Originally published as:

Taylor & Francis
Taylor & Francis Group

Assessing the quality of geoscientific simulation models with visual analytics methods – a design study

Doris Dransch[a]*, Patrick Köthur[a], Sven Schulte[b], Volker Klemann[a] and Henryk Dobslaw[a]

[a]*Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany;* [b]*Department of Architecture, Facility Management and Geoinformation, University of Applied Science Anhalt, Dessau, Germany*

Simulation models are essential means of scientific knowledge building and also the basis for decision-making. Because of their relevance, they have to be assessed thoroughly with respect to their quality. Simulation model assessment comprises two challenges: (a) modelers have to create a comprehensive mental image of the model's quality despite the massive multidimensional, multivariate, and often heterogeneous data; and (b) the model assessment process should be as efficient as possible. We face these challenges with a visual analytics approach. We aim at developing interactive visual representations which, in combination with present computational analysis methods, support the scientist's reasoning process to enhance the assessment of simulation models. In a design study, we analyzed two exemplary reasoning processes which cover the main model assessment procedures: the evaluation of the internal coherence of the model's structure and behavior and the assessment of its empirical validity. The analysis was conducted by means of a user- and task-centered approach which combines several knowledge elicitation techniques and task analysis concepts. We derived domain tasks as well as cognitive actions and developed and implemented interactive visualization components which supplement the statistical analysis methods already used. An informal qualitative user study shows that our visual analytics approach and tools help gain a more detailed mental image and hence a better understanding of the data and the underlying simulation model and allow for a faster and more comprehensive assessment of the simulation model.

Keywords: geovisualization; geovisual analytics; task analysis; model assessment; reasoning

1. Introduction

1.1. *Assessment of simulation models*

Simulation models allow us to gain insight into environmental systems and their characteristics. Thus, they are a means of scientific knowledge building and also basis for administrative and political decisions. Often they are the only source of information about environmental systems and their behavior. Because of their relevance in science and decision-making, models have to be assessed thoroughly with respect to their quality (Bossel 1994, Birta and Arbez 2007, Smith and Smith 2007). A main objective in this

---

*Corresponding author. Email: dransch@gfz-potsdam.de

process is to discover areas and components in the model which are not consistent and which differ from the real-world process. For that reason, scientists examine the internal coherence of the model as well as the accordance of the model's output with the real world.

Simulation models must be assessed with respect to the following criteria (Bossel 1994):

- Validity of model behavior: Does the simulation model show the same temporal behavior as the real-world system?
- Validity of model structure: Does the model's structure resemble the real-world system's basic structure?
- Empirical validity: Do the predicted values correspond to observed values?
- Validity for its intended purpose: Does the model fulfill the user's requirements?

These aspects give information about the accuracy of the simulation model, its behavior, and the influence of the different model components.

Simulation model assessment comprises two challenges: The first challenge deals with the output of the simulation model and its analysis. Modelers have to create a comprehensive mental image of the model's quality from the data. The second challenge covers the assessment process itself. Model assessment should be as efficient as possible; the ratio of time and cost investment to quality improvement must be appropriate.

Environmental simulation models, especially large-scale simulation models with detailed spatial and temporal resolution, produce huge amounts of multidimensional and multivariate data. The output of climate simulation models can reach petabytes of data. These data have to be examined in their spatial, temporal, and attribute dimension to detect outliers or inconsistencies with respect to a model's behavior over time, its components and structure, and its empirical validity. A further issue that has to be faced especially during the empirical validation is the heterogeneity of predicted and observed data. The predicted data are calculated on a discrete grid that covers a geographic region homogeneously. They can also be calculated for any time interval of interest. Besides, all data are computed in the same way and are, therefore, of well-defined accuracy. In contrast, the collected data are mostly available only at single points in space and at single points in time. In addition, they are often of different accuracy especially if the data are not measured directly but derived from indices. In spite of these restrictions, scientists have to create knowledge about the consistency of predicted and collected data and about the model's quality variation in space, temporal behavior, and internal structure.

The second challenge is related to the assessment process itself. Scientists have to detect the areas and components where their models need improvement. They have to analyze the models' output in a systematic way to catch all outliers and inconsistencies. Furthermore, they have to find reasons for the deficiencies by identifying causes, judging them with their expert knowledge, drawing conclusions, and adjusting the model accordingly. This process of assessment and adjustment is performed in an iterative loop. As the process can become quite time consuming and computationally intensive, especially with complex models, it is necessary to support it efficiently.

Methods and work environments applied at present do not meet these challenges comprehensively (Hey *et al.* 2009). In the model assessment process various statistical methods are applied. Quantitative approaches calculate a variety of difference measures and correlation coefficients (Smith and Smith 2007). However, these methods have limitations with regard to heterogeneous data or the simultaneous depiction of statistical, spatial, and temporal dimensions. Geovisualization can cope with these limitations. Despite this potential, little visualization is used so far. A common graphical approach in environmental

modeling is plotting and regressing predicted versus observed values in a scatter plot and comparing the regression to the 1:1 line (Piñeiro *et al.* 2008). Further graphical means are maps or map iterations which depict a model's result in its spatial distribution and temporal behavior (Dobslaw 2007, Smith and Smith 2007). The potential of advanced visualization techniques such as interactive visualization or linked and multiple views is rarely used. Additionally, scientists have to deal with work environments which are not related to their reasoning process. They have to switch back and forth between multiple software tools with limited scope and also between various data subsets. For example, statistical programs have shortcomings in graphical representations and visualization tools often lack calculation and analysis functionality. This results in an inefficient work process, limited support of the reasoning process, and probably unsatisfying results. Accordingly, enhanced technology is required which facilitates the reasoning process comprehensively and allows for linking all necessary functionality (Gahegan 2005).

1.2. *Visual analytics – a means for assessing scientific simulation models*

Visual analytics is an expanding field of interest to cope with the fast growing amount of data collected from sensors and provided by databases. It focuses on the 'analytical reasoning facilitated by interactive visual interfaces [and develops and provides methods and tools] to synthesize information and derive insight from massive, dynamic, ambiguous and often conflicting data' (Thomas and Cook 2005, p. 4). The goal of visual analytics is to facilitate high-quality human judgment with a limited investment of the analysts' time. The research challenge of visual analytics is to develop enhanced concepts and technologies which combine knowledge and methods from various disciplines: visualization, interaction, computer-based data analysis, analytical reasoning, cognition, and perception (Thomas and Cook 2005, VisMaster Consortium 2010).

Initially, visual analytics was described in the context of homeland security. However, its potential goes far beyond this and is advantageous for all data-intensive application fields. One of these fields is the natural sciences. In a recent article published in *Science* (Bell *et al.* 2009) and in a recent book called '*The fourth paradigm: Data-intensive scientific discovery*' (Hey *et al.* 2009), a new era of science has been defined. The authors distinguish four evolving states of science: experimental science describing natural phenomena; theoretical science using generalizations such as Keppler's or Newton's law; computational science, which simulates complex phenomena; and finally data-intensive science, which explores data collected from instruments, experiments, simulation models, and databases. Scientists 'now do not actually look through telescopes. Instead, they are 'looking' through large-scale, complex instruments which relay data to data centers and only then do they look at the information on their computer' (Gray 2009). All authors argue that new methods and tools are necessary to support data analysis, facilitate the scientists' analytical reasoning process, and improve reproducibility of the process as well as the results.

The process of analytical reasoning is central to the scientists' task of developing meaning from the data by applying human judgments to reach conclusions from a combination of evidence and assumptions based on the data (Thomas and Cook 2005). The core of the process is the analytical discourse which can be described as interactive, computer-mediated process of applying human judgment to assess an issue. It is an iterative and evolutionary process where the strength of both computer system and the human are harnessed to improve the analysis process (Thomas and Cook 2005). 'An important aspect of the science of analytical reasoning is to support the analytical discourse by creating ways to represent data in forms that afford interaction and enable thought processes to translate

from data to information, information to meaning, and meaning to understanding' (Thomas and Cook 2005, p. 50). Interaction must be central in the data representation concept because it influences how information is considered. It allows taking a second and subsequent look as well as viewing the same information from different perspectives. Thus, it supports key components in reasoning and knowledge building. The representations of data, in our case the visual representations, can be considered as external memory aids which enable scientists to cope with the volume and complexity of information as Norman describes in his book '*Things that make us smart*' (Norman 1999).

The assessment of scientific simulation models can benefit from visual analytics methods and technology. The assessment process can be regarded as a scientist's analytical discourse with the simulated data, the observed data, and the model itself. For this process, appropriate interactive visual representations are required as external memory aids, which support data analysis, the identification of causes for nonfitting parts of the model, and the judgment of findings and conclusions which are applied to the model to improve it.

1.3. Rationale for this study

In our research, we address the specified challenges from model assessment and visual analytics. We combine concepts from analytical reasoning, data processing and analysis, as well as geovisualization and relate them to the requirements of simulation model assessment. The objective of this article is to develop interactive visual representations which, in combination with present computational analysis methods, support the scientists' analytical discourse during the assessment of simulation models. We explore the following questions: How does the reasoning process in the case of simulation model assessment work? And which visual representations are suitable to facilitate the scientists' reasoning process and the analytical discourse between predicted data, collected data, and the model? We developed and implemented concepts for interactive visual representations which supplement the statistical analysis methods already used.

The article is structured as follows: in Section 2, we outline the methodical approach of our study. In Section 3, we analyze two exemplary reasoning processes which are related to the assessment of the internal coherence of the simulation model's structure and behavior, and to the model's empirical validity. Section 3 also describes the interactive visualization components which were developed based on (a) the reasoning process and (b) on the requirements of the huge amounts of spatiotemporal, multivariate, and heterogeneous data. The results of an informal qualitative user evaluation are given in Section 3, too. The article ends with a discussion of the results and conclusions for further research (Section 4).

2. Methodology

2.1. General methodological approach

For our investigation we combined several methods and concepts. We applied a user- and task-centered approach to create suitable visualizations with respect to the data as well as the users' reasoning process. This methodological approach follows the theoretical concept of activity theory (Nardi 1996, Dransch 2002) which forms the basis for the task-oriented approach applied in human–computer interaction (HCI) and modern geovisualization (Andrienko and Andrienko 2006) to achieve a suitable design of computer and visualization artifacts. Artifacts are tools which people create as external aids to fulfill a task. According to activity theory, each activity is described by objectives, which direct an activity, and by

tasks, which are necessary to reach the objectives. The tasks are operationalized by actions into executable components. Activity theory is applied in many disciplines to study and formalize physical as well as mental activities (Hacker 1978, Dörner 1983, Werlen 1997). Existing task taxonomies were analyzed and compared concerning their suitability for our purpose (Köthur 2009). We followed Knapp's (1995) methodology because of its comprehensive approach to task analysis which considers user needs as well as data characteristics. To use this methodology, detailed knowledge about the users' reasoning process, workflow, and work environment is necessary. The users or scientists should be able to verbally express what they are doing, why they are doing it, and how they are doing it. When this information is available, Knapp offers an approach to task analysis that derives specific visualization designs from domain tasks, related cognitive actions, and data characteristics. Knapp's methodology consists of three main components: task structure, task model, and design model. The *task structure* is established through an initial interview and organizes the user's domain tasks. These tasks are further specified in a *task model* which comprises six components:

- Task: 'What' is to be accomplished?
- Goal: 'Why' it is to be accomplished?
- Physical actions: 'How' it is to be accomplished?
- Mental actions: Thought process while accomplishing it.
- Data: The data set with which it will be accomplished.
- Visual operators: Cognitive actions for visual interaction with display.

For the definition of the visual operators, Knapp draws on the work of Wehrend and Lewis (1990). She identifies 4 of their 11 visual operators as being most important for the visual process: *identify*, *locate*, *compare*, and *associate*. Lastly, for each visual operator, the *design model* describes the relevant information components and relationships and determines their visual representation.

Although we are guided by Knapp, we introduce several changes regarding the methodological approach and terminology. Following the concepts of activity theory, we split Knapp's task model into the domain task model and the cognitive action model. The domain task model consists of all domain-specific descriptions of the tasks. The cognitive action model describes the cognitive actions for visual interaction with the display, which are called visual operators in Knapp's terminology. We made this differentiation to emphasize the two different levels which have to be regarded during the task analysis process. Additionally, we distinguished the cognitive actions according to the taxonomy introduced by Wehrend and Lewis (1990). We found this taxonomy to be more detailed and less ambiguous than Knapp's. As a further modification, we introduced two groups of cognitive actions with differing granularity: cognitive actions and classes of cognitive actions. We noticed that many domain tasks have quite similar cognitive actions in terms of action and data type (e.g., 'identify areas of similar standard deviation values' versus 'identify areas of similar difference values'). Where appropriate, we grouped them into general *classes of cognitive actions* (e.g., 'identify areas of similar quantitative values'). The visualization design model developed for these general classes is also applicable to its instantiations, the specific cognitive actions.

Summing up, our user- and task-centered approach to visualization design comprises three steps: (a) analyzing the reasoning process and creating a domain task model, (b) deducing cognitive actions and combining them to classes, and (c) selecting suitable visualization and interaction techniques and combining them with computational analysis
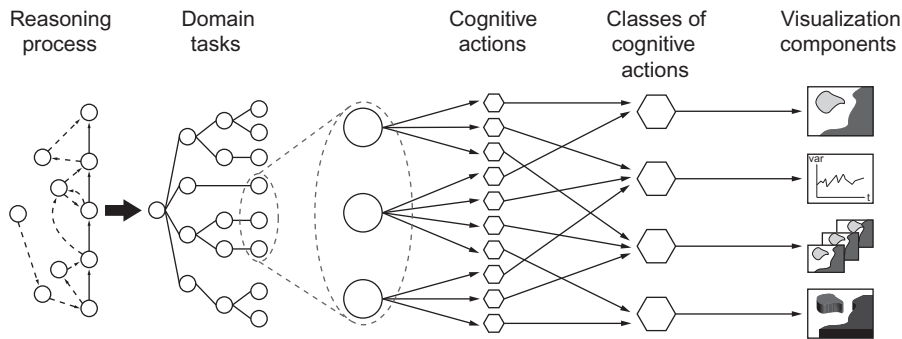
Figure 1. Methodological approach: analysis of the reasoning process and derived domain tasks, cognitive actions, and visualization components.

methods. Figure 1 illustrates our approach and the relations between reasoning process, domain tasks, cognitive actions and their classes, and visualization components (VCs).

2.2. Task analysis

The starting point of our case study was the request of two scientists who approached us with their respective simulation model assessment problems, hoping to get some advice on how to use visualization to facilitate the assessment process. The scientists are both experts in geoscientific modeling and involved in all three aspects: developing, evaluating, and improving their respective simulation model. These two scientists participated in the task analysis. Two experts are surely not sufficient to derive a complete task model of simulation model assessment, but they will provide a good basis for such a model which has to be completed by conducting further task analyses.

We began our case study with a task analysis in which we used unstructured and semi-structured interviews to elicit the required knowledge from the domain experts. These techniques provide more flexibility than completely structured interviews, which is especially useful at the beginning of the task analysis process when the task analyst has little domain knowledge (Cooke 1994, Bortz and Döring 1995, Jonassen *et al.* 1999). Each participant took part in four interviews lasting from 1 up to 3 hours. First, we conducted an unstructured interview, starting with a short briefing in which we explained our goal of grasping the participant's reasoning process and establishing a hierarchy of relevant domain tasks. The general opening question was 'What do you do to assess your simulation model? Please take us through the process.' Domain tasks mentioned by the participant were written on index cards and card sorting was applied to have the domain expert build a task tree. After the initial interview, we used the information provided by the interviewee to add Knapp's (1995) modified task model components to each domain task in the derived task structure. However, the information did not always suffice to create detailed descriptions of goal, physical actions, mental actions, and data, much less cognitive actions. Hence, subsequent semi-structured interviews focused on the discussion of the derived task structure, specifically identifying the domain experts' physical and mental actions as well as relevant data subsets. To elicit the physical and mental actions we used think-aloud technique.

The resulting domain task model provides a comprehensive view on the domain problem and its related domain task structure. Yet, at this point in the task analysis process the domain task model alone did not suffice because it lacked information about the cognitive actions of the user (visual operators in Knapp's (1995) terms). This information, however, is essential

for effective visualization design. Therefore, the next step was to determine relevant cognitive actions. The domain experts' descriptions of goals and mental actions provided valuable insight. For example, one interviewee said that it is important to find areas of high temporal standard deviation values of sea surface heights because the geophysical properties of these areas may point to possible numerical causes in the model. We concluded that related cognitive actions are, among others, to describe the spatial distribution of standard deviation values and to locate high standard deviation values in geographical space.

2.3. Visualization design

For each class of cognitive actions we defined a VC. During the design we were guided by the following principle: the cognitive actions determine the information which has to be presented. Bertin's (1974) concept of graphical semiology as well as cartographic and information visualization methodology and theories are the basis for creating suitable presentation methods. User preferences assist to make the best choice among several visualization options. We designed each VC according to three issues: first, we identified the spatial, temporal, and statistical information necessary for the user to fulfill his task. Second, we determined the visualization design which suitably depicts the required information. We considered data characteristics (e.g., statistical scale, geometrical dimension, and spatial or temporal distribution), graphic variables (Bertin 1974), cartographic map types and chromatics (Imhof 1972, Arnberger 1977, MacEachren 1995), interactive and linked views from information visualization (Tufte 1997, Spence 2007), as well as user preferences. Third, we determined if interaction related to the data or graphical design is required and which type and design of interactivity would best support the requirements (Zetie 1995, Preim 1999). In a final step, the VCs were combined in a way to support the indicated domain tasks.

2.4. Evaluation

We evaluated our visual analytics concepts in an informal qualitative user study. We chose a user-based, result-oriented, and formative evaluation form. The two scientists who initiated our work by asking for support in the model assessment process and who already participated in the task analysis took the role of evaluators of our visual analytics tools. They assessed the tools iteratively during the whole design process. While demonstrating our tools to the scientists, we asked them to evaluate the information, graphical design, and interactivity of each VC with regard to their ability to facilitate their tasks and reasoning process. We discussed the advantages and disadvantages in an informal interview. The results of the discussion were considered in the tool design which was improved iteratively.

3. Results

3.1. The reasoning process of simulation model assessment – two examples

As the basis for appropriate visual analytics concepts and tools for simulation model assessment is a sound understanding of the related reasoning process, we explored two exemplary model assessment processes. The first process deals with the assessment of the internal coherence of the simulation model's structure and behavior, the second one examines the empirical validity of the simulation model.

The *assessment of the internal coherence of the simulation model's structure and behavior* was studied in the context of the Ocean Model for Circulation and Tides (OMCT). This globally discretized numerical model simulates the general ocean circulation in response to atmospheric forcing as well as the dynamics of ocean tides generated by the attracting forces of sun and moon. Originally developed to assess nonlinear interactions among circulation and tides, the model is presently used for a variety of geophysical applications that include the correction of short-term mass signals in satellite gravimetry observations, the interpretation of regional changes in sea level, or the assessment of the impact of ocean dynamics on changes in the Earth's rotation rate (Dobslaw 2007). The OMCT is discretized on a globally defined regular grid with $1875°$ spatial resolution. The third dimension is realized with 13 layers of varying thickness ranging between 20 m close to the surface and 1200 m in the deep ocean basins. Ocean state variables, as, for example, three-dimensional distributions of temperature, salinity, and current velocities, are calculated every 30 minutes, selected output fields such as sea surface height variations or changes in ocean bottom pressure are stored every 6 hours for subsequent analyses. The output data of the model can therefore be characterized as high-dimensional, multivariate, and huge in their quantity, for example, 240 GB for a 20-year time series. To cope with the amount of data, scientists focus on the dimensions and parameters which have proven to be most relevant to the assessment of the OMCT, for example, sea-level variations and bottom pressure changes.

The assessment process of new simulations is twofold. First, the plausibility of the simulation output is assessed. For this purpose, the reduced spatiotemporal and multivariate data set has to be analyzed systematically in its spatial, temporal, and attribute dimension to detect outliers or inconsistencies. This is done in two ways. On the one hand, the temporal variability of attribute values is used as a means to find outliers. The domain expert calculates the standard deviation and studies the spatial location and distribution of outliers. He further analyzes the temporal dynamics of attribute values in areas with high standard deviation and deduces possible causes from the areas' spatial location and their geophysical characteristics. On the other hand, the time-averaged OMCT data are compared to climatologies. High differences between both data sets give the first clues on potential weaknesses of the simulation. The scientist analyzes the spatial location and distribution of particularly high differences as well as the magnitude and sign of these differences. The areas' spatial location and their geophysical characteristics, again, point toward possible causes.

The second task for assessing the model's internal coherence is the evaluation of the influence of certain modifications in the model related to, for example, changes in the atmospheric forcing fields or parameterization refinements of individual physical processes on the model output. For this purpose, the domain expert compares the values of two simulation runs searching for areas of high standard deviation of the differences. Next, he studies the temporal variability of the differences in these areas. Lastly, he infers possible geophysical and/or numerical reasons for the high variability in context with the local conditions of the area and the nature of the temporal variations.

The empirical validity of the simulation model, as our second example, was studied in the context of a glacial isostatic adjustment (GIA) model. GIA means the deformational motion of the solid Earth and the corresponding sea-level variations related to surface-mass changes like the ice-sheet dynamics during the last glacial cycles. The GIA, for example, explains the uplift of northern Sweden and partially the subsidence of the Netherlands. The numerical model describes the dynamic coupling of the water redistribution stored in the oceans and the cryosphere with the deformation processes inside the solid Earth which are mainly influenced by the mantle viscosity and the flexure of the lithosphere. It delivers in addition to sea-level variations the kinematics and gravity variations at the Earth surface. The model is

applied to reconstruct sea-level variations during the past 100,000 years and to predict its future trends as well as to correct geodetic observables like the Earth's gravity field variations determined by GRACE satellites. The model's quality is assessed and adjusted because of the process of sea-level variation during the past 20,000 years reflected by sea-level indicators (SLIs). SLIs represent fossils such as shells, bones, or plants with evidences for the relative sea level related to the time when the SLI settled or was fixed to its present position. About 14,000 SLIs are available for the past 60,000 years and are stored in a relational database system. They are described by 25 quantitative and qualitative attributes; the most important are location (x, y coordinates), height with respect to present sea level, age, type, and quality. They are irregularly distributed in space as well as in time. According to the character of the SLI – for example, if it is a species or some material washed to the former shore – its evidence for relative sea level is often indirect or imprecise.

In the assessment process, the modeler has to compare precise model predictions of sea-level variations, which are regularly distributed fields in space and time, with the SLIs which are distributed irregularly in space and time and have only an interval-related value for the sea-level height. Because of the heterogeneous accuracy of the SLIs, using statistical methods to calculate quality measures for the consistency of data at single points is not appropriate. Other analytical procedures have to be chosen. The SLIs are processed according to a fuzzy logic approach (Klemann and Wolf 2007) which enables the inference of imprecise data based on a well-defined mathematical formalism (Zadeh 1978). This formalism assigns an approximated evidence of sea-level height to each SLI which then can be compared to the predicted value from the model. A decision about the model consistency can be made. To analyze the model quality, SLIs have to be validated according to their fit to the model predictions. To find the reasons for deviations, the SLIs are studied with respect to their spatial location and height, their classification in terms of sea-level evidence, or whether the fuzzy logic approach has to be modified. Besides, the investigation of single points, spatial areas, time periods, and SLI categories are traced for clustering of fitting or nonfitting SLIs. These findings are investigated by the scientist, usually in an iterative process, for a better interpretation of the SLIs and the adjustment of the fuzzy methods applied.

From these reasoning processes we derived hierarchically structured domain task models which show the tasks and their relations (Figures 2 and 3). For each task, we defined the related cognitive actions. An example is given in Table 1. Finally, all specific cognitive actions were grouped into general classes of cognitive actions (Table 2).

3.2. Visualization components to facilitate the reasoning process

For both assessment processes we developed and implemented visualization components (VCs) comprising appropriate visualization and interaction techniques. Each VC supports one class of cognitive actions. For the design of each VC we considered (a) the information required to perform an action; (b) data characteristics, graphical variables, as well as cartographic and information visualization methods; and (c) user preferences. We applied visualization techniques preferred in the application domain if they appropriately support information presentation.

VCs to assess the model's internal coherence (e.g., OMCT) are as follows:

VC A1
Class of cognitive actions: Identify areas of similar quantitative values. Describe the spatial distribution of quantitative values.
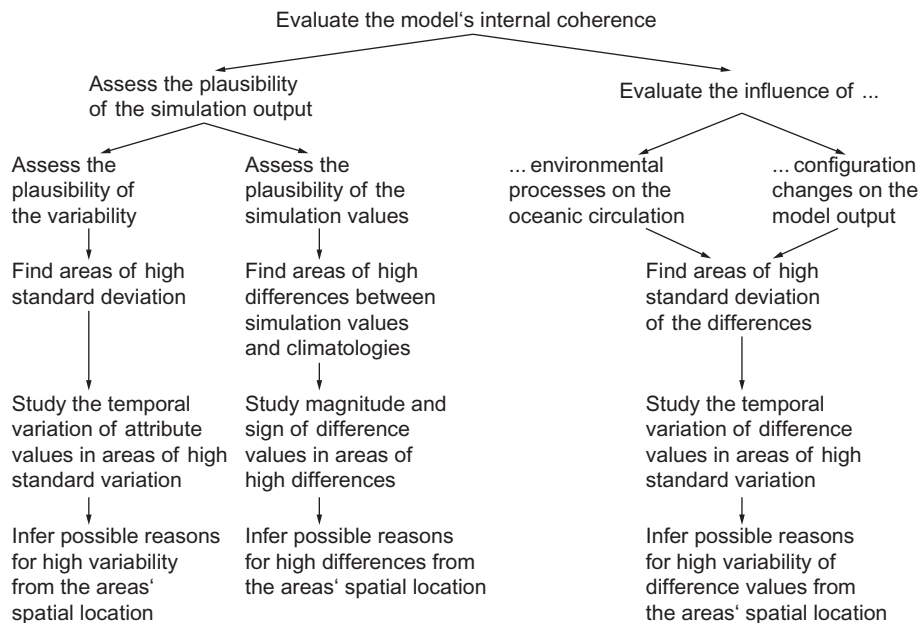
Figure 2.   Exemplary domain task model for assessing the internal coherence of the simulation model's structure and behavior (e.g., OMCT).
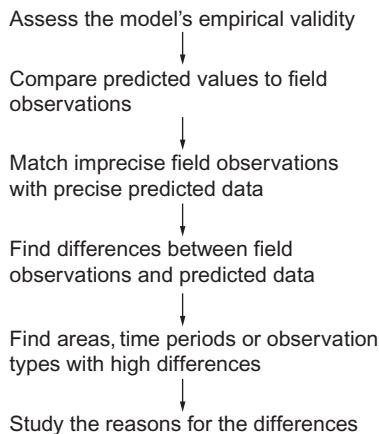

Figure 3.   Exemplary domain task model for assessing the empirical validity of a model (e.g., GIA model).

Visualization design: The information which has to be presented is the spatial distribution of quantitative values. We used a cartographic isopleth map with continuous tint to show the continuous spatial distribution of the values. Color coding following the color cycle (in our case from light yellow to dark red) in combination with lightness was chosen to present the quantitative data. The combination of both graphical variables was applied to get a suitable interval of discriminable color values. Three-dimensional representations are rejected because of their disadvantages in depicting spatial distribution because of occlusion. Although the OMCT data have three spatial dimensions, isopleth maps are a suitable 2D visualization technique for the cognitive action we identified. Scientists reduce the 3D

Table 1. Domain tasks and related cognitive actions for assessing the plausibility of the variability (example: OMCT).

Domain tasks	Cognitive actions
Find areas of high standard deviation.	Describe the spatial distribution of standard deviation values.
	Identify standard deviation values at spatial and thematic locations.
	Identify areas of similar standard deviation values.
	Identify high standard deviation values in attribute space.
	Locate high standard deviation values in geographical space.
Study the temporal variation of attribute values in areas of high standard deviation.	Describe the spatiotemporal dynamics of attribute values in areas of high standard deviation.
	Identify periods of high spatiotemporal variability.
	Describe temporal dynamics of attribute values at geospatial locations quantitatively.
Infer possible reason for high variability from the areas' spatial location.	Associate areas of high variability with their geophysical characteristics.

Table 2. Classes of cognitive actions for assessing geoscientific simulation models.

Classes of cognitive actions for assessing the internal coherence of the simulation model's structure and behavior (e.g., OMCT)

Identify areas of similar quantitative values.
Identify high quantitative values in attribute space.
Identify quantitative values at spatial, temporal, and thematic locations.
Identify periods of high spatiotemporal variability.
Locate high quantitative values in geographical space.
Describe the spatial distribution of quantitative values.
Describe the spatiotemporal dynamics of quantitative values.
Describe temporal dynamics of values at geospatial locations quantitatively.
Associate areas with their geophysical characteristics.

Classes of cognitive actions for assessing the model's empirical validity (e.g., GIA model)

Describe the spatial distribution of field observations.
Identify single field observations with outlying local position.
Cluster field observations according to spatial characteristics.
Compare field observations of one cluster with respect to type, time period, and quantity.
Correlate field observations of one cluster to predicted data with respect to observation type, time period, and quantity.
Identify spatial areas with high differences between field observations and predicted data.
Identify time periods with high differences between field observations and predicted data with respect to their spatial location.
Identify classes of field observations with unsuitable categorization.
Describe the pattern of differences between field observations and predicted data for all clusters with respect to type, time, and quantity.

OMCT data set to discretized layers within the ocean where they study particular parameters. A further reason for this type of map was the user preference; scientists are well trained in using and reading this type of map. Interactivity was not necessary in this case (Figure 4, upper left view).
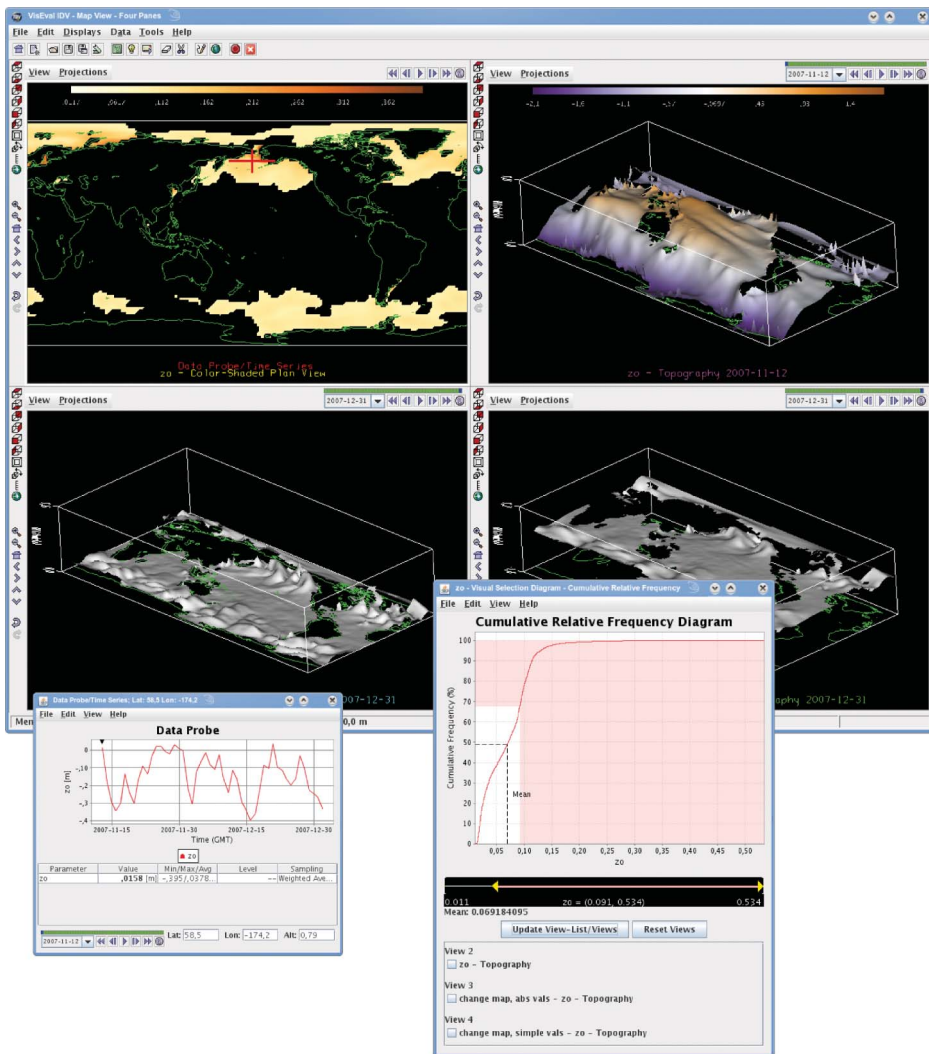
Figure 4.　Visualization components for the domain task of assessing the plausibility of the variability of the OMCT output. VC A1 (upper left view) shows the spatial distribution of standard deviation values. VC A2 (lower right diagram) allows for the identification of high values in attribute space as well as for filtering of areas through sliders (VC A5). VC A6 (upper right view) visualizes spatiotemporal dynamics of the underlying values. VC A4 (lower right and left views) depicts absolute or simple spatiotemporal variability values. The lower left line chart (VC A7) allows for an exact quantification of the dynamics of values at a spatial point location which is indicated by a red cross in VC A1.

VC A2

Class of cognitive actions: Identify high quantitative values in attribute space.

Visualization design: The information which has to be presented in this VC is the statistical distribution of the quantitative data in the attribute space; especially the range of high attribute values should be apparent. Therefore, we created a cumulative relative frequency diagram to show the distribution of values in attribute space. Scientists are familiar with this representation form. A graphical marker of the mean

value allows for a better identification of high values and a colored square denotes the range of data which has been selected. To support interactive, dynamic selection of a range of values an interactive slider was implemented in the diagram (Figure 4, lower right diagram).

VC A3

Class of cognitive actions: Identify quantitative values at spatial, temporal, and thematic locations.

Visualization design: The information required for the cognitive action is the exact parameter value at single points in spatial, temporal, and thematic space. Interactive mouse pointing combined with other VCs is implemented to meet this requirement.

VC A4

Class of cognitive actions: Identify periods of high spatiotemporal variability of an attribute.

Visualization design: This VC has to show the following information: (a) the temporal variability of an attribute which means the difference – increase or decrease – of the attribute values between subsequent time steps; (b) the temporal behavior/pattern of the variability in a defined time period; and (c) the spatial pattern of the variability in a time interval. To present the absolute or simple variability of the attribute values we used the cartographic method of change map to show the difference between two values of subsequent time steps. The quantitative variability data are mapped to the graphical variable size (height) which best presents ratio scaled data and which makes outliers and gradients most salient. Because of the data's continuous spatial distribution, a thematic surface was chosen as cartographic representation method. Although we have chosen an isopleth map with continuous tint to show quantitative values in VC 1 we decided on the 3D representation in this case because it increases the saliency of outliers and gradients in the quantitative values. Peaks become more apparent than in continuous 2D color representations. To make the spatial as well as the temporal pattern of the variability in a defined time period apparent, we decided to apply animation. The strength of animation is to enable users to recognize structures and patterns which are not apparent in static representations as studies have shown (Johansson 1973, Moellering 1976, Ullmann 1979, Braddick 1980) (Figure 4, lower left and right views).

VC A5

Class of cognitive actions: Locate high quantitative values in geographical space.

Visualization design: To show the required information of locations with high attribute values, the cumulative relative frequency diagram (VC A2) is combined and linked with map views (e.g., VC A1). The diagram provides a view on the attribute space, the map on the geographical space. Dynamic selection of high values in attribute space – realized through interactive sliders in the frequency diagram – facilitates the detection of areas with high attribute values in the linked and dynamically updated map view where the corresponding spatial locations are filtered (Figure 4).

VC A6

Class of cognitive actions: Describe the spatiotemporal dynamics of values.

Visualization design: This VC is similar to VC A4. The difference is the required information and, therefore, the applied cartographic method. VC A4 uses change maps to depict the difference of a value (variability) between two time steps in combination with a thematic surface and animation method. VC A6 adopts only thematic surface and

animation as cartographic methods because underlying attribute values and not differences between values have to be depicted. The criteria for choosing these methods are already explained in VC A4. In Figure 4, upper right view, we additionally applied a bipolar color scale to provide further information about the progression outward from a zero point of the attribute range.

VC A7

Class of cognitive actions: Describe the temporal dynamics of values at geospatial locations quantitatively.

Visualization design: The temporal behavior of attribute values at single spatial points is the information which has to be shown by this VC. We have chosen a line chart with time axis. Charts are good means to represent quantitative values; line charts are used to denote continuous data such as temporal data. Charts are also used in cartography to depict quantitative values at single points (Figure 4, lower left line chart). A marker in a linked map view indicates the corresponding spatial point location (Figure 4, upper right view). The line chart updates instantaneously when dragging the marker across the map.

VCs to assess the model's empirical validity (e.g., GIA model) are as follows:

VC B1

Class of cognitive actions: Describe the spatial distribution of field observations. Identify single field observations with outlying local position.

Visualization design: The spatial distribution of single qualitatively different field observations is the information which has to be represented in the VC. A dot map is applied to show the spatial distribution of field observations. The dots are color coded to represent the different qualitative observation type (Figure 5a).

VC B2

Class of cognitive actions: Cluster field observations according to spatial characteristics.

Visualization design: A convex boundary box is calculated to group the field observations according to the scientist's requirements. This clustering is necessary because the scientist compares larger areas and not single field observations (points) in the first evaluation stage. The area inside the boundary box is marked with transparent color to make the spatial dimension of the cluster apparent (Figure 5a–d).

VC B3

Class of cognitive actions: Compare field observations of one cluster with respect to type, time period, and quantity.

Visualization design: The required information is a qualitative and quantitative synopsis of several field observations differentiated according to type, time period, and quantity. A chart map was chosen to depict the quantitative data and their spatial position. We used a horizontally stacked bar chart in combination with color to differentiate the field observation types. The horizontal axis of the diagram represents the number of field observations of a certain type. An additional vertical axis denotes the time period. The further the bar is from the origin, the older the observations are (Figure 5c).

VC B4

Class of cognitive actions: Correlate field observations of one cluster to predicted data with respect to observation type, time period, and quantity.
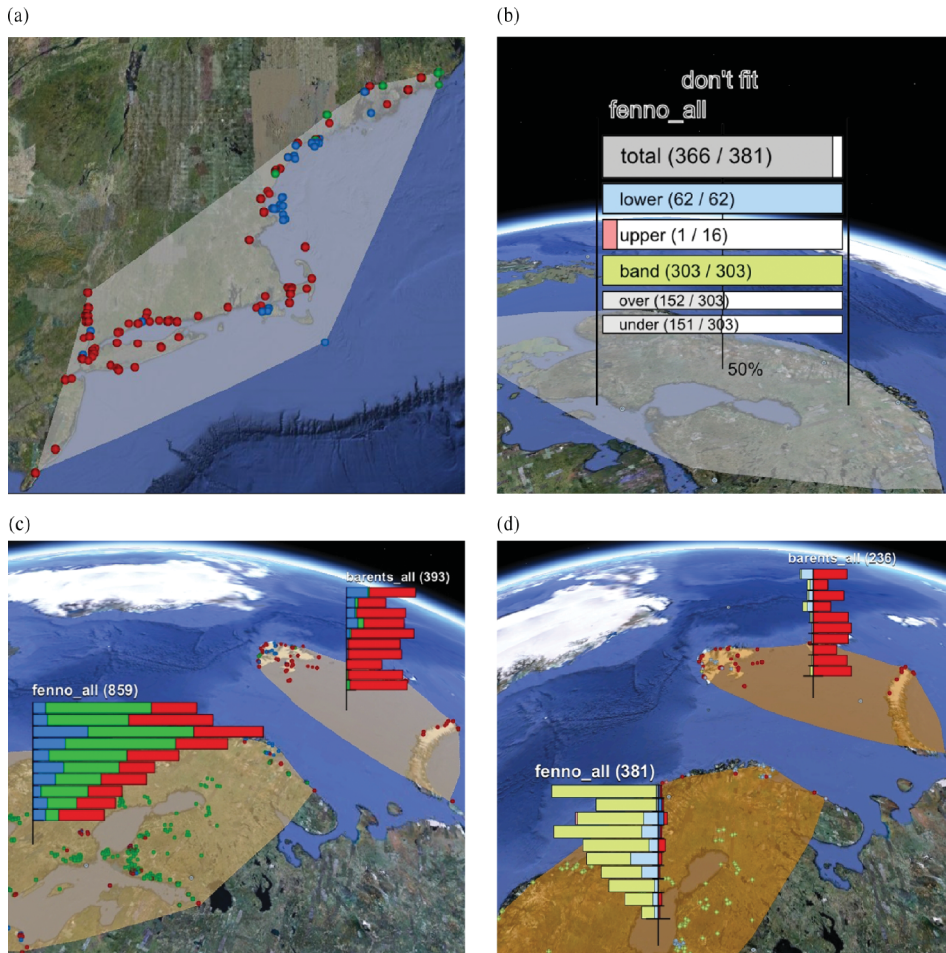
Figure 5.   (a) A color-coded dot map depicts the spatial distribution of field observations within a spatial cluster presented as gray area. (b) The quantities of nonfitting field observations per observation type in a cluster are presented in the diagram. (c) The bar diagrams show the summarized values of the clustered field observations characterized by type (color), quantity (bar length), and age (distance from origin). (d) The diagram depicts the relation between fitting and nonfitting field observations. Nonfitting observations are grouped left from the diagram axis and presented with more brightness.

Visualization design: This VC has to depict the information of VC B3 supplemented by an additional information component which shows the fitting or nonfitting of field observations and predicted data. Therefore, the basic graphical means of VC B3 and B4 are identical. However, because of the requirements of comparing two quantitative data sets with respect to their fitting, a second stacked bar chart is added. The bar chart at the right side of the vertical axis shows the observations which fit with the predicted data, the bar chart at the left side shows the nonfitting ones. The nonfitting observations are presented with more brightness for better differentiation (Figure 5d). Brightness is used to show the ranking of the data (fitting, good quality; nonfitting, poor quality).

VC B5

Class of cognitive actions: Identify spatial areas with high differences between field observations and predicted data.

Visualization design: The spatial pattern of nonfitting values has to be presented in this VC. The dot map of VC B1 is applied to show the spatial distribution of single fitting and nonfitting points. To differentiate fitting and nonfitting points, the brightness introduced in VC B4 is added to the dots. Besides, the clusters of VC B2 are differentiated by color if the ratio of all clustered fitting and nonfitting points is higher than a chosen threshold. This representation facilitates the spatial comparison on the cluster level. To investigate the reasons for high variations of fitting and nonfitting values, a diagram with numerical information related to field observation type (VC B7) can be created from the database and added interactively.

VC B6

Class of cognitive actions: Identify time periods with high differences between field observations and predicted data with respect to their spatial location.

Visualization design: The spatiotemporal pattern of fitting and nonfitting observations has to be presented in this VC. To depict the spatial pattern, we combined the dot map of VC B5 and the convex boundary box of VC B2. To depict the temporal pattern, we chose an interactive slider to select continuous time periods which determine the dots presented in the map. The continuous temporal sequence of dot maps with varying dot distributions and fitting or nonfitting ratios facilitates the identification of high variations and therewith time periods with high differences of observed and predicted data. In agreement with the Earth scientist who is the user of the visual analytics tool, we chose this type of representation instead of a diagram because it shows the spatial as well as the temporal dimension. This VC can be used supplementary to VC B4 to investigate the spatiotemporal pattern of nonfitting observations in one cluster in more detail.

VC B7

Class of cognitive actions: Identify classes of field observations with unsuitable categorization.

Visualization design: This VC has to depict classes of field observations in a cluster which have a high number of nonfitting observations. We used a bar chart to depict the quantitative data. The chart depicts the ratio of fitting and nonfitting values for each observation type. The length of the bars denotes 100% of nonfitting values. The colored segment of the bar represents the actual nonfitting value. Color is applied analogous to VC B4 to denote the observation type and the fitting or nonfitting category. Additionally, more precise information about positive and negative deviations of observed and predicted data of one class can be added interactively to the chart in the form of additional bars (Figure 5b).

VC B8

Class of cognitive actions: Describe the pattern of differences between field observations and predicted data for all clusters with respect to type, time, and quantity.

Visualization design: A holistic view on the variation in total and relative number of observations, in temporal distribution, in fitting and nonfitting ratio as well as in spatial distribution on a cluster level is necessary to fulfill this cognitive task. To achieve these requirements, we choose a diagram map which presents the charts of VC B4 for all clusters to facilitate a comprehensive overview of the overall patterns.

We grouped the interactive VCs in a way that they facilitate the domain tasks which were identified in the analysis of the reasoning process. These groups of VCs form the modules of our visual analytics tools. They can be combined with statistical and data processing methods additionally applied in the model assessment process. When a domain task is selected in the user interface, the tool initializes the required module accordingly. Figure 6 gives an example of the combination of interactive VCs supporting a specific domain task.

The following use case demonstrates how scientists can apply the VCs in their reasoning process to fulfill the domain task presented in Figure 6. The domain task of assessing the plausibility of the variability of a realization of the OMCT consists of several domain subtasks: finding areas of high standard deviation; studying the temporal variation of attribute values in areas of high standard deviation; and inferring possible reason for high variability from the areas' spatial location. A number of cognitive actions must be performed to carry out these domain subtasks. Each cognitive action is supported by a VC.

To find areas of unusually high temporal standard deviation, scientists must first get a general idea of the spatial distribution of standard deviation values. An isopleth map with continuous tint (VC A1) showing these values enables the domain experts to identify approximately similar values and to qualitatively describe their geospatial distribution. However, the identification of outlying attribute values also requires a closer look at their distribution in attribute space. Scientists may use a cumulative relative frequency diagram (VC A2) to assess the distribution of standard deviation values in attribute space and to identify and interactively select unusually high values using range sliders. A link between the diagram and the map (VC A5) enables the users to filter geographic areas according to their standard deviation values, that is, locate outliers in geographical space. When a range of
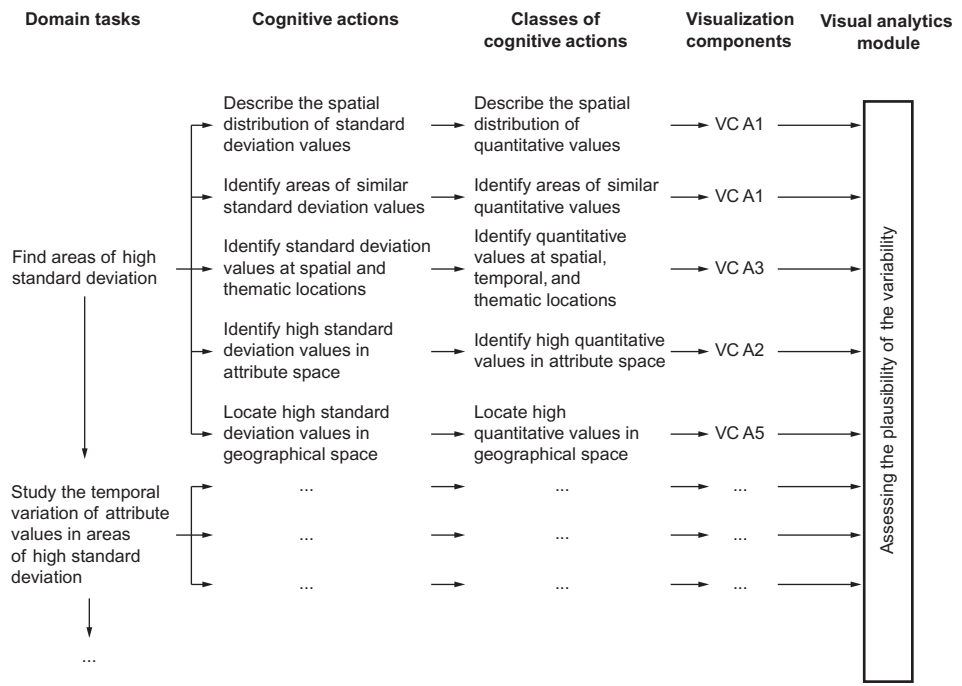

Figure 6. Example of combined visualization components to support a specific domain task.

high values is selected in the diagram, the corresponding locations are instantaneously filtered in the map.

Now that the areas of interest are known to the scientists, they must study the nature and characteristics of the spatiotemporal variation of attribute values in these areas to determine their significance. For this purpose, we provide several VCs. Animated change maps (VC A4), which map changes between subsequent time steps to height, allow for identifying periods of high variability. The mapping of values to height makes outliers and gradients salient. Whenever the experts visually perceive an interesting feature, they may use the animation controls to go back and look at relevant time steps or periods in detail. An animated thematic surface (VC A6) visualizes the spatiotemporal dynamics of the under-lying sea surface heights. To quantitatively describe temporal dynamics of sea surface height values at geospatial locations, the users are provided with a line chart with time axis (VC A7). A marker in a linked map view indicates the corresponding spatial point location. The line chart updates instantaneously when dragging the marker across the map. At any time in the assessment process, the experts may also use mouse pointing (VC A3) to identify exact values at spatial, temporal, or thematic locations.

The domain task of inferring possible reason for high variability from the areas' spatial location is important for the assessment of the OMCT. On a cognitive action level, it requires associating these areas with their geophysical characteristics. However, this cognitive action is not supported by any of our VCs because it requires additional domain knowledge which cannot be derived from the OMCT data.

We implemented the VCs applying various software packages. The OMCT example uses the Integrated Data Viewer (IDV) from Unidata as well as Java^{TM} programming and Python^{TM} scripting. The GIA example dealing with the validation of the simulation model with field observation employs Google^{TM} Earth, KML, and SVG to visualize the data, PostgreSQL and Java^{TM} Database Connectivity (JDBC) for data management, and Java^{TM} to connect statistical and fuzzy data processing.

3.3.   *Evaluation of the visual analytics concepts*

The results of our informal user study can be summarized as follows: For the scientists, visually perceiving the data in its spatiotemporal context helps gain a more detailed mental image and hence a better understanding of the data and the underlying simulation model. The combination of different interactive and linked visual components facilitates the detection of outliers and inconsistencies. The ability to readily study multiple model parameters makes the inference of possible causes more efficient. For example, the visualization helped discover a bias in the simulation output of the OMCT in certain geographical areas. The bias became apparent when the scientist examined an animated thematic surface of sea surface height values (VC A6). It was not discovered earlier because he only used static isopleth maps with continuous tint to show the output in its spatial extent. Maps depicting different time steps were juxtaposed to also account for the temporal dimension. This is not suitable for the detection of outliers and/or biases for at least two reasons: (1) spatially scattered outliers have very low saliency in a static map depicting the OMCT's more than 13,000 data points, and (2) juxtaposition is only appropriate for a limited number of maps, which leads to a coarse temporal resolution when assessing long time series (Andrienko *et al.* 2003). Mapping values to height in VC A6 clearly increased the saliency of outliers. A bias of these outliers in certain geographical areas was discovered by examining the entire time series in an animation. The domain expert used his domain knowledge to infer that it was caused by deficiencies in the relevant forcing data because of interpolation. Improving the

quality of the forcing data eliminated the bias. In the case of assessing the GIA model, the tool helped to gain a highly differentiated picture of spatial, temporal, and categorical variations of field observations and predicted data. The quality measure of the GIA model could be improved iteratively. Besides, the explicit support of the scientists' workflow and reasoning process allows for a faster and more comprehensive assessment of the simulation model. Because of this efficiency gain, the tool also encourages the exploration of rather marginal features as well as the verification of additional hypotheses, steps which normally would have been very time consuming.

4. Discussion

In our research, we have exemplarily studied the reasoning process in the application field of simulation model assessment in Earth science. We derived exemplary task models and related cognitive actions. Additionally, we developed VCs to facilitate the tasks. The results contribute to a generic task model and visual analytics tool for simulation model assessment which is our long-term research goal. In our future research we would like to continue with these analyses and derive classes of domain tasks, cognitive actions, and VCs to complete our current findings.

The methodical approach we applied in this study, which combines various methods and concepts, has proven to be suitable and, furthermore, essential for developing geovisual analytics concepts and tools supporting the reasoning process. It can be applied in similar studies.

So far, our visualization strategy supports the detection of outliers and inconsistencies as well as an initial inference of possible causes. For verifying these inferences, in-depth analyses – for example, frequency analysis of the underlying attribute values – with more sophisticated statistical tools such as MATLAB^{TM} are necessary. The framework should incorporate these analyses to support the entire assessment and reasoning process. This is a further aspect which should be addressed in our future work.

The informal evaluation indicates that the developed visual analytics strategy allows for a faster and more comprehensive assessment of the two exemplary simulation models. However, a broad and sophisticated user study should be conducted to gain further insight, especially as to where the geovisual analytics components could be improved.

The results demonstrate that our geovisual analytics approach and tools meet the challenges related to simulation model assessment mentioned at the beginning: (a) to analyze the multidimensional, multivariate, heterogeneous data and (b) to find and infer possible causes for model inconsistency more comprehensively and efficiently. They contribute to enhance model assessment and model improvement with respect to the validity of model behavior, validity of model structure, as well as the empirical validity of a model.

Acknowledgements

References

Andrienko, N. and Andrienko, G., 2006. *Exploratory analysis of spatial and temporal data: a systematic approach*. Berlin, Germany: Springer.
Andrienko, N., Andrienko, G., and Gatalsky, P., 2003. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing*, 14 (6), 503–541.

Arnberger, E., 1977. *Thematische kartographie*. Braunschweig: Westermann.

Bell, G., Hey, T., and Szalay, A., 2009. Beyond the data deluge. *Science*, 323 (5919), 1297–1298.

Bertin, J., 1974. *Graphische Semiologie: Diagramme, Netze, Karten*. Berlin: Walter de Gruyter.

Birta, L.G. and Arbez, G., 2007. *Modelling and simulation: exploring dynamic system behaviour*. London: Springer.

Bortz, J. and Döring, N., 1995. *Forschungsmethoden und Evaluation*. Springer-Lehrbuch. 2nd ed. Berlin: Springer.

Bossel, H., 1994. *Modellbildung und Simulation: Konzepte, Verfahren und Modelle zum Verhalten dynamischer Systeme; ein Lehr- und Arbeitsbuch*. 2nd ed. Braunschweig: Vieweg.

Braddick, O.J., 1980. Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London*, Series B (209), 137–151.

Cooke, N.J., 1994. Varieties of knowledge elicitation techniques. *International Journal of Human-Computer Studies*, 41 (6), 801–849.

Dobslaw, H., 2007. *Modellierung der Allgemeinen ozeanischen Dynamik zur Korrektur und Interpretation von Satellitendaten*. Thesis (PhD). Technische Universität Dresden.

Dörner, D., 1983. Kognitive Prozesse und die Organisation des Handelns. *In*: W. Hacker, W. Volpert, and M. Cranach, eds. *Kognitive und motivationale Aspekte der Handlung*. Bern: Huber, 26–37.

Dransch, D., 2002. *Handlungsorientierte Mensch-Computer-Interaktion für die kartographische Informationsverarbeitung in Geo-Informationssystemen*. Berliner geowissenschaftliche Abhandlungen, Reihe C: Kartographie. Berlin, Germany: Selbstverlag Fachbereich Geowissenschaften – Freie Univertität Berlin.

Gahegan, M., 2005. Beyond tools: visual support for the entire process of GIScience. *In*: J. Dykes, A.M. MacEachren, and M.-J. Kraak, eds. *Exploring geovisualization*. Amsterdam: Elsevier, 83–99.

Gray, J., 2009. Jim Gray on eScience: a transformed scientific method. *In*: T. Hey, S. Tansley, and K. Tolle, eds. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research, xvii–xxxi.

Hacker, W., 1978. *Allgemeine Arbeits- und Ingenieurpsychologie: Psychische Struktur und Regulation von Arbeitstätigkeit*. Bern: Huber.

Hey, T., Tansley, S., and Tolle, K., eds., 2009. *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research.

Imhof, E., 1972. *Thematische Kartographie. Lehrbuch der Allgemeinen Geographie*. 1st ed. Berlin, Germany: Walter de Gruyter.

Johansson, G., 1973. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14, 201–211.

Jonassen, D.H., Tessmer, M., and Hannum, W.H., 1999. *Task analysis methods for instructional design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Klemann, V. and Wolf, D., 2007. Using fuzzy logic for the analysis of sea-level indicators with respect to glacial-isostatic adjustment: an application to the Richmond-Gulf region, Hudson Bay. *Pure and Applied Geophysics*, 164 (4), 683–696.

Knapp, L., 1995. A task analysis approach to the visualization of geographic data. *In*: T.L. Nyerges *et al.*, eds. *Cognitive aspects of human-computer interaction for geographic information systems*. Dordrecht: Kluwer Academic Publishers, 355–371.

Köthur, P., 2009. *Visualisierung zur Evaluation numerischer geowissenschaftlicher Simulationsmodelle am Beispiel des Ozeanmodells für Zirkulation und Gezeiten*. Thesis (Diploma). Humboldt-Universität zu Berlin.

MacEachren, A.M., 1995. *How maps work: representation, visualization, and design*. Paperback edition 2004. New York, NY: The Guilford Press.

Moellering, H., 1976. The potential uses of a computer animated film in the analysis of geographical patterns of traffic crashes. *Accident Analysis and Prevention*, 8 (4), 215–227.

Nardi, B.A., ed., 1996. *Context and consciousness: activity theory and human-computer interaction*. Cambridge, MA: The MIT Press.

Norman, D.A., 1999. *Things that make us smart: defending human attributes in the age of the machine*. 10th ed. Reading, MA: Perseus Books.

Piñeiro, G. *et al.*, 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216, 316–322.

Preim, B., 1999. *Entwicklung interaktiver Systeme: Grundlagen, Fallbeispiele und innovative Anwendungsfelder*. Berlin: Springer.

Smith, J. and Smith, P., 2007. *Introduction to environmental modelling*. Oxford, NY: Oxford University Press.

Spence, R., 2007. *Information visualization: design for interaction*. 2nd ed. Harlow: Pearson Prentice Hall.

Thomas, J.J. and Cook, K.A., eds., 2005. *Illuminating the path: the research and development agenda for visual analytics*. 1st ed. Los Alamitos, CA: IEEE.

Tufte, E.R., 1997. *Visual Explanations: images and Quantities, Evidence and Narrative*. 5th printing, with revisions, January 2002. Cheshire, CT: Graphics Press.

Ullmann, S., 1979. *The interpretation of visual motion*. Cambridge, MA: The MIT Press.

VisMaster Consortium, 2010. *The European visual analytics research roadmap*. Unpublished manuscript. Version: April 6, 2010.

Wehrend, S. and Lewis, C., 1990. A problem-oriented classification of visualization techniques. In: *IEEE proceedings of the first IEEE conference on visualization: visualization '90*. IEEE, 139–143.

Werlen, B., 1997. *Gesellschaft, Handlung und Raum 1997*. Stuttgart: Franz Steinder.

Zadeh, L.A., 1978. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1 (1), 3–28.

Zetie, C., 1995. *Practical user interface design: making GUIs work*. London: McGraw-Hill.