Fuzzy Spreadsheet: Understanding and Exploring Uncertainties in Tabular Calculations

Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Eduard Gröller, Marc Streit

Abstract—Spreadsheet-based tools provide a simple yet effective way of calculating values, which makes them the number-one choice for building and formalizing simple models for budget planning and many other applications. A cell in a spreadsheet holds one specific value and gives a discrete, overprecise view of the underlying model. Therefore, spreadsheets are of limited use when investigating the inherent uncertainties of such models and answering what-if questions. Existing extensions typically require a complex modeling process that cannot easily be embedded in a tabular layout. In Fuzzy Spreadsheet, a cell can hold and display a distribution of values. This integrated uncertainty-handling immediately conveys sensitivity and robustness information. The fuzzification of the cells enables calculations not only with precise values but also with distributions, and probabilities. We conservatively added and carefully crafted visuals to maintain the look and feel of a traditional spreadsheet while facilitating *what-if analyses*. Given a user-specified reference cell, Fuzzy Spreadsheet automatically extracts and visualizes contextually relevant information, such as impact, uncertainty, and degree of neighborhood, for the selected and related cells. To evaluate its usability and the perceived mental effort required, we conducted a user study. The results show that our approach outperforms traditional spreadsheets in terms of answer correctness, response time, and perceived mental effort in almost all tasks tested.

Index Terms—Uncertainty visualization, tabular data, spreadsheet augmentation.

✦

1 INTRODUCTION

S PREADSHEETS, due to their intuitive and direct approach to computation, are one of the most widely used tools for building and formalizing models in areas as diverse as science, finance, and business [2], [29]. However, the precision implicit in traditional spreadsheet cells [42] often hinders users in performing calculations with uncertain values. Established spreadsheet tools—such as Microsoft Excel [34], Google Sheets [20], and Apache OpenOffice Calc [51] do not take uncertainty into account at all or lack features for exploring how uncertainties propagate and combine throughout calculations.

Integrated handling of variability would be helpful, for instance, to model estimated future values [49] or to plan a budget. To cope with the lack of built-in support for working with uncertain information, users often simplify their problems by calculating alternative scenarios, for instance, covering the worst, best, and average cases. This oversimplification of the problem results in a limited understanding of the complete scenario. To better support users who want to properly include uncertainty in their calculations, a tool must communicate effectively how individual components influence the final result of the calculations— in terms of both expected value and of uncertainty. To be integrated into existing workflows, such a tool should retain the fa-

miliar spreadsheet interface and enhance it with additional information about uncertainty.

Our **primary contribution** is the Fuzzy Spreadsheet approach, which augments well-established spreadsheets with compact in-cell visualizations. These visualizations allow users to assess probability distributions and trace computational relationships directly within the spreadsheet cells. Users can analyze how uncertainty propagates through the values and explore the effects of hypothetical changes. As a **secondary contribution** we present the results of a user study, which indicates that Fuzzy Spreadsheet is effective in working with uncertain information in tabular calculations.

2 CHARACTERIZING UNCERTAINTY IN SPREADSHEETS

To discuss the limitations of traditional spreadsheets and to introduce our terminology, we present an example of a spreadsheet that models the resources required for the maintenance of cars (see Figure 1). We use this spreadsheet as a guiding example throughout the paper. Let us assume that a classic car enthusiast wants to manage the predicted costs for two of her cars. For the first car, she sums the cost of the checkup and the cost of replacing its expendable parts. For the second car, she plans to upgrade the engine. To reflect this, she enters the cost of buying a new engine and subtracts the selling price of the old one, which results in the total cost for Car 2. To obtain the Grand Total Cost, she combines the summed costs of each car.

To make the underlying network of cells more transparent and to understand which cells influence other cells, we use a computational graph [57]. Figure 2 shows the

- *Vaishali Dhanoa is with Pro2Future GmbH and Johannes Kepler University Linz. E-mail: vaishali.dhanoa@pro2future.at.*
- *Conny Walchshofer, Andreas Hinterreiter, and Marc Streit are with Johannes Kepler University Linz. E-mail: {conny.walchshofer, andreas.hinterreiter, marc.streit}@jku.at.*
- *Eduard Gröller is with TU Wien and VRVis Research Center. E-mail: groeller@cg.tuwien.ac.at.*

Manuscript received xxx

Car Maintenance

	Cost (in 1000 €)	Probability	Time Spent (in hours)	Standard Deviation
Car 1				
Checkup	3	1	5	1.5
Spare Part A	2	0.75	2	1
Spare Part B	1.5	0.5	3	0.5
Total	6.5	0.38	10	1.87
Car 2				
New Engine (SP)	3	1	3	1
Old Engine (CP)	1	0.8	1	0.5
Total	2	0.80	4	1.12
Grand Total	**8.5**	0.30	**14**	2.18

Fig. 1. A simplified spreadsheet for planning the maintenance of two classic cars, containing costs and time required for the maintenance of each car and uncertainty values provided in the adjacent columns.

computational graph of the example above that models the relationships between cells and highlights the functional expressions that combine them.

Based on experience with Car 1 in the previous years, she knows that each spare part has a different chance of failing within a year. This makes the cost uncertain, as it will only be incurred if the spare part fails. She models such costs with a Bernoulli distribution, indicating the cost expected in the case of failure along with the probability of failure. For the second car, her friend shows an interest in buying the engine for the price specified. As the deal has not yet been closed, she captures this uncertainty in the same way. She computes the total cost by summing the input costs and multiplying the result with the product of the individual probabilities. She observes that, according to the spreadsheet (shown in Figure 1), the Total Cost (Car 1) is 6.5 k Euros with a probability of 38 %. However, since such a calculation covers only one of many possible paths in a probability tree, she receives no indication of the value(s) for the remaining 62 % of the cases. In order to gain a complete overview of the actual cost, all the influencing costs must be combined correctly into a distribution. Furthermore, the contribution of each influencing cost on the distribution of the Total Cost (Car 1) can be analyzed if each influencing cost is treated as a distribution rather than a scalar value.

As a traditional spreadsheet lacks this ability to treat a cell as a distribution, she must somehow keep track of
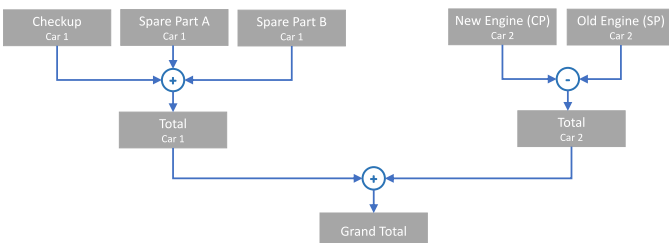

Fig. 2. Computational graph of the example in Figure 1, showing the relationships between cells.

the underlying distributions herself. The distribution of an independent cell, such as Spare Part A (Car 1), which is not influenced by other cells, may still be easy to understand. However, as more of these cells are combined by using formulas to produce cells such as Total Cost (Car 1) and Grand Total, the calculation becomes increasingly complex. The functional expressions in these cells, in this case "sum", no longer perform computations on scalar values, but on distributions. Hence, the Total Cost (Car 1) is then obtained as the sum of individual distributions. In the absence of additional visual cues, it becomes difficult to estimate how the uncertainty propagates from independent cells (Spare Parts A & B (Car 1)) to the cells which depend on them (Total Cost (Car 1)). Furthermore, the car owner decides to include the time it would take her to inspect and later fix or remove each part to avoid paying a professional car mechanic. She enters the time required for each part. However, the time spent on each part varies. Since she needs to spend time on inspecting parts regardless of a potential failure, she wants to model the time spent on each part with normal distributions (as opposed to the costs, which she modeled with Bernoulli distributions). She does so by specifying a standard deviation for each part in a column adjacent to the estimated mean. The Total Time spent is then computed by adding the influencing distributions. She now has to monitor not only the uncertainty in cost but also in time. Using only the basic functionality provided by typical spreadsheet programs, it becomes difficult to keep track of the distributions and how they propagate throughout the spreadsheet. The Fuzzy Spreadsheet approach supports users in achieving these tasks.

In some cases, it may be relevant to analyze the impact a cell has on another one, for instance, to identify highly influential cells such as Total Cost (Car 1) and its impact on the Grand Total. Conceptually, a cell's impact can be defined as how much its value contributes to the value of another cell. In some cases, it is sufficient to simply look at the values of the cells to identify those with the highest impact. However, determining the exact impact a cell has on another one requires calculation. We compute this impact as a percentage of how much the value x_i of cell i contributes to the value x_j of cell j. If i is a first-degree influencing neighbor of j, i.e., $i \in \mathcal{N}_1(j)$, then we define its impact as:

$$\mathscr{I}(i \rightarrow j) = \frac{\sigma \cdot x_i}{\sum_{k \in \mathcal{N}_1(j)} |x_k|}. \tag{1}$$

Here, $\sigma = +1$ if i contributes to j via a summation, and $\sigma = -1$ if i contributes to j via a difference. The total relative sum of the impact percentages of each contributing cell c_i must be 100 %. This direct computation involving simple addition or subtraction is straightforward, but longer calculations involving intermediate results, such as a difference of sums, can make assessing the impact challenging. We categorize the nature of this impact as positive or negative based on the sign and relationship of cell j with i. In a spreadsheet, an additional column may be created to compute the impact that value x_i in cell i would have on value x_j of cell j. For a cell i, as its impact value influences j, j must be re-evaluated every time i changes.

Finally, it can often be helpful to explore the results of hypothetical changes to the spreadsheet to answer questions

TABLE 1
Terminology used for describing uncertainty and other relevant concepts.

Term	Definition
Bernoulli probability	Probability $p = P(X = A)$ that a variable X modeled via a Bernoulli distribution assumes the value A
Relationship	Relationship between a cell and other cells based on the direction of influence (input/output) and the degree of neighborhood
Degree of Neighborhood (DoN)	Distance between two cells in the computational graph
Impact	Contribution of a cell value to the value of another cell, expressed as percentage

such as "What would happen to the Grand Total Cost if I were to change the chance of failure of Spare Part A of Car 1 from 0.75 to 0.5?" The model must be able to incorporate the new uncertainty for Spare Part A of Car 1 and propagate the change through the computational graph. The analytical concept of a *what-if analysis* often involves changing the certainty of values to create and analyze different scenarios. Such an analysis requires considerable mental effort if intermediate changes are not tracked and no visual clues are provided. With Fuzzy Spreadsheet, we aim to support users in maintaining their mental models during calculations involving uncertainty. Brief descriptions of the key terms introduced in this section are given in Table 1.

3 RELATED WORK

Uncertainty is inherent in data collection, processing, and sensemaking. A recent survey showed a consensus among visualization researchers on the importance of using visualization to communicate not only the data itself, but also the underlying uncertainty [25]. Exploring and understanding variabilities may have a considerable impact on the outcome of decision-making [3], [28]. Research into how uncertain values can be visualized effectively has been ongoing for centuries [4], [40]. Various uncertainty visualization techniques have been applied in a diverse range of fields, such as medicine [32], geoscience [47], business intelligence [54], and fishery [43]. As technology develops rapidly, tools are proliferating, but current applications focus on showing complex information and do not make use of the familiar spreadsheet layout to present different scenarios—such as worst, best, and expected case—in a way that is easy to understand and visualize [41], [44]. With our work, we aim to deepen the understanding of uncertain values and their use in tabular calculations.

3.1 Uncertainty Visualization

Uncertainty can be visualized for zero-, one-, or higher-dimensional data [44]. Sanyal et al. [46] introduced a framework for data dimensionality (0D, 1D, 2D, 3D), visualization approaches (scalar, vector, tensor), and uncertainty visualization techniques (e.g., blurring, transparency, noise). Typical techniques for showing zero-dimensional data (after decoupling of the temporal dimension) include changes in glyph size to indicate a variation in the data points [5]. As with one-dimensional data, line blurring and transparency changes are common techniques to indicate the level of certainty [46]. Further, probability density functions [9], [13]

are used to represent uncertainty, for instance, by line and bar charts [48], [53], dotplots [14], violin/boxplots [23], and heatmaps [24].

3.2 Spreadsheet-based Tools

We investigated multiple applications, such as Web-Charts [15] and Google Visualizations [8], which allow users to create and use web applications in spreadsheets. However, we decided to limit our discussion to spreadsheet-based tools that address the encoding and propagation of uncertainty information.

Streit et al. [49] proposed a technique for augmenting a spreadsheet with uncertain information. To introduce uncertainty in the calculations, users can specify an interval or range in a cell. The underlying theoretical concept is known as interval arithmetic. The uncertainty information is then propagated to other cells that use these ranges as inputs. To convey to the user that—as an effect of the propagation—these cells now contain intervals, affected cells are highlighted by shading. Finally, these intervals are plotted in a chart that represents the uncertainty.

The FuziCalc tool [11], intended for modeling under uncertainty, followed a similar approach. FuziCalc allowed the user to enter fuzzy inputs in the form of shapes, such as triangles representing best, worst, and expected cases. It then generated a fuzzy output from fuzzy input cells. This way, it reduced the mental effort for users by not exposing them to the complex underlying computations. Since FuziCalc was developed as an independent spreadsheet tool, it missed fundamental features of conventional general-purpose spreadsheet tools. In most cases, it was used in addition to conventional spreadsheet tools and not as a replacement, which resulted in a cumbersome workflow. Since this commercial software is no longer available, it could not be tested.

More recent tools for uncertainty visualization in spreadsheets are Palisade @Risk [39], Oracle Crystal Ball [37], and SIPmath [45]. These tools use Monte Carlo simulation to compute possible outcomes of uncertain events [16]. They are commercially available as extensions for Excel, and allow users to explore different scenarios and provide results in the familiar Excel charts. They share the concept that the user can input a distribution for the cells by using either a formula or a graphical user interface. The user can choose distributions for cells that are not influenced by other cells and compute the model for a *reference cell* (which is usually a cell that is influenced by input cells).

In the case of @Risk and Oracle Crystal Ball, the cells are then colored to indicate that they no longer contain exact but fuzzy values based on probability distributions. More information on a distribution appears upon mouse-over action and click-based interaction. The user can interact with the distribution to analyze the probability of finding a value in a particular range. However, these tools lack an overview of the underlying uncertainty across the spreadsheet, and require a complex modeling process of the distributions for each cell. Furthermore, for users who are not familiar with probability distributions, it is challenging to define and interpret them [19]. In the case of SIPmath, histograms in the form of sparklines are added in the cells to represent the underlying distribution. However, the histograms are stretched in both directions to fill out the cell, which makes it difficult to compare distributions across cells, especially if the values on the horizontal axis are not the same.

A tool that avoids the complexity of the modeling process is Guesstimate [21]. It was developed as an independent web-based spreadsheet tool that allows the user to easily input a fuzzy number as a range, an interval, or a distribution, and immediately provides an output based on the relationship. Relationships between the cells are shown as links connecting the related cells. Similar to Palisade @Risk, Oracle Crystal Ball, and SIPmath, it uses Monte Carlo simulation to compute the propagation of uncertainty. In the course of our investigation, we contacted one of the main contributors of Guesstimate. He informed us that users generally prefer staying close to traditional spreadsheet-based approaches and primarily rely on normal distributions due to a lack of knowledge about applying other distributions, even though further ones are supported. This confirmed our findings in the context of the FuziCalc software that users are less likely to model their use cases in an unfamiliar environment. We found that existing tools for uncertainty exploration with *what-if analyses* are either too complex for a novice user or lack the wide range of features available in general-purpose spreadsheet tools. This inspired us to develop an approach that (1) retains the layout of a traditional spreadsheet and (2) is augmented with compact visualizations to communicate uncertainty.

4 TASK CHARACTERIZATION

We collected the most prevalent user tasks that are supported by various related tools (see Section 3 for a survey) and put these tasks into the context of existing analysis task frameworks [1], [7]. For the purpose of discussing Fuzzy Spreadsheet, we categorize user tasks in two different ways.

First, we assign the tasks to one of two *phases* of the typical spreadsheet workflow: *Authoring* and *Analysis*. In the Authoring phase, users set up the content and structure of the spreadsheets. In the Analysis phase, users seek to understand the results of calculations and draw conclusions for their applications. Complex problems may require users to go through multiple iterated Authoring and Analysis phases.

Second, we label user tasks by the *requirements* they impose on Fuzzy Spreadsheet. In order to facilitate decision-making in the face of uncertainty, Fuzzy Spreadsheet must extend traditional spreadsheets in two ways. To reflect on

the extensions necessary, we define the following three task requirements:

Basic spreadsheet functionality is required for all tasks that can be readily performed in a typical spreadsheet tool without any extensions, such as inputting numbers and referencing cells in formulas.

Computational graph parsing is required for all tasks that involve extracting relationship information from the computational graph underlying the spreadsheet.

Fuzzification of the spreadsheet is required whenever tasks cause a transition from exact numbers to probability distributions.

Not all user tasks need to be assignable to exactly one of these requirements. Some involve only basic spreadsheet functionality, but may lead to fuzzification or enable improved analysis of the computational graph. Other tasks may be equivalent to those performed in non-augmented spreadsheets, but may acquire a new meaning in the context of uncertainty. Table 2 summarizes the user tasks we identified for spreadsheets with uncertainty propagation and lists their requirements and to which phase we assigned them. The following sections describe the user tasks in more detail.

4.1 Authoring Tasks

In the authoring phase, users set up the initial content and structure of the spreadsheets. We extracted three tasks that are essential in this phase. First, users **Tabulate Numbers**, that is, they input their data as exact numbers.

Each value is a potential node in the spreadsheet's underlying computational graph. Second, users **Set Up Relations** by inputting formulas, which introduce links between nodes in the computational graph. Third, users **Specify Uncertainties & Introduce Alternatives**. In this step, essential for the fuzzification of spreadsheets, users specify which cells should no longer be treated as exact, but as fuzzy values. To the users, this task can be equivalent to a *Tabulate Numbers* task, with the important conceptual difference that the input numbers are given a special meaning in the subsequent analysis phase. In particular, input values adjacent to "fuzzified" cells are interpreted as the parameters of underlying probability distributions (see Section 5.1).

4.2 Analysis Tasks

While the authoring tasks are important for setting up content and structure of the spreadsheet, Fuzzy Spreadsheet focuses mainly on enabling new tasks in the analysis phase. The analysis tasks are extracted from the rationale-based tasks identified by Amar et al. [1] and extended to incorporate tasks that played a role in the tools we surveyed. Additionally, we also cover the four search tasks (Lookup, Locate, Browse, Explore) and two query tasks (Identify and Compare) from the Multi-Level Task Typology framework of Brehmer and Munzner [7]. Our discussion of the visual encodings used in Fuzzy Spreadsheet and of the user study results are based on the following analysis tasks:

T1 Look up Values

The most basic analysis task when working with spreadsheets is to look up values in cells. In the same way

TABLE 2
User tasks for Fuzzy Spreadsheet. Tasks are divided into Authoring and Analysis phases, and labeled depending on the requirements.

▦	Tasks requiring basic spreadsheet functionality	
⛓	Tasks requiring parsing of the computational graph	
ⅲ	Tasks requiring fuzzification	
Authoring	• Tabulate Numbers	▦
	• Set Up Relations	▦⛓
	• Specify Uncertainties & Introduce Alternatives	▦ⅲ
Analysis	**T1** Look up Values	▦ⅲ
	T2 Trace Relations	⛓
	T3 Assess Impact	⛓
	T4 Expose Uncertainty	ⅲ
	T5 Formulate Cause and Effect	▦⛓ⅲ
	T6 Compare Probability Distributions	ⅲ

that input tasks can acquire a special meaning if they introduce uncertainty to the spreadsheet, so do lookup tasks that retrieve these special values.

T2 Trace Relations ⛓

Tracing relationships between cells in a spreadsheet is important for understanding the "flow" of computations. As discussed in Section 2, this flow is described by the spreadsheet's underlying computational graph. In classic spreadsheet tools, the computational graph is, for the most part, hidden from the user. Given a selected cell, only the immediate inputs used in formulas in that cell can be readily retrieved. However, *Trace Relations* is a more general, bidirectional task that also includes finding cells that are influenced by other cells. Neither *influencing* nor *influenced by* cells are limited to direct neighbors in the computational graph, but can be extended to a higher degree of neighborhood (DoN).

T3 Assess Impact ⛓

Having located cells that influence another cell, users often need to gain a better understanding of the individual contributions of each of these influencing cells. In Section 2, we introduced the *impact* as a direct measure of how strongly the value of one cell contributes to the result in another cell. In simple cases, such as the addition of multiple values, assessing the impact can be as straightforward as looking up several values and comparing them. If more elaborate formulas are used, and/or in the case of tracing longer paths in the computational graph, assessing impacts is challenging in standard spreadsheet tools.

T4 Expose Uncertainty ⅲ

In the authoring phase, users *Specify Uncertainties & Introduce Alternatives*. As a result, computations in the spreadsheet can no longer be based on exact numbers but require probability distributions. Users must be able to assess these distributions, preferably directly

in the spreadsheet cells without auxiliary calculations. Depending on the level of detail with which they want to analyze these probability distributions, users must perform a number of low-level visualization tasks, such as assessing shapes, finding extrema, and estimating areas under the curve.

T5 Formulate Cause and Effect ▦⛓ⅲ

While adding uncertainty to spreadsheets allows users to make more informed decisions, it also makes interpreting the relationships between cells in a spreadsheet more challenging. As described in Section 2, a *what-if analysis* is a powerful approach that allows users to safely explore the results of hypothetical changes in an attempt to better understand which values are important. To enable this, users must be able to change the spreadsheet reversibly. When users end their exploratory analysis, they can decide to either keep the new results or revert back to the initial values. Hypothetical changes typically include changing the parameters of distributions while keeping the topology of the computational graph unaltered. Given the density of data in typical spreadsheets, making reversible changes might not be enough for an effective *what-if analysis*. Users must be able to see how a change in one cell propagates through the computational graph and how each change leads to additional, implicit changes in other cells. Explicitly addressing this subtask by visualizing changes prevents users from having to *Trace Relations* (**T2**) repeatedly in the course of a *what-if analysis*.

T6 Compare Probability Distributions ⅲ

In later stages of the analysis, most of the intermediate and final results of calculations in the spreadsheet are modeled by probability distributions. This means that comparing two or more cells requires comparing two or more probability distributions. We refer to this as *between-cell* comparison. Additionally, hypothetical changes in a cell during a *what-if analysis* may entail comparing old and new results within a single cell. We refer to this as *within-cell* comparison.

Fuzzy Spreadsheet addresses most of these analysis tasks by introducing compact visualizations directly in the spreadsheet cells, with additional detailed information in a separate side panel. The *between-cell* comparison is supported by in-cell encodings. Fuzzy Spreadsheet allows users to compare probability distributions of two or more cells by means of individual encodings presented within the cells. Detailed information about each cell, such as the probability distribution with computed mean and standard deviation, can be viewed in the side panel upon selecting a cell. During a *what-if analysis*, the encodings of the old and the new probability distributions are stacked within a cell, which facilitates a *within-cell* comparison. The within-cell changes can also be viewed in the side panel.

5 FUZZY SPREADSHEET TECHNIQUE

We designed Fuzzy Spreadsheet as an extension to traditional spreadsheet programs, users can incorporate our solution into their familiar workflows. Fuzzy Spreadsheet comprises three parts: (1) functionalities for parsing the
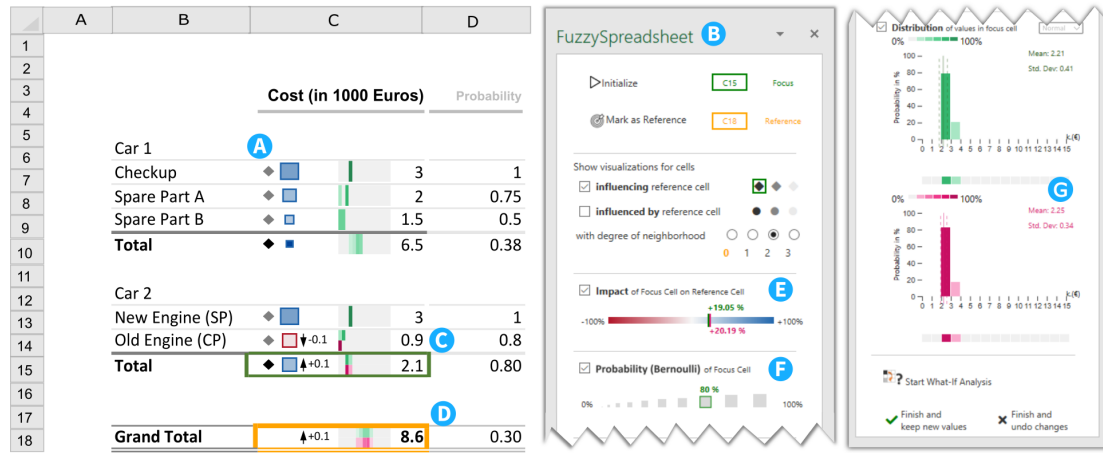
Fig. 3. Fuzzy Spreadsheet applied to a subset of the car-maintenance example in Figure 1. Visualizations are embedded directly in the cells (A) to indicate impact ■■□□■, Bernoulli probability ▪▪■■■, relationship ◆◆◇ ●●○, and probability distributions ▮▮ ▮. The side panel (B) allows users to control the visualizations. Users select a cell (C) and explore its influence on a reference cell (D). The active legends (E & F) indicate the original values encoded by the impact and Bernoulli visualizations (green markers) and new values (dark pink markers) obtained during a *what-if analysis*. Here the user reduced the value of *Old Engine* by 0.1. The detailed distribution charts (G) for the original and the new values of (C) are also displayed, and the probability ranges can be viewed by hovering over them.

spreadsheet to extract the presence and propagation of uncertain information; (2) carefully chosen, compact visualizations that are embedded directly in the spreadsheet cells; and (3) a side panel with the user interface and additional information. Figure 3 shows an example spreadsheet with some of the Fuzzy Spreadsheet encodings (the traditional spreadsheet version is shown in Figure 1). We first discuss the pre-processing needed for calculating the propagation of uncertainty information, and how it relates to the authoring tasks. We then describe how users can control the in-cell visualizations from the side panel, and explain in detail the design of the visual components.

5.1 Computational Graph Parsing and Uncertainty Calculations

Before relational and uncertainty information can be encoded and interacted with, the relevant content must be extracted from the spreadsheet. To this end, Fuzzy Spreadsheet parses the spreadsheet and its underlying computational graph. Fuzzy Spreadsheet automatically treats cells as fuzzy if they are the result of an "average" calculation. Other cells can be made fuzzy by wrapping their content with Excel's "average" function. Fuzzy Spreadsheet then looks for additional uncertainty information in the adjacent cells. By default, the cell itself is treated as the mean of a normal distribution, and the cell to the right as the standard deviation. In addition, Fuzzy Spreadsheet provides a shortcut for multiplying a value with the result of a Bernoulli distribution by looking for the corresponding probability in the next but one cell. We focused on these two distributions due to a discussion with the creator of the Guesstimate software [21] in which he mentioned that typical users almost exclusively choose normal and Bernoulli distributions for their models. Advanced users can also switch to other distributions via the side panel, which results in the adjacent cells to be interpreted differently depending on the distribution chosen. In addition to normal and Bernoulli distribution, Fuzzy Spreadsheet also supports uniform and Poisson

distributions. We discuss the advantages and limitations of letting users *Specify Uncertainties & Introduce Alternatives* via a simple *Tabulate Numbers* task in Section 9.

In order to perform an analysis task (see Section 4.2), a user needs to choose a *reference cell*—typically a cell of interest whose value depends on other cells. The Fuzzy Spreadsheet extension stores the address of this cell and uses it as a reference for all future computations. This gives the user the possibility of selecting another cell and viewing its details with respect to the reference cell. Thus, Fuzzy Spreadsheet allows the user to focus on two cells at any given time: a *reference cell* and a *selected cell*.

For the analysis tasks, the extension performs the computations required for obtaining impact values and probability distributions, as explained in Section 2. The tool then performs a Monte Carlo simulation to obtain the distributions. Due to computational limitations, we use 100 samples for each simulation. We briefly discuss the scalability of Fuzzy Spreadsheet in Section 9. During a *what-if analysis*, after a cell value is changed, an automatic parsing of the computational graph is performed to derive updated results. If the newly obtained results differ from those of the initial parsing process, the cell encoding for the probability distribution and the side-panel visualization for all selected options is updated.

5.2 Side Panel

The *side panel*, as shown in Figure 3(B), provides the control interface for our extension. To start working with Fuzzy Spreadsheet, users need to trigger the parsing and the creation of the computational graph by clicking on the *Initialize* button. Users must then select a cell in the spreadsheet and mark it as the reference cell. The reference cell is automatically highlighted with an orange border in the spreadsheet (see Figure 3(D)), and its address is shown in the same color in the upper-right corner of the side panel.

Next, users select which kind of cells they are interested in: cells *influencing* the reference cell or cells that are *influ-*

enced by the reference cell. Subsequently, they choose the degree of neighborhood (DoN) up to which they want to explore related cells. A DoN of zero refers to the reference cell itself. A DoN of one corresponds to the direct neighbors (connected by links in the computational graph) of the reference cell, excluding the reference cell itself. A DoN of two corresponds to the direct neighbors of the direct neighbors, including the cells with a DoN of one. Up to this point (DoN of zero), no visual changes have appeared in the spreadsheet. To indicate that a DoN of zero limits the selection to the reference cell itself, the zero option is colored in the same orange hue that is used for the borders of the reference cell. This neighborhood-selection dialog simultaneously serves as a legend for the relationship encoding (see Section 5.3.1). We refer to this dialog/legend hybrid as *active legend*. In our Excel extension prototype (see Section 6), we use a dark green shade for all information related to the selected cell to keep the design consistent with Excel's default color setting for the border of the selected cell (see Figure 3 (C)).

Based on this information, users can display visualizations of impact, Bernoulli probability, and distribution for the selected and related cells on demand. At any point, they can change the DoN, which controls the overall amount of information displayed. Furthermore, users can start a *what-if analysis* by activating a button at the bottom of the side panel. We explain for each visualization separately how this analysis mode affects the encoding.

5.3 Compact Visual Encoding

Given the small cell size in traditional spreadsheets, we had to evaluate encodings that work effectively in this constrained space. Sparklines and glyph-based encodings [5] are one option. Nobre et al. [36] introduced an aggregated layout in tabular cells that could be adapted to show fuzzy values within a cell. Box plots, heatmaps, sparklines, histograms, and dotplots used in Taggle by Furmanova et al. [17] are also possible choices for in-cell encodings. All these visualizations inspired our final encodings. The cell encodings are designed to give overview information about the reference cell and its related cells—serving as an initial guide to identifying critical cells. We justify our final choice of encoding for each of the suggested visualizations in individual sections below.

5.3.1 Relationship Encoding ◆ ◆ ◆ ● ● ●

To enable *Trace Relations* tasks (**T2**), each cell of interest is equipped with a relationship marker. The shape of this marker encodes the direction of influence with respect to the reference cell. A diamond marker indicates a cell that *influences* the reference cell, and a disk indicates a cell that is *influenced by* the reference cell. We chose these shapes to avoid cross-talk with the impact and Bernoulli encodings. The side panel also emphasizes this information by outlining the marker in dark green if a related cell is selected. Additionally, the DoN is encoded in the brightness of the marker.

In the design process, we considered two alternatives to the categorical relationship markers. First, we considered drawing lines that connect related cells, which would have

represented the underlying node-link structure of the computational graph. In Excel, users already have the option to turn on this kind of encoding as an overlay. Jannach et al. [26] also mention research works which used a similar representation for the relationship encoding, such as the work of Chen et al. [10], where differently colored arrows indicated the varying degree of neighborhood. Hermans et al. [22] used a flow diagram to represent the underlying relationship structure directly in the spreadsheet. However, we found that this encoding caused visual clutter and was prone to drastic changes depending on the layout (i.e., positioning) of the cells chosen by the user in the Authoring phase. Second, we considered small arrow-like glyphs or stubs [30] pointing towards or away from the reference cell. While this encoding was easy to understand in simple cases, we discarded it because identical glyphs would have acquired different meanings depending on their locations in the spreadsheet. The current encoding has the advantage that it is fully independent of the spreadsheet layout chosen by the user in the Authoring phase, and the markers make it easy to quickly spot the cells of interest.

5.3.2 Impact Encoding ■ ■ □ ■ ■

Impact markers enable users to *Assess Impact* (**T3**). We add boxes of fixed size on the left side of a cell and color them according to how strongly the value in the cell influences the reference cell (see Figure 3(A)). We use a diverging color scale from dark red (strong negative impact) to white (no impact) and dark blue (strong positive impact) to help users to quickly identify the sign of a cell's contribution to the reference cell. Furthermore, the dark colors at each end of the scale make high-impact cells more salient, while cells with an (almost) white impact marker can be safely disregarded by users when gaining an overview.

The side panel also serves as an active legend for the impact. The active legend consists of a red/blue color scale (see Figure 3(E)). Using a dark green marker, the legend shows—for a selected cell—either its impact on the reference cell or the impact of the reference cell on it. Using the same dark green which is used as the border color of the selected cell, the legend also shows the value of the impact in percent. The position of this marker changes dynamically when the selected cell is changed. To support *Formulate Cause and Effect* tasks (**T5**) during a *what-if analysis*, a second, dark pink marker is shown in the active legend if any of the hypothetical changes affect its value, as indicated in the legend in Figure 3(E).

5.3.3 Bernoulli Encoding ▪ ▪ ▪ ■ ■ ■

Based on our decision to facilitate the use of Bernoulli distributions (see also Section 5.1), we decided to include an additional indicator to highlight values in cells that are the results of "binary scenarios". The area of this marker encodes the probability that all binary outcomes leading to this result have an outcome of one rather than zero. If only the Bernoulli but not the impact visualization is switched on, we show gray markers with corresponding areas. If both Bernoulli and impact visualizations are switched on, we encode the Bernoulli probability in the size of the impact marker and use the color from the impact encoding.

The side panel also functions as an active legend for the Bernoulli probability. The active legend for the Bernoulli probability consists of grey squares placed in ascending order of their area. For a selected cell, we show the probability value (in percent) in the legend and draw a dark green frame around the corresponding square, as seen in Figure 3(F). As for the impact encoding, a second, dark pink marker in the active legend indicates values that changed during *what-if analyses*.

5.3.4 Probability Distribution Encoding

Users can switch on the probability visualization in the side panel to show small, in-cell distribution heatmaps attached to the cells of interest (see Figure 3). The heatmaps serve as a first visual aid for *Expose Uncertainty* tasks (**T4**). Each heatmap is based on a binned histogram of samples drawn from a continuous probability distribution $p(x)$. The value x increases from left to right, and the heatmap is divided into a fixed number of intervals I (i.e., the bins). The range of the heatmap is based on the minimum and maximum cell values present in the spreadsheet, and is the same for all cells to ensure an unbiased comparison. Due to the limited space, the range cannot be shown numerically or quantitatively within a cell. However, as the vertical space in the side panel is not limited, we indicate the range of the heatmap on a shared x-axis with a bar chart for the detailed distribution plot, as shown in Figure 3(G). The horizontal orientation of the distribution heatmaps facilitates comparison between values of cells in the same column (**T6**) and makes optimal use of the space provided by typically sized spreadsheet cells, which is an advantage over SIPmath [45]. The color scale for the heatmaps is based on the color selected (by default dark green in Excel), and each interval is colored according to the binned probability $P(x)$, where $x \in I$. Thus, higher probabilities are represented by darker shades of the base color. Zero is represented by a neutral grey, which lets users distinguish easily between improbable (light-colored) and impossible (gray) values.

Similarly, we color each bar of the detailed distribution plot in the side panel with the same shade of the base color as in the heatmap, and align a copy of the heatmap with the bottom of the bars. This allows users to easily match the heatmap encoding in the spreadsheet with the bar-chart encoding in the detailed view. Mouse-over interactions let users assess exact probability values for each interval. Additionally, we add the mean and standard deviation of the shown samples as numbers and indicate these values by vertical lines in the bar chart. The distribution bar chart in the side panel enables *Expose Uncertainty* tasks (**T4**), for which the heatmap encoding cannot provide sufficient detail.

In the course of a *what-if analysis*, changes made by the users often affect the probability distributions of cells. To let users assess these changes at a glance, we replace each heatmap with a stacked heatmap (see Figure 3(C) or thumbnail at the beginning of this section). The stacked heatmaps allow users to *Compare Probability Distributions* (**T6**) and *Formulate Cause and Effect* (**T5**) directly in the spreadsheet. We show the original distribution heatmap in green in the top half of the cell, and the updated distribution in pink in the bottom half. Both heatmaps have the same value

range as the original heatmap. Similarly, we change the detailed view in the side panel from a single bar chart to two bar charts that represent the original and the updated distributions (see Figure 3(G)). The base hue used in the updated distribution heatmap and bar chart is the same shade of pink that indicates updated values in the active legends for the impact and Bernoulli markers.

In the design process, we also considered using sparklines to visualize the probability distributions, as done in SIPmath [45]. We chose the heatmap encoding instead for two reasons. First, the vertical space in the cells is constrained. As heatmaps encode the values in the color channel, they require little vertical space compared to bar charts or sparklines, which rely on the vertical position for encoding the values. Additionally, during a *what-if analysis* two distributions must often be shown for a cell at the same time. We found that a stacked heatmap encoding uses the space more effectively and with less visual clutter than overlaid sparklines. Second, heatmaps are particularly suited to comparing distributions (**T6**) across vertically neighboring cells [18].

5.3.5 Change Indicators

Not all *Formulate Cause and Effect* tasks (**T5**) necessarily involve tracking the changes of probability distributions in detail. In many cases, it might be sufficient to simply see how the (mean) value of a cell is affected by changes made elsewhere. During a *what-if analysis*, we display small arrows to indicate increases (upward arrow) or decreases (downward arrow) as a result of the changes (see Figure 3(C) or thumbnail at the beginning of this section). Additionally, the value difference is specified next to the arrow.

6 IMPLEMENTATION

For the implementation, we investigated the most commonly used spreadsheet tools and evaluated them according to their official API extension support and feature list. Microsoft Excel emerged as our tool of choice because it provides the Office JavaScript API. We used the Host Specific Office API model for Excel [35], which allows our extension to be loaded as a so-called *add-in* in both the desktop and the web version. The add-in currently supports Office 365 and the Microsoft Edge Web Browser. We implemented the prototype in TypeScript, HTML, and CSS. The source code is available at https://github.com/jku-vds-lab/fuzzy-spreadsheet. A demo version can be accessed at https://jku-vds-lab.at/fuzzy-spreadsheet/. Our prototype uses external libraries, the most notable ones for sampling discrete probability distributions [33], jstat [55] for performing advanced statistical operations, maths.js [12] for flexible expression parsing, and D3.js [6] for creating customized visualizations.

7 USAGE SCENARIO

To demonstrate the applicability of Fuzzy Spreadsheet, let us assume that conference organizers have been assigned the task of planning the budget for an annual conference. Using a simple model, the organizers predict the revenues, expenses, and the balance for the next year as the average of
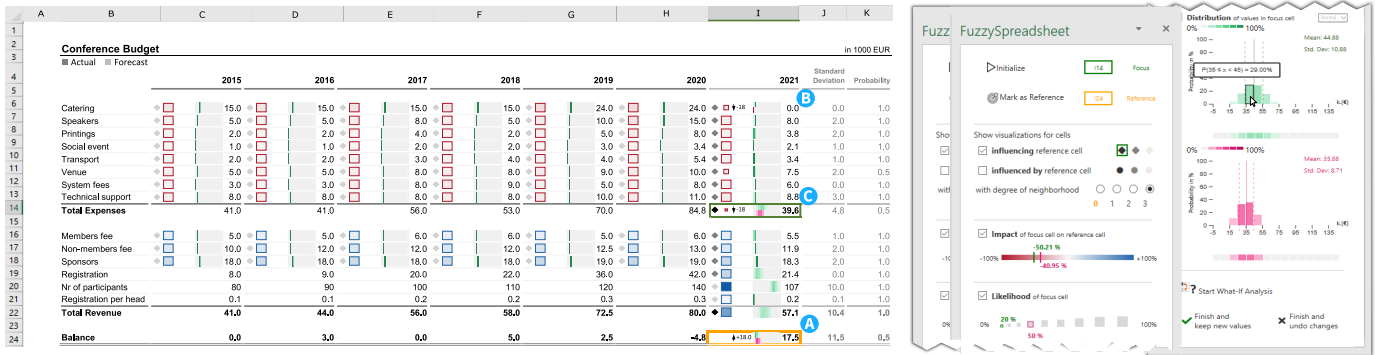
Fig. 4. Conference planning usage scenario with Fuzzy Spreadsheet encodings. The user selects a reference cell (A) and controls the visualization using the side panel. In the *what-if analysis* mode, they reduce the value of Catering 2021 to zero (B) and analyze its effect on the focus cell (C). More detailed information about the focus cell is shown in the side panel (right).

the previous six conferences (see Figure 4). Since planning is prone to uncertainty, decisions on the variation of all uncertain values must be made and conveyed in the spreadsheet. Since the spreadsheet itself does not offer much support for dealing with uncertain data, the organizers decide to use the Fuzzy Spreadsheet extension. Due to its easy modeling requirements, they simply create two additional columns to enter the uncertainty parameters, namely the standard deviation for normally distributed values and the Bernoulli probability to model binary events. The organizers expect a probability of one (100 %) for all uncertain values except for Venue and Catering. These two costs will only occur if the conference takes place on site. Since unexpected worldwide events have recently led to travel restrictions, they estimate the probability for an on-site conference—and therefore for the Venue and Catering costs—as 50 %.

To visually analyze the effect of uncertainty on the balance and the intermediate calculations, they load the Fuzzy Spreadsheet extension. Conference organizers tend to aim for a slightly positive balance, using any surplus for the next conference or better catering. The organizers start the analysis by clicking on the *Initialize* button, which automatically parses the computational graph of the spreadsheet, and mark the summed balance as the reference cell (see Figure 4(A)).

The organizers can now decide which visual analysis support they want to activate in the side panel: relationship, neighborhood, impact, Bernoulli probability, and/or distribution. To determine which cells influence the uncertainty of the balance (**T4**), they first display the influencing cells (**T2**) and change the DoN to two. The glyphs that appear within the cells indicate the direction of influence by shape and the DoN by color (◆). Subsequently, to analyze the probability ranges of expenses, the organizers activate the distribution visualizations. Related cells then show the heatmaps of the distribution for the expenses. Upon analyzing the distribution, they realize that the value of the balance as given in the spreadsheet is not very likely, and the computed mean—based on correct uncertainty propagation—is negative.

To evaluate how changes in the probabilities for the expenses affect the balance, they enter the *what-if* analysis mode and choose a DoN of three to view the changes in all influencing cells. Their goal is to compare the scenario of an on-site conference with the alternative of a virtual event.

To reflect the expected budget changes in the case of a virtual conference, the organizers start by reducing the expenses of venue and catering to almost zero with a high probability (see Figure 4, where the value of Catering is reduced to zero). This new value causes changes in the distributions of Catering, Total Expenses, and Balance. They inspect the new distribution in the lower stack of the heatmap present in each of these cells and keep an eye on the side panel, as shown in Figure 4(C). To reflect the increased technical support needed for running the event virtually, they increase the corresponding cost. Since no local advertising is necessary and venue/catering expenses are reduced, conference sponsors will also pay less. Based on this information, the organizers receive a more reliable (i.e., narrower) range for the balance. Additionally, registration costs per person are also to be reduced. After decreasing sponsorship amount and registration cost, the organizers try to increase the number of participants to make sure that the balance lies within a positive range, while being insensitive to minor fluctuations. Once the organizers are satisfied with their adapted planning, they decide to keep the updated values and save the changes in a separate file. The in-cell visualizations are preserved in the new file. They may choose to share this file with their colleagues, who can locally view the encodings even in the absence of the Fuzzy Spreadsheet add-on.

Once familiar with these encodings, such as the relationship encodings for the influencing cells, it becomes easy to grasp the underlying computational graph. The impact encoding helps to identify the extent and nature of the influence a cell has on the reference cell. The changes made during the *what-if analysis* present in the form of change indicators provide an overview of what has changed and by how much. The stacked heatmap encodings in changed cells indicate how a value has become more certain. After sharing the findings, the organizers continue the analysis of the on-site conference scenario in the original file. In this case, the probability for expenses such as Venue and Catering to occur can be increased, and the deviation in the amount of sponsorship can be decreased. This changes the probability distribution of the balance, which in turn makes the balance number more certain.

This usage scenario highlights the benefits of using Fuzzy Spreadsheet: hypothetical scenarios can be analyzed based on immediate visual and numerical feedback. This is especially important in the case of high-stakes investments, where changes are crucial to ensuring a positive balance. The *what-if analysis* helps to determine sensitive values and to obtain a stable range of values for the final outcome without limiting the number of possible scenarios that can be investigated. Thus, Fuzzy Spreadsheet can improve the outcome of the decision-making process.

8 EVALUATION

To evaluate the usefulness of Fuzzy Spreadsheet, we conducted a user study. The goal of this study was to compare the performance of a standard spreadsheet tool with that of the Fuzzy Spreadsheet approach for the six analysis tasks summarized in Table 2. We chose Microsoft Excel as a representative spreadsheet tool, assuming that alternatives such as Google Sheets and Apache OpenOffice Calc would perform similarly. As Fuzzy Spreadsheet is built for casual users and is not meant to compete with specialized tools for expert analysis, such as Oracle Crystal Ball and @Risk, we did not choose them for the user study. Our main objective is to keep the familiar spreadsheet layout and to add uncertainty exploration. Therefore, we also dropped Guesstimate as a comparison tool, as it lacks spreadsheet features and focuses on the uncertainty analysis only. To avoid a learning effect, we used a between-subject design for comparing the two conditions (i.e., Excel with and without our extension). We conducted three pilot studies to test the functionality of the Fuzzy Spreadsheet, to fine-tune the levels of difficulty of the tasks, and to estimate the time needed for each participant. We incorporated the feedback regarding functionality and usability from the pilot into our prototype. To reduce the time per subject, we decided to reduce the number of questions to two per task, with additional two questions for the more high-level task **T5**. A dataset was chosen that could easily be understood by all participants. The context of the dataset was conference planning, as for the dataset described in Section 7. It contained three columns of forecast values with their corresponding Bernoulli probabilities and standard deviation values and fewer parameters than described in Section 7. We conducted the main study remotely with 14 participants (P1–P14; Gender: m = 10, f = 4; Age: $M = 28.21$, $SD = 4.32$), seven of whom were assigned to Excel and seven to the Fuzzy Spreadsheet prototype. Two instructors were present throughout—one moderator guiding the participant through the experiment and a second one for taking notes. Note that these studies were performed with some slightly different choices for the visual encodings. The initial design of the change indicators consisted of red and green glyphs that—despite of their different shapes—could be hard to distinguish for people with color vision deficiency, which is why we changed them for the current version. Additionally, we fixed a minor error in the color schemes of the heatmaps.

We used a mixed-method approach to compare how both tools perform at each task in terms of three quantitative measures: (1) the ratio of correct answers in percent; (2) the response time in seconds, and (3) the mental effort

required [38], measured on a seven-point Likert scale ranging from one (very low mental effort) to seven (very high mental effort). In addition, we assessed the System Usability Scale (SUS) score based on ten questions developed by Brooke et al. [27, p. 189ff]. Furthermore, we analyzed the thinking-aloud protocol, the feedback from questionnaires, and the follow-up interviews.

To assess the usability of Fuzzy Spreadsheet, we formulated four hypotheses, as listed in Table 3. H1 investigated the differences between a traditional spreadsheet in Excel and a Fuzzy Spreadsheet with regard to answer correctness. We expected Fuzzy Spreadsheet—with its in-cell visualizations and the active legend—to show a higher answer correctness than Excel (H1). H2 evaluated the influence of our prototype on the response time. We presumed that, with the additional support provided by Fuzzy Spreadsheet, a shorter response time can be achieved. With H3, we tested whether Fuzzy Spreadsheet or traditional Excel requires a higher mental effort. Given that our approach uniquely combines visual support and calculation of changes (*what-if analyses*), we hypothesized that the mental effort required is lower with Fuzzy Spreadsheet than with Excel. With H4, we examined statistical differences between the two systems regarding the SUS score. Based on these hypotheses, we expected Fuzzy Spreadsheet to achieve a higher SUS score than Excel.

8.1 Procedure

To make the results comparable, we followed a similar procedure for both conditions, as illustrated in Figure 5. On enrollment in the study, each participant received an explanation of the study goals and the procedures, and was given access to the dataset loaded into the spreadsheet. This introduction also incorporated the nomenclature and the formulas used. Note that for the user study we referred to the probability parameter of the Bernoulli distribution as "likelihood", because we found that the name "Bernoulli" was not commonly known (despite participants being familiar with the concept of modeling binary scenarios with probabilities). Participants from both conditions received similar introductions, including, for instance, how to perform a second DoN task. The slides we used for onboarding are part of the supplementary material [52]. For the Fuzzy Spreadsheet condition, we provided an additional explanation of the specific visual encodings of Fuzzy Spreadsheet in the side panel and the in-cell visualizations for relationship, impact, Bernoulli probability, and distribution. Further, for the Fuzzy Spreadsheet participants, we installed our extension on the participants' computers and verified its functionality by means of a test file. After successful set-up and introduction, we asked the participants to perform the six analysis tasks. To avoid a selection bias, tasks were assigned in random order. In the course of the study, we recorded (1) externalized knowledge in a thinking-aloud protocol and (2) the mental effort required for each task. After completion of the tasks, we gathered data for the SUS score and demographics in a survey, and performed semi-structured interviews to gain further insights into the usage of uncertain calculations. We closed each session with a debriefing.

TABLE 3
Overview of the hypotheses tested in our user study, with ✓ and ✗ indicating accepted and rejected hypotheses, respectively.

| #	Hypothesis	Accept / Reject
H1	Fuzzy Spreadsheet shows a statistically significantly higher answer correctness than the traditional spreadsheet.	✓
H2	Fuzzy Spreadsheet shows a statistically significantly lower response time than the traditional spreadsheet.	✓
H3	Fuzzy Spreadsheet requires a statistically significantly lower mental effort than the traditional spreadsheet.	✓
H4	Fuzzy Spreadsheet shows a statistically significantly higher SUS score than the traditional spreadsheet.	✓
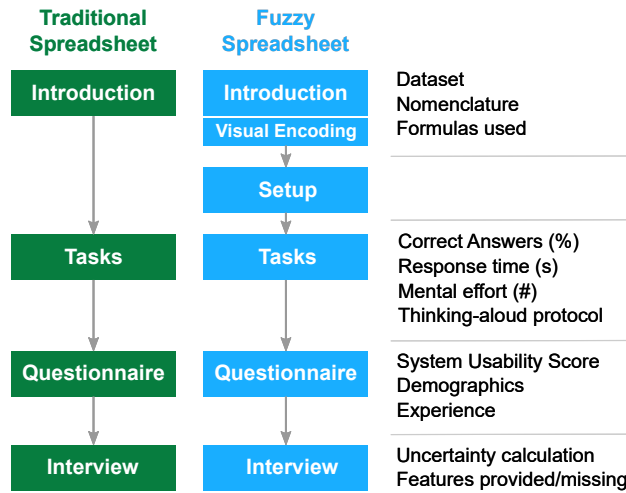
Fig. 5. Study procedure for both systems, ■ Excel and ■ Fuzzy Spreadsheet, starting with an introduction followed by task completion, a questionnaire, and an interview.

8.2 Setup

As both conditions were based on Excel, we created one worksheet per question in a single file. For a fair comparison, we developed a Visual Basic for Applications (VBA) script that recorded the answers, checked for correctness, and recorded the time until answer submission. Response-time tracking started with the activation of a sheet and ended when a participant submitted an answer. In addition, the script stored the Likert scale value for the mental effort required. Since this was a remote laboratory experiment, we performed and recorded each session using the video conference platform Zoom [58]. The moderator shared the screen to introduce the prototype features to the participants. During the study, we asked the participants to share their screens so we could observe their actions. If the installation of the extension did not work properly due to incompatible system requirements, participants were asked to connect to the moderator's computer via TeamViewer [50]. We employed the online and open-source survey tool LimeSurvey [31] for scoring the ten items from the System Usability Scale, for capturing the demographics, and for checking the participants' levels of experience with spreadsheet tools, data visualization, and data science.

8.3 Study Results

To assess differences between the two conditions, we used a χ^2 independence test for our dichotomous variable an-

swer correctness. The effect size was ascertained using Cramér's V. Further, to analyze differences in the response time, in the mental effort captured with a seven-point Likert scale and in the SUS scores, we applied a Student's t-test and use eta-squared (η_p^2) for the effect size. The results at task level were examined using a Mann–Whitney U test for answer correctness and a one-way Analysis of Variance (ANOVA) for both response time and mental effort required. Further, we analyzed behavioral observations from the user studies and the thinking-aloud protocol.

8.3.1 Tool Comparison

First, we analyzed the significance of differences between the two tools in terms of answer correctness. We ascertained that answer correctness with Fuzzy Spreadsheet was statistically significantly higher than that of the standard Excel condition ($\chi^2(1) = 21.448$, $p = .000$, $V = .334$); as a result we can accept H1. Based on our analysis of response times, we can also accept H2. With Fuzzy Spreadsheet, the mean response time was 102.21 s ($SD = 57.90$ s), which is significantly shorter than the response time for the Excel tool ($M = 170.60$ s, $SD = 142.54$ s, $t(184) = -4.286$, $p = .000$, $\eta_p^2 = .091$).

Subsequently, we took a closer look at the subjective measure of mental effort required and investigated the difference between the two tools. There was a significant difference in mental effort required for the Fuzzy Spreadsheet ($M = 2.94$, $SD = 1.23$) and the traditional spreadsheet ($M = 3.44$, $SD = 1.77$) conditions ($t(188) = -2.260$, $p = .025$, $\eta_p^2 = .026$), leading us to an acceptance of H3. These results coincide with our observation that Fuzzy Spreadsheet was also used as an affirmative support tool. Participants first sought to calculate the results in their heads before confirming them with the extension (e.g., P2 for **T6**; P4 for **T1**). Thus, the mental effort was perceived to be lower with Fuzzy Spreadsheet than with Excel. In general, participants in both groups tried to rely on prior knowledge to answer questions about the financial model without recourse to the extension. Further, participants from the Excel control group made supplementary calculations, either directly in Excel or with external tools, such as calculators or Wolfram Alpha [56]. Participants using Fuzzy Spreadsheet scored statistically significantly higher on the System Usability Scale ($M = 78.57$, $SD = 15.27$) than participants using a standard spreadsheet ($M = 48.57$, $SD = 15.93$, $t(12) = 3.597$, $p = .004$, Cohen's $d = 0.19$). Hence, we can also accept H4. According to the SUS scores, Fuzzy Spreadsheet was classified as a *good* and the traditional Excel as an *awful* tool for the given tasks.

Fig. 6. Differences in answer correctness, response time, and mental effort between ■ Excel and ■ Fuzzy Spreadsheet for each analysis task (**T1**–**T6**).

In summary, Fuzzy Spreadsheet had a higher response accuracy (with a large effect size), shorter response time, required less mental effort overall, and was rated with a higher SUS score than Excel.

8.3.2 Task Comparison

We noticed that some tasks were performed more easily with one tool than with the other one. Thus, we decided to assess the applicability of our spreadsheet augmentation in a fine-grained analysis. Based on our quantitative measures, we determined differences between the six tasks.

As indicated in Figure 6, Fuzzy Spreadsheet outperformed the traditional spreadsheet in all tasks. Especially the low response accuracy below 50 % for **T2** and **T4** in the Excel group is noticeable. On closer examination, we found that participants took statistically significantly longer ($M = 351.47$ s, $SD = 196.84$ s) for **T3** with the traditional spreadsheet than for the other tasks ($F(25, 13, 597) = 18, 540$, $p = .000$, $\eta_p^2 = .329$). This is consistent with our observations, the thinking-aloud protocol, and the large effect size. Participants using Excel without Fuzzy Spreadsheet stated that an impact in percent "is difficult to compute, especially in the case of higher DoNs" (P6–P10). We also determined—by means of a post-hoc Tukey HSD test—that *Assess Impact* (**T3**) was not only the task that showed the longest response time, but also the task which required the highest mental effort for Excel users ($M = 5.46$, $SD = 1.61$). All participants in the control group used auxiliary calculations to answer questions related to this task, while participants from the Fuzzy Spreadsheet condition used the side panel as an additional confirmatory tool. In contrast, Look up Value (**T1**) tasks were perceived as the least demanding by Excel users.

Overall, Fuzzy Spreadsheet led to higher response accuracies, shorter response times, and required lower mental effort (except for **T1**).

8.3.3 Behavioral Patterns

To further describe differences between the Excel and the Fuzzy Spreadsheet condition, we analyzed the thinking-aloud protocols and the interviews. We discuss behavioral patterns that we observed while participants performed the tasks.

Observations of the control group showed behavioral patterns with the traditional spreadsheet, which we had anticipated and addressed with Fuzzy Spreadsheet. For instance, we observed that all participants from the Fuzzy Spreadsheet condition activated the relationship and neighborhood encoding for each task by default. In contrast, participants who used the Excel tool had "difficulties in incorporating the probability and further determining the impact" (translated from German) (**T3**, but also true for **T4** and **T6**). In particular, P10 mentioned that it was unclear "how one can determine the impact of a neighbor by taking both probability and standard deviation into account". This can be traced back to a generally reduced understanding of probability calculations (P8, P9) [19] or understanding problems of uncertainty propagation (P1, P4, P5, P6, P9, P11) [57]. "If the likelihood rises, the value will be more certain [...] but how about the mean value? If I think about the residual probability being equal" to zero, I assume that the mean will increase (P11 for **T6**). The encoding selected to indicate *influencing* and *influenced by* cells—glyphs with a gray value for the DoN—was received particularly well and considered to be very helpful. However, some participants (P1, P2, P14) were confused by the placement of the glyphs on the far left edges of the cells. This behaviour was observed especially for **T1** and **T2**. If only the relationship glyphs were displayed and no other visual support was enabled, participants tended to associate the glyphs with the cell to the left rather than the selected cell itself. Interestingly, observations from the Excel control group revealed that two participants (P8 and P11) introduced similar highlighting for related cells themselves. They used different shades for the DoN to facilitate fast detection of coherent cells. P11 from the Excel group even colored reference and selected cells in a consistent color scheme across all tasks, which is reminiscent of our highlighting approach. Participants who did not mark important cells needed longer to complete the task. These observations support our design choice of introducing distinct colors for the different cell types.

We further gained insight into how Excel users performed *Formulate Cause and Effect* tasks (**T5**). In the course of the analysis, two participants (P8 and P10) realized that they forgot to write down the initial values. "Well, now I should know what was there before I changed a value" (P10). Participants typically took notes next to the financial dashboard to reliably track and compare value changes (P8, P9, P10) and to summarize their insights (P8, P9). This way, they were able to compare an initial value with a newly calculated one. Furthermore, in most cases participants entered initial values in hard-coded form (P6, P7, P8), regardless of the level of familiarity with Excel. This static note-taking approach made it difficult for participants to compare and track changes, particularly for tasks that included secondary calculations or even an increased degree of neighborhood.

8.4 Summary of Findings

Fuzzy Spreadsheet makes it easier to understand the information about the underlying relationships between cells and to explore the propagation of uncertainty. Most importantly, the user study indicates that the correctness of answers can be influenced significantly by computational and

visual support. Necessary functionalities that we anticipated in our Fuzzy Spreadsheet approach and that are available only to a limited extent in traditional spreadsheets were well received. In particular, the interviewees considered the glyph encodings applied to the impact (P1, P2, P4, P5), directed relationships with neighborhood information (P2, P4), and the *what-if analyses* for assessing variations (P3, P5) as the key features of Fuzzy Spreadsheet. Thus, information is immediately and transparently shown to the user, which otherwise remains hidden. In summary, learning from our results, we recommend holding on to the familiar tools of the spreadsheet and augmenting these established designs judiciously and with great care to support the encodings of visuals.

9 DISCUSSION

In this section, we discuss conceptual limitations of the overall approach and technical limitations of the prototype implementation.

Supported Operations: For this work, we focused on making the visual encoding and the uncertainty propagation easy to understand. We concentrated on supporting the basic mathematical operations of addition, subtraction, multiplication, and average, and we plan to extend this set of functions in future work. For certain functions, such as division or the square root, a "fuzzy" implementation requires treatment of potentially diverging or undefined results caused by individual samples. More elaborate Excel functions, such as those used in regression modelling, would require completely new implementations to correctly account for uncertainty. Additionally, Fuzzy Spreadsheet currently does not support users in tracing relations (**T2**) across worksheets (i.e., when a calculation depends on cells from a different tab within the same file).

Uncertainty Authoring: As described in Section 4.1, the current extension assigns specific meaning to the cells to categorize them as uncertain cells, with some default settings that facilitate the use of normal distributions and Bernoulli distributions. One advantage of this input scheme is that relevant uncertainty parameters are always visible to the users and can easily be edited [2]. Furthermore, fuzzified spreadsheets can easily be shared between users without having to resort to additional configuration files or automatically created helper sheets. However, this simple input scheme restricts the users' freedom during the authoring phase and may necessitate considerable edits when they want to fuzzify existing spreadsheets. We did not address these shortcomings, since we first decided to focus on the analysis phase with given pre-authored spreadsheets. An obvious future improvement of Fuzzy Spreadsheet would be a more advanced authoring scheme in which the uncertainty parameters are directly attached to a single cell and stored in the background by the extension (as it is the case in Oracle Crystal Ball, @Risk, and Guesstimate). In this case, retrieval and editing of parameters could be incorporated into the side panel.

Scalability: The fact that Fuzzy Spreadsheet is implemented as a Microsoft Excel extension leads to particular limitations regarding its scalability. The main concern in terms of computational complexity is the number of samples used to estimate the probability distributions. The computation time per fuzzy cell increases linearly from ~ 0.5 ms for 100 samples to 6 ms for 10,000 samples. This means that for large spreadsheets with several thousands of cells, even a rough estimate of the distributions may take several seconds. Another performance issue is related to the way Excel draws the visualizations. If the distribution visualization is switched on for several hundred fuzzy cells at once, Excel can become unresponsive until all visualizations have been drawn. This can take up to several seconds, but can typically be avoided by setting the DoN properly. Both the computational performance and the drawing issues become worse in the online version, where the auto-save feature interferes with the extension. The parsing step is typically of no concern: extracting all fuzzy values and relationships from a spreadsheet with 1400 cells takes approximately half a second using the Excel desktop version. All timing experiments were performed on a standard laptop computer.

Fuzzy Spreadsheet currently provides a responsive and smooth experience for small to medium spreadsheets with up to several hundreds of cells. The usability for larger spreadsheets would not only benefit from computational optimization, but perhaps also from additional visual encodings. Identifying suitable strategies to quickly provide users with a visual overview of uncertainty remains an important challenge for future work.

10 CONCLUSION

Tracing uncertainty in spreadsheets is a challenging task that arises in many application contexts. In this paper, we have presented Fuzzy Spreadsheet, an augmentation approach that adds in-cell visualizations to communicate sensitivity and robustness information while staying as close to the familiar spreadsheet layout as possible. Fuzzy Spreadsheet allows users to track the propagation of uncertainty information through a spreadsheet and to compare alternative scenarios as part of *what-if analyses*.

To evaluate the efficacy of our solution, we performed a small-scale user study that compared the Fuzzy Spreadsheet approach with traditional spreadsheets in terms of answer correctness, response time, mental effort, and usability. In summary, the results indicate that Fuzzy Spreadsheet outperforms traditional spreadsheets and empowers users to carry out tasks related to tracking and exploring uncertain information more effectively.

the Federal State of Upper Austria (Human-Interpretable Machine Learning). This paper was partly written in collaboration with the VRVis Competence Center. VRVis is funded by BMVIT, BMWFW, Styria, SFG and Vienna Business Agency in the scope of COMET—Competence Centers for Excellent Technologies (854174) which is managed by FFG.

REFERENCES

[1] R. Amar and J. Stasko. A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. In *IEEE Symposium on Information Visualization*, pp. 143–150, 2004. Citation Key Alias: amar_knowledge_2004. doi: 10.1109/INFVIS.2004.10

[2] L. Bartram, M. Correll, and M. Tory. Untidy Data: The Unreasonable Effectiveness of Tables. *arXiv:2106.15005 [cs.HC]*, 2021.

[3] J. Bilcke, P. Beutels, M. Brisson, and M. Jit. Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models: A Practical Guide. *Medical Decision Making*, 31(4):675–692, 2011. doi: 10.1177/0272989X11409240

[4] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and State-of-the-Art of Uncertainty Visualization. In C. D. Hansen, M. Chen, C. R. Johnson, A. E. Kaufman, and H. Hagen, eds., *Scientific Visualization*, pp. 3–27. Springer, London, 2014. Series Title: Mathematics and Visualization.

[5] R. Borgo, J. Kehrer, D. H. S. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. *Eurographics 2013 - State of the Art Reports*, 2012. doi: 10.2312/CONF/EG2013/STARS/039-063

[6] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '11)*, 17(12):2301–2309, 2011. doi: 10.1109/TVCG.2011.185

[7] M. Brehmer and T. Munzner. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions Visual. Comput. Graphics*, 19(12):2376–2385, 2013. doi: 10.1109/TVCG.2013.124

[8] Charts API. Google charts. https://developers.google.com/chart. Accessed: Sep. 1, 2021.

[9] H. Chen, S. Zhang, W. Chen, H. Mei, J. Zhang, A. Mercer, R. Liang, and H. Qu. Uncertainty-Aware Multidimensional Ensemble Data Visualization and Exploration. *IEEE Transactions on Visualization and Computer Graphics*, 21(9):1072–1086, 2015. doi: 10.1109/TVCG.2015.2410278

[10] Y. Chen and H. C. Chan. Visual Checking of Spreadsheets. *arXiv:0805.2189 [cs.HC]*, 2008.

[11] R. A. Choudhury, W. F. Mahaffee, N. McRoberts, and W. D. Gubler. Modeling Uncertainty in Grapevine Powdery Mildew Epidemiology Using Fuzzy Logic. *BioRxiv*, 264622, 2018. doi: 10.1101/264622

[12] J. de Jong. math.js – an extensive math library for JavaScript and Node.js. https://mathjs.org/, 2019. Accessed: Sep. 1, 2021.

[13] D. Feng, L. Kwock, Y. Lee, and R. M. Taylor. Matching Visual Saliency to Confidence in Plots of Uncertain Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):980–989, 2010. doi: 10.1109/TVCG.2010.176

[14] M. Fernandes, L. Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*, pp. 1–12. ACM Press, 2018. doi: 10.1145/3173574.3173718

[15] D. Fisher, S. M. Drucker, R. Fernandez, and S. Ruble. Visualizations everywhere: A Multiplatform Infrastructure for Linked Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1157–1163, 2010. doi: 10.1109/TVCG.2010.222

[16] N. French and L. Gabrielli. Discounted cash flow: accounting for uncertainty. *Journal of Property Investment & Finance*, 23(1):75–89, 2005. doi: 10.1108/14635780510575102

[17] K. Furmanova, S. Gratzl, H. Stitz, T. Zichner, M. Jaresova, M. Ennemoser, A. Lex, and M. Streit. Taggle: Scalable Visualization of Tabular Data through Aggregation. *Information Visualization*, 2019. doi: 10.1177/1473871619878085

[18] A. Gogolou, T. Tsandilas, T. Palpanas, and A. Bezerianos. Comparing Similarity Perception in Time Series Visualizations. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '18)*, 25(1):523–533, 2019. doi: 10.1109/TVCG.2018.2865077

[19] D. G. Goldstein and D. Rothschild. Lay understanding of probability distributions. *Judgment and Decision Making*, 9(1):14, 2014.

[20] Google LLC. Google Sheets. https://www.google.com/sheets/about/. Accessed: Sep. 1, 2021.

[21] Guesstimate Inc. Guesstimate. https://www.getguesstimate.com/, 2019. Accessed: Sep. 1, 2021.

[22] F. Hermans, M. Pinzger, and A. van Deursen. Supporting professional spreadsheet users by generating leveled dataflow diagrams. In *Conference on Software Engineering (ICSE '11)*, p. 451. ACM Press, 2011. doi: 10.1145/1985793.1985855

[23] J. L. Hintze and R. D. Nelson. Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2):181–184, 1998. doi: 10.1080/00031305.1998.10480559

[24] C. Holzhüter, A. Lex, D. Schmalstieg, H.-J. Schulz, H. Schumann, and M. Streit. Visualizing Uncertainty in Biological Expression Data. In *Visualization and Data Analysis*, p. 82940O. International Society for Optics and Photonics, 2012. doi: 10.1117/12.908516

[25] J. Hullman. Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, 2020. doi: 10.1109/TVCG.2019.2934287

[26] D. Jannach, T. Schmitz, B. Hofer, and F. Wotawa. Avoiding, finding and fixing spreadsheet errors – A survey of automated approaches for spreadsheet QA. *Journal of Systems and Software*, 94:129–150, 2014.

[27] P. W. Jordan, ed. *Usability evaluation in industry*. CRC Press, 1996.

[28] M. Knauff and A. G. Wolf. Complex cognition: the science of human reasoning, problem-solving, and decision-making. *Cognitive Processing*, 11(2):99–102, 2010. doi: 10.1007/s10339-010-0362-z

[29] A. Kohlhase. Human-Spreadsheet Interaction. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler, eds., *Human-Computer Interaction – INTERACT 2013*, Lecture Notes in Computer Science, pp. 571–578. Springer, 2013. doi: 10.1007/978-3-642-40498-6_47

[30] A. Lex, C. Partl, D. Kalkofen, M. Streit, S. Gratzl, A. M. Wassermann, D. Schmalstieg, and H. Pfister. Entourage: Visualizing Relationships between Biological Pathways using Contextual Subsets. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '13)*, 19(12):2536–2545, 2013. doi: 10.1109/TVCG.2013.154

[31] LimeSurvey GmbH. Professionelle Online-Umfragen mit LimeSurvey. https://www.limesurvey.org/, 2020. Accessed: Sep. 1, 2021.

[32] C. Lundström, P. Ljung, A. Persson, and A. Ynnerman. Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation. *IEEE Transactions on Visualization and Computer Graphics (Vis '07)*, 13(6):1648–1655, 2007. doi: 10.1109/TVCG.2007.70518

[33] J. Menick. Sampling – JavaScript for sampling from a handful of discrete probability distributions. https://github.com/jacobmenick/sampling, 2016. Accessed: Sep. 1, 2021.

[34] Microsoft Corporation. Microsoft Excel. https://www.microsoft.com/en-us/microsoft-365/excel/. Accessed: Sep. 1, 2021.

[35] Microsoft Corporation. Understanding the Office JavaScript API. https://docs.microsoft.com/en-us/office/dev/add-ins/develop/understanding-the-javascript-api-for-office/, 2020. Accessed: Sep. 1, 2021.

[36] C. Nobre, M. Streit, and A. Lex. Juniper: A Tree+Table Approach to Multivariate Graph Visualization. *Transactions on Visualization and Computer Graphics (InfoVis '18)*, 25(1):544 – 554, 2019. doi: 10.1109/TVCG.2018.2865149

[37] Oracle Corporation. Oracle Crystal Ball. https://www.oracle.com/applications/crystalball/, 2020. Accessed: Sep. 1, 2021.

[38] F. G. W. C. Paas, J. J. G. van Merriënboer, and J. J. Adam. Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills*, 79(1):419–430, 1994. doi: 10.2466/pms.1994.79.1.419

[39] Palisade. @RISK. https://www.palisade.com/risk/, 2020. Accessed: Sep. 1, 2021.

[40] A. T. Pang, C. M. Wittenbrink, and S. K. Lodh. Approaches to Uncertainty Visualization. *The Visual Computer*, 13(8):370–390, 1997.

[41] A. Parush, A. Hod, and A. Shtub. Impact of visualization type and contextual factors on performance with enterprise resource planning systems. *Computers & Industrial Engineering*, 52(1):133–142, 2007. doi: 10.1016/j.cie.2006.11.002

[42] C. V. Phillips and L. M. LaPole. Quantifying errors without random sampling. *BMC Medical Research Methodology*, 3(1):9, 2003. doi: 10.1186/1471-2288-3-9

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3119212, IEEE Transactions on Visualization and Computer Graphics

15

[43] T. Pitcher. Beverton and Holt Equations: Spreadsheet functions and Uncertainty, 1999.

[44] K. Potter, P. Rosen, and C. R. Johnson. From Quantification to Visualization: A Taxonomy of Uncertainty Visualization Approaches. In *Uncertainty Quantification in Scientific Computing*, pp. 226–249. Springer, London, 2012.

[45] Probability Management. Probability Management. https://www.probabilitymanagement.org/sipmath. Accessed: Sep. 1, 2021.

[46] J. Sanyal, Song Zhang, G. Bhattacharya, P. Amburn, and R. Moorhead. A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1209–1218, 2009. doi: 10.1109/TVCG.2009.114

[47] J. Sanyal, Song Zhang, J. Dyer, A. Mercer, P. Amburn, and R. J. Moorhead. Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, 2010. doi: 10.1109/TVCG.2010.181

[48] D. Spiegelhalter, M. Pearson, and I. Short. Visualizing Uncertainty About the Future. *Science*, 333(6048):1393–1400, 2011. doi: 10.1126/science.1191181

[49] A. Streit, Binh Pham, and R. Brown. A Spreadsheet Approach to Facilitate Visualization of Uncertainty in Information. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):61–72, 2008. doi: 10.1109/TVCG.2007.70426

[50] TeamViewer Germany GmbH. TeamViewer. https://www.teamviewer.com/, 2020. Accessed: Sep. 1, 2021.

[51] The Apache Software Foundation. Apache OpenOffice Calc. https://www.openoffice.org/product/calc.html. Accessed: Sep. 1, 2021.

[52] Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Eduard Gröller, Marc Streit. Supplementary material for fuzzy spreadsheets. https://osf.io/8ypux/. Accessed: Sep. 1, 2021. doi: 10.17605/OSF.IO/8YPUX

[53] D. van der Laan, E. de Jonge, and J. Solcer. Effect of Displaying Uncertainty in Line and Bar Charts – Presentation and Interpretation:. In *Conference on Information Visualization Theory and Applications (IVAPP '15)*, pp. 225–232, 2015. doi: 10.5220/0005300702250232

[54] Z. Vosough, D. Kammer, M. Keck, and R. Groh. Visualizing uncertainty in flow diagrams: a case study in product costing. In *Symposium on Visual Information Communication and Interaction (VINCI '17)*, pp. 1–8. ACM Press, 2017. doi: 10.1145/3105971.3105972

[55] M. Williams and T. Norris. jStat – JavaScript statistical library. https://github.com/jstat/jstat, 2020. Accessed: Sep. 1, 2021.

[56] Wolfram Alpha LLC. WolframAlpha. https://www.wolframalpha.com/. Accessed: Sep. 1, 2021.

[57] Y. Wu, G.-X. Yuan, and K.-L. Ma. Visualizing Flow of Uncertainty through Analytical Processes. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '12)*, 18(12):2526 –2535, 2012. doi: 10.1109/TVCG.2012.285

[58] Zoom Video Communications, Inc. Zoom. https://zoom.us/, 2020. Accessed: Sep. 1, 2021.

Andreas Hinterreiter is a PhD student at the Institute of Computer Graphics, Johannes Kepler University (JKU) Linz, Austria. His research interests include dimensionality reduction and explainable AI. He recently spent a year at the Biomedical Image Analysis Group of the Imperial College London. He received his Diplomingenieur (MSc) in Technical Physics from JKU.

Eduard Gröller is Professor at the Institute of Visual Computing & Human-Centered Technology (VC & HCT), TU Wien, where he is heading the Research Unit of Computer Graphics. He is a scientific proponent and key researcher of the VRVis research center (http://www.vrvis.at/). The center performs applied research in visualization, rendering, and visual analysis. Dr. Gröller is Adjunct Professor of Computer Science at the University of Bergen, Norway. His research interests include computer graphics, visualization, and visual computing.

Vaishali Dhanoa is a researcher at Pro2Future GmbH and PhD student at the Institute of Computer Graphics of the Johannes Kepler University (JKU) Linz, Austria. Her research interests include visual analytics, comprehensible approaches to user onboarding, and developing performance optimized interactive tools. In the past, she worked as a software developer at Intel GmbH Linz, Austria. She received her Diplomingenieur (MSc) in Informatics from JKU.

Marc Streit is a Full Professor of Visual Data Science at the Johannes Kepler University (JKU) Linz, Austria. He finished his PhD at the Graz University of Technology in 2011. His scientific areas of interest include visualization, visual analytics, and biological data visualization. He is also co-founder and CEO of datavisyn, a spin-off company developing data visualization solutions for applications in pharmaceutical and biomedical R&D. For more information see http://marc-streit.com.

Conny Walchshofer is a PhD student at the Institute of Computer Graphics of the Johannes Kepler University (JKU) Linz, Austria. In her prior research, she focused on the perception and handling of multidimensional visualizations. She applies an interdisciplinary approach to judge cognitive load during the interpretation of visual representations by using physiological measurement methods (e.g., eye-tracking, heart rate variability).