Visual Exploration of Patterns in Multi-run Time-varying Multi-field Simulation Data Using Projected Views

Vladimir Molchanov
Jacobs University
Campus Ring 1
28759 Bremen, Germany
v.molchanov@jacobs-university.de

Lars Linsen
Jacobs University
Campus Ring 1
28759 Bremen, Germany
l.linsen@jacobs-university.de

ABSTRACT

In the fields of science and engineering, it is common to run hundreds of simulations to investigate the dependence of the modeled process on various simulation and input the parameters. We propose a comprehensive approach for the visual analysis of such multi-run data to detect patterns and outliers. We use dimensionality reduction algorithms to generate a visual representation that exhibits the distribution of the simulation results under varying parameter settings. Each field (or even multi-field) of every time step and every simulation run is represented as a point in a 2D space, where the 2D layout conveys similarity of the scalar fields. Points corresponding to consecutive time steps of one run are connected by line segments, such that each simulation run is represented as a polyline. Consequently, the multi-run data are visually encoded as a set of polylines. Variations of hue, saturation, opacity, and shape allow for distinguishing groups of simulations and depicting various characteristics of runs. The user can interactively change these settings, while further interaction mechanisms allow for selection, refinement, zooming, requesting textual information, and brushing and linking to coordinated (or embedded) views of physical and attribute space visualizations. We apply our approach to two applications with significantly different data structure: a multi-run climate simulation over a 2D regular grid and a multi-run binary star evolution simulation with unstructured 3D particles evolving over time. We demonstrate the contribution and impact of our visualization method for the interactive visual analysis of the multi-run data by identifying meaningful groups of simulations, detecting global patterns, and finding interesting outliers.

Keywords

Time-varying Data, Multi-field, Multi-modal and Multi-variate Data, Scalar Field Data, Dimensionality Reduction

1 INTRODUCTION

Multi-run (or ensemble) simulations in the fields of science and engineering serve to investigate dependence of modeled time-varying phenomena on various simulation and input parameters. Typical goals when generating and analyzing multi-run simulations are to explore the entire space of possible scenarios, to find optimal or to detect critical parameter settings, and to investigate how sensitive outputs are to certain input parameters. Obviously, a large number of simulations need to be performed in order to obtain a sufficiently dense parameter space sampling. The analysis of the resulting datasets is extremely challenging, since the data often describe time-varying multi-fields, which when computed for multiple runs quickly reach extremely large sizes.

Novel visualization approaches can be very helpful for researches to simplify post-processing of multi-run datasets. This stage may include fast navigation over the data, selection of an individual simulation (or a time step) or a subset thereof for detailed comparison, visual encoding of various characteristics of runs, and easy creation of animations. However, the first step after the dataset is generated is to throw a glance at the dataset in order to roughly estimate its homogeneity, to build a notion of a typical simulation behavior (common pattern) and to identify those simulations whose behavior significantly differs from the pattern (outliers). We propose a technique to produce an overview representation of a complete multi-run dataset, equip it with a number of interaction tools based on coordinated views, and demonstrate the effectiveness of the resulting system on two application examples.

The proposed framework for interactive visual analysis of multi-run data is based on coordinated views of numerous aspects of multi-run data. Pattern and out-

lier detection in multi-run simulations is supported by a projected view of an attribute space. The attribute space is derived from the data space by defining characteristic properties of a simulation time step. According to that attribute space representation of each time step of each simulation run, we compare different time steps of the same or different simulation runs with each other. We create a 2D layout that depicts similarity or dissimilarity of time steps using dimension reduction methods. Each time step is represented as a point in this 2D layout and runs are shown as polylines connecting those points. In such a layout, runs of similar behavior have a similar polyline trajectory. Hence, patterns can be observed quickly and the same is true for trajectories that do not adhere to the pattern, i.e., outliers.

Although this similarity view is most important for analyzing multiple simulation runs, it is equally important that one can relate detected outliers or otherwise interesting runs to the parameter space and the multiple fields that are simulated. Our framework is based on coordinated views of parameter space visualization, spatial data visualization of the fields, visualization of the derived attribute space, and the projected views.

We present the overall framework of our work (Section 3) and demonstrate its effectiveness and intuitiveness by deploying it to two rather different instances (Section 4). The first instance is a grid-based climate simulation, where the attribute space can be trivially derived from the data space and the derived projection patterns follow our intuition. The second instance is a particle-based astrophysical simulation, where a non-trivial computation of the attribute space was necessary and where we show how our framework allows for interactive analysis of multi-run data. In this context, we also apply multiple coordinated projected views.

2 RELATED WORK

Increasing complexity and heterogeneity of generated and registered scientific data becomes a hot topic in modern researches in the field of data mining and visualization. Kehrer and Hauser [KH13] wrote recently a survey on visualization methods for *multi-faceted* data. Multi-faceted is here a general term to emphasize the fact, that scientific data may have spatio-temporal structure and/or multi-variate format, may consist of parts stemming from different modalities, may be multi-run and/or multi-model data.

User interaction plays a crucial role when inspecting the interplay of initial conditions and simulation results. Brushing and linking embedded into a coordinated multiple views framework is probably the most popular tool in visualization systems aimed to handle multi-run data. Examples of interactive exploration of parameter space can be found in [MAB+97, Ma99, MOD00, SPK+07]. Pattern recognition and annotation of clusters and outliers can be helpful for researchers leading multi-run simulations in order to plan further tests, re-sample parameter space to obtain detailed local information and define critical parameter settings. A recent grid-density based technique to complete these tasks was proposed by Kandogan [Kan12].

Supporting uncertainty in the visual analytics process is of great importance in systems aimed to help the user to find optimal parameter settings. A framework for uncertainty modeling, propagation and aggregation is proposed by Correa et al. [CCM09]. Possible solutions for graphical representation of uncertainty are discussed in Cooke and Noortwijk [CvN00].

Bruckner and Möller [BM10] developed a visualization system helping the user to find simulation parameters corresponding to the desired outcome animation. Their approach is based on a spatio-temporal clustering technique producing an overview of achievable final states and their trajectories in the phase space. Another novel approach for mapping of deforming mesh sequences to the visual domain is given by Cashman and Hormann [CH12]. The user is allowed, in particular, to create an arbitrarily-shaped path in the abstract representation domain, from which a corresponding mesh animation can be reconstructed.

We are the first who developed a framework for analyzing multi-run time-varying multi-field data. In our work, we propose to use dimensionality reduction algorithms [JJ09] including the classical Principle Component Analysis (PCA) [Pea01] and Multidimensional Scaling (MDS) [BG05] to produce 2D similarity layouts between time steps of multiple simulation runs. This set-up allows for an interactive visual detection of patterns and outliers in the multi-run data, especially when embedded in our coordinated views framework of parameter space visualization, spatial data visualization, and visualization of a derived attribute space in addition to the projected view visualization.

3 INTERACTIVE VISUAL ANALYSIS FRAMEWORK

Multi-run data result from a repeating procedure of picking an instance from the parameter space *PS* and running a simulation for selected input parameters. For each chosen set of parameters, the simulation typically creates a time-varying multi-field data set, i.e., a spatial phenomenon is simulated using a multitude of fields that change over time. To effectively capture or sample the parameter space, the simulation has to be run many times for different initial conditions (or parameter settings). As a result, one obtains a large amount of time-varying multi-field simulation results. When changing the parameter setting, the simulation result changes in

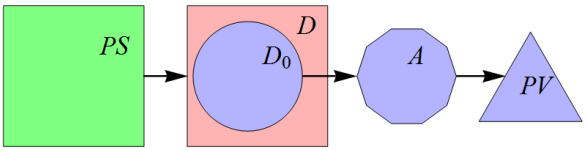


Figure 1: Steps of processing multi-run simulation data. Data D is generated for a dense subset of parameter space PS. Subset D_0 of dataset D is mapped to attribute space A. Projected view PV is an image of a dimensionality reduction algorithm applied to A and, therefore, it is an abstract representation of D.

some way. In order to understand, how much a parameter change influences the simulation result, one has to find a means to compare simulation results. Our framework operates on the individual time steps of the multiple simulation runs and compares those time steps, where time steps are typically multi-fields. In order to do so effectively, we have to describe each time step (or multi-field) using appropriate attributes. Such an attribute space shall be significantly smaller than the multi-field space, but still capture the main properties of that time step. Moreover, it shall allow for a direct comparison of two multi-fields by confronting their attribute values.

Even when defining a suitable attribute space, the collection of attributes of all time steps of all simulation runs is typically too large to allow for an efficient processing. Hence, one typically applies a pre-processing step that subsamples the multi-run simulation space D. This subsampling can be applied in many ways: (i) One may subsample in the parameter space, i.e., reduce the number of runs to a certain subgroup of runs; (ii) one may subsample in the time domain, i.e., only consider a certain reduced number of time steps of each simulation run; (iii) one may subsample the variable space, i.e., out of the multi-field data set, one only considers a certain subset of fields; or (iv) one may subsample the spatial domain, i.e., reduce the spatial resolution of each field. We want to denote by D_0 the subsampled space of D. Our notations are summarized in Figure 1.

The next step is a generation of a set of descriptors for all elements $d \in D_0$, i.e., data from D_0 is mapped to an attribute space A. Element d in this context is a multi-field. The generated attributes serve to compactly characterize d. It is required that A is a metric space, so that similarity of its elements can be measured. This induces a metric in the original space D_0 and allows for comparing its instances, which is not always possible to do directly.

We note that the construction of a proper space A is an application-specific and not uniquely defined procedure. However, we believe that in most cases, experts who performed simulations can give a good advice. Besides that, the number of alternatives is significantly reduced when trying to balance simplicity of the structure of A and its ability to reflect main characteristics of simulations. In the subsequent sections, we provide two examples, where attribute spaces are constructed for two structurally different simulations from two different application areas. Although there is no unified recipe on how to best define an attribute space for any multi-run simulation, we generally believe that it is possible for any simulation output and, thus, feel that our framework is generally applicable.

Once the attribute space is defined, we generate visualizations based on dimensionality reduction to map the attribute space A to a 2D visual space, which we denote as projected view PV. The visual encoding of PV uses one point for each multi-field d and generates polylines for each (time-varying) simulation run.

The main contribution of our work is to propose the general framework and, in particular, the use of a projected view based on a suitable attribute space. However, the full effectiveness of our approach is given by the interactive exploration of the various spaces in the context of coordinated views. Brushing and linking between suitable visualizations of parameter space PS, multi-field data space D_0, attribute space A, and projected view PV provides means for an effective visual analysis tool to detect patterns, trends, and outliers in multi-run time-varying multi-field simulation data.

3.1 Parameter Space Visualization

Our parameter space visualization is based on 1D plots, 2D plots, or 2D plot matrices depending on the number of input simulation parameters. Sampling of the parameter space is shown as points, such that each point corresponds to one simulation run. A color-coding of the samples in the parameter space may result from a classification given by the application (e.g., chemical composition of modeled stars), from theoretical prediction of simulations (e.g., parameters which determine laminar and turbulent flows) or may depict the outcome of an interactive analysis process (e.g., it may show which simulation runs were selected by the user in any of coordinated widgets).

3.2 Data Space Visualization

The data space visualization requires the visualization of multi-field data. Here, any spatial data visualization methods can be used. In case of two spatial dimensions, a transfer function can be applied to generate 2D visualizations. In case of three spatial dimensions, volume visualization methods such as direct volume rendering or isosurface extraction can be applied. The multi-field challenge can be tackled via an integrated visualization or via multiple (coordinated) views each visualizing one field. In our examples, we used the latter approach.

 ISBN 978-80-86943-71-8

3.3 Attribute Space Visualization

As mentioned above, the definition of the attribute space is application-specific. Its visualization depends on the respective choice made within the application. In some cases, attribute space visualization can be embedded in the projected view (e.g, in form of pictograms) to allow visual correlation of the projected view layout with descriptive attributes.

3.4 Projected View Visualization

After data subset D_0 is mapped to attribute space, it can be projected to the visual domain of the projected view based on dissimilarities of the elements in A. Again, it depends on the choice of attribute space A, which projection method is most suitable. We want to provide some guidelines: If A is isomorphic to an Euclidean space, the range of available methods is large and includes such algorithms as PCA. Otherwise, if only dissimilarity measures can be computed, good candidates are MDS, Stochastic Proximity Embedding, Part-linear Multidimensional Projection, and their variants. The choice of a proper method should be based on two main criteria: First, efficiency of the algorithm should allow for fast processing of large data, and second, similarity of projected elements should be conserved as much as possible. The latter criterion means that the distribution of samples in attribute space shall be matched as closely as possible in the projected view. Besides that, if the algorithm constructs a projection operator in explicit form, it is possible to apply it later to the whole dataset D. This is, in particular, the case if the projection acts as a linear operator.

An image of D_0 on the projected view is a set of points, each representing a certain time step of a simulation from the dataset. Since it makes sense to analyse simulation trajectories rather than separated time steps, it is helpful to connect points representing consecutive time steps from the same simulation run by line segments.

Based on shape and location of the drawn trajectories, it can be easily observed, which simulation runs behave similarly. This facilitates detecting patterns and general trends. Moreover, outliers can be identified, where outlier detection relates to both finding trajectories following abnormal or remarkable paths and single extreme nodes denoting special states of the system. We want to note that outliers, if any, may refer to a phenomenon detected for a special parameter setting as well as to the settings, for which the underlying physical model becomes invalid or applied numerical methods fail.

3.5 Interaction and Coordinated Views

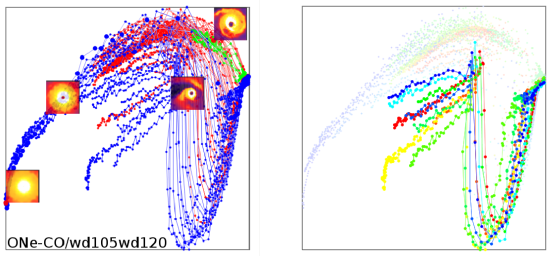
All four views on the data contribute unique information about the multi-run simulation. Most effective are they, if they are coordinated such that any selections made in one view are shown and can be further refined in all the other views. As such, one can, for example, detect and select outliers in the projected view, investigate which parameter settings they relate to in the parameter space visualization, select a subset of the outliers by brushing in the parameter space, investigate respective time steps of that run in the data space visualization, and relate them to the attributes space to observe what attributes made them to be outliers.

Moreover, one may generate multiple attribute spaces per element $d \in D_0$, e.g., one per data field. Then, multiple projected views are generated and coordination of these views allows for interactive analysis of runs with respective to different fields. Examples are given in the subsequent sections. The effectiveness of the coordinated interaction is best observed in the accompanying video.

We also provide further interaction mechanisms with the projected views. When investigating many runs or many time steps simultaneously, opening too many coordinated views may become confusing at some point. To provide a better overview of which point of the projected view belongs to which kind of observed phenomena, we provide embedded views of data or attribute space visualizations as shown in Figure 2. The user may enable an option of showing small pictograms near selected points characterizing the current state of the system (Figure 2a). Here, the pictograms are low-resolution visualizations of the attribute space with a user-defined transfer function applied. We used the same color scheme in all renderings of data or attribute spaces, see Figure 2c. In addition, a textual information about a single selected element is shown in the lower-left corner of the screen. This textual information provides the necessary information for the user to uniquely identify the run and time step that is under investigation. Moreover, the trajectories in the projected view are color coded. The user may choose, which information is color coded, e.g., showing a given classification (Figure 2a) or assigning to every simulation gets a different color. However, since the number of simulations is typically large, there is the perceptual issue that humans can only distinguish well a certain amount of colors with different hues. Therefore, if a group of trajectories is selected, it is possible to redistribute colors among these elements to better distinguish individual trajectories (Figure 2b).

4 INSTANCES OF FRAMEWORK

We produce two instances of our general framework by applying our technique to two datasets of very different structure. The first one represents a 100-year time series of a quasi-equilibrated (“persisting”) preindustrial climate state. The data are sampled on a 2D regular grid, which makes the construction of attribute space trivial. We process these easy-to-handle data to illustrate how



(a) Embedded attribute space (b) Selected trajectories

(c) Color scheme

Figure 2: Projected view constructed for SPH dataset (temperature field, MDS with l_2-distance computation). (a) Elements of attribute space visualizations are embedded into the layout as pictograms showing temperature profile with user-defined transfer function applied. Textual information about selected simulation is given in the lower-left corner of the layout. Polylines denoting simulations are colored according to the chemical composition of modeled stars (3 classes). (b) Redistribution of colors among selected simulations to better distinguish the trajectories. (c) Color scheme used to render attribute spaces in all experiments.

intuitively expectations match with the projected view configuration, to show basic tools of user interaction, and to validate the general concept of the proposed approach. The second dataset comes from a Smoothed Particle Hydrodynamics (SPH) binary system simulation. It is given in the form of volumetric unstructured data, which makes the construction of the mapping to the attribute space a non-trivial task. Here, we investigate how different multidimensional projection methods affect the resulting projection and demonstrate the functionality of multiple coordinated views.

4.1 Gridded Climate Simulation

Simulation results in the field of weather prediction and climate research are usually very sensitive to initial conditions. Therefore, it is a common practice to run an ensemble of simulations with varied initial parameters. Results are then presented in the form of a set of outcomes ranging from the most pessimistic to the most optimistic prognosis. Climate-related simulation data exploration was recently investigated by Hibbard et al. [HBSB02] and by Nocke [NFB07].

4.1.1 Data description

The data represent a climate simulation which is a spin-up of a preindustrial climate state (approx. 1850 AD). It covers a period of 100 years with monthly mean data written into files. Thus, the dataset consists of 1,200 time steps at which many scalar fields either volumetric or at the Earth surface are recorded. We picked up a number of 2D fields sampled at a regular (longitude-latitude) grid of size 96×48, among them surface pressure, surface temperature, ice thickness, ice cover, evaporation, total precipitation, cloud cover, and snowfall.

We interpret each of the 100 simulated years as a separate run with 12 time steps. Thus, the ordinal number of the year serves as the only discrete simulation parameter. Due to the nearly cyclic character of climate processes within each year, we expect to obtain their representation in the projected view in form of a pendulum trace or a circle-like curves. Of interest is to discover trajectories deviating from the mean pattern at most, which would correspond to the years with extreme climate state.

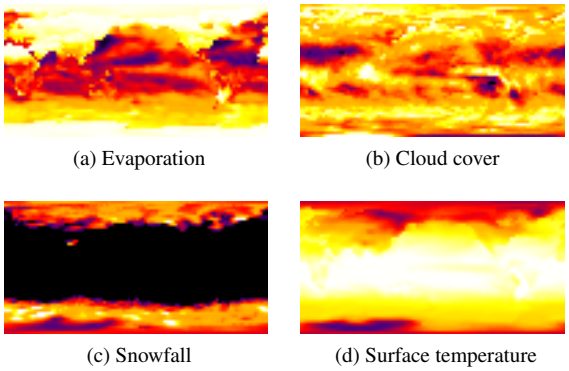
4.1.2 Definition of Attribute Space

Since the data have a fixed spatial grid structure and resolutions are manageable, we decided to use the identity for mapping from data to attribute space. Hence, the attribute vector of an element $d \in D_0$ is simply the vector with the scalar entries over the 2D grid. Thus, it is straightforward to introduce a metric for gridded scalar fields f and g using the l_p-norm as follows:

$$\|f-g\|_{l_p}^p = \sum_{i,j} |f_{ij} - g_{ij}|^p,$$

where indices i and j vary over the grid. In practice, it is common to restrict consideration to cases $p = 1, 2$.

For the given dataset, we observed that the selected scalar fields can be grouped into four categories as presented in Figure 3. Their values are either distributed approximately uniformly over the whole domain or the field behaves very differently over sea and land. Moreover, large values are either concentrated in the equatorial zone or close to the poles. Figure 3 shows examples of these four categories.

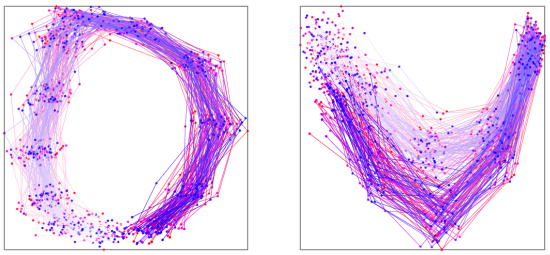


(a) Evaporation (b) Cloud cover

(c) Snowfall (d) Surface temperature

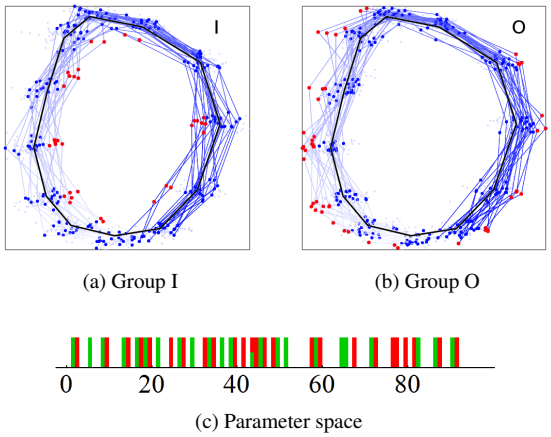
Figure 3: Four types of scalar field distributions in climate simulation dataset. Field magnitudes can be very different over sea and land as in (a) or can be almost irrelevant to this factor (b). Fields' maximum values can be reached in the polar zones (c) or near the equator (d).

 ISBN 978-80-86943-71-8



(a) Cloud cover (b) Snowfall

Figure 4: Two types of process representations on projected view: Circular (a) and pendular (b). Color value increases according to time.



(a) Group I (b) Group O

(c) Parameter space

Figure 5: Analysis of climate simulation data by means of projected view constructed for multi-field data with four scalar fields. The common pattern is a cyclic curve. The mean path is shown as a black polyline. Extreme trajectories deviating inside ((a), denoted as I) and outside ((b), denoted as O) from the mean path are highlighted. Deviation points are shown in red. 1D parameter space with depicted selections is presented in (c). Years from group I are marked in green, years from group O in red.

4.1.3 *Multi-field Exploration in Projected Views*

Due to the seasonal character of the modeled process, it was expected to obtain representations of the projected views in one of the forms shown in Figure 4. They are circular and pendular polylines and the results support the intuition: Circular trajectories intuitively represent fields where all four seasons have different characteristics, while pendular trajectories represent fields where two seasons (e.g., spring and fall) have similar characteristics and the other two seasons (e.g., summer and winter) represent the extrema.

We merged the four attribute representations shown in Figure 3 to produce a projected view based on multi-field data. Here, the attribute vector of the four merged fields is just a concatenation of the four attribute vectors of the four individual fields, i.e., since the number of grid nodes is $96 \cdot 48$, the attribute vector has length $96 \cdot 48 \cdot 4$ and is filled with values of evaporation, cloud cover, snowfall, and surface temperature fields. PCA algorithm places the items along circular trajectories. We selected two groups of outliers based on their distance from the imaginary mean path. The first group of polylines (further denoted by I) passes through points significantly deviating towards the inner part of the mean path. We counted 23 such trajectories, see Figure 5a. The second group of simulations (further denoted by O) deviates outside the domain confined by the mean path. There are 25 selected simulations in Figure 5b.

Comparing which simulations belong to which group, we found 10 pairs corresponding to two consecutive years belonging to different identified groups. When ordered according to time, 8 of them are of type I-O, i.e., the first year was of type I and the subsequent second year of type O, and only two of type O-I. Also, we detected two patterns of the form O-I-O, i.e., three consecutive years were found with the first year of type O, the second of type I, and the third again of type O, and two combinations over four years of type I-O-I-O (the groups of three and four years were not counted towards the groups of two). Interestingly, there was only one single year (43rd year of the simulation, shown as a two-colored bar) that had strong deviations from the mean path both towards the inside and the outside. We show the statistics in Figure 5c. We present these results to show how quickly and intuitively one can draw conclusions with our interactive analysis tool. As a next step, we plan on taking these findings back to the data providers to investigate whether our findings give new insights to them or to further refine our analysis from the application point of view.

4.2 Astrophysical Particle Simulation

SPH is a Lagrangian numerical method, which has a number of built-in conservation properties, e.g., conservation of total mass, total energy, linear and angular momenta, which is crucial for modeling of some hydrodynamical and especially astrophysical processes. The key idea of the approach is to represent objects as a set of *particles* – spatial samples changing their positions during simulation. They carry a number of scalar (density, temperature, etc.) and vector (velocity, force, etc.) values, which may either evolve in time or remain constant (e.g., mass). Interaction forces between particles, which move them from one time step to another, are defined by non-negative radial kernels. The same kernels are used for reconstruction of continuous fields from values at disconnected particle locations. We refer the reader to the recent review by Rosswog [Ros09] for more details on basic principles of SPH and to Price [Pri12] for a comprehensive exposition of SPH approximation techniques.

 ISBN 978-80-86943-71-8

4.2.1 Data Description

The dataset consists of simulations' outputs of the merger process of two White Dwarfs (WD). The stars are bound by gravity, which triggers a mass transfer from the lighter star (donor) to the heavier one (accretor). This process affects the orbital evolution, the angular momentum and the thermodynamical state of the final object, the associated gravity wave signal, and whether or not a particular binary merges at all. Simulation of binary systems is of great importance in astrophysics, see, for example, grid- [MFTD07] and particle-based simulations [DRGRR11]. In the grid-based calculations, the mass transfer continues for tens of orbits, whereas particle simulations showed that the stars merge after a few orbital periods. This led to discussions about impact of accurate initial conditions on the stability of the modeled process.

The main input parameters are the masses of donor and accretor stars, denoted by M_1 and M_2 ($M_1 \leq M_2$), correspondingly. The main goals of the study are:

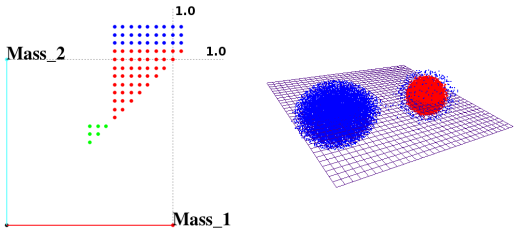
- understand the evolution of WD systems from the onset of mass transfer until the merger,
- identify accurately the parameter space region where detonations may occur in the lead-up to the merger or at surface contact,
- compare against previous SPH and grid-based calculations, in particular, examine the duration of stable mass transfer phase.

Both M_1 and M_2 (measured in units of the solar mass) take real values. It is assumed that the simulation output continuously depends on the governing parameters.

The available part of the dataset consisting of 78 simulations can be classified according to the chemical composition of WDs: double Carbon-Oxygen (CO-CO), double Helium-Carbon-Oxygen (HeCO-HeCO), and Oxygen-Neon / Carbon-Oxygen (ONe-CO) binaries, see Figure 6a. Every simulation involves $40k$ particles ($20k$ particles for each star) and includes two vector fields (velocity and gravity), six scalar fields (particle radius, internal energy, mass, density, temperature, and electron density), and many chemical units records. The simulations are stopped after a timescale of more than three initial orbital periods from the merger momenta, which results in different number of time steps for different simulations ranging from 600 to $1,600$. The total size of the dataset exceeds 0.5TB.

4.2.2 Choosing Data Space

To build up a representative subset D_0 of the whole data, which is possible to operate in reasonable time, we took every tenth recorded time step from each simulation and skipped the part of data related to chemical components of stars. The resulting data consist of $6,231$ time steps taking 32GB of disk space.



(a) Parameter space (b) Attributes computation

Figure 6: (a) Sampling of the parameter space in SPH dataset. Masses of two stars are given in units of solar mass. Colors encode chemical composition of WDs: red stands for double CO binaries, simulations involving HeCO stars are shown in green, and ONe-CO systems marked as blue. (b) Construction of attribute vector for SPH dataset. Shown is an initial stage of mass transfer between donor (blue) and accretor (red) WDs. A regular grid is built in the plane containing orbits of stars. The center of mass of the accretor is placed at a certain position on the grid. After a scalar field of interest is approximated at grid nodes, these values are used to characterize the current state of the system and compare it against other time steps.

4.2.3 Definition of Attribute Space

It is not possible to compare different time steps from the dataset directly due to the unstructured character of the spatial data. Particle positions change significantly from one time step to another within a simulation. Their locations in different simulations are not related in any way. Thus, a re-sampling of fields of interest on a common sampling pattern fixed in space is necessary.

Though the simulated phenomena are volumetric, they are supposed to be symmetric with respect to the plane parallel to the stars' orbits and passing through the centers of mass of the stars. In the following, we call this plane a *cross-section*. Footprints of fields on this plane are usually shown by astrophysicists in their demonstrations to illustrate findings. Thus, we decide to construct a 2D regular grid on the cross-section and interpolate scalar fields at the grid nodes using native SPH approximation. Such data arrays serve as elements of the attribute space in our algorithm. Their comparison can be handled as done for the climate simulation data.

When orbiting, stars change their absolute position in space, whereas their relative position may remain nearly the same. It is desired to perform analyses which are invariant to the absolute positions of stars. Our solution relies on the fact that, while the donor star is being disrupted, the accretor remains a relatively stable object. Thus, the vector pointing out of the center of total mass towards the center of mass of the accretor defines the orientation of the system. Therefore, we align the x-direction of the constructed grid with this vector, shift it in order to place the center

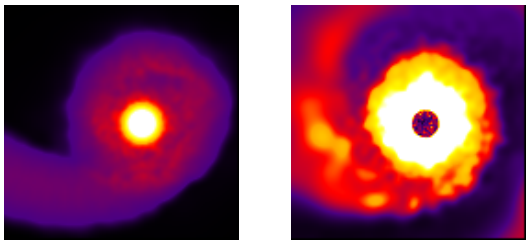


Figure 7: Cross-sections with approximated density (left) and temperature (right) field of a ONe-CO binary with masses $M_1 = 0.65$ and $M_2 = 1.1$. A dark (cold) region inside the accretor star is due to the cold and isothermal initial WD approximation used in the simulation. The heating processes take place at the boundary of the accretor star when merging.

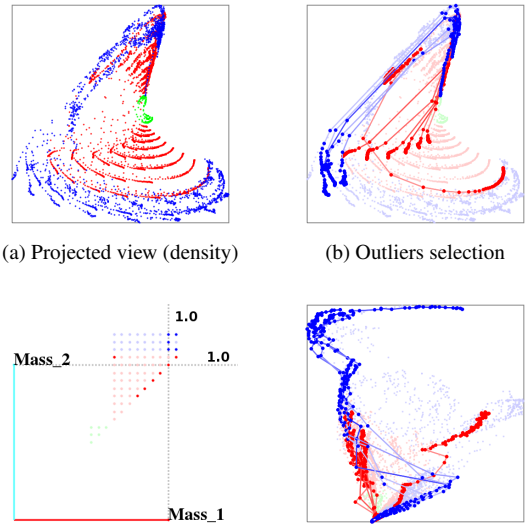
of the heaviest star at the point $(0.75, 0.0)$ (relative grid coordinates), and perform no scaling. The extents of the grid are chosen to cover the space occupied by stars in any simulation. An example is shown in Figure 6b.

In our experiments, we worked with grids of size 128^2. For example, cross-sections with approximated density and temperature field of a ONe-CO binary with masses $M_1 = 0.65$ and $M_2 = 1.1$ are shown in Figure 7.

4.2.4 *Visual Exploration with Projected Views*

Projected views are helpful for the analysis of multi-run simulation data, especially, in tasks of outlier detection and interactive classification of results. We started our analysis of the multi-run SPH simulation dataset with the generation of projected views for density and temperature fields using the PCA algorithm. The layout for the density field (see Figure 8) shows a good localization of chemical composition classes, a clear pattern of simulation evolution in form of arc-like paths, and a good layer-wise separation of trajectories according to their input parameters. At the same time, several trajectories not following the common pattern can be easily distinguished. Some are highlighted in Figure 8b and the corresponding parameter settings are highlighted in the parameter visualization in Figure 8c.

We detected three groups of outliers. The first one consists of 6 simulations from the red group (double CO-star systems), in which both stars have same masses (diagonal elements in the parameter space corresponding to $M_1 = M_2 = 0.8, \ldots, 1.05$). Their trajectories start in the middle part of the projected view in Figure 8b and turn to the top-right direction instead of following arc-like paths. Since both stars have equal masses, there are no donor and accretor roles in these systems. Both stars should merge symmetrically and that explains the different shape of their trajectories. We asked our collaborators who provided the data why the rest three CO-systems with $M_1 = M_2 = 0.65, 0.7, 0.75$ do not show any deviation from the pattern behavior. We were told that high total mass of the binary $(M_1 + M_2)$ leads to



(a) Projected view (density) (b) Outliers selection

(c) Selection in parameter space (d) Selection in projected view (temperature)

Figure 8: Outliers detection on projected view built for density field from SPH dataset. Simulation trajectories follow an arc-like paths in the clockwise direction in the layout (a). There is a good localization of chemical composition classes and a clear distribution of simulations among different layout regions. Some trajectories do not follow a common pattern. Their selection is shown in (b). Corresponding selection in the parameter space is presented in (c). Not all of the detected outliers can be identified when using projected view built for temperature field (d).

fast disruption of stars. For fast processing, its symmetry can be easily broken by numerical noise. That means that for the selected six outlier simulations, the numerical algorithm failed to preserve their symmetry by producing physically unreliable results. These simulation were already on the "black list" of our collaborators, since their number of orbits did not fit the overall picture, which made them suspicious. For slow evolving systems with low total mass (smaller than 1.6), we observed that the results still show physical soundness.

The second group of outliers consists of five simulations belonging to the ONe-CO group (shown in blue color). Their paths reach the most top-right region that is occupied by nodes and then go vertically down. Analysis of the parameter space in Figure 8c shows that these outliers have large values of total mass. This fact affects the physics, which can be better explained when looking to the temperature field, see below.

Finally, the last outlier, which is the red trajectory in Figure 8b that starts at the lower-right corner of the projected view, shows a strange behavior in the last part of its evolution. Its trajectory goes in reverse direction and then stabilizes around a certain position. Our collaborators explained us that this simulation was not stopped

 ISBN 978-80-86943-71-8

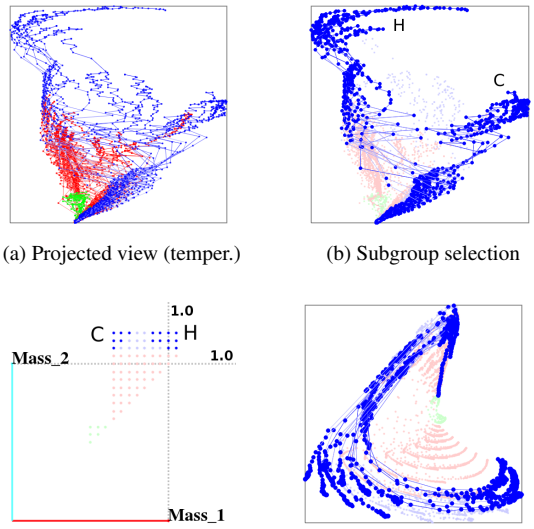
after three initial orbital periods from the merger momentum. It was decided to run computations longer, since the remnant object proved to be stable. Using our projected view, we were able to easily identify this run.

Now, we make use of coordinated interaction between multiple projected views. The outliers identified above are shown in Figure 8d on the projected view built for the temperature field. It is worth to note that two of the three outlier groups identified above can also be identified in this layout. However, the equal masses outlier group, which was the first one we discussed, does not look odd. This means that developed instability of computations mostly affected the density distribution and not so much the temperature field.

Now, we investigate the entire picture of the projected view built for the temperature field, see Figure 9a, and start a similar analysis as for the projected density view. Here, we identified two subgroups for further investigations, namely the two groups of blue polylines marked as "H" and "C" in (b). Group "H" includes the blue outlier group identified above. The groups differ in the mass of the donor star, as follows from the parameter space visualization (c). Then, we looked at the attribute space visualizations to find reasons for different trajectory shapes. We figured out that the inner part of the accretor remains cold along paths from group "C", similar to the case presented in Figure 7 (right). The remnants from group "H", on the other hand, show intensive heating up of the internal region. A physical explanation relies on the fact that massive donors deliver more matter to the accretor, which shifts the location of the maximal temperature inside the star and leads to its heating up. The two groups detected are hard to distinguish on the projected view constructed for the density field (d), as they show a similar behavior. Hence, it is important to use multiple coordinated projected views to detect and analyze all possibly interesting trajectories.

We note that simulations involving stars with He in their composition (green group) do not show big evolutionary changes as opposed to the binaries with a CO companion star. The trajectories of the green group are bound to a small space in the view, which requires zooming in for further analysis. These systems have very low total mass and, thus, the magnitudes of density and temperature changes are very low, which is reflected in the respective projected views we generated. In fact, direct comparison of binaries with He- and CO-WDs is not fair. Much lower temperature value is needed to trigger nuclear burning for He-stars than for CO-stars. The systems from the green group more frequently end their evolution with explosion. Thus, such binaries should be processed and analyzed separately.

When computing the attribute space elements, the 2D grid was rotated and translated according to current positions of stars. Looking to the rotation angle history, it



(a) Projected view (temper.) (b) Subgroup selection

(c) Selection in parameter space (d) Selection in projected view (density)

Figure 9: Subgroups selection on projected view for temperature field from SPH dataset (a). Two subgroups of the ONe-CO simulation class (blue) show different behavior (b). They correspond to the parameter settings presented in (c). Same subgroups exhibit a common pattern on the projected view built for density field (d), thus, they can not be identified there.

is easy to estimate the number of orbits resolved in each simulation and record the found values as simulation attributes. Figure 10 shows a projected view created by means of MDS algorithm with l_2-distance measure for the density field. Color coded are the number of orbits represented in each simulation ranging from small (red, rapid merging with intensive heating) to high (blue, gradual merging). Colors in the parameter space match the colors for trajectories. We observe that the number of orbits is more sensitive to variation of M_1 (mass of the donor star) than to changing the second parameter (mass of the accretor). However, since there was no strict criterion to stop simulation and different runs lasted different number of periods from the merger momenta, the results in Figure 10 characterize the simulation performance rather than the modeled processes.

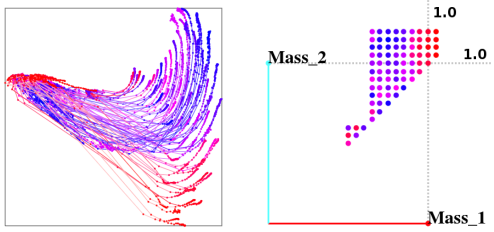


Figure 10: Projected view (left) and parameter space (right) color coded according to the number of orbits within each simulation. Red stands for high and blue for low values.

 ISBN 978-80-86943-71-8

5 CONCLUSION

We have presented a framework for the visual analysis of multi-run time-varying multi-field data that allows for effective and intuitive detection of patterns and outliers in multi-run data. We applied our framework to two instances of different character (gridded climate simulations and astrophysical particle simulations) to show its general applicability. The most excessively used component of our framework is the projected view, which is based on similarity within a metric attribute space. However, linked views to parameter space visualization and spatial data visualization allowed for a comprehensive visual analysis.

Currently, we apply our framework in a post-processing analysis setting. In future work, we plan to extend it such that we can interpolate and extrapolate in parameter space to guide the application scientists towards interesting parameter setting. Hence, our framework could serve as a computational steering tool.

6 ACKNOWLEDGMENTS

The authors wish to thank Marius Dan and Stephan Rosswog for sharing the SPH dataset. We also thank Sergey Lisitsyn for sharing his library `tapkee` (`http://tapkee.lisitsyn.me/`). This work was supported in part by a DFG grant LI 1530/6-2.

7 REFERENCES

[BG05] I. Borg and P.J.F. Groenen. *Modern multidimensional scaling: Theory and applications*. Springer, 2005.

[BM10] S. Bruckner and T. Moller. Result-driven exploration of simulation parameter spaces for visual effects design. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1468–1476, 2010.

[CCM09] C. Correa, Y.-H. Chan, and K.-L. Ma. A framework for uncertainty-aware visual analytics. In *Visual Analytics Science and Technology*, pages 51–58, 2009.

[CH12] T. J. Cashman and K. Hormann. A continuous, editable representation for deforming mesh sequences with separate signals for time, pose and shape. *Computer Graphics Forum*, 31(2):735–744, 2012.

[CvN00] R. M. Cooke and J. M. van Noortwijk. Graphical methods for uncertainty and sensitivity analysis. In Satelli, Chan, and Scott, editors, *Sensitivity Analysis*, pages 245–266. 2000.

[DRGRR11] M. Dan, S. Rosswog, J. Guillochon, and E. Ramirez-Ruiz. Prelude to a double degenerate merger: The onset of mass transfer and its impact on gravitational waves and surface detonations. *The Astrophysical Journal*, 737(2):89, 2011.

[HBSB02] B. Hibbard, M. Böttinger, M. Schultz, and J. Biercamp. Visualization in earth system science. *SIGGRAPH Comput. Graph.*, 36(4):5–9, 2002.

[JJ09] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):993–1000, Nov. 2009.

[Kan12] E. Kandogan. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *Visual Analytics Science and Technology (VAST)*, pages 73–82, 2012.

[KH13] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Trans. Vis. Comput. Graph.*, 19(3):495–513, 2013.

[Ma99] K.-L. Ma. Image graphs – a novel approach to visual data exploration. In *IEEE Visualization 1999*, pages 81–88, 1999.

[MAB+97] J. Marks, B. Andalman, P. A. Beardsley, W. T. Freeman, S. Gibson, J. K. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. M. Shieber. Design galleries: a general approach to setting parameters for computer graphics and animation. In *SIGGRAPH*, pages 389–400, 1997.

[MFTD07] P. M. Motl, J. Frank, J. E. Tohline, and M. C. R. D'Souza. The stability of double white dwarf binaries undergoing direct-impact accretion. *The Astrophysical Journal*, 670(2):1314, 2007.

[MOD00] M. Monks, B. M. Oh, and J. Dorsey. Audioptimization: Goal-based acoustic design. *IEEE Computer Graphics and Applications*, 20(3):76–91, 2000.

[NFB07] T. Nocke, M. Flechsig, and U. Böhm. Visual exploration and evaluation of climate-related simulation data. In *Winter Simulation Conference, WSC 2007*, pages 703–711, 2007.

[Pea01] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

[Pri12] D. J. Price. Smoothed particle hydrodynamics and magnetohydrodynamics. *J. Comput. Phys.*, 231(3):759–794, Feb. 2012.

[Ros09] S. Rosswog. Astrophysical smooth particle hydrodynamics. *New Astronomy Reviews*, 53(4-6):78–104, 2009.

[SPK+07] R. C. Smith, R. R. Pawlicki, I. Kókai, J. Finger, and T. Vetter. Navigating in a shape space of registered models. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1552–1559, 2007.