

# Reliable Visual Analytics, a Prerequisite for Outcome Assessment of Engineering Systems

Ekaterina Auer<sup>a</sup>, Wolfram Luther<sup>b</sup>, and Benjamin Weyers<sup>c</sup>

## Abstract

Various evaluation approaches exist for multi-purpose visual analytics (VA) frameworks. They are based on empirical studies in information visualization or on community activities, for example, VA Science and Technology Challenge (2006-2014) created as a community evaluation resource to “decide upon the right metrics to use, and the appropriate implementation of those metrics including datasets and evaluators”<sup>1</sup>. In this paper, we propose to use evaluated VA environments for computer-based processes or systems with the main goal of aligning user plans, system models and software results. For this purpose, trust in VA outcome should be established, which can be done by following the (meta-)design principles of a human-centered verification and validation assessment and also in dependence on users’ task models and interaction styles, since the possibility to work with the visualization interactively is an integral part of VA. To define reliable VA, we point out various dimensions of reliability along with their quality criteria, requirements, attributes and metrics. Several software packages are used to illustrate the concepts.

**Keywords:** reliable visual analytics, evaluation, verification and validation assessment, quality criteria, metrics

## 1 Introduction

With the advance of ubiquitous computing, the Internet of things or cloud based technologies, ambient intelligence (AmI) and smart environment software gain more and more importance for supporting mobile users in all areas of their daily lives. To provide efficient and meaningful support, the developers of such software have to deal with quite a few challenges, for example, managing large amounts of heterogeneous input/output data and high system complexity. This calls for innovative analytic approaches such as visual or/and collaborative ones.

---

<sup>a</sup>Hochschule Wismar, Department of Electrical Engineering, D-23966 Wismar, Germany, E-mail: [ekaterina.auer@hs-wismar.de](mailto:ekaterina.auer@hs-wismar.de)

<sup>b</sup>University of Duisburg-Essen, Department of Computer Science and Applied Cognitive Science, D-47048 Duisburg, Germany, E-mail: [luther@inf.uni-due.de](mailto:luther@inf.uni-due.de)

<sup>c</sup>University of Trier, Department IV, D-54296 Trier, Germany, E-mail: [weyers@uni-trier.de](mailto:weyers@uni-trier.de)

<sup>1</sup>cf. [59]

In particular, the emerging area of visual analytics (VA) has been shown to offer a solution to these challenges [68]. Its main strength lies in the ability to engage in the analytical process the whole of human perceptual and cognitive capabilities augmented by advanced computations [10]. In [63], the authors remark that VA “employs interactive visualizations to integrate users’ knowledge and inference capability into numerical algorithmic data analysis processes. Visual Analytics Science and Technology (VAST) is an active research field that has applications in many sectors, such as security, finance, and business” as well as healthcare, natural sciences and engineering. VA “will foster the constructive evaluation, correction and rapid improvement of our processes and models and – ultimately – the improvement of our knowledge and our decisions”, as stated in [35]. Already in 1990, Healy [29] has suggested that “an informative visualization technique that allows rapid and accurate visual analysis would decrease the amount of time needed to complete the analysis task.” VA hardware and software architectures serve to assess and visualize important system/process parameters, descriptors and uncertain environment entities. Therefore, a working definition of VA could be as follows.

**Definition 1.** *VA is a science of analytical reasoning facilitated by interactive visual interfaces [64]. It is a multidisciplinary field merging analytical reasoning techniques with data representation approaches and (interactive) visualization theories. In other words [35], “VA combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets”.*

As an example of a (statistical) method aided by relatively simple visualization, let us consider correlation analysis, often of high practical relevance in engineering. The correlation denotes the relationship between two random variables, fuzzy numbers or simply two (interval) sets characterizing, for instance, the input and output of a system. Depending on the distribution of the variables, specific correlation coefficients are defined to evaluate the strength of this relationship, for example, the Pearson coefficient or the Spearman rank correlation [56]. Correlation analysis is supported by graphical techniques [78] such as scatter plots, scatter plot matrices, heat maps and others. In particular, a scatter plot is, as a rule, a good visual aid to access quickly whether or not two variables have any linear correlation and to give an indication about its direction (positive/negative) in case there is a correlation. Apart from built-in routines for scatter plots or matrices within such general-purpose environments as MATLAB, correlation analysis is supported by further visualization tools, for example, `corrplot`<sup>2</sup> package in R or CI Thermometer [78]. More typical VA applications with complex visualizations come from such diverse areas as neuroscience, artificial intelligence, healthcare, finance or environmental sciences (e.g., meteorology).

As with any computer-based approach, the issues of reliability and comparability (and thus, standardization) play an important role for VA. Although developing sets of quality standards and assessment methods has received a fair amount

---

<sup>2</sup>[cran.r-project.org/web/packages/corrplot/](https://cran.r-project.org/web/packages/corrplot/)

of attention in such areas as (big) data or software management, the corresponding research for VA is still at its beginnings. For example, a universal, two-layer standard for big data quality assessment proposed in [19] considers not only reliability but also availability, usability, relevance and presentation aspects. Few publications explicitly introduce the term 'reliable visual analytics' (RVA) or propose guidelines for assessment of VA frameworks/methodologies and of their applicability and efficiency. Several authors focus on device-dependent transformation, accurate understanding of outcome using reliable mapping algorithms and standardized procedures to automatically select, analyze, refine and combine visual data [59]. Sometimes accuracy and reliability are explicitly or implicitly addressed in the context of uncertain data acquisition, aircraft and power plant safety, risk assessment and healthcare monitoring and management [61]. However, reliability of VA software is an important prerequisite for its use in the context of (visual) assessment of (the outcome of) other computer-based systems or processes from engineering.

Reliable VA frameworks need guidelines and regulations for all stages in their development cycle based on real world use cases, benchmarks, and formal and laboratory studies. One further generic requirement might take into account ethical considerations. Besides, specifications are needed for interaction styles, for example, for those using virtual reality 3D devices [75], since interactivity is an integral part of any visual analysis. Moreover, collaboration styles also play an important role. A typical example for this comes from real-life healthcare applications, which often need group or pair (visual) analytics. Sessions with experts of various domains can be analyzed using the joint action theory protocol analysis and pair analytics methods [1], the purpose of which is to prove the emergence of common ground, that is, mutual, general or joint knowledge, beliefs and assumptions among the involved parties (stakeholders), a precondition for solving problems collaboratively with the help of VA.

These developments imply that human-centered paradigms have become an important feature within a workflow for designing, modeling, and implementing various real life processes and Aml environments. For example, human-centered paradigms are pointed out in [41] in the context of a formal verification and validation (V&V) assessment not only for code/result verification, uncertainty management, validation and evaluation, but also for user interaction, recommender techniques and VA. In turn, that means that the human issues have to be taken into account also while ensuring the reliability of VA architectures in order to apply them for formal V&V assessment within the workflow of the modeling and simulation cycle.

In [73, 74], Weyers presents a tentative conceptual framework for characterization of reliability in VA, which uses three major *dimensions* (visual integrity, user interface, interaction process) assessed by means of three *quality criteria* (accuracy, adequacy and efficiency) that reflect different levels in the analysis. Here, accuracy refers to a low-level (data) type correctness measure, whereas efficiency denotes the quality of the work process and the task the VA tool is used for. Each dimension and criteria pair is rated using a set of *metrics*, for example, the lie

factor proposed by Tufte [68] that quantifies the mismatch between the visually represented effect or value and the actual effect. Here, we use the word ‘metrics’ not in the mathematical sense, but in its meaning ‘quality measure’ [6], the value of which is determined by optimizing the above mentioned quality criteria over given parameters and specifications. This quantification-based assessment needs to be complemented by empirical studies, at least when it comes to the investigation of complex interaction and analysis scenarios. Additionally, Weyers et al. [75] introduce a formal component to the assessment by compiling an overview of formal methods in human-computer interaction (HCI) including V&V approaches to interactive systems. Finally, Sun et al. complement the discussion by pointing out in [63] that “uncertainty modeling and visualization play a critical role in ensuring the reliability and trustworthiness of the analytics process”.

In this contribution, we point out existing standards and evaluation suggestions for V&V assessment as well as quality criteria and metrics in the context of visual analytics (Section 2). In accordance with the new IEEE Std 1012<sup>TM</sup>-2016 norm [30], we advocate a broad approach to V&V assessment that allows its users to

- refer reliability to data, design strategies, processes, software and outcome analysis;
- define requirements, quality criteria and metrics for the outcome of the considered process or task and its analysis at the early stages in its development cycle;
- choose the appropriately evaluated tools (e.g., for data mining, visualization, analysis, decision) in dependence on the balance between costs and risks or get them recommended.

While the early drafts for V&V assessment did not consider uncertainty treatment as worthy an explicit role in the 1990s, human factors still do not take a significant place in the overall procedure. Where HCI is concerned, the approach discussed in this paper and extended in [41] goes beyond the norm [30] in considering error avoidance not only for interfaces between technical systems, but also for those between a human and a computer [75, 77]. Bearing in mind the methodologies from the neighboring fields, we discuss the possibility of a multilayer quality assessment procedure for VA, similar to that from data analytics, concerning reliability, accuracy, performance, efficiency, group activity monitoring as well as validation and evaluation. This leads us to formulating a tentative definition for reliable visual analytics. In Section 3, we present use cases corroborating the ideas from Section 2 together with their assessed tasks. Conclusions are in the last section.

## 2 Reliable VA: A Tentative Definition

Experts can use VA environments for a variety of tasks within the broad area of output data analysis. In particular, it is possible to employ them for validating computer-based processes or systems, which requires aligning users’ requirements

with tool and problem domains. Additionally, questions need to be answered about whether the model is right and the program is built well for the intended use, that is, if it solves the problem properly and is correctly applied [9]. Regardless of the later application, it is necessary to evaluate VA environments w.r.t. various aspects in order to be sure that the visual analysis is correct. In this context, correctness can be defined by two major components: first, the technical correctness of the transfer of data into their visual representations and, second, the correctness of the mental understanding a user gathers about the data by using the VA tool. The technical correctness is a necessary condition for the correct mental understanding, which in turn is a necessary condition for a correct interpretation of the presented data. Ultimately, the interpretation leads to the user's decision and so defines the potential impact the VA tool might have on the user and the user's environment.

Between 2006 and 2014, the VAST challenge<sup>3</sup> prompted its participants to decide “upon the right metrics to use, and the appropriate implementation of those metrics including datasets and evaluators” [58]. Based on the results [59], we can now define guidelines and rules that are used inside VA environments to assess input and output data and (inter-)action logic of computer-based systems and processes, the subsequent result analysis, and the follow-up activities.

In this section, we point out the place of VA within the general modeling and simulation cycle in engineering [55] first. This material also outlines how to assess the modeling and simulation process systematically at each stage to arrive at reliable and trustworthy results, visual or otherwise. After that, we review the literature on the topic. Finally, we summarize this information to present a tentative definition of *reliable* VA along with the corresponding dimensions, criteria and metrics with a focus on interaction styles.

## 2.1 RVA Within the General Modeling and Simulation Cycle in Engineering

In Figure 1, our proposal for assessment of computer-based environments or processes from engineering is outlined. It includes the use of RVA tools which can themselves be evaluated according to the same principles (possibly excluding VA this time to avoid recursion). Detailed information on the definition of RVA and specific techniques for evaluation of VA environments are in Section 2.3.

The entries in the first column in Figure 1 show stages of the modeling and simulation cycle, which can be reiterated. We start by transforming the mental model for a given engineering process into a (formal) description using, for example, a modeling language. It is then translated into a computer program that simulates the process, implements the user's intention, and solves the specified task. We assume that the goal of the modeling phase is to develop a computer program corresponding as far as possible to the user's mental model. Finally, the outcome of the computer program can be analyzed and the whole process possibly repeated, for example, to improve/simplify the previous design or identify unknown parameters.

---

<sup>3</sup><http://www.cs.umd.edu/hcil/varepository/>

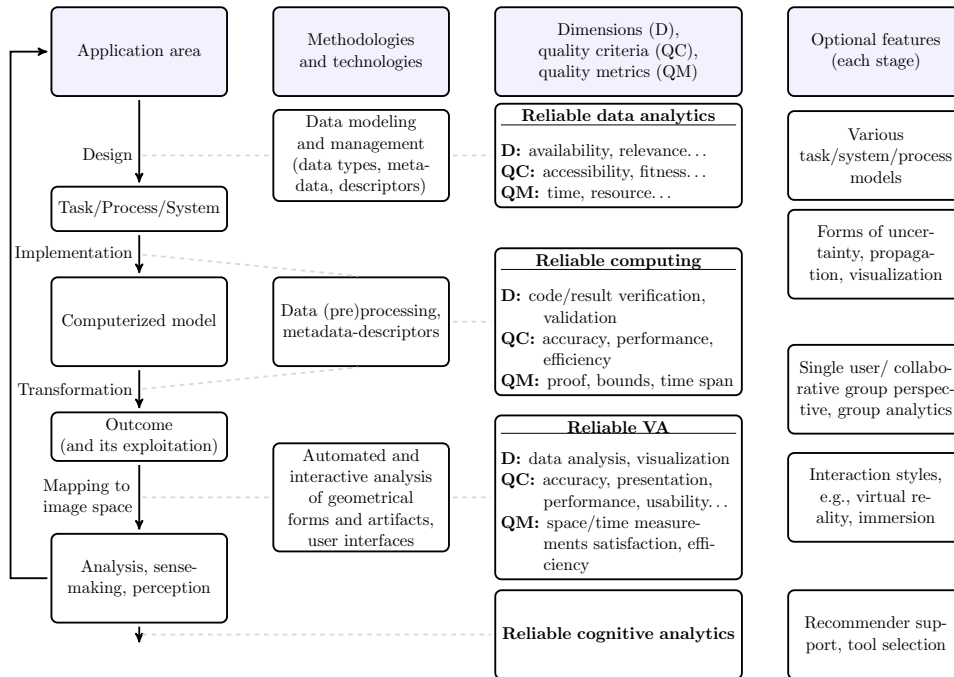


Figure 1: A scheme for a V&V approach to assess an environment from its modeling to its outcome analysis stage.

The loop on the outmost left also covers aspects of validation of physical models under uncertainty (as described, e.g., in [24]) and appropriate experiments design since ‘outcome’ in the figure can also mean validation results (a dimension of reliable computing) with the optional feature ‘uncertainty’ taken into account (cf. the text about Column 4 later on). An example of a visual aid for experiments design is a Gaussian process regression allowing one to perform sensitivity analysis on complex computational models using a limited number of model evaluations. The Gaussian process approximation error can be propagated to the sensitivity index estimates, in this way allowing us to visualize the main effect of a group of variables and the uncertainty of its estimate. The book [24] provides quite a few references to the software mapping the outcome of such uncertainty analysis and quantification tasks to an appropriate image space in contexts of simulation, design exploration, process-based sensitivity analysis and Bayesian model calibration along with imprecise probability. Further tools for uncertainty quantification having extensive visualization components are COSSAN<sup>4</sup>, UQLAB<sup>5</sup> or UQPY<sup>6</sup>. Such tools support various visual analysis techniques and inspection modes in the context

<sup>4</sup>[www.irz.uni-hannover.de/en/research/software-projects/](http://www.irz.uni-hannover.de/en/research/software-projects/)

<sup>5</sup>[www.uqlab.com](http://www.uqlab.com)

<sup>6</sup>[github.com/SURGroup/UQpy](https://github.com/SURGroup/UQpy)

of the highlighted quality criteria, such as accuracy, adequacy, efficiency, to detect relationships between the models, input variables, parameters and outcomes.

We use the term ‘mental model’ to denote “the image of the world around us, which we carry in our head”, a definition attributed to Jay Wright Forrester [22]. Note that this ‘image’ encompasses not only static objects but also such aspects as our understanding of interrelations between objects, their actual and expected state and dynamics as well as their representations on various levels of abstraction. A formal model is (ideally) a theoretical representation of the user’s mental model and, therefore, a description of a real world process, its functioning and effects as a virtual execution of a plan or fulfillment of a task. The term ‘computer program’ or ‘computerized system’ denotes (mental/formal) models implemented on a computer. The computer program transforms the input data to the output data to be interpreted and exploited automatically or by humans. The outcome analysis can be facilitated by preprocessing, reformatting, mapping and rendering content to visual items that will be scrutinized, perceived and manipulated using an appropriate interface.

In the second column, methodologies and technologies used for the corresponding transitions are listed. The focus is on raw data and data types, metadata and descriptors as well as the outcome analysis via appropriate visual interfaces.

The third column describes assessment options along with reliability dimensions, quality criteria and their metrics. These concern input and output data modeling, where reliable data analytics should be used, then the stage of data processing by the computerized model made reliable through code and numerical result verification, and finally the stage of system/process validation where reliable VA environments and technologies can be employed. Additionally, reliable cognitive analytics can be used to improve or check human decision-making.

Although we assign typical places for these options inside the general modeling and simulation cycle, they are not the only possibilities to employ a given technique inside the overall process. For example, reliable data analytics can be used for both input and output data, a peculiarity reflected in the Figure by the possibility to reiterate (the leftmost arrow). As regards the possibilities offered by RVA inside V&V assessment, we can think of meaningful employment at practically each step:

1. For input/output (big) data analysis
2. For validation of the outcome of the computer program (e.g., comparisons with experimental results)
3. For formal verification (e.g., visualization of the steps in the proof)
4. For result verification (e.g., comparisons with analytical solutions or benchmark results)
5. For code verification (e.g., visualizing execution paths or interconnections between blocks)
6. For improving the human grasp of uncertainty influence in an application
7. For software comparisons

#### 8. For justifying software recommendations

Note that this list just clarifies the employment possibilities. We do not always change the place of RVA in the overall cycle but rather the perspective on the kind of tool we apply this cycle to (e.g., input data preprocessing tools).

The fourth column in Figure 1 deals with meta-design principles and system design aspects. Those might or might not take into account such issues as uncertainty representation in data types, its propagation and visualization; group analytics; immersion with the help of virtual reality; or automated tool selection through suggestions by recommender systems. Nowadays, experts agree that awareness of underlying uncertainty is crucial for the whole design and modeling process to build trust and confidence in the outcome of a computerized system. We can deal with the aspects of its representation, propagation and visualization by using, for example, interval methods [43], as opposed to working with computer arithmetics based on crisp data types. Various task models, resources and interaction methodologies necessary to carry out systematic analyses also use collaboration possibilities (or group analytics) in an organized way so that multiple-user modes can be an integral part of the computerized model. Moreover, users can interact with the program via WIMP<sup>7</sup> or post-WIMP interfaces. WIMP interfaces utilize mouse and keyboard-based interaction on screens and are well suited for presenting and manipulating 2D content. Post-WIMP interfaces enable new interaction paradigms for navigation and manipulation using, for example, 3D virtual reality environments and visualizations. That is, users can navigate, select objects and manipulate items with the help of 3D devices such as elastic arms and virtual hands [14]. In this way, it is possible to move around items and detect interesting viewpoints or areas similarly to physical interaction. Finally, users can be supported while selecting tools or quality criteria by recommender platforms [5, 6].

Our focus is on assessment possibilities in column three. To assess a given computer program, geometric or statistic descriptors along with reliable analysis tools need to be selected. The mentioned tools implement algorithms from various fields, for example, data assimilation/mapping/mining, numerical analysis or statistics. Sensitivity analysis allows us to reduce data or problem dimensions and to map results and their artifacts such as uncertainties to visual spaces. Possible assessment dimensions characterizing data analytics are reliability, availability, usability, relevance, and presentation quality as proposed in [19]. For example, as relevant criteria for availability, the authors suggest accessibility and timeliness assessed with the help of such measures as existence of the access interface, data arrival on time, regularity of updates, and meeting time constraints for collecting data and preparing its processing. A discussion of further dimensions, quality criteria and metrics can be found in [19].

At the stage of implementation, dimensions characterizing reliability are code verification and numerical result verification of the computerized model. The term verification means that we need to ensure that the model is implemented correctly. That is, the major question to be answered by verification is whether “the pro-

---

<sup>7</sup>Windows, Icons, Menus, Pointer



gram is implemented right”. The reliability of the output data produced using the computerized model can be characterized by validation. Validation addresses the purpose of computerized model and defines various requirements and metrics for comparing the outcome with experimental measurements, alternative simulations or other approaches [30]. That is, the major question to be answered by validation is whether “the right program is implemented”.

Important quality criteria are accuracy, performance and efficiency. Accuracy means in this context that the data used or provided were correctly expressed by the chosen data types. To assess this criterion, we need ground truth, a reference or guaranteed bounds. Possible quality metrics encompass the use of computer-based proofs, analytic solutions, algorithms based on interval or other set-based arithmetics, computation of guaranteed error bounds, sensitivity analysis or simply consistent employment of a standardized finite precision arithmetic. Performance is a generic term for successful task completion and includes efficiency and effectiveness, where efficiency rates resource usage and effectiveness assesses the speed of task completion. It can be quantified using, for example, the time span needed to complete a certain task.

After the given data transformation by the computerized model is validated, that is, after users consider the outcome trustworthy, reliable VA environments can help to analyze it. However, if a VA environment is reliable, it can be also used on the previous stage of the cycle for validation. Relevant reliability dimensions and quality criteria serve to assess such tasks as the outcome analysis, knowledge discovery and management, decision making, and reporting. In this paper, we develop a tentative definition of RVA and formulate how to assess VA environments in Section 2.3. Note that, compared to the areas of reliable computing and reliable data analytics, the corresponding definitions and techniques for VA are just beginning to emerge and need systematization.

## 2.2 Assessment of (VA) Environments: Literature Overview

VA methods are used more and more heavily nowadays to assess different aspects of computer-based processes (e.g., their outcome). Therefore, the need to ensure that VA environments are reliable becomes eminent. Here, uncertainty plays an important role, since failing to take it into account often leads to wrong interpretation of analysis results. With the goal to embed our definition of RVA into existing work, we concentrate on relevant aspects from the third and the fourth columns of Figure 1 and point out current research directions in this section. We discuss V&V norms, solutions and approaches in data processing, representation and manipulation with a focus on uncertainty management. Additionally, we highlight the existing work on VA assessment leading to a better understanding of how RVA can be defined.

### 2.2.1 General V&V Assessment

First and foremost, the IEEE Standard for System, Software, and Hardware Verification and Validation (IEEE Std 1012<sup>TM</sup>-2016) should be mentioned. It defines how to assess systems and tasks using quality criteria and metrics [30]. Additionally, reliability and trust in the outcome of a simulation or a VA program can be achieved using the numerical verification approach proposed by the first two authors in 2009 and extended in [3]. There, the degree of verification of a system or process from engineering is assessed with the help of a four-tier numerical verification and validation taxonomy in dependence on the use of standardized floating point or interval arithmetic data types, of sensitivity analysis and of uncertainty quantification (with verified or stochastic methods) or of algorithms with automatic result verification. The objective is to support users and developers of a numerical software project as early as during the stage of goal and process flow definition for it. This approach complements the already existing V&V methodologies by making use of result verification technologies. For dealing with uncertainty, important advances have been made in the recent years by combining verified (interval) methods with stochastic approaches [49, 79]. A comprehensive study on quality assessment for big data is in [19].

Meta-design principles that support system evaluation w.r.t. tasks, resources and methodologies are necessary to carry out systematic analyses and to choose collaboration assets in an organized way. This includes selecting, for example, domain specialists or users for testing the considered (VA) system. Moreover, group building strategies need to be chosen in a methodical way to support both analysts' interaction via appropriate interfaces and their cooperation for knowledge discovery. An example of using collaborative VA is given in [34]. Here, a complete VA system and a collaborative touch-table application are designed and evaluated for solving real-life tasks with two integrated components: a single-user desktop and an extended system suitable for a collaborative environment. As further characteristics, perceptual and cognitive issues should be assessed from the point of view of psychology to determine confidence, speed, and accuracy of judgments under uncertainty.

The next issue within a collaborative setting is to develop efficient data fusion strategies supporting high quality decision making [27]. The JDL/DFIG model<sup>8</sup> defines a six level approach for this purpose consisting of source preprocessing and subject assessment; object, situation, impact assessment; process refinement; and user (cognitive) refinement. The last level is necessary to overcome the HCI bottleneck in the information process fusion [51]. The important aspects are

**Cognitive aids** that provide functions to aid and assist human understanding and exploitation of data

**Negative reasoning enhancement** that helps to overcome the human tendency to seek for information which supports their hypothesis and ignore negative information

---

<sup>8</sup>developed in the 1980s by Joint Directors of Laboratories who formed Data Fusion and Information Group

**Uncertainty representation** methods that are necessary to improve quantification, visualization and, with that, the understanding of uncertainty

**Time compression/expansion** replay techniques that can assist in understanding of evolving tactical situations, on account of human capabilities to detect changes

**Focus/defocus of attention** techniques that can assist in directing the attention of an analyst to different aspects of data

**Pattern morphing** methods that can translate patterns of data into forms that are easier to interpret for a human

Information fusion strategies mentioned above need to be supplemented by evaluation of uncertainty visualization techniques for them. This is due to the fact that “huge quantities of (higher dimensional) data from several sources carrying various forms of uncertainty” need to be represented “on a two or three dimensional device” [51], which can only be done in a reliable way if this uncertainty is properly translated using generally accepted perceptual and cognitive principles.

Automated recommender platforms support users in selecting appropriate software frameworks, interfaces, and interaction styles. For reliable methods, several recommendation frameworks were developed in [3, 8, 41]. Visualization tools or techniques and metrics can be recommended depending on the data category [35] and requirements for the quality criteria.

### 2.2.2 Visualizing Uncertainty

Information about uncertainty has been found to play a crucial role for establishing trust in results of a computer simulation or in the analytics process as such [63]. To understand and capture the impact of uncertainty with the help of a VA environment, eight guidelines are formulated in [53]. However, they can be applied more broadly for any application in engineering:

- Quantify uncertainties in each component (or in each process step, respectively)
- Propagate and aggregate uncertainties
- Visualize (or make known otherwise) uncertainty information
- Enable interactive uncertainty exploration
- Make the (VA) systems functions accessible
- Support the analyst in uncertainty aware sense-making
- Analyze human behavior in order to derive hints on problems and biases
- Enable analysts to track and review their analysis

Taxonomies for visualizing uncertainty were published, for example, in [11, 47, 38]. Uncertainties in perception and cognition are addressed in [18, 11, 42, 65]. In particular, a typology is developed for geospatially referenced data in [65] that is considered to be general enough to be applied to reasoning under uncertainty (a claim

which still needs to be substantiated by further studies). According to this typology, uncertainty visualization can express additional information about: accuracy/error (difference between observation and reality); precision (exactness of measurement); completeness (extent to which information is comprehensive); consistency (extent to which information components agree); lineage (conduit through which information passed); currency/timing (temporal gaps from information collection); credibility (assessment of information source); subjectivity (amount of judgment included); and interrelatedness (source independence). Similarly, MacEachren et al. [42] define the following seven goals for uncertainty visualization:

1. Understanding the components of uncertainty and their relationships to domains, users, and information needs
2. Understanding how knowledge of information uncertainty influences information analysis, decision making, and decision outcomes
3. Understanding how (or whether) uncertainty visualization aids exploratory analysis
4. Developing methods for capturing and encoding analysts' or decision makers' uncertainty
5. Developing representation methods for depicting multiple kinds of uncertainty
6. Developing methods and tools for interacting with uncertainty depictions
7. Assessing the usability and utility of uncertainty capture, representation, and interaction methods and tools.

Actual application of such general rules, especially for the case of probabilistic representation of uncertainty, is illustrated, for example, in [26, 28, 46] or in the overview papers [15, 47]. In addition to that, if mixed interval-probabilistic techniques are used to represent uncertainty, tools and theories described in, for example, [21, 48] can be used for visualization. In particular, the use of Dempster-Shafer or p-box theories described therein allows one to work with and visualize uncertain distributions by defining belief and plausibility functions (or lower and upper bounds on probability). A further example is given in [49] and described in more detail in Section 3: Dempster-Shafer theory is employed here for uncertain localization. In this context, the joint probability density function usually needs to be simplified by using either independence assumptions or dependency models to avoid multivariate, often parametric distributions (such as Gaussian or Weibull). By using the Dempster-Shafer theory, a (multivariate) probability density function can be replaced by a joint basic probability assignment with similar simplification possibilities (decomposition into marginal distributions) and easier visualization.

An important aspect to deal with while visualizing outcomes of systems with uncertain parameters is specifying how the constraints of the theory we treat the uncertainty with influence these outcomes. For example, if the uncertain parameters are represented by intervals and propagated through an engineering system using methods with result verification, the ranges for the simulation outputs are usually more conservative than the real ones would be (the so-called "outer enclosure",

mathematically proven to contain the exact result). This can lead to ambiguities negatively influencing the overall analysis so that users need to be alerted to the possibility. For methods with result verification, this can be dealt with by providing “inner enclosures” along with the outer ones [2, 25]. Roughly speaking, outer enclosures are supersets of the set’s true image by a function (or an operator), whereas inner ones are subsets of this image. Another example where these considerations play an important role is reliable object discovery and classification in safety-critical systems, which is one of the key challenges in artificial “vision” applications (e.g., autonomous driving). Here, application of Bayesian neural networks (a combination of Bayesian inference methods and neural networks), recently proposed by several authors [39, 67], can lead to a clear separation of the influences of different categories of uncertainty. The VA aspects of these theories are a topic for the ongoing research [36].

### 2.2.3 Assessing VA Environments

Although the general V&V techniques described in 2.1 and 2.2.1 can (and should) be used for assessing VA environments, there are several aspects specific to visualization that need a separate mention, first and foremost evaluation of graphical design. As early as in the 1970s, Bertin [7] provides general guidelines and rules for graphical representations. Zuk et al. [80, 81] discuss basic graphical design principles. Tufte [69] formulates principles for graphical excellence: clarity, precision, and efficiency. Ware [71] focuses on preattentive processing and Gestalt laws (e.g., proximity or connectedness). In the following, we summarize additionally the literature on classical VA evaluation along with formalizations and heuristics for metrics and quality criteria. At the end of the section, we touch upon works concerned with scenario-based evaluation.

The assessment goals, dimensions, criteria and relevant guidelines, rules and measures to assess a VA tool’s system model and its application context are discussed in [20, 23, 57, 58, 59]. A taxonomy of tasks presented there helps to structure important steps in outcome analysis, group building, interaction and collaboration for knowledge discovery and management, aggregation of expert judgments and group decisions. It encompasses the following aspects: data quality assessment, uncertainty management and tool quality assessment (cf. Figure 1), which lacks giving attention to human factors assessment. Besides, general guidelines, rules, heuristics and recommendations are formulated there for assessing the mapping<sup>9</sup> and visual presentation of data under uncertainty. In [6], the authors address quality metric formalization (based on the data categories established in the information visualization<sup>10</sup>) and requirements for quality criteria.

Evaluation of VA environments is often based on heuristics. For example, Zuk et al. [80, 81] deal with the selection of perceptual and cognitive heuristics by considering

---

<sup>9</sup>data objects to visual objects or data to geometrical descriptors

<sup>10</sup>multi- and high-dimensional, relational, sequential, geospatial and text data

**Shneiderman's** information seeking mantra: Overview first – Zoom and filter – Details on demand – Relate – Extract history [60]

**Amar and Stasko's** knowledge and task-based framework: Expose uncertainty – Concretize relationships – Determine domain parameters – Give multivariate explanation – Formulate cause and effect – Confirm hypotheses [62]

Recent assessment approaches come from the area of scenario-based VAST evaluation. Important task work and evaluation goals are addressed with such subgoals as, for example, task allocation and completion, accuracy, and efficiency. Usefulness, efficiency, and intuitiveness are important characteristics of known or innovative metrics which help to assess such aspects as analytical reasoning, visualization methodologies, interaction and collaboration within a formalized sense-making and result reporting process [45]. Lam et al. [31] describe a scenario-based approach to evaluation in information visualization. Seven scenarios evaluating visual data analysis and reasoning tools, environments and work practices, communication through visualization, collaborative data analysis, user performance, user experience, performance and quality of visualization algorithms are derived through an extensive review of over 800 visualization publications. These scenarios distinguish various study goals and types of research questions and are illustrated through example studies. However, numerical reliability, uncertainty issues and input/output data quality standards are not addressed.

### 2.3 RVA Definition

In the previous sections, we pointed out, on the one hand, the role of VA in the overall modeling and simulation cycle and general techniques for V&V assessment of computer programs. On the other hand, we indicated assessment possibilities for VA environments shown in recent literature. In this section, we first apply V&V analysis to VA and summarize how to evaluate VA environments to arrive at a tentative RVA definition. Then we outline the possibilities offered by RVA inside the general V&V assessment procedure.

Reliable visual analysis requires a complete evaluation of all components that are to be used inside the V&V assessment process of a software system and its outcome. The first step in this direction is to understand what the term RVA means. Our tentative definition is as follows.

**Definition 2.** *Reliable VA is formed by a set of reliability dimensions, quality criteria and useful, efficient, and intuitive metrics for which reliability is ascertained (or which are already evaluated, for example, using techniques described in Sections 2.2.1, 2.2.3).*

The purpose of RVA is to assess not only visualization (cf. Section 2.2.3) but also analytic processes, interaction (cf. Section 2.4), collaboration, sense-making, and result reporting taking place in a given VAST environment. RVA rates the formal strength of the computer-based process or system model descriptions from this environment as an implementation of a mental model/user plans w.r.t. the quality criteria of

**accuracy:** fidelity of mapping, consistency, integrity, grasp of uncertainty;  
**usability:** presentation quality, navigation/interaction, readability, recommendation, security, privacy, confidence;  
**adequacy:** correct resources used for correct purposes;  
**efficiency, performance and intuitiveness** of the environment, analytical process, interaction, presentation.

Meeting the quality criteria is assessed using requirements, rules, standards, laws and ethical regulations based on metrics, benchmarks or equivalent solutions taking into account the specified user tasks, for example, interaction with visual items, outcome analysis, sense-making/data fusion, knowledge creation, reporting. Moreover, human factors and subjective preferences need to be addressed in addition to such objective characteristics as accuracy, efficiency and fidelity. Group building, interaction and collaboration are further important assessment issues. More details on interaction and collaboration styles aiding RVA are in Section 2.4.

In this context, *accuracy* means that the output data are correctly represented using the chosen visual objects. Its sub-criterion fidelity can be assessed based on (semi-)formal object descriptions and a methodological framework. It measures realism or degree of similarity. Mapped objects must preserve properties; descriptors should be equally perceived and rated. *Consistency* rates the logical relationship between correlated terms and items. Additionally, consistency confirms that such a logical relationship actually exists. The next component, *integrity*, evaluates the appearance of an item and depends on the context. For example, the item should correspond to formal description and fulfill predefined standards. A further requirement can be that the descriptors are not modified during mapping and visual depiction or rendering. Finally, the *grasp of uncertainty* requires that data types deal with uncertain values and algorithms quantify and propagate uncertainty. To assess this aspect, a generally approved notation and taxonomy for uncertain data visualization are necessary. They should also reflect the degree up to which the interface supports interactive exploration and decision making under uncertainty.

The next criterion is *usability*, which rates user satisfaction and, in particular, HCI's efficiency and effectiveness. For the aspect of visualization, it is crucial to assess the presentation quality. This assessment is based on guidelines for data visualization taking into account display format, color, contrast, position, size, style, labels. Further metrics relying on time and memorability can be introduced. With these formalisms as a starting point, dedicated recommender components can be developed to choose the best visualization technique for a given task [6].

*Adequacy* is a meta criterion assessing, for a specific aspect of VAST, its suitability for a given purpose and its use of resources to fulfill the requirements. The requirements should be appropriately chosen (e.g., error limits or computation times should be realistic, access to results of comparable processes fast, the visual space for ensemble data configurable [59]). The quality criteria of *efficiency, performance* and *intuitiveness* concentrate on describing how fast or effortless the intended tasks can be carried out.

## 2.4 Interaction Styles Aiding VA

An integral part of VA is the possibility to work with the visualization interactively, for example, by executing various operations to manipulate visualization parameters, the data preprocessing or both. Reliable interaction is aided by such concepts as the already mentioned Shneiderman Mantra [60] or other approaches such as multiple-coordinated views [52]. The general support for interactivity in VA is provided by the user interface (UI), based either on classic WIMP or non-standard interaction methodologies (e.g., virtual reality). Designing user interfaces, formal modeling, simulation and re-configuration can be realized using the UIEditor tool [75].

User interfaces for VA environments have to consider the user, the task and the overarching goals (and context) in which the interaction takes place [75]. The data analysis task, that is, the VA workflow applied to a visually represented data set, specifies the exploration space. Here, the exploration space characterizes the set of all possible changes in a given visualization that can be initiated by the user. Thus, the potential exploration space relates directly to operations available to the user via the user interface [72]. The set of interrelations between the available operations can be denoted as interaction logic.

Obviously, UI development benefits from the analysis of the task and the process addressed by it. In accordance with the criteria mentioned in the previous section, the quality of UIs for VAST can be assessed using the criteria of AC (accuracy, the potential for error prevention offered by VA UI), AD (adequacy, the level of suitability of the UI for the given analysis question), and EF (efficiency, UI performance for solving the given analysis question) [73, 74].

These quality criteria can be fulfilled by a VA tool to a high degree if, first, its user interface and interaction logic are modeled with the help of formal descriptions and methods. This mainly addresses AC by allowing for formal validation of user interface against formally described tasks, requirements, and specifications. Second, empirical measures can be applied similarly to studies of usability and user experience (addressing AD and EF by user involvement). For this, user studies have to be designed carefully as discussed, for example, in [70].

There are many publications in which formal modeling methods are demonstrated to support AC in the development of interactive tools. In [75], a broad variety of formal methods is presented for modeling interactive systems and, specifically, user interfaces. For instance, Weyers [72] presents a visual modeling language that enables (interactive) description of interaction logic and algorithmic transformation into Petri-net based and executable representation of a user interface. Bowen et al. [13] use Z-based specifications to describe interaction processes, which offers formal verification capabilities and helps to identify erroneous implementations, as the authors demonstrate in the context of safety critical scenarios. Another example is the use of a Petri-net based modeling approach proposed by Navarre et al. [44] addressing user interfaces and interaction in airplane cockpits. They strongly focus on verifying interaction processes for controlling an airplane.

AD and EF of a VA user interface can be evaluated empirically by conducting



*user studies* for quantifying various types of measures [73]. There are measures for usability (e.g., SUS [16]) and for user experience (e.g., UEQ [40]) based on questionnaires. Additionally, qualitative methods can be applied. For example, users can give feedback in *semi-structured interviews* about how well a VA tool can be employed for a certain task after trying it out for some time. *Think-aloud protocols* [32] allow users to phrase their thoughts about the application during its use. Similarly to this approach, *cognitive walkthroughs* [50] foster design decisions for development of VA environments. During a cognitive walkthrough users are asked to imagine the employment of a tool for solving a specific task and then to describe this verbally.

### 3 V&V Assessment – Various Examples

In this section, we discuss how (reliable) VA can be used in such varied areas as engineering, data analysis, teaching, and co-curation in virtual museums. Several software packages, initially developed at the chair of computer graphics and scientific computing at the University of Duisburg-Essen and now hosted by the owners, are summarized in Table 1. The focus of this summary is on the assessment options and features from in Columns 3 and 4 of Figure 1. It can be seen from the table that the majority of the considered tools implement VA options (which are at least partially assessed), reliable computing, uncertainty quantification, and adaptable interaction styles. Refer to the given literature for details about each of the features from the table.

First, we describe applications in which three of the relevant aspects/features are addressed. VIACOBi is an extensively evaluated interactive teaching and learning system for computer graphics. It accurately and efficiently implements geometric object rendering algorithms and visualizes them in a variety of user-driven ways. In particular, the reliability is ascertained in the following way. The implementation deals with class KAF of correctly computed functions defined on image matrices (with  $n$ -tuple values of  $k$  digit binary or base  $b$  numbers). Class KAA of the correctly implemented algorithms computes functions in KAF. KAA are numerical algorithms with result verification and accurate rendering (e.g., the Bresenham algorithm for a line with the integer start/endpoints or a circle with the integer midpoint and the square of the radius). Although uncertainty, group analytics and recommendations are not addressed explicitly, the program is highly interactive, that is, the interface and visualization can be adapted by the users. The next application in this group is given in [12]. It deals with a Petri-net based implementation of a procedural process model (a control room of KSG/GfS Essen Kupferdreh) featuring automatic HCI supervision. The author considers a part of a dynamic overall process of a nuclear power plant and the necessary interaction between the operator and the system by using formal situation operator models. The created process simulation runs in parallel with the operating process in a guided experiment hosted by the industrial partner and can be validated since operating errors are recorded and classified. A further example that features code

Table 1: Overview of the use cases. AArea stands for the intended application area of the tool, the third column reflects the use of assessment options (reliable) data analytics (RDA), (reliable) computing (RC), (reliable) visual analysis (RVA), and the fourth column shows whether the optional features uncertainty (U), group analytics (GA), interactivity (I, e.g., with VR) and recommender (R) are considered

Tool	AArea	RDA/RC/RVA	U/GA/I/R
VIACoBi [33]	interactive learning	-/+/+	-/-+/-
[12]	automatic HCI supervision	-/+/+	-/-+/-
OLSIM [17]	traffic simulation	+/+/-	+/-/-/-
UNIVERMEC [37]	geometric computations	+///+	+/-/-/-
PROREOP[4]	biomechanics	+///+	+/-/-/-
VERICOMP [5]	software comparison	+/-/+	+/-/-/+
HoR [76]	risk communication	-/-/+	+///+//+
[49]	GIS	+///+	+/-/+/-
SILENOS [66]	steel inclusions	+/-/+	+///+/-
VIMEDEAS [54]	virtual museums/labs	+///+	+///+//+

verification and model validation is a microscopic traffic modeling and simulation system from [17]. Additionally, a mechanism for analyzing uncertainty in the given data is developed there.

The next three applications address four of the aspects given in Columns 3 and 4 of Figure 1. For example, UNIVERMEC is an integrated framework for verified geometric computations. Users can specify an application problem in a standardized V&V environment. This allows them to use different verified solution techniques, to enter object data, solution quality requirements and links to algorithms, and to visualize results with the help of formalized interfaces. They can develop metrics for efficiency comparison of the employed algorithms, calculate performance parameters as well as connect various existing or newly developed tools within the framework to significantly simplify problem solving. The next application from the table is described in [4], where a number of techniques aiding femur prosthesis surgery are presented. They allow for data grabbing or reliable modeling and visualization with superquadrics. A complete classical V&V assessment of the process has been carried out. The last tool from this group, VERICOMP, is devised within an academic setting, but can also be of use for industry. It is a web-based platform for comparing verified initial value problem solvers for systems of ordinary differential equations. For users to be able to decide at a glance what solver

is the best for a given problem or to compare the general performance of different solvers for a certain class of problems, VERICOMP uses a number of visual aids such as work-precision diagrams (WPDs). WPDs help users to assess the accuracy of the verified solution provided by a particular solver (that is, its ability to provide tight bounds), its performance and its sensitivity to different characteristics (e.g., problem parameters, certain option settings, etc.). Although WPD construction itself is accurate, further work is necessary to assess the adequacy, usability and intuitiveness of this data representation. Additionally, VERICOMP provides a formalism for recommending a verified tool for the specific user's task, the process which also has yet to be assessed.

Five of the features from Table 1 are addressed in the conceptual HOUSE OF RISK (HoR) [76] which is devoted to the reliable communication of individual threats, thematically classified and placed in an indoor or outdoor context by using reliable visual representations of this data. The presented information is meant to inform experts but also the broader audience, which supports the optional feature of group analytics. HoR will address public threats and macrocatastrophes such as volcanic eruptions. Being inspired by virtual museums, HoR can be facilitated by VR technology, which also includes the (visual) representation of uncertainty. In general, HoR can be used either to visually evaluate, for example, evacuation plans, or to communicate these plans to the public. Additionally, a suggestion about potential areas of risks or information relevant for evacuation plans can be generated using a scientific recommender. In this project, key aspects of RVA and all of the optional features are addressed.

A further application addressing five of the features is from the area of reliable geographic information systems (GIS). It takes into account uncertainty during traffic localization and network planning. In [49], the authors present a verified model of uncertainty in GPS-based location systems based on the Dempster-Shafer theory with two-dimensional and interval-valued basic probability assignments. Applications that use GPS location information often neglect the fact that GPS signals are subject to uncertainty originating from such physical factors as weather conditions that influence the transmission. The authors propose visual representations and rendering methods to allow the user to investigate the induced uncertainty and assess its impact on the precision of the location. The main benefit this approach offers for GIS applications is a workflow concept using Dempster-Shafer models that are embedded into an ontology-based semantic querying mechanism accompanied by 3D visualization techniques. A 3D visualization of the position and direction uncertainty reflects the three-dimensional nature of the underlying data completely, in contrast to such 2D forms as ellipses, triangles, interval curves or tubes in the current literature [47]. To achieve this, the 2D position data is shown jointly with its mass assignment along the third axis. The developed visualization component is capable of generating layered presentations of single measurements as well as Dempster-Shafer results, for example, textured height maps or 3D box plots using EBNF based input and the Web3D visualization frameworks X3D and X3Dom [49]. Reliable computing and visualization requirements including interactive means of querying uncertain GIS models are employed throughout the workflow of this tool.

SILENOS deserves a separate mention since this practical application from the area of steel production analyzes (big) data collected about non-metallic inclusions and other defects in steel samples. It features image processing, a particle detecting and analysis system as well as an inclusion processing framework viewer IPF 2.0 [66]. It takes into account process parameters such as intentional settings or measurements taken during monitoring of various steel grades and their metadata; defect parameters, descriptors and volume data for each defect; isoperimetric shape factors such as volume, surface area, mean curvature; sample parameters such as milling machine slices of the steel surface; and statistical descriptors of the defects such as the sample cleanliness. It performs 3D reconstruction of cracks, non-metallic inclusions or pores and a trend/sensitivity analysis answering the question of how the defect data (positions, sizes, types, number) change depending on process parameters. This tool was assessed w.r.t. effectiveness, user satisfaction, learnability ( $\rightsquigarrow$  ensemble analysis); adoption rate, usability, reliability, trustability ( $\rightsquigarrow$  task work); utility, scalability, learnability of the visualization engine ( $\rightsquigarrow$  repeated multiple views); as well as w.r.t. performance, optimal visualization parameters, and accuracy of the incremental approximation.

Finally, the multipurpose system VIMEDEAS addresses all features mentioned in Table 1. It enables dynamic generation and publication of arbitrary room designs and generates virtual museum (VM) environments according to given parameters and metadata designs specified in the VM modeling language VIMCOX. It was used to implement a virtual version of the Leopold Fleischhacker Museum (LFM) within a four-year crowdsourcing project [8]. A virtual version of LFM consists primarily of annotated photographs and reconstructed tombstones. It hosts about 200 pictorial exhibits and their 3D assets in 13 rooms and a virtual cemetery area. Visitors can work with four versions of the LFM, each of which proposes a specific way to navigate through the exposition areas and various degrees of interaction. A knowledge and rule-based evaluation was carried out to deal with software stability in accordance with either the ISO/IEC 9126 or the ISO/IEC/IEEE 29119 norm, with failure-free system operation over a specified time, with stress tests for fluent navigation and display, and with the confirmation of complete and correct realization of the curator's content specifications.

## 4 Conclusions

In this contribution, we aimed at widening the focus of the scientific computations community to a broad human-centered system modeling approach and validation design. Bearing in mind the methodologies from the neighboring fields, we discussed possibilities to define a multilayer quality assessment procedure (similar to that from data analytics) concerning reliability, accuracy, performance, efficiency, group activity monitoring as well as validation and evaluation. This included various interaction/collaboration methodologies and mixed reality platforms where scientists of different disciplines could interact with each other, with data and with information.

Reliable visual analytics can be a part of such an enhanced V&V management within a workflow for designing, modeling, implementing, and analyzing various processes and their outcomes. We introduced a tentative RVA definition and illustrated the general ideas with the help of use cases implementing relevant parts of the proposed enhanced V&V assessment. Various dimensions of reliability and quality criteria, task model and interaction styles, metrics, rules and requirements were discussed. However, the final definitions are still missing.

To summarize, the following techniques have been suggested so far to ensure VAST reliability:

- characterizing big amounts of heterogeneous data by applying various quality criteria with the corresponding metrics,
- dealing with uncertainty by choosing the appropriate data types and algorithms allowing for V&V assessment through the whole process,
- visualizing uncertainty in the outcome by using geometrical forms/glyphs, colors, textures or statistical descriptors such as moments,
- using both automated and interactive data mining techniques as well as letting verified algorithms perform only a partial analysis in difficult situations, supervised and supplemented by a human,
- providing a choice of assessed mappings and visual presentation of data (or information) for systems versus experimental or simulation outcomes, combined with good structuring options,
- supporting the user in the choice of a reliable technique based on normalized values for selected quality criteria depending on the task with the help of a scientific recommender which maximizes a multi-objective utility function as an overall quality measure,
- providing a platform for data fusion, (collaborative) sense and decision making and reports with actual assessment of the suggested quality recommendations and guidelines.

Guidelines with benchmarks and measures to assure auditability and to rate mental and computer-based models are our future work.

## References

- [1] Al-Hajj, S., Fisher, B., Smith, J., and Pike, I. Collaborative visual analytics: A health analytics approach to injury prevention. *International Journal of Environmental Research and Public Health*, 14(9):1056, 2017. DOI: 10.3390/ijerph14091056.
- [2] Alefeld, G. and Mayer, G. Interval analysis: Theory and applications. *J. Comput. Appl. Math.*, 121(1-2):421–464, September 2000. DOI: 10.1016/S0377-0427(00)00342-3.

- [3] Auer, E. *Result Verification and Uncertainty Management in Engineering Applications*. Verlag Dr. Hut, 2014. Habilitation Monograph.
- [4] Auer, E., Luther, W., and Cuypers, R. Process-oriented verification in biomechanics. In Deodatis, G., Ellingwood, B. R., and Frangopol, D. M., editors, *Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures*, pages 391–398, London, 2014. DOI: 10.1201/b16387.
- [5] Auer, E. and Rauh, A. VERICOMP: a system to compare and assess verified IVP solvers. *Computing*, 94(2):163–172, Mar 2012. DOI: 10.1007/s00607-011-0178-4.
- [6] Behrisch, M., Blumenschein, M., Kim, N. W., Shao, L., El-Assady, M., Fuchs, J., Seebacher, D., Diehl, A., Brandes, U., Pfister, H., Schreck, T., Weiskopf, D., and Keim, D. A. Quality metrics for information visualization. *Computer Graphics Forum*, 37(3):625–662, 2018. DOI: 10.1111/cgf.13446.
- [7] Bertin, J. and Barbut, M. *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Gauthier Villars, 1973.
- [8] Biella, D., Pilz, Th., Sacher, D., Weyers, B., Luther, W., Baloian, N., and Schreck, T. Crowdsourcing and co-curation in virtual museums: A practice-driven approach. *Journal of Universal Computer Science*, 22(10):1277–1297, oct 2016. DOI: 10.1007/978-3-319-45550-1\_15.
- [9] Bingue, W. P. and Cook, D. A. A practical approach to verification and validation. A talk, 2014.
- [10] Blytt, M. Big challenges for visual analytics: Assisting sensemaking of big data with visual analytics. Technical report, Norwegian University of Science and Technology, 2013.
- [11] Bonneau, G.-P., Hege, H.-Chr., Johnson, Chr. R., Oliveira, M. M., Potter, K., Rheingans, P., and Schultz, Th. Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization*, volume 37 of *Mathematics and Visualization*, pages 3–27. Springer, London, 2014. DOI: 10.1007/978-1-4471-6497-5\_1.
- [12] Boussairi, H. Petrinetz-basierte Implementierung eines verfahrenstechnischen Prozessmodells – SOM-basierte automatische Überwachung der Mensch-Maschine-Interaktion. Master’s thesis, University of Duisburg-Essen, 2008.
- [13] Bowen, J. and Reeves, S. Combining models for interactive system modelling. In *The Handbook of Formal Methods in Human-Computer Interaction*, pages 161–182. Springer, 2017. DOI: 10.1007/978-3-319-51838-1\_6.
- [14] Bowman, D., Kruijff, E., LaViola Jr, J. J., and Poupyrev, I. P. *3D User interfaces: theory and practice, CourseSmart eTextbook*. Addison-Wesley, 2004.

- [15] Brodlie, K., Allendes Osorio, R., and Lopes, A. A review of uncertainty in data visualization. In Dill, J., Earnshaw, R., Kasik, D., Vince, J., and Wong, P. Ch., editors, *Expanding the Frontiers of Visual Analytics and Visualization*, pages 81–109, London, 2012. Springer London. DOI: 10.1007/978-1-4471-2804-5\_6.
- [16] Brooke, J. SUS – A quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [17] Brüggmann, J., Schreckenber, M., and Luther, W. A verifiable simulation model for real-world microscopic traffic simulations. *Simulation Modelling Practice and Theory*, 48:58–92, 2014. DOI: 10.1016/j.simpat.2014.07.002.
- [18] Brunnhuber, M., May, M., Traxler, Chr., Hesina, G., Glatzl, R. W., and Kontrus, H. Using different data sources for new findings in visualization of highly detailed urban data. In Schrenk, M., Popovich, V. V., Zeile, P., Elisei, P., and Beyer, C., editors, *REAL CORP 2017 – PANTA RHEI – A World in Constant Motion. Proceedings of 22nd International Conference on Urban Planning, Regional Development and Information Society*, pages 637–646, 2017.
- [19] Cai, L. and Zhu, Y. The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14(2):1–10, 2015.
- [20] Carpendale, S. Evaluating information visualizations. In Kerren, A., Stasko, J.T., Fekete, J.-D., and North, C., editors, *Information Visualization: Human-Centered Issues and Perspectives*, pages 19–45. Springer, 2007. DOI: 10.1007/978-3-540-70956-5\_2.
- [21] Person, Scott, Kreinovich, Vladik, Hajagos, Janos, Oberkampf, William, and Ginzburg, Lev. Experimental uncertainty estimation and statistics for data having interval uncertainty, 2007. DOI: 10.2172/910198, Sandia report SAND2007-0939.
- [22] Forrester, J. W. Counterintuitive behavior of social systems. *Theory and Decision*, 2(2):109–140, Dec 1971. DOI: 10.1007/BF00148991.
- [23] Forsell, C. and Johansson, J. An heuristic set for evaluation in information visualization. In *Proceedings of the International Conference on Advanced Visual Interfaces, AVI '10*, pages 199–206, New York, NY, USA, 2010. ACM. DOI: 10.1145/1842993.1843029.
- [24] Ghanem, R., Owhadi, H., and Higdon, D. *Handbook of uncertainty quantification*. Springer, 06 2017. DOI: 10.1007/978-3-319-12385-1.
- [25] Goldsztejn, A. and Hayes, W. Rigorous inner approximation of the range of functions. In *12th GAMM - IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics (SCAN 2006)*, page 19, Sep. 2006. DOI: 10.1109/SCAN.2006.38.

- [26] Grant, R. *Data Visualization*. Chapman and Hall/CRC, New York, 2018. DOI: 10.1201/9781315201351.
- [27] Hall, D. L. and Llinas, J. *Handbook of Multisensor Data Fusion*. CRC Press, June 2001.
- [28] Harris, R. L. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, Inc., New York, NY, USA, 1999.
- [29] Healey, Chr. G. Visualization of multivariate data using preattentive processing. Master's thesis, University of Waterloo, 1990.
- [30] IEEE Std 1012-2016 (revision of IEEE Std 1012-2012/ incorporates IEEE Std 1012-2016/cor1-2017), IEEE standard for system, software, and hardware verification and validation, Sept 2017. DOI: 10.1109/IEEESTD.2017.8055462.
- [31] Isenberg, P., Bertini, E., Lam, H., Plaisant, C., and Carpendale, S. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18:1520–1536, 09 2012. DOI: 10.1109/TVCG.2011.279.
- [32] Jääskeläinen, R. Think-aloud protocol. *Handbook of translation studies*, 1:371–374, 2010. DOI: 10.1075/hts.1.thi1.
- [33] Janser, A. *Entwurf, Implementierung und Evaluierung des interaktiven Lehr- und Lernsystems ViACoBi für die Visualisierung von Algorithmen der Computergraphik und Bildverarbeitung*. PhD thesis, University of Duisburg-Essen, 1998.
- [34] Jeong, D. H., Ji, S.-Y., Suma, E. A., Yu, B., and Chang, R. Designing a collaborative visual analytics system to support users' continuous analytical processes. *Human-centric Computing and Information Sciences*, 5(1):5, Feb 2015. DOI: 10.1186/s13673-015-0023-4.
- [35] Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melançon, G. Visual analytics: Definition, process, and challenges. In Kerren, A., Stasko, J. T., Fekete, J.-D., and North, Ch., editors, *Information Visualization: Human-Centered Issues and Perspectives*, pages 154–175, Berlin, Heidelberg, 2008. Springer. DOI: 10.1007/978-3-540-70956-5\_7.
- [36] Kendall, A. and Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 5574–5584. Curran Associates, Inc., 2017.
- [37] Kiel, St. *UNIVERMEC – A Framework for Development, Assessment and Interoperability of Verified Technics*. PhD thesis, University of Duisburg-Essen, 2014.



- [38] Kinkeldey, Ch., MacEachren, A. M., and Schiewe, J. How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386, November 2014. DOI: 10.1179/1743277414y.0000000099.
- [39] Kwon, Y., Won, J.-H., Kim, B. J., and Paik, M. Cho. Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics and Data Analysis*, 142:106816, 2020. DOI: 10.1016/j.csda.2019.106816.
- [40] Laugwitz, B., Held, Th., and Schrepp, M. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and Usability Engineering Group*, pages 63–76. Springer, Berlin, Heidelberg, 2008. DOI: 10.1007/978-3-540-89350-9\_6.
- [41] Luther, W. and Auer, E. Towards human-centered paradigms in verification and validation assessment. In Hajian, A., Luther, W., and Vinck, A. J. H., editors, *Collaborative Technologies and Data Science in Smart City Applications (Proceedings of CODASSCA Workshop 2018)*, pages 68–81, Berlin, 2018. Logos Verlag.
- [42] MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., and Hetzler, E. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, July 2005. DOI: 10.1559/1523040054738936.
- [43] Moore, R. E., Kearfott, R. B., and Cloud, M. J. *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, 2009. DOI: 10.1137/1.9780898717716.
- [44] Navarre, D., Palanque, Ph., Ladry, J.-F., and Barboni, E. ICOs: A model-based user interface description technique dedicated to interactive systems addressing usability, reliability and scalability. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 16(4):18, 2009. DOI: 10.1145/1614390.1614393.
- [45] Pirolli, P. and Card, S. The sense making process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of 2005 International Conference on Intelligence Analysis*, McLean, VA, 2005.
- [46] Potter, K., Kirby, M., Xiu, D., and Johnson, Ch. R. Interactive visualization of probability and cumulative density functions. *International Journal for Uncertainty Quantification*, 2(4):397–412, 2012. DOI: 10.1615/Int.J.UncertaintyQuantification.2012004074.
- [47] Potter, K., Rosen, P., and Johnson, C. R. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In Dienstfrey, A. M. and Boisvert, R. F., editors, *Uncertainty Quantification in Sci-*

- entific Computing*, pages 226–249. Springer, Berlin, Heidelberg, 2012. DOI: 10.1007/978-3-642-32677-6\_15.
- [48] Rebner, G., Auer, E., and Luther, W. A verified realization of a Dempster-Shafer based fault tree analysis. *Computing*, 94(2-4):313–324, March 2012. DOI: 10.1007/s00607-011-0179-3.
- [49] Rebner, G., Sacher, D., Weyers, B., and Luther, W. Verified stochastic methods in geographic information system applications with uncertainty. *Structural Safety*, 52(Part B):244–259, 2015. DOI: 10.1016/j.strusafe.2014.05.002.
- [50] Rieman, J., Franzke, M., and Redmiles, D. Usability evaluation with the cognitive walkthrough. In *Conference companion on Human factors in computing systems*, pages 387–388. ACM, 1995. DOI: 10.1145/223355.223735.
- [51] Riveiro, M. Evaluation of uncertainty visualization techniques for information fusion. In *10th International Conference on Information Fusion*, pages 1–8, 2007. DOI: 10.1109/ICIF.2007.4408049.
- [52] Roberts, J. C. State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*, pages 61–71. IEEE, 2007. DOI: 10.1109/CMV.2007.20.
- [53] Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G. P., and Keim, D. A. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Trans. Vis. Comput. Graph.*, 22(1):240–249, 2016. DOI: 10.1109/TVCG.2015.2467591.
- [54] Sacher, D. *A Generative Approach to Virtual Museums Using a New Metadata Format. A Curators', Visitors' and Software Engineers' Perspective*. PhD thesis, University of Duisburg-Essen, 2017.
- [55] Schlesinger, S. Terminology for model credibility. *Simulation*, 32(3):103–104, 1979. DOI: 10.1177/003754977903200304.
- [56] Schober, P., Boer, Ch., and Schwarte, L. A. Correlation coefficients: Appropriate use and interpretation. *Anesthesia and analgesia*, 126(5):1763–1768, 2018. DOI: 10.1213/ANE.0000000000002864.
- [57] Scholtz, J. Beyond usability: Evaluation aspects of visual analytic environments. In *Proceedings of the IEEE Symposium On Visual Analytics Science And Technology (VAST)*, pages 145–150, November 2006. DOI: 10.1177/1473871613490290.
- [58] Scholtz, J. Developing guidelines for assessing visual analytics environments. *Information Visualization*, 10(3):212–231, 2011. DOI: 10.1177/1473871611407399.

- [59] Scholtz, J., Plaisant, C., Whiting, M. A., and Grinstein, G. G. Evaluation of visual analytics environments: The road to the visual analytics science and technology challenge evaluation methodology. *Information Visualization*, 13(4):326–335, 2014. DOI: 10.1177/1473871613490290.
- [60] Shneiderman, B. The eyes have it: A task by data type taxonomy for information visualizations. In *In IEEE Symposium on Visual Languages*, pages 336–343, 1996.
- [61] Shneiderman, B., Plaisant, C., and Hesse, B. W. Improving health and health-care with interactive visualization methods. Technical report, HCIL, 2013.
- [62] Stasko, J. T. and Amar, R. A. Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11:432–442, 07 2005. DOI: 10.1109/TVCG.2005.63.
- [63] Sun, G., Wu, Y., Liang, R., and Liu, S.-X. A survey of visual analytics techniques and applications: State-of-the-art research and future challenges. *Journal of Computer Science and Technology*, 28(5):852–867, 2013. DOI: 10.1007/s11390-013-1383-8.
- [64] Thomas, J. J. and Cook, K. A. *Illuminating the Path: The R&D Agenda for Visual Analytics*. IEEE, National Visualization and Analytics Center, 2005.
- [65] Thomson, J., Hetzler, E., MacEachren, A., Gahegan, M., and Pavel, M. A typology for visualizing uncertainty. In *Proc. SPIE and IS&T Conf. Electronic Imaging: Visualization and Data Analysis*, volume 5669, pages 146–157, 2005. DOI: 10.1117/12.587254.
- [66] Thureau, M., Buck, Chr., and Luther, W. IPFViewer: Incremental, approximate analysis of steel samples. In *Proceedings of SIGRAD 2014, Visual Computing, June 12-13, 2014, Göteborg, Sweden*, pages 1–8, 2014.
- [67] Truong Le, M., Diehl, F., Brunner, T., and Knol, A. Uncertainty estimation for deep neural object detectors in safety-critical applications. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3873–3878, Nov 2018. DOI: 10.1109/ITSC.2018.8569637.
- [68] Tufte, E. R., Goeler, N. H., and Benson, R. *Envisioning information*, volume 126. Graphics press Cheshire, CT, 1990.
- [69] Tufte, Edward R. *Beautiful Evidence*. Graphics Press, Cheshire, CT, 2006.
- [70] Wallraven, Chr. and Cunningham, D. W. *Experimental design: From user studies to psychophysics*. AK Peters/CRC Press, 2011.
- [71] Ware, C. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

- [72] Weyers, B. *Reconfiguration of user interface models for monitoring and control of human-computer systems*. PhD thesis, Universität Duisburg-Essen, Fakultät für Ingenieurwissenschaften, 2012.
- [73] Weyers, B. Toward collaborative virtual environments for reliable visual analytics. In Hajian, A., Luther, W., and Vinck, A.J.H., editors, *Collaborative Technologies and Data Science in Smart City Applications (Proceedings of CO-DASSCA Workshop 2018)*, pages 59–67, Berlin, 2018. Logos Verlag.
- [74] Weyers, B., Auer, E., and Luther, W. The role of verification and validation techniques within visual analytics. *Journal of Universal Computer Science*, 25(8):967–987, 2019.
- [75] Weyers, B., Bowen, J., Dix, A., and Palanque, Ph. *The Handbook of Formal Methods in Human-Computer Interaction*. Springer International Publishing, 1st edition, 2017. DOI: 10.1007/978-3-319-51838-1.
- [76] Weyers, B. and Luther, W. Risk communication and perception in low- and high-immersion virtual environments. In Huang, H.W., Li, J., Zhang, J., and Chen, J.B., editors, *Proceedings of the 6th Asian-Pacific Symposium on Structural Reliability and its Applications*, pages 577–582, Shanghai, China, May 2016.
- [77] Wise, J.A., Hopkin, V.D., Stager, P., and Division, North Atlantic Treaty Organization. Scientific Affairs. *Verification and Validation of Complex Systems: Human Factors Issues*. International Association of Geodesy Symposia. Springer, Berlin, Heidelberg, 1993.
- [78] Wnuk, A., Debski, K. J., and Kozak, M. CI thermometer: Visualizing confidence intervals in correlation analysis. *IEEE Computer Graphics and Applications*, 37(6):103–108, November 2017. DOI: 10.1109/MCG.2017.4031058.
- [79] Zio, E. and Pedroni, N., editors. *Literature review of methods for representing uncertainty*. Cahiers de la Scurit Industrielle, Toulouse, 2013-03.
- [80] Zuk, T. and Carpendale, S. Theoretical analysis of uncertainty visualization. In *Proceedings of SPIE: Visualization and Data Analysis*, volume 6060, pages 66–79, 2006. DOI: 10.1117/12.643631.
- [81] Zuk, T. D. *Visualizing Uncertainty*. PhD thesis, University of Calgary, 2008.