

## INTERACTIVE VISUAL EXPLORATION OF SIMULATOR ACCURACY: A CASE STUDY FOR STOCHASTIC SIMULATION ALGORITHMS

Martin Luboschik  
Stefan Rybacki  
Roland Ewald  
Benjamin Schwarze  
Heidrun Schumann  
Adeline M. Uhrmacher

Albert Einstein Str. 22  
University of Rostock  
18059 Rostock, GERMANY

### ABSTRACT

Visual Analytics offers various interesting methods to explore high dimensional data interactively. In this paper we investigate how it can be applied to support experimenters and developers of simulation software conducting simulation studies. In particular the usage and development of approximate simulation algorithms poses several practical problems, e.g., estimating the impact of algorithm parameters on accuracy or detecting faulty implementations. To address some of those problems, we present an approach that allows to relate configurations and accuracy visually and exploratory. The approach is evaluated by a brief case study, focusing on the accuracy of Stochastic Simulation Algorithms.

### 1 INTRODUCTION

Simulation studies, particularly those that aim at exploring a model's dynamics, usually generate large amounts of output data. Such data can be analyzed in various ways, for example with statistics tools like R or plotting utilities like GNUPLOT. While, e.g., GNUPLOT provides a presentation of the data, it only shows the data in a static way allowing no further interaction. Moreover, information on how the data was generated is missing (e.g., the parameter configuration used), as are suitable means to interactively explore the relation between data and the process of generation.

Here, visual analytics promises to be a valuable tool, as it generally facilitates an effective dynamic exploration of large data amounts. It can be used to visually relate the data to the data generation process such as the different configurations that lead to the data as well as to effectively compare different configurations and therefore the influence of different parameter settings, different components etc. on the simulation output. However, the application of visual analytics techniques must not be perceived as just another presentation approach for simulation output, but instead becomes an integral part of the simulation study itself. Thereby, we move from a static presentation of data and its analysis towards a dynamic exploration of data and the data generating process.

In this paper we will focus on *inaccuracies* introduced by *non-exact* stochastic simulation algorithms. The algorithms that we will analyze belong to the family of Doob-Gillespie algorithms (also in the following simply called stochastic simulation algorithm (SSA)). These algorithms, first applied to biochemical reaction networks by Gillespie (Gillespie 1977), essentially model a chemical reaction network as a continuous-time Markov chain and generate a sample trajectory. Since then, many variants have been developed to improve the efficiency of the original algorithms. Some of those are *exact*, in the sense that by generating an arbitrary number of trajectories, one is able to approximate the underlying chemical master equation arbitrarily close

(e.g., see Gillespie 2007). Others, like  $\tau$ -leaping (Gillespie 2001), are *non-exact*. They typically outperform the exact methods, sometimes by several orders of magnitude (Jeschke and Ewald 2008). Since  $\tau$ -leaping variants trade speed for accuracy, this raises another question: *how much accuracy has been lost, and is the accuracy still sufficient for a specific simulation study?* Answering this question is quite hard, for several reasons. First,  $\tau$ -leaping can be parametrized so that both speed and accuracy are affected. Second,  $\tau$ -leaping speed and accuracy additionally depend strongly on the simulation model at hand. Finally, the *goal* of the simulation study may also be relevant, e.g., it could matter whether a specific time delay between important events (such as gene expressions) shall be observed, or just the average amount of a species over a large period of simulation time.

We address these problems by providing an interactive visual exploration of simulation accuracies for several  $\tau$ -leaping configurations. Due to the large amount of different configurations, this in particular means to visualize thousands of accuracy information along with the data. Moreover, appropriate interaction techniques are required to enable the exploration of the accuracy information. This way, our visual analytics approach supports the exploratory selection of a promising configuration in terms of a given task. However, this implies that the task at hand is similar to the task for which the  $\tau$ -leaping accuracy has been explored, i.e., that the model used for generating the accuracy data exhibits similar dynamics. Since determining the similarity of tasks (in terms of expected accuracy) is beyond the scope of this paper, the research prototype presented in the following should be considered a first step towards a more comprehensive solution.

Although there are successful solutions for single aspects of exploratory simulation studies as well as for accuracy visualization (cp. Sect. 2), yet there is no visual approach for an accuracy exploration that facilitates a guided configuration of a simulation study. For this reason, we contribute a novel procedure that integrates visual analytics into the overall experimentation process to support experimenters in interactively solving the simulation algorithm selection problem (see Ewald (2010), for example). As detailed in Sect. 3, our approach comprises the computation of accuracy information for several  $\tau$ -leaping configurations (Sect. 3.1), the visualization of that accuracy information along with the simulation output (Sect. 3.2), the interactive visual exploration of the accuracy information (Sect. 3.3) as well as the selection of a configuration with a desired accuracy and the transfer of that configuration to the current experiment for subsequent usage (Sect. 3.4). We use this procedure within the JAMES II framework (Himmelspach and Uhrmacher 2007) to configure subsequent simulation experiments. Section 4 demonstrates our approach and Section 5 concludes the paper.

## 2 BACKGROUND & RELATED WORK

By introducing visual analytics to the selection of simulation algorithm configurations, our work becomes interdisciplinary and relates to the field of stochastic simulation algorithms as well as the field of accuracy visualization. Hence, the following discussion on related work is separated according to those fields.

### 2.1 Stochastic Simulation Algorithms

The study of stochastic simulation algorithms (SSAs) is particularly relevant because the family of execution algorithms that interpret biochemical reaction networks as continuous-time Markov chains is large. In particular, the last decade has seen the adaptation and development of modeling languages for modeling and simulating biochemical systems whose semantics is predominantly rooted in continuous-time Markov chains, and thus rely on an efficient SSA for execution. To those belong stochastic process algebras for systems biology (e.g., Priami 1995) or rule-based modeling languages (e.g., Maus et al. 2011). Hence, all these developments may profit from insights into SSA performance.

While we focus on a single, parameterizable variant of  $\tau$ -leaping in this paper, i.e., the one presented by Cao et al. (2006), there are many other approaches that could be compared and analyzed in the same manner.  $\tau$ -leaping improves on the execution speed of exact methods by coalescing reaction events, i.e., it *leaps* over several reactions within a certain time period at once.  $\tau$ -leaping variants differ from each other

in the way the leap value  $\tau$  is calculated, and also how specific corner cases are handled. For example, in some approaches, it is possible to arrive at *negative* amounts of species, since multiple reactions are executed in conjunction (see Sandmann (2009) for a more detailed overview).

There are only few comparative performance studies dedicated to  $\tau$ -leaping algorithms so far. Jeschke and Ewald (2008) compare  $\tau$ -leaping to various exact SSA configurations. While  $\tau$ -leaping was consistently faster than (or equal to) the 162 exact SSA configurations it was compared with, its runtime was also shown to be sensitive with respect to its parameters (nine  $\tau$ -leaping configurations were evaluated). The in-accuracies of  $\tau$ -leaping could be shown by a statistical test, but this was only done for very few time points and a single  $\tau$ -leaping configuration (with default parameters). Similar results regarding  $\tau$ -leaping accuracy are also reported in (Cao, Gillespie, and Petzold 2006). To the best of our knowledge, no error bounds (of practical relevance) can be computed for any  $\tau$ -leaping method, given an arbitrary chemical reaction network. We therefore have to restrict ourselves to heuristic approaches, i.e., letting experimenters estimate the accuracy beforehand.

## 2.2 Accuracy Visualization

The aim of our approach, visually exploring the accuracy of several different simulation configurations, is closely connected to the field of uncertainty visualization in terms of visual analytics: Several investigations on the definition and characteristics of uncertainties for example identify accuracy as one aspect of uncertainty (MacEachren et al. 2005; Griethe and Schumann 2006; Skeels et al. 2010). Thereby, the visualization of uncertainties depends on the underlying data (e.g., volume data, architectural data, abstract data...) and differs accordingly. Nevertheless, those visualizations have in common to visualize confidence information along and closely related to the data. To differentiate both—data and uncertainties—it is a common approach to use a fuzzy visual representation for uncertainties (MacEachren 1992; Luboschik et al. 2010). But there are also explicit representations especially for stochastic measures. In (Potter et al. 2010) classical box plots are extended to communicate accuracy measures along a single line plot. An application to several line plots would result in visual clutter. Other approaches like (Holzhüter et al. 2012) enable the visualization of a single accuracy value per data item. Appropriate techniques for the visualization of several accuracy measures for hundreds of different simulation configurations are missing.

Different parameter settings result in different accuracy values. Hence, our approach is also related to parameter visualization. Previous work in this field focuses on different aspects concerning parameters and their relation to the output data, whereas a major aspect addressed in this domain is to understand the influence of parameters or parameter changes to the output data (Tweedie et al. 1994; Tweedie et al. 1996; Kerren 2006; Waser et al. 2010; Matkovic et al. 2010). Other work focuses on complex parameterizations and thus facilitates the visual exploration of large parameter spaces and the clustering of parameters (Pretorius et al. 2011) as well as the interaction with such spaces (van Wijk and van Overveld 2003). The approaches in (Marks et al. 1997; Bruckner and Möller 2010) also relate parameter settings to output data, but their main focus is on the exploration of the resulting output data to set parameters accordingly. The latter also makes sense in our simulation study case, as we focus on the simulation results as well as the associated accuracies in order to find concrete parameter settings. However, there are no methods yet that handle hundreds of simulation trajectories along with their accuracy information.

Integrating visual feedback onto responsible parameters with an exploratory data and accuracy analysis means to combine uncertainty visualization and parameter visualization. Although there are successful approaches that relate simulation output to uncertainties or related parameters, there are no methods yet for combining both aspects, and thus allowing for an accuracy-driven configuration of simulation experiments.

## 3 METHOD

The main goal of our work is to support experimenters in choosing an SSA configuration that is likely to have a desired accuracy in subsequent experiments. Accuracy measures can be calculated easily, but the

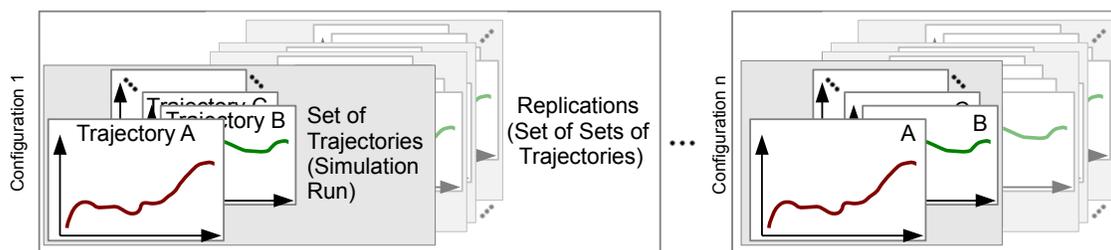


Figure 1: Trajectories and their participation in different sets.

complex interplay of parameters, simulation output, and the associated accuracies — especially regarding the semantics of different accuracy measures — inhibits an automatic configuration selection. Thus, an exploratory analysis is required that handles a large number of different configurations and may provide insight into the interrelation between accuracies, simulation output, and configuration parameters. For this purpose, we use a visual analytics approach and integrate it into the simulation study. Hence, our configuration selection procedure comprises four steps:

1. Computation of accuracy information for several  $\tau$ -leaping configurations.
2. Visualization of that accuracy information, the simulation output and used parameters.
3. Interactive visual exploration with respect to the *accuracy* information.
4. Selection of a setting providing a desired accuracy and transfer of the associated configuration to subsequent experiments.

Combining the different aspects of a simulation study (parameters, output, accuracies) to select a certain configuration in this manner is a novelty. By this, our procedure rudimentarily breaks up the “black box” behavior of the used simulator. Our approach is discussed in more detail in the following.

### 3.1 Data Generation

There are several kinds of data generated in a simulation study. First, there is simulation output data. The measurement of interest in our context, chemical reaction networks, is the amount of particles of a certain species at a given point in time. Hence, for each species of interest, the observed output data consists of a list of integers: one value for each point in simulation time that shall be observed. This list of time-stamped integers is referred to as a *trajectory* in the following. As there can be more than one species of interest, a *set* of trajectories is generated for each simulation run (see Fig. 1).

Second, there is configuration data, i.e., data that allows the computation of the trajectories. This includes the model, the simulation algorithm, and the parameters of both the model and the simulation algorithm. The configuration data is sufficient to replicate the simulation run an arbitrary number of times, hence generating a *set of sets* of trajectories representing *replications* for a configuration. Having more than one configuration will result in a *set of replications*.

Third, there are aggregated accuracy measurements that are defined across replication sets. These measurements are computed after all configurations are finished, and may vary in their level of detail, i.e., how much they aggregate the data (see Sect. 3.1.2).

Trajectories, trajectories sets and replication sets and their associated accuracy data as well as their corresponding configuration data are collected and referred to as *records* in the following.

#### 3.1.1 Experiment Setup

The data generation process relies on the modeling and simulation framework JAMES II and uses its SSA implementations (detailed in Jeschke 2010). As the main goal of the experiment is to estimate the accuracy

of different  $\tau$ -leaping configurations with respect to exact SSAs, we generate sets of replications for two exact SSA configurations and several hundred  $\tau$ -leaping configurations. Having data from two exact SSA configurations allows us to contrast the results with the inherent variation of simulation results even for exact approaches. These variations are due to the finite size  $n$  of the replication sets, as exact SSAs only solve the chemical master equation in the limit  $n \rightarrow \infty$ .

The overall experiment takes a considerable amount of time, even though each model is simulated for a relatively small interval of simulation time and hence an individual trajectory is generated quickly. But as a substantial number of configurations shall be explored — each requiring many replications — the experiment results in millions of trajectories altogether. We expect experimenters to use a suitable SSA configuration they identified for similar experiment tasks, e.g., involving larger simulation time intervals or models with similar dynamics. In such settings, the generated data could still allow for a relatively accurate comparison of different configurations, provided that the models at hand are sufficiently similar (see Sect. 3.4) and the simulation time interval was not too small to observe characteristic model dynamics.

### 3.1.2 Data Aggregation

We aggregate the observed simulation output along two dimensions: over points of simulation time and over trajectories. For a given species  $S$ , one could, for example, aggregate over all its trajectories in a replication set, i.e., all trajectories of  $S$  that were generated with the same configuration. These trajectories could now be aggregated to an averaged trajectory, i.e., a trajectory that contains the average amount of  $S$  for each time point. On the other hand, one could average over all time points of a trajectory, to arrive at the amount of  $S$  at any time. Finally, one could combine both aggregations and calculate the average amount of  $S$  at any time and in all trajectories of  $S$  in the given replication set.

As long as we restrict data aggregation to a single replication set, we are merely able to quantify model-inherent uncertainty, since all trajectories pertain to the same simulation configuration.

To quantify accuracy, i.e., the difference between exact and approximative SSA variants, we need to compare a measure of an approximative variant with the same measure of an exact variant. Differences in these measures will hint at a loss of accuracy, since the results of the exact variant are 'true' approximations of the chemical master equation. For example, one could quantify the difference between the averaged trajectory generated by an exact variant with one generated by a  $\tau$ -leaping variant.

## 3.2 Visual Design

The determination of accuracy measures is the first step to support experimenters in choosing a parameter configuration. But since each accuracy measure provides hundreds of accuracy values along each trajectory, we provide a visual analytics approach to ease this process. Our main contribution is to combine the visualization of simulation output with the visualization of associated accuracies as well as the visual feedback into the parameter space. This novel approach represents different types of data and helps the experimenter to get an immediate feedback on his configuration selection concerning these data properties.

### 3.2.1 Requirements

To reach this goal, the visualization has to meet several requirements, evolving from the need to *support* experimenters in terms of an interactive exploration process:

**Visualizing the Simulation Output ( $R_1$ ):** The initial inspection of an applied simulation configuration is generally done by roughly inspecting the simulation output. This way, for instance, the existence or absence of peaks within a line plot is a first clue for accepting or rejecting the current configuration. Thus, a basic requirement is to visualize the simulation output itself.

**Relating Accuracy to Data ( $R_2$ ):** To find a sufficient accuracy for the user's specific simulation study, the investigated configuration has to be judged by considering accuracy information. This requires

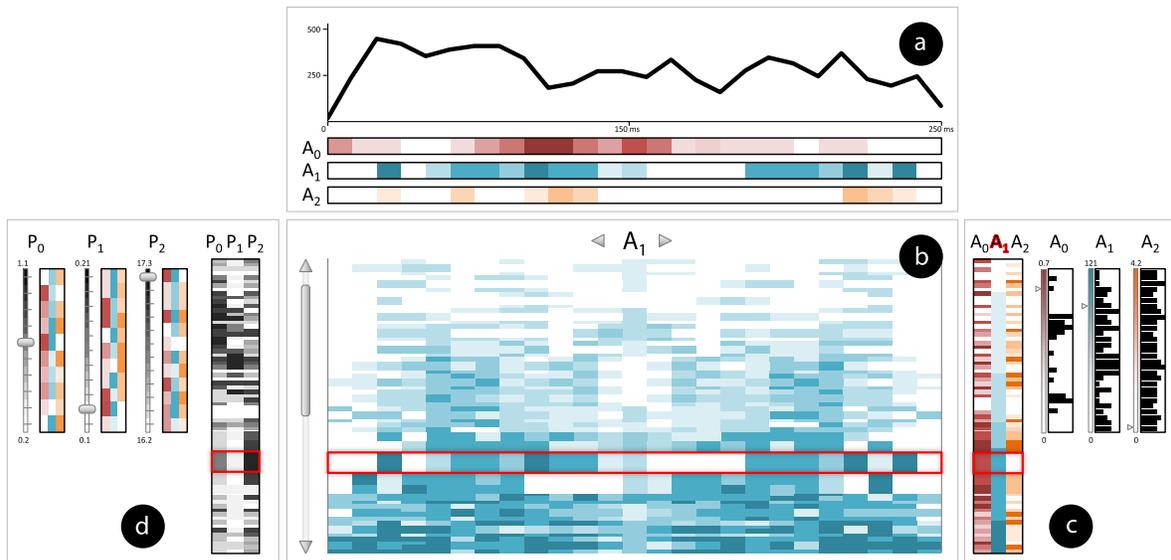


Figure 2: For an accuracy driven exploration of configurations, we use a multiple view approach consisting of (a) the *record view*, (b) the *accuracy view*, (c) the *accuracy map* and (d) the *parameter map*.

to visually associate the accuracy values directly to the simulation output. Since different accuracy measures are generated, this connection has to be shown for several measures simultaneously.

**Guidance to Desired Accuracies ( $R_3$ ):** The above requirements do not suffice to find alternative configurations in terms of a desired accuracy. Realizing a guiding facility leaving the experimenter in full control, the accuracy values of alternative configurations need to be depicted in a comparative way. In addition, it requires to show into which direction of exploration the accuracy is likely to increase or decrease and how alternative accuracy measures behave.

**Feedback into Parameter Space ( $R_4$ ):** While the simulation output and accuracies are in focus of the exploration process, the experimenters nevertheless need feedback on how these results were obtained. This is not only due to the need to use the found parameter setting in subsequent experiments, but also to confirm the found configuration. For example, the simulation output and the accuracy might fit the experimenters task at hand, but result from an unacceptably small step size. Thus, judging the applicability of a parameter setting at least requires to show which parameters induce the current simulation and accuracy result.

### 3.2.2 Visualization Approach

We use a multiple view approach — schematically illustrated in Figure 2 — to meet the requirements  $R_1 \dots R_4$ .

**Record View** The *record view* (Fig. 2a) is dedicated to represent the simulation output of *one* record to be investigated along with *all* accuracy values associated. In the simulation domain, it is a common approach to use line plots for visualizing trajectory data. Hence, we use them in the record view, thereby fulfilling requirement  $R_1$ .

The record view also shows various accuracy measures associated with the given trajectory. These measures may be characterized by strongly differing ranges and hence, a visualization of accuracies by additional lines or extended box plots (see Potter et al. 2010) could lead to visual clutter. Therefore, we show the accuracy values by color coding. Each accuracy measure is represented by a separate ribbon (called *accuracy plot*) and a separate color scale. In this way, different accuracy measures can be distinguished by different colors, at which we choose color scales from white for high accuracies to a second saturated color for low accuracies. The used color scales are interactively replaceable.

The record view in Figure 2 exemplarily shows three accuracy plots. By aligning them to the line plot, the accuracy values can be associated with the corresponding parts of the output data ( $R_2$ ). For instance, the accuracy plot for accuracy measure  $A_0$  communicates a low accuracy in the middle of the line plot (dark red) and an increased accuracy at its end (white).

**Accuracy View** This view shows the accuracy values of *one* specific accuracy measure for *all* records right beneath and aligned to the record view (Fig. 2b). The accuracy view is composed of stacked entries that are the *accuracy plots* of all records. This way, it facilitates an overview and comparison of one specific accuracy measure—especially towards the current trajectory shown in the record view. Its row is therefore highlighted in the accuracy view. The accuracy measure currently examined and thus shown within the accuracy view is interactively exchangeable.

The alignment of the accuracy view and the record view allows to associate all accuracy values of different trajectories to the corresponding regions of the currently visualized trajectory in the record view ( $R_2$ ). Showing the accuracy values of *all* records in the accuracy view already gives an overview of alternative configurations. By sorting the entries according to their accuracy values, the accuracy view also shows which record will produce more or less exact results ( $R_3$ ). For instance, in Figure 2 all entries above the currently selected represent configurations with higher accuracies concerning  $A_1$ . Further details on sorting options are given in Section 3.3.

The accuracy view shows only *one* accuracy measure. To preserve the global context, we enhance this view: Within an *accuracy map* to the right, we show further information on accuracies, whereas a *parameter map* to the left gives details on the used parameter settings.

**Accuracy Map** This view shows further information on accuracies to help judging the currently selected record in a global manner. That *accuracy map* (Fig. 2c) is split itself into two parts. The left part visualizes the record-wise *aggregated* accuracy values per accuracy measure. It uses the same color coding known from the accuracy plots and is aligned to each entry of the *accuracy view*. Using *aggregated* accuracy values, it indicates how alternative accuracy measures behave and hence allows the experimenter to keep track of unshown accuracies in surrounding records ( $R_3$ ). In Figure 2c for example, the entry right below the selected record shows the same  $A_1$  accuracy, but differs strongly in  $A_2$ .

The right part of the accuracy map provides a vertical histogram for each accuracy measure. These histograms display how many parameter settings result in the same aggregated accuracy. In doing so, the experimenter may find single configurations with outstanding accuracies or may determine which accuracies are not covered by the simulation study and subsequently reconfigure his setup (see Sect. 3.1). Cursors at each histogram indicate the aggregated accuracy values of the currently selected record (see Fig. 2c). Those histograms are independent of and not aligned to the accuracy view.

**Parameter Map** The *parameter map* (Fig. 2d) lets the experimenter gain insights into the parameter settings that were used while generating the data and is also split into two parts. Its right part is aligned to the accuracy view and depicts the corresponding parameter values for *each* record ( $R_4$ ). For this purpose, we use the color coding approach again with a gray color scale from white for low to black for high parameter values. This gray scale has been applied to each parameter ( $P_1 \dots P_3$  in Fig. 2), but it is adapted to each's value range. In doing so, we limit the use of further colors but enable the user for instance to look for differing configurations that show a similar behavior in the accuracy view.

The left part of the parameter map encodes parameter settings by position. For each parameter an associated slider is shown, whereas their handles reflect the values of the currently selected record ( $R_4$ ). Inspired by Willett et al. 2007, we additionally show the aggregated accuracy values of each measure along their track by color coded bars. Increasing or decreasing parameter values influences the aggregated accuracy values and thus the enhanced sliders communicate their interplay. Note that the number of parameters in the *parameter map* is independent of the visualized accuracy measures in the *accuracy map*.

Concluding, our multiple view approach considers all requirements  $R_1 - R_4$ . In this way, we contribute a novel visualization approach that combines the visualization of simulation output, associated accuracies,

and responsible parameter settings. The next step is to interactively explore the accuracy information using the introduced views to find a desired configuration.

### 3.3 Interactive Exploration

The procedures introduced in Sect. 3.1 and 3.2 acquire and visually communicate a simulation preview along with associated accuracy and parameter information. But as a static presentation will not suffice to efficiently find a configuration that fits the experimenter's task at hand, the next step is to enable the user to interact with the provided information to find a desired accuracy as well as to confirm it, using the simulation output and related parameters.

First, this means to enable an easy navigation through the accuracy view (global exploration) as well as to access accuracy view entries according to specific characteristics within a single entry (local exploration). Due to the fact that the accuracy view may quickly exceed the available screen space, scrolling is a common option to overcome this problem (like, e.g., in large documents). It allows to quickly overview the whole accuracy view and makes every entry accessible. Moreover, sorting the accuracy entries supports finding configurations of less or more accuracy in a global manner. It is plausible to offer a sorting according to the currently shown accuracy measure. In addition, the sorting can be changed interactively, using other accuracy measures or even parameters. This is realized by interactively selecting the corresponding labels in the *accuracy* or *parameter map* respectively (red contoured  $A_1$  in Fig. 2c).

To support local exploration, a given sorting can be adjusted. If a certain region of the simulation output like a single peak or the steady state phase is of concern, the experimenter is enabled to interactively select that region. Subsequently, the sorting in the accuracy view is adapted by only considering accuracies within that region. This interaction, for instance, allows finding entries that comply with the reference trajectory in that region. In this regard, an applied local fish-eye lens emphasizes the currently selected entry and facilitates a more detailed view (cp. Fig. 2, 3).

To support accepting or rejecting a certain configuration, the *record view*, the *accuracy view*, the *accuracy map* as well as the *parameter map* are closely interconnected. Using *brushing and linking* (Buja et al. 1991), a changed selection in the *accuracy view* automatically updates all other views to show the current record. In return, the slider-equipped parameter map can be used to determine a specific parametrization and investigate the related record in global and local aspects. Simultaneously, all views are updated and the corresponding entry of the accuracy view is automatically selected and panned to. Moreover, changing one parameter updates the visualized aggregated accuracies at the other sliders.

### 3.4 Experiment Configuration

Once an experimenter has explored the configurations and decided on a promising configuration for a subsequent simulation experiment, the corresponding simulation algorithm parameters are used to configure JAMES II. Note that feeding suitable parameters back into JAMES II is advisable only if the new experiment is defined on a model relatively similar to the one for which the data was generated, e.g., in terms of reaction network topology and species amounts. To alleviate this restriction, the approach presented here could be repeated with *several* models, each having distinct characteristics and being representative for some well-established kind of model. As mentioned in Section 1, we regard this problem as future work.

## 4 APPLICATION

### 4.1 Accuracy Experiments

We generated accuracy data for the *linear chain system* (LCS) from (Cao et al. 2004), a benchmark model that consists of a single chain of unary reactions,  $S_0 \rightarrow S_1, \dots, S_{n-1} \rightarrow S_n$ , each with a reaction rate constant of 1. LCS was configured to contain 601 species and 600 reactions ( $n = 600$ ). It is simulated until 50 seconds of simulation time and initialized with 10.000 of  $S_0$ , all other species amounts being set to zero.

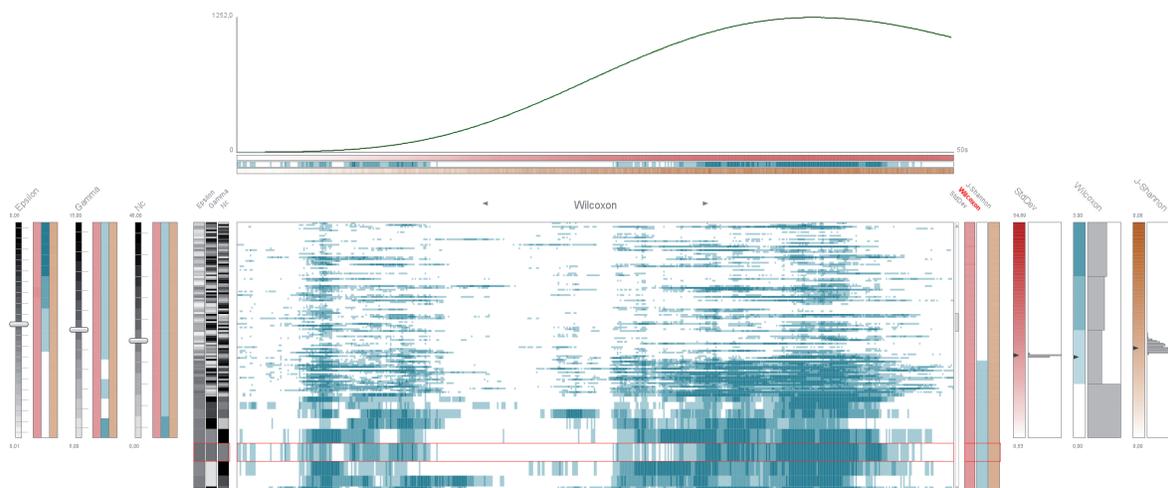


Figure 3: Applying our interactive approach supports finding configurations with a desired accuracy behavior (e.g., white regions). For this purpose, the records sorting can be adapted by restricting it to a certain region of the simulation time interval. This way, one can easily determine the configuration with the highest  $\epsilon$ -value that is exact at the trajectories peak — out of 2200 setups.

Only the amount of  $S_{25}$  is observed, sampled at 500 equidistant time points. We chose the LCS for this case study because it is a simple and well-known benchmark model for which  $\tau$ -leaping accuracy has already been briefly investigated in (Jeschke and Ewald 2008).

To investigate the accuracy of  $\tau$ -leaping, we let each configuration simulate 1.000 replications. We varied the parameter  $\epsilon$ , which directly affects the size of the selected  $\tau$  leap, the parameter  $\gamma$ , which controls the lower threshold for the selected  $\tau$  (if  $\tau$  is too small, the simulator falls back to an exact SSA variant), and  $n_c$ , which controls what reactions are considered critical and hence are treated differently. We vary the parameters around the values suggested in (Cao et al. 2006), where  $\epsilon = 0.03$ ,  $\gamma = 10$ , and  $n_c = 10$ :  $\epsilon \in \{0.01, 0.012, \dots, 0.048\}$ ,  $\gamma \in \{5, 6, \dots, 15\}$ , and  $n_c \in \{0, 5, \dots, 45\}$ , amounting to  $20 \cdot 11 \cdot 10 = 2.200$   $\tau$ -leaping setups altogether. To generate the reference data with an exact SSA, we used two configurations of the next reaction method (Gibson and Bruck 2000), relying on a heap-based event queue and a sorted list event queue, respectively. As with  $\tau$ -leaping, 1.000 replications were generated for both configurations.

All in all, more than two million replications have been simulated, resulting in about 4.5 GB of simulation output data to be aggregated. For each set of replications (see Sect. 3.1), we calculate the mean and standard deviation for each time point, so that we arrive at an averaged trajectory and also have a measure for the model-inherent uncertainty. We measure the accuracy by comparing the replication set of each  $\tau$ -leaping configuration (and the additional next reaction configuration) to the reference data. For each time point, we compare the empirical probability distribution (i.e., the 'histogram') generated by the given configuration with the reference data, by using the paired Wilcoxon rank-sum test (Sheskin 2007, p. 513) and the Jensen-Shannon divergence (Lin 1991).

## 4.2 Exploration of Accuracy Data

As introduced above, our visualization approach has to handle 2,200 different configurations each with 500 accuracy values along the simulation output. In the following, we briefly describe the different views of our approach — all related to Figure 3 that shows the aforementioned data.

The *record view* depicts the line plot of the currently selected configuration along with the exact reference simulation run (green). The starting configuration for exploration is set by the center of each parameter's range (see sliders in Fig. 3). The related simulation output shown in the line plot (black) nearly behaves like the exact solution. Below the line plot, the visualized accuracy measures (Wilcoxon rank-sum

test result, Jensen-Shannon divergence) and the standard deviation, communicate the accuracy over the simulation time interval. Especially the Wilcoxon rank-sum test result indicates a lack of accuracy at the beginning and near the peak of the plot (dark cyan regions). For that specific measure, we distinguish three significance levels of rejecting  $H_0$  ( $\alpha \in \{0.05, 0.01, 0.001\}$ ) (discrete cyan levels) and the case where  $H_0$  cannot be rejected (white).

Due to the uneven distribution, the Wilcoxon rank-sum test result is shown within the *accuracy view*. The sorting is based on the aggregated Wilcoxon measure (median of the sorted significance levels, indicated by the red contoured label). Thus, the entries above the current configuration will be more accurate, whereas entries below will be less accurate.

The *parameter map* reveals that the neighboring entries within the *accuracy view* result from similar values of parameter  $\varepsilon$  whereas parameters  $\gamma$  and  $n_c$  strongly fluctuate. Scrolling through the *accuracy view* reveals the correlation of  $\varepsilon$  and the accuracy (w.r.t. significance levels). Consequently, the *accuracy map* histogram shows that the current configuration accordingly results in a mid-range accuracy concerning the Wilcoxon rank-sum. Nevertheless, the histogram also shows that more configurations contribute to more or less significant deviations from the reference data than to the mid range.

Since  $\varepsilon$  has crucial impact on run time—the higher  $\varepsilon$ , the smaller the execution time (Cao et al. 2006)—the experimenter is now able to choose a configuration that presents a suitable trade-off between execution time and accuracy requirements. For instance, the experimenter may be interested in exact values concerning the significance at the trajectory’s maximum and simultaneously looks for a configuration with short execution time. Starting with the configurations with most significant deviation (w.r.t. reference data) that typically performs best (high  $\varepsilon$ ), the user now may simply scroll through the accuracy view to find an entry that at first shows a white region along the corresponding trajectory region. A sorting according to that specific region is a further option to find that specific configuration. Visualizing alternative accuracy measures in the *record view* and the associated parameters in the *parameter map* finally supports accepting or rejecting that configuration.

## 5 CONCLUSIONS

The approach presented in this paper illustrates the potential of visual analytics when it comes to facilitating the configuration of simulation algorithms.

In our case study, we develop a visualization to explore the accuracy of over 2.000  $\tau$ -leaping configurations, by considering data that is aggregated from over two million simulation runs. Our approach simultaneously visualizes the output data, the associated accuracies and the responsible parameters. Appropriate interactions let the user get an overview and insight into the interplay of different simulation study aspects and supports selecting a configuration that fits the task at hand. This configuration can then be used in subsequent simulation experiments with similar models.

However, note that we did not yet address the question of *how* similar two models need to be so that accuracy results can be transferred from one model to the other. Answering this question empirically requires a lot of effort, as we need to investigate many models in this manner, and also have to compare the accuracy results of a single configuration *across* these test models. On the other hand, results for additional models could help simulation users to build an intuition for  $\tau$ -leaping accuracy.

We plan to extend the presented visualization approach by displaying additional data that could influence a user’s selection, e.g., the average execution time of each configuration. Finally, there are many potential measures for accuracy, and we only considered two of them in this paper (see Sect. 4.1). Evaluating the usefulness of other accuracy measures is therefore also a subject of future work.

## ACKNOWLEDGMENTS

This research is supported by the DFG (German Research Foundation): the overall approach has been developed in VASSiB (part of SPP 1335), simulation data was generated by ALESIA (EW 127/1-1).

## REFERENCES

- Bruckner, S., and T. Möller. 2010. "Result-Driven Exploration of Simulation Parameter Spaces for Visual Effects Design". *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1467–1475.
- Buja, A., J. A. McDonald, J. Michalak, and W. Stuetzle. 1991. "Interactive Data Visualization using Focusing and Linking". In *Proc. of IEEE Visualization (Vis'91)*, 156–163: IEEE Computer Society.
- Cao, Y., D. T. Gillespie, and L. R. Petzold. 2006, January. "Efficient step size selection for the tau-leaping simulation method.". *J Chem Phys* 124 (4).
- Cao, Y., H. Li, and L. Petzold. 2004. "Efficient formulation of the stochastic simulation algorithm for chemically reacting systems". *The Journal of Chemical Physics* 121 (9): 4059–4067.
- Ewald, R. 2010. *Automatic Algorithm Selection for Complex Simulation Problems*. Ph. D. thesis, University of Rostock.
- Gibson, M. A., and J. Bruck. 2000. "Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels". *The Journal of Chemical Physics* 104:1876–1889.
- Gillespie, D. T. 1977. "Exact Stochastic Simulation of Coupled Chemical Reactions". *Journal of Physical Chemistry* 81 (25): 2340–2361.
- Gillespie, D. T. 2001. "Approximate accelerated stochastic simulation of chemically reacting system". *Journal of Chemical Physics* 115:1716–1733.
- Gillespie, D. T. 2007. "Stochastic simulation of chemical kinetics.". *Annual review of physical chemistry* 58 (1): 35–55.
- Griethe, H., and H. Schumann. 2006. "The Visualization of Uncertain Data: Methods and Problems". In *Proc. of the Conference on Simulation and Visualization*, 143–156. Magdeburg, Germany.
- Himmelspach, J., and A. M. Uhrmacher. 2007. "Plug'n simulate". In *Proc. of the 40th Annual Simulation Symposium*, 137–143: IEEE Computer Society.
- Holzhüter, C., H. Schumann, A. Lex, D. Schmalstieg, H.-J. Schulz, and M. Streit. 2012. "Visualizing Uncertainty in Biological Expression Data". In *Proc. of the Conference on Visualization and Data Analysis*, edited by P. C. Wong, D. L. Kao, M. C. Hao, C. Chen, R. Kosara, M. A. Livingston, J. Park, and I. Roberts, Volume 8294, 1–14.
- Jeschke, M. 2010, October. *Efficient non-spatial and spatial simulation of biochemical reaction networks*. Ph. D. thesis, University of Rostock.
- Jeschke, M., and R. Ewald. 2008, October. "Large-Scale Design Space Exploration of SSA". In *Proc. of Computational Methods in Systems Biology (CMSB)*, edited by M. Heiner and A. Uhrmacher, 211–230: Springer.
- Kerren, A. 2006. "Improving Strategy Parameters of Evolutionary Computations with Interactive Coordinated Views". In *Proc. of the IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP'06)*, 88–93: ACTA Press.
- Lin, J. 1991, January. "Divergence measures based on the Shannon entropy". *IEEE Transactions on Information Theory* 37 (1): 145–151.
- Luboschik, M., A. Radloff, and H. Schumann. 2010. "Using Non-Photorealistic Rendering Techniques for the Visualization of Uncertainty". In *IEEE Information Visualization (Poster)*.
- MacEachren, A. M. 1992. "Visualizing Uncertain Information". *Cartographic Perspectives* 13:10–19.
- MacEachren, A. M., A. Robinson, S. Hopper, S. Gardner, R. Murray, and M. Gahegan. 2005. "Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know". *Cartography and Geographic Information Science* 32 (3): 139–160.
- Marks, J., B. Andelman, P. A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber. 1997. "Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation". In *Proc. of the Conference on Computer Graphics and Interactive Techniques*, 389–400: ACM Press.

- Matkovic, K., D. Gracanin, M. Jelovic, A. Ammer, A. Lez, and H. Hauser. 2010. "Interactive Visual Analysis of Multiple Simulation Runs Using the Simulation Model View: Understanding and Tuning of an Electronic Unit Injector". *IEEE TVCG* 16 (6): 1449–1457.
- Maus, C., S. Rybacki, and A. Uhrmacher. 2011. "Rule-based multi-level modeling of cell biological systems". *BMC Systems Biology* 5 (1): 166+.
- Potter, K., J. Kniss, R. Riesenfeld, and C. R. Johnson. 2010. "Visualizing Summary Statistics and Uncertainty". In *Computer Graphics Forum*, Volume 29, 823–831.
- Pretorius, A. J., M.-A. Bray, A. E. Carpenter, and R. A. Ruddle. 2011. "Visualization of Parameter Space for Image Analysis". *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2402–2411.
- Priami, C. 1995, July. "Stochastic pi-Calculus". *The Computer Journal* 38 (7): 578–589.
- Sandmann, W. 2009, December. "Streamlined Formulation of Adaptive Explicit-Implicit Tau-Leaping with Automatic Tau Selection". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 1104–1112. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Sheskin, D. J. 2007, January. *Handbook of Parametric and Nonparametric Statistical Procedures, Fourth Edition*. Chapman & Hall/CRC.
- Skeels, M., B. Leeb, G. Smith, and G. G. Robertson. 2010. "Revealing uncertainty for information visualization". *Information Visualization* 9 (1): 70–81.
- Tweedie, L., R. Spence, H. Dawkes, and H. Su. 1996. "The Influence Explorer – a tool for design". In *Conference Companion on Human Factors in Computing Systems*, 390–391: ACM.
- Tweedie, L., R. Spence, D. Williams, and R. Bhogal. 1994. "The Attribute Explorer". In *Conference Companion on Human Factors in Computing Systems*, 435–436: ACM.
- van Wijk, J. J., and C. W. A. M. van Overveld. 2003. "Preset Based Interaction with High Dimensional Parameter Spaces". In *Data Visualization: The State of the Art*, 391–406.
- Waser, J., R. Fuchs, H. Ribicic, B. Schindler, G. Blöschl, and E. Gröller. 2010. "World Lines". *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 1458–1467.
- Willett, W., J. Heer, and M. Agrawala. 2007. "Scented Widgets: Improving Navigation Cues with Embedded Visualizations". *IEEE Transactions on Visualization and Computer Graphics* 13 (6): 1129–1136.

## AUTHOR BIOGRAPHIES

**MARTIN LUBOSCHIK** holds a diploma in Computer Science from the University of Rostock and pursues a PhD at the Computer Graphics Group at the University of Rostock.

**STEFAN RYBACKI** holds a diploma in Computer Science from the University of Rostock and pursues a PhD at the modeling and Simulation Group at the University of Rostock.

**ROLAND EWALD** holds a PhD in Computer Science from the University of Rostock and works at the modeling and Simulation Group at the University of Rostock.

**BENJAMIN SCHWARZE** holds a bachelor degree in Computer Science from the University of Rostock and currently pursues a Masters degree.

**HEIDRUN SCHUMANN** is a Professor at the Department of Computer Science at the University of Rostock and head of the Computer Graphics and Visualization Group.

**ADELINDE M. UHRMACHER** is a Professor at the Department of Computer Science at the University of Rostock and head of the Modeling and Simulation Group.