the initial study, higher numeracy correlated with better performance, and clutter and aesthetics influenced perceived difficulty and success.

Our user studies, aligned with recent research on uncertainty visualization evaluation, revealed key insights for designing dashboards with time series predictions. While we focused on epidemiological forecasts, our guidelines are not limited to a specific domain. Instead, they aim to aid decision making by providing necessary information with minimal cognitive effort. In summary, our findings offer practical design guidance to visualization authors to maximize the utility of the information they have access to but within the constraints they face.

Future research should explore interactive tools to enhance comprehension of uncertainty visualizations, particularly for users with varying levels of numeracy. Additionally, further studies could investigate the impact of standardized terminology and consistent exposure to uncertainty visualizations to improve overall interpretability and usability across diverse user groups.

References

[ALX*22] AJANI K., LEE E., XIONG C., KNAFLIC C. N., KEMPER W., FRANCONERI S.: Declutter and focus: Empirically evaluating design guidelines for effective data communication. *IEEE Transactions on Visualization and Computer Graphics 28*, 10 (2022), 3351–3364. doi:10.1109/TVCG.2021.3068337. 7

[BID*22] BARBAZZA E., IVANKOVIĆ D., DAVTYAN K., POLDRUGOVAC M., YELGEZEKOVA Z., WILLMINGTON C., MEZA-TORRES B., BOS V. L., ÓSCAR BRITO FERNANDES, ROTAR A., NUTI S., VAINIERI M., CARINCI F., AZZOPARDI-MUSCAT N., GROENE O., NOVILLO-ORTIZ D., KLAZINGA N., KRINGOS D.: The experiences

of 33 national covid-19 dashboard teams during the first year of the pandemic in the world health organization european region: A qualitative study. *DIGITAL HEALTH 8* (2022), 20552076221121154. doi:10.1177/20552076221121154. 2

[BSG*23] BETZ P. K., STOLL J., GRAPPENDORF V., GILG J., ZEUMER M., KLITZ M., SPATARO L., KLEIN A., ROTHENHAEUSLER L., BOHNACKER H., KRAEMER H., MEYER-HERMANN M., SOMOGYI S., GERNDT A., KÜHN M. J.: Esid: Exploring the design and development of a visual analytics tool for epidemiological emergencies. In *2023 IEEE VIS Workshop on Visualization for Pandemic and Emergency Responses (Vis4PandEmRes)* (2023), pp. 8–14. doi:10.1109/Vis4PandEmRes60343.2023.00007. 2, 3, 4

[CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79*, 387 (1984), 531–554. doi:10.1080/01621459.1984.10478080. 5

[CM07] CAWTHON N., MOERE A. V.: The effect of aesthetic on the usability of data visualization. In *2007 11th International Conference Information Visualization (IV '07)* (2007), pp. 637–648. doi:10.1109/IV.2007.147. 5

[CQHP22] CASTRO S. C., QUINAN P. S., HOSSEINPOUR H., PADILLA L.: Examining effort in 1d uncertainty communication using individual differences in working memory and nasa-tlx. *IEEE Transactions on Visualization and Computer Graphics 28*, 1 (2022), 411–421. doi:10.1109/TVCG.2021.3114803. 4

[DR04] DANCEY C., REIDY J.: *Statistics Without Maths for Psychology*. 01 2004. 5

[EPX22] ELHAMDADI H., PADILLA L., XIONG C.: Using processing fluency as a metric of trust in scatterplot visualizations. *arXiv* (2022). arXiv:2209.14340. 3

[GSKG21] GNANVI J. E., SALAKO K. V., KOTANMI G. B., GLÈLÈ KAKAÏ R.: On the reliability of predictions on covid-19 dynamics: A systematic and critical review of modelling techniques. *Infectious Disease Modelling 6* (2021), 258–272. doi:10.1016/j.idm.2020.12.008. 1

[HMRW14] HOEKSTRA R., MOREY R. D., ROUDER J. N., WAGENMAKERS E.-J.: Robust misinterpretation of confidence intervals. *Psychon. Bull. Rev. 21*, 5 (Oct. 2014), 1157–1164. 1

[HQC*19] HULLMAN J., QIAO X., CORRELL M., KALE A., KAY M.: In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 903–913. doi:10.1109/TVCG.2018.2864889. 4

[Hul20] HULLMAN J.: Why authors don't visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2020), 130–139. doi:10.1109/TVCG.2019.2934287. 2, 3

[Ins] INSTITUTE FOR HEALTH METRICS AND EVALUATION: Ihme: Covid-19 projections. https://covid19.healthdata.org/ (Accessed: June 19, 2024). 2

[JLJ20] JEWELL N., LEWNARD J., JEWELL B.: Caution warranted: Using the institute for health metrics and evaluation model for predicting the course of the covid-19 pandemic. *Annals of Internal Medicine 173* (04 2020). doi:10.7326/M20-1565. 2

[KT79] KAHNEMAN D., TVERSKY A.: Prospect theory: An analysis of decision under risk. *Econometrica 47*, 2 (1979), 263–291. URL: http://www.jstor.org/stable/1914185. 7

[Mac86] MACKINLAY J.: Automating the design of graphical presentations of relational information. *ACM Trans. Graph. 5*, 2 (apr 1986), 110–141. doi:10.1145/22949.22950. 5

[NAS88] Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Human Mental Workload*, Hancock P. A., Meshkati N., (Eds.), vol. 52 of *Advances in Psychology*. North-Holland, 1988, pp. 139–183. doi:10.1016/S0166-4115(08)62386-9. 4

[OM02] OLSTON C., MACKINLAY J.: Visualizing data with bounded uncertainty. *Proc. of the IEEE Symposium on Information Visualization 2002* (07 2002). doi:10.1109/INFVIS.2002.1173145. 2, 4

[OTMH19] OGNJANOVIC S., THÜRING M., MURPHY R. O., HÖLSCHER C.: Display clutter and its effects on visual attention distribution and financial risk judgment. *Applied Ergonomics 80* (2019), 168–174. doi:10.1016/j.apergo.2019.05.008. 5

[PFCB23] PADILLA L., FYGENSON R., CASTRO S. C., BERTINI E.: Multiple forecast visualizations (mfvs): Trade-offs in trust and performance in multiple covid-19 forecast visualizations. *IEEE Transactions on Visualization and Computer Graphics 29*, 1 (2023), 12–22. doi:10.1109/TVCG.2022.3209457. 7

[PHSD07] PETERS E., HIBBARD J., SLOVIC P., DIECKMANN N.: Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health affairs (Project Hope) 26* (05 2007), 741–8. doi:10.1377/hlthaff.26.3.741. 7

[SSK*16] SACHA D., SENARATNE H., KWON B. C., ELLIS G., KEIM D. A.: The role of uncertainty, awareness, and trust in visual analytics. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 240–249. doi:10.1109/TVCG.2015.2467591. 1, 2, 4

[SZB*09] SANYAL J., ZHANG S., BHATTACHARYA G., AMBURN P., MOORHEAD R.: A user study to compare four uncertainty visualization methods for 1d and 2d datasets. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1209–1218. doi:10.1109/TVCG.2009.114. 4

[TT10] TAN S. H., TAN S. B.: The correct interpretation of confidence intervals. *Proceedings of Singapore Healthcare 19*, 3 (2010), 276–278. doi:10.1177/201010581001900316. 1

[TTvE14] TAK S., TOET A., VAN ERP J.: The perception of visual uncertaintyrepresentation by non-experts. *IEEE Transactions on Visualization and Computer Graphics 20*, 6 (2014), 935–943. doi:10.1109/TVCG.2013.247. 2

[TvEBvB19] TOET A., VAN ERP J. B., BOERTJES E. M., VAN BUUREN S.: Graphical uncertainty representations for ensemble predictions. *Information Visualization 18*, 4 (2019), 373–383. doi:10.1177/1473871618807121. 2

[WDT*13] WELLER J. A., DIECKMANN N. F., TUSLER M., MERTZ C. K., BURNS W. J., PETERS E.: Development and testing of an abbreviated numeracy scale: A rasch analysis approach. *Journal of Behavioral Decision Making 26*, 2 (2013), 198–212. doi:https://doi.org/10.1002/bdm.1751. 4

[Wor] WORLD HEALTH ORGANIZATION: Who/europe and ihme sign agreement cementing collaboration on forecasting of health data. https://www.who.int/europe/news/item/07-02-2022-who-europe-and-ihme-sign-agreement-cementing-collaboration-on-forecasting-of-health-data (Accessed: June 19, 2024). 2

[YS12] YEUNG N., SUMMERFIELD C.: Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. Lond. B Biol. Sci. 367*, 1594 (May 2012), 1310–1321. doi:10.1098/rstb.2011.0416. 3