Author(s): Piccolotto, Nikolaus; Bögl, Markus; Muehlmann, Christoph; Nordhausen, Klaus; Filzmoser, Peter; Schmidt, Johanna; Miksch, Silvia

Title: Data Type Agnostic Visual Sensitivity Analysis

Year: 2024

Version: Published version

Please cite the original version:

Piccolotto, N., Bögl, M., Muehlmann, C., Nordhausen, K., Filzmoser, P., Schmidt, J., & Miksch, S. (2024). Data Type Agnostic Visual Sensitivity Analysis. IEEE Transactions on Visualization and Computer Graphics, 30(1), 1106-1116. https://doi.org/10.1109/tvcg.2023.3327203

Data Type Agnostic Visual Sensitivity Analysis

Nikolaus Piccolotto (iD), Markus Bögl (iD), Christoph Muehlmann (iD), Klaus Nordhausen (iD), Peter Filzmoser (iD), Johanna Schmidt (iD), and Silvia Miksch (iD)
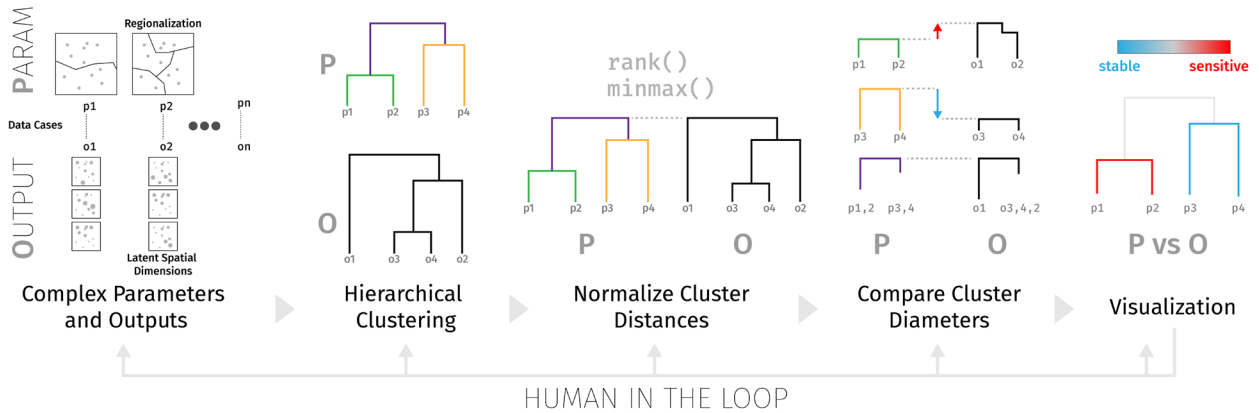
Fig. 1: *Illustration of the visualization pipeline used for the Discrepancy Dendrogram. The key idea of it is to use cluster diameters as a measure for variation in complex parameter settings and outputs. Given dissimilarity measures for each, we perform hierarchical clustering separately in each space. Distances are normalized by ranking or min-max normalization for comparison. For every cluster obtained through hierarchical clustering, we evaluate the difference in its diameter and visualize that by color.*

Abstract—Modern science and industry rely on computational models for simulation, prediction, and data analysis. Spatial blind source separation (SBSS) is a model used to analyze spatial data. Designed explicitly for spatial data analysis, it is superior to popular non-spatial methods, like PCA. However, a challenge to its practical use is setting two complex tuning parameters, which requires parameter space analysis. In this paper, we focus on sensitivity analysis (SA). SBSS parameters and outputs are spatial data, which makes SA difficult as few SA approaches in the literature assume such complex data on both sides of the model. Based on the requirements in our design study with statistics experts, we developed a visual analytics prototype for data type agnostic visual sensitivity analysis that fits SBSS and other contexts. The main advantage of our approach is that it requires only dissimilarity measures for parameter settings and outputs (Fig. 1). We evaluated the prototype heuristically with visualization experts and through interviews with two SBSS experts. In addition, we show the transferability of our approach by applying it to microclimate simulations. Study participants could confirm suspected and known parameter-output relations, find surprising associations, and identify parameter subspaces to examine in the future. During our design study and evaluation, we identified challenging future research opportunities.

Index Terms—Visual analytics, parameter space analysis, sensitivity analysis, spatial blind source separation

◆

1 INTRODUCTION

In many domains, data analysis requires dealing with multivariate measurements in space. For instance, mining corporations and public agencies may analyze geochemical soil samples for mine prospecting or investigating environmental pollution, respectively. Depending on the specific goal and application, various tasks, e.g., dimension reduction or finding meaningful linear combinations of variables, must be carried out on such datasets. Spatial blind source separation (SBSS) [2, 45, 46] is designed explicitly for multivariate spatial data and reveals linear combinations of such data. SBSS offers various benefits compared to alternative methods, e.g., it keeps the well-known loadings-scores

scheme from principal component analysis and adequately accounts for spatial dependence due to its model-based approach. Therefore, latent dimensions identified with SBSS often correspond to the physical reality where data was collected, making it an excellent analysis tool for spatial data. A detailed description of SBSS is out of scope for this paper, and we refer interested readers to [45, 46, 51]. SBSS has been successfully applied to a geochemical dataset [46] and may be potentially used in any application domain that involves multivariate quantitative measurements at different locations.

SBSS requires setting two complex tuning parameters: A partition of the spatial domain in non-overlapping regions (regionalization) and a ring-shaped point neighborhood (kernel). On the other side of the model (Fig. 2), SBSS yields a set of latent spatial dimensions (i.e., maps), where each is a linear combination of original dimensions with weights (loadings) given by the unmixing matrix. Consequently, parameter space analysis tasks [59] become relevant. Previous work [51] focused on the *optimization* task, but *sensitivity analysis* (SA) is considered equally important for SBSS. SA compares the relative variation in parameter settings and output of the model, thus highlighting relevant/irrelevant parameters and their stable/sensitive ranges. This analysis is essential to obtain and communicate reliable results, i.e., those not a consequence of luck and coincidence. SA is especially important for SBSS as it lacks so far any goodness-of-fit criteria; hence deciding between alternative parameter settings is challenging. SA can help with

---

- *Nikolaus Piccolotto, Markus Bögl, Christoph Muehlmann, Peter Filzmoser and Silvia Miksch are with TU Wien, Austria. E-mail: {firstname}.{lastname}@tuwien.ac.at.*
- *Klaus Nordhausen is with University of Jyväskylä, Finland. E-mail: klaus.k.nordhausen@jyu.fi.*
- *Johanna Schmidt is with VRVis GmbH, Austria. E-mail: johanna.schmidt@vrvis.at.*
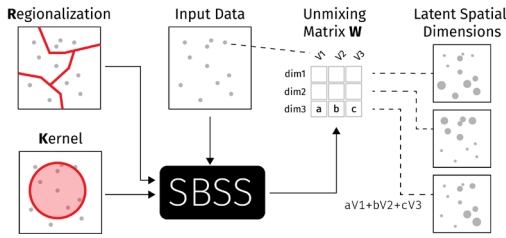
Fig. 2: *SBSS [45, 46, 51] takes a regionalization (R) and a kernel (K) as parameters and outputs a linear combination of input variables (latent spatial dimensions), described by the unmixing matrix (W).*

this decision as in prior work on blind source separation [49, 51], analysts noted that they find stable parameter settings more trustworthy and associated outputs more likely to be the "real" solution. SA may thus further strengthen the outcome of an *optimization* task and, additionally, inform geostatistical modeling: If, e.g., the regionalization parameter barely influences the output, analysts might reasonably suspect that the input dataset is spatially stationary (a geostatistical modeling decision).

SBSS is interesting for the visualization community primarily because of the mentioned affordances of its parameters and outputs: Parameter settings and outputs are spatial objects or otherwise complex in a way that a multivariate representation does not do them justice. While the literature contains many examples of visual parameter space exploration [50, 59], to the best of our knowledge, none of them support complex parameters and outputs without resorting to multivariate representation or feature derivation (Section 2). However, these requirements are not specific to SBSS, as many examples exist for models with complex parameters *and* outputs. For instance, spatial or time-varying inputs and outputs can arise in microclimate simulations [64]. They predict meteorological variables (e.g., air temperature, humidity, or wind speed) in a small area, typically for a single street or building.

We intend to close this gap with our paper. The core idea of our proposal is illustrated in Fig. 1: We take a cluster's diameter as a measure of variation for the contained parameter settings or associated outputs (referred to as data cases, respectively). Then we can enable SA for SBSS in the following way. Given appropriate dissimilarity measures for data cases, we compute pairwise distances in each space (parameter and output), based on which a hierarchical clustering is produced. After normalizing distances, we compute the diameter difference of all clusters between one space and the another. This information is then presented in our main visualization, the Discrepancy Dendrogram. Supporting visualizations complete required user tasks. In particular, the contributions of our design study are that we

- propose a task abstraction for SA in the context of SBSS (Section 3);

- based on SBSS requirements, develop a visualization that supports SA and works on any data type (Fig. 1, Section 4);

- integrate this and other visualizations in a visual analytics prototype (Section 5);

- evaluate the prototype with experts in visualization (Section 6.1) and SBSS (Section 6.2);

- show the transferability to other problems by applying our approach to microclimate simulations (Section 6.3).

2 RELATED WORK

2.1 Sensitivity Analysis

Sensitivity Analysis (SA) is "the study of how the uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input [55, p. 1]." SA allows analysts to determine how variations in the input influence the output. A broad distinction between various SA methods can be drawn at whether they are *local* or *global* [57]. Local methods are applicable when the model is linear as they yield, e.g., a partial derivative according

to one parameter. An example of such local methods is the one-at-a-time approach, where one parameter is varied while the others are kept fixed. Global methods, on the other hand, are applicable to non-linear models, too. A well-known example is the Sobol index [62], a variance-based global SA method. Several surveys exist [7, 9, 24, 28, 29, 56] that collect and discuss both local and global methods. Methods covered in these surveys mainly consider models with multivariate parameters, e.g., the output scalar y is a function of an input vector $\mathbf{x}$: $y = f(\mathbf{x})$. Spatially-varying parameters [36, 52] or outputs [35, 42] have been considered as well. However, these methods do not fit to SBSS (Fig. 2).

2.2 Visual Parameter Analysis

Visual parameter analysis (VPA) has a long history in the visualization literature, with seminal works published in the 1990s, like Design Galleries [41] or spreadsheet interfaces [30]. Sedlmair et al. [59] provided a common data flow model and a task taxonomy, such as optimization, uncertainty, or SA. Piccolotto et al. [50] surveyed user interfaces and visualizations that support visual parameter space exploration. Several examples of VPA for multivariate parameters can be found in the literature [5, 15, 22, 32, 48, 72]. However, these approaches do not apply to SBSS parameters.

Many approaches have been used when it comes to visualizing parameter-output relations [50]. When parameters are multivariate, visualizations that show correlations and trends can be used to carry out SA, such as histograms, scatterplots, or PCPs [4, 14, 66]. These visualizations are often *juxtaposed* and linked, such that selections in one view highlight the same data in other views [43]. Another option is to *embed* parameters and outputs in the same visualization, e.g., by encoding them as axes in the same PCP [63] or by color-coding a 3D model [18]. A consequence of juxtaposition is that general visualization-independent approaches may be used together. E.g., first grouping data cases by similarity, then inspecting properties of individual groups [1, 8, 25] is popular. Orban et al. [47] devised two linked dimensionally-reduced (DR) scatterplots, an approach that can generally be extended to complex data and SBSS parameters/outputs. However, our target users struggled with DR scatterplots in previous work [49]. The difficulty was that the DR spatializations looked like scatterplots but did not show the same information and required a different way of reading, which was unintuitive to them. Therefore, we developed an alternative approach. A more specific form of juxtaposition is to *align* data cases in useful ways that highlight dependencies between parameters and outputs, e.g., as part of a spreadsheet [19, 38, 39]. The idea is that dependencies become visible when the spreadsheet is sorted by multiple columns. However, it requires a compact visual representation. *Superposition* may be possible if parameter and output refer to the same space, such as particle trajectories and their initial position [21]. *Sequential Superposition* leverages a system's interactivity. The analyst may rapidly browse between parameter/output pairs, and sudden visual jumps in the emerging animation point to sensitive parameter ranges [26, 54, 58]. Parameter and output visualizations may also be *integrated* with explicit links drawn between them. E.g., a trapezoid that connects parameter and output histograms shows sensitivity by the relative length of horizontal segments [68]. Another option for composite visualizations of parameters and outputs for SA is *nesting*, i.e., putting visualizations inside the marks of another, like correlation matrices in an interval tree [19].

Data mining methods may also support visual SA. E.g., if regression analysis between parameter and output is possible, that information can be shown in the parameter visualization in the spirit of scented widgets [17, 33, 69]. Correlation analysis between parameters and derived output features may also be done if they lend themselves to it [19]. Developing a surrogate model augmenting the original model with fast but inaccurate output predictions for new parameter settings is standard practice in VPA [59]. It may be possible to extract information from the surrogate to support SA, such as parameters in linear regression [43], or partial derivatives in neural networks [25].

Generally, in existing work, either the parameter (by multivariate representation) or the output (by feature derivation) must have multivariate characteristics. Our contribution to visual sensitivity analysis

enables it in situations where both parameter and output are of complex data types, e.g., spatial objects.

2.3 Visual Cluster Analysis and Clustering Comparison

Clustering is an essential wide-spread class of data analysis methods, and various flavors were proposed over time [71]. Generally, clusterings partition data cases into coherent groups according to a distance function. Visual inspection of these groups may reveal previously hidden patterns. To visualize the whole clustering, nowadays, color-coded dimensionally-reduced scatterplots are commonly employed [12,34,70]. However, these scatterplots are only approximate, as they contain projection errors [31], and may require specialized knowledge to interpret [67]. Glyph-based visualizations [11] were proposed in the context of geospatial data. Dendrograms [20,61] commonly depict hierarchical clusterings. Blanch et al. [6] proposed the Dendrogramix, a combination of dendrogram and matrix visualization. The clustering outcome depends on the specific algorithm and parameters, so visualizations were proposed to compare these. However, they focus on the analysis of cluster members [12], comparison of clusterings concerning parameters [12] or algorithms [34,40]. I.e., the definition of distance between data cases is fixed. Our work may be seen as comparing clusterings with alternative distances (Fig. 1).

3 USERS & TASK ABSTRACTION

As in previous work on SBSS [51], our primary users are experts in statistics. We anticipate our user base to eventually include domain experts, e.g., from geochemistry. We conducted an extensive literature review [50] to understand how visual VPA and, consequently, visual SA work in other contexts. Based on that, we distilled generic SA subtasks to enable SA on the SBSS-specific complex parameters with our clustering-based approach (T1–T5). We presented and discussed them with our collaborators (statistics/SBSS experts who are co-authors of this paper) to ensure their suitability. Based on these tasks, we developed the main visualization (Section 4).

Tasks.   First, to start the analysis, analysts must *compare the association between parameters and outputs* (T1). Pairs of highly associated parameters and outputs are less interesting to investigate. For any given parameter/output, they must *assess its overall variation* (T2) to learn about contained similarity structures and outliers. Furthermore, analysts must *identify groups of data cases with low/high variation in a parameter/output* (T3) in order to *compare variation between parameters and outputs, both overall and for a group of data cases* (T4). To support analysts in reasoning why this variation happens, they must be able to *view individual data cases* (T5).

Guidelines.   In addition to user tasks, we formulate three design guidelines for the visualizations. These were informed by evaluations conducted in our past work [49,51] and by widely used visualization guidelines. First, *visual marks of similar values should be adjacently arranged* (D1). This visual requirement suggests continuity that scalars exhibit naturally, but complex objects do not. It will make it easier to perceive stable/sensitive parameter ranges. *Occlusion must be avoided* (D2) to not clutter the display. The visualization should, if possible, *resemble a familiar graphic* (D3) that our target users are familiar with.

4 DISCREPANCY DENDROGRAM

We describe in this section how our main visualization, the Discrepancy Dendrogram, is constructed (also compare Fig. 1). The complete VA prototype will be discussed in the following section. We aim for a visual-interactive approach for two reasons. First, we did not find numerical SA approaches that are applicable to our data (Section 2.1). Second, our approach needs configuration (e.g., Section 4.2 or Section 4.3), where each choice highlights different patterns (compare Fig. 10), impacting the conclusions to draw. Thus, in an interactive setting, the analyst can quickly change between those configurations and thoroughly compare them (see, e.g., Section 6.2).

The core of SA is to compare the relative variation in parameter settings and outputs. It can readily be quantified for numbers (cf. variance-based SA approaches), but measuring *variation* for complex

objects, like the spatial SBSS parameters, is not straightforward. Our proposal's core idea (Fig. 1) is to consider cluster diameters for that purpose: A cluster gets wider the more dissimilar contained data cases are. Conversely, the cluster diameter is zero when all contained data cases are the same. There are advantages to that approach. First of all, a clustering can be obtained when only pairwise similarity information (Section 4.1) is available. Thus a formal notion of variation need not exist for the data type at hand. Second, cluster analysis generally supports tasks T2 and T3 when one investigates global cluster structures (e.g., how many exist, how many data cases they contain) and local structures (e.g., finding outlier cases). Hence we propose to augment a visualization of cluster structures with the information required for SA, i.e., whether clusters shrink or expand when applying another dissimilarity measure to the data cases. This approach can be seen as *orienting* guidance [13] that points analysts to interesting data cases. The major available choices at this point are i) the type of visualization and ii) how to compute the augmenting information. The two choices are independent, and we focus on the latter before discussing the former in Section 4.5.

Sampling.   Any parameter space analysis task requires a reasonable set of (parameter setting, output) tuples. Common desired sampling properties are that it is uniform and spans a large part of the parameter space, which is achieved via automated sampling techniques. These are hard problems for SBSS, where two random parameter settings are not a-priori equally reasonable. Domain knowledge critically informs parameter selection in SBSS [51]. Single-execution runtimes measured in minutes or hours further complicate the issue. Thus, following study participants' current practices in SBSS and microclimate simulations, we rely on a few dozen, mostly manually selected, parameter settings and limit SA insights to that subspace. While not solving everything at once, our approach still improves their current situation.

4.1 Dissimilarity Measures

Dissimilarity measures, considerably the basic requirement for any analysis, exist for many data types. A dissimilarity measure is a function $d(\cdot,\cdot) \to \mathbb{R}^+$ that quantifies how similar two objects are. Generally, we expect that $d(a,b) = 0$ iff $a = b$ and that $d(a,b)$ is strictly monotonically increasing with the differences between a and b. We assume such a dissimilarity measure for every model parameter and output.

4.2 Hierarchical Clustering

Flat partitioning cluster algorithms, like k-means, divide the dataset into an a-priori specified number of groups while minimizing intra-group distances. On the other hand, hierarchical clustering algorithms retain all cluster structures in the dataset and, therefore, do not require a k parameter. Hierarchical clustering is thus preferable because it will contain all possible clusters the analyst might be interested in, and we can enumerate them. We chose a clustering by agglomerative nesting (AGNES) [53] because bottom-up hierarchical clustering is easier to think about and, thus, easier to explain to analysts than the top-down variant. Further, many current alternatives, such as HDBSCAN [10], require Euclidean distances and can not be used with just dissimilarities. The main parameter of AGNES is the linkage criterion, i.e., how to compute the distance between two clusters. Only some linkage criteria *can* be used in our case. E.g., centroid-based variants like Ward's method are not applicable as the concept of a centroid may not exist for complex data types, such as regionalizations. Consequently, we provide complete and average linkage as user-selectable hierarchical clustering parameters.

4.3 Normalize Cluster Distances

We aim to evaluate whether a given cluster shrinks or expands when an alternative dissimilarity measure $d_A()$ is applied. The obvious problem here is that $d()$ and $d_A()$ might have differing images, i.e., one maps to the unit interval $[0,1]$ while the other maps onto $[0,1312]$. We propose ranking or min-max normalization to solve this issue. Both operations work on a distance matrix. Ranking replaces values in all cells by their rank, while min-max normalization maps values onto the unit interval. When comparing ranks, the focus will naturally be on ordinal

Data: Cluster C of data cases, normalized distance matrices D_1
and D_2, cluster diameter definition $diam()$.

```
1 Function index(C, D₁, D₂, diam) is
2     D₁[C] ← upperTri(select(D₁,C))
3     D₂[C] ← upperTri(select(D₂,C))
4     return diam(D₁[C]) − diam(D₂[C])
5 end
```

Algorithm 1: Pseudocode of sensitivity index computations.

changes, ignoring magnitude. Min-max normalization, on the other hand, preserves magnitude. The analyst can switch between the two as both approaches have advantages and drawbacks (compare Fig. 10).

4.4 Compare Cluster Diameters (Sensitivity Index)

Finally, we require a way to measure a cluster's diameter, which roughly corresponds to the linkage criterion in Section 4.2. To find candidates, we turn to internal clustering validation measures [37], as no external information exists in our case. These usually incorporate the *compactness* of clusters, which measures the variation within a cluster. Based on the selected linkage criterion, we use the largest distance between any two elements (complete linkage) or the average distance between all elements (average linkage).

Given two distance-normalized hierarchical clusterings P and O (e.g., one with distances of parameter settings and one with output distances) and a cluster diameter definition, we can compute by how much a cluster in P shrinks or expands in O, or the other way around, as P and O cluster the same data cases. We evaluate the *index()* function (Alg. 1) for every cluster, i.e., every horizontal line in a dendrogram. $D_{\{1,2\}}$ are the respective distance matrices of P and O. The subroutine `upperTri` returns the upper triangle of a square matrix, and `select` selects specified rows and columns of a square matrix. The function can be seen as a sensitivity index as it quantifies how much the variation differs between the parameter and output space.

4.5 Visualization

Two established visualization idioms for clusterings are dimensionally-reduced scatterplots and dendrograms. As our target users (statistics experts) found the former approach in previous work [49] rather unintuitive, we chose the latter for our context, fulfilling design guideline D3 (Section 3). Additionally, a dendrogram supports many other guidelines and user tasks. The leaves are juxtaposed (D2), and similar leaves, which are joined into clusters earlier than dissimilar leaves, naturally appear adjacent (D1). Optimal leaf orderings may be used [3]. Lines encode the diameter of every possible cluster that could be interesting (T2–T3). These lines do not overlap (D2). The open challenges are encoding the sensitivity index (Section 4.4) in the dendrogram (T4) and ensuring that visualizations of data cases are visible (T5).

The free visual channels in a dendrogram we could use to support T4 are line color (hue, saturation), line texture (e.g., dashed or dotted), and line thickness. We encoded the sensitivity index in color hue (compare Fig. 1). The index diverges with 0 at the center. Hence, the direction is as important as the magnitude. Two-directional encodings are standard for color hue (diverging scales) but very uncommon for the other attributes and likely confusing for our target users. We use two diverging scales dependent on the choice of distance normalization (Section 4.3): Red–blue (ranked) and purple–green (min-max). By default, the color scale spans the whole theoretically possible index interval, but the analyst may instead use the interval as found in the dataset to highlight small-scale patterns.

To support task T5, we show customized space-efficient visualizations as leaves of the Discrepancy Dendrogram (Fig. 5-A, bottom). There is little available space when the dendrogram shows many data cases. We combat this issue with several strategies. First, clusters of the dendrogram can be hidden. Second, when leaves are clicked, a tooltip containing a more detailed visualization appears. I.e., we show the regionalization parameter of SBSS as flat polygons in the dendrogram and as an interactive Leaflet map in tooltips. Any cluster can be selected

Fig. 3: *Parameter assessment changes depending on the assignment to primary and alternative distance in the Discrepancy Dendrogram. Glyphs in the document show the color of wider output clusters.*
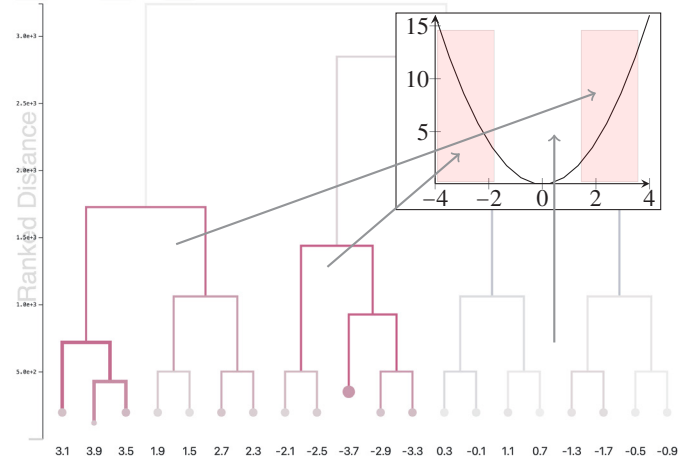

Fig. 4: ●*XY Discrepancy Dendrogram of the function $y = x^2$ (inset top right), with some clusters collapsed for readability. Red color highlights clusters that are wider in Y than X (=sensitive parameter ranges, i.e., marked parabola arms in inset).*

to be shown in the Gallery (Fig. 5-B). More interactions are described in Section 5.1.

4.6 Interpretation, Notation and Example

The choice of the color scale's orientation is arbitrary. We decided that red (purple) highlights an expanded cluster while blue (green) marks shrunk clusters in the alternative distance (O in Fig. 1). Consequently, interpretations regarding stability or sensitivity depend on how parameters and outputs are assigned to primary and alternative distances (Fig. 3). E.g., sensitive parameter settings are associated with wider clusters in the output space compared to the parameter space, which can appear as blue (parameter as primary distance) or red (parameter as alternative distance). In the remainder of the paper, we will use appropriate glyphs to denote the direction of sensitive parameters. A ●XY Discrepancy Dendrogram will thus i) compare X and Y, ii) show a dendrogram of clusters in X, iii) mark data cases with sensitive parameter settings as red.

Fig. 4 shows a ●XY Discrepancy Dendrogram for the function $y = x^2$ sampled uniformly in the interval $[-4, 4]$. The dendrogram separates the parameter space into three clusters with X 1.4 to 4, −4 to −1.9, and −1.8 to 1.3 (from left to right). The lines' hue may be interpreted as the absolute gradient: Red lines mark wider clusters in Y (high) while the right-most cluster is gray (low). When plotted as a line chart, these patterns would refer to the parabola arms (red clusters) and the part between them (gray) as visible in the inset.

5 VISUAL ANALYTICS PROTOTYPE

To facilitate SA of SBSS parameters and outputs, we propose a visual analytics prototype (Fig. 5). We developed it in a user-centered design process in collaboration with statistics experts, who are co-authors of this paper. Links to a web version of the software are available in the supplemental material.

5.1 Discrepancy Dendrogram (T2–T5, D1–D3)

We discuss the construction of the Discrepancy Dendrogram in Section 4 and focus here on interactions. We provide several interactions

Fig. 5: *Screenshot of our prototype showing 48 SBSS parameters and outputs (Section 6.2). Components: (A) Discrepancy Dendrogram (Section 4, Section 5.1), (B) Gallery (Section 5.2), (C) Subset Sensitivity View (Section 5.3), (D) Shepard Matrix (Section 5.4), (E) tooltip.*

with the Discrepancy Dendrogram to allow detailed investigation of clusters and to scale it to larger datasets. First of all, the user may choose between ranked and normalized distances (Section 4.3) and select the bounds of the color scale (Fig. 5, left top). Further, they may choose the linkage criterion for the dendrogram (Section 4.2), which also affects cluster diameter computations (Section 4.4). Second, there might be multiple parameters and outputs in a given dataset. The analyst can thus select which parameter/output to build the dendrogram with (primary distance), which parameter/output to compare it to in the sensitivity index calculation (alternative distance), and which parameter/output to show in the dendrogram leaves (Fig. 5-A1 to A3). As color hue is not a precise visual channel, we encode a cluster's diameter difference additionally in the length of a vertical line segment next to the dendrogram's Y-axis legend. Other dendrogram interactions are more concerned with scalability. It is possible to collapse a cluster (shift + click), collapse all other clusters (meta + click), and collapse all clusters below a user-defined height (click on the Y-axis). These interactions free up additional display space for an area of interest. Colored circles replace collapsed clusters. The circle size is proportional to the amount of data cases in the cluster, while the color corresponds to the clicked line's color. The data cases of a collapsed cluster are replaced by a cluster representative. Finally, a cluster can be selected, after which contained data cases are shown in the Gallery (Fig. 5-B).

5.2 Gallery (T5)

The Gallery shows data cases of a selected cluster in a grid (Fig. 5-B). The number of columns and their width can be selected by the analyst, as can the sort order of data cases and which parameter or output they should show. It is possible, e.g., to sort parameter visualizations by output similarity, as is often done in visual parameter space analysis [19,39]. Thus, the Gallery can show complex patterns.

5.3 Subset Sensitivity View (T4)

The Gallery shows data cases of a selected cluster in a grid (Fig. 5-B). The number of columns and their width can be selected by the analyst, as can the sort order of data cases and which parameter or output they should show. It is possible, e.g., to sort parameter visualizations by output similarity, as is often done in visual parameter space analysis

[19,39]. We obtain the sort order by a 1D multidimensional scaling projection. Thus, the Gallery can show complex patterns.

5.4 Shepard Matrix (T1)

We want to give analysts a way to judge which parameter-output relations to investigate (T1). To this end, we use a Shepard diagram [16] showing all pairwise distances of data cases in a scatterplot. Each axis is the distance according to one measure. A diagonal line in a Shepard diagram thus means a perfect correspondence between two distance measures, and a dispersed Shepard scatterplot may be more interesting to investigate. We use the same color hue as in the Discrepancy Dendrogram for dots in a Shepard diagram, i.e., the further away from the diagonal, the more color hue is used. As the dataset usually has more than two parameters/outputs, we adapt the scatterplot matrix to Shepard diagrams to show all possible combinations (Fig. 5-D).

6 EVALUATION

We evaluated our visualizations heuristically and with expert interviews. The TU Wien pilot ethics board assessed our methods. Thus, our research adheres to the highest ethical standards. Specifically, our research questions were:

- (RQ1) Does our visualization design allow efficient and effective SA for SBSS parameters/outputs?

- (RQ2) Is our designed guidance effective?

- (RQ3) Does our visualization design transfer to other contexts than SBSS?

For RQ1 and RQ2, we conducted a heuristic evaluation with five visualization experts (Section 6.1). Two SBSS experts used our visualizations on their own data (Section 6.2), which also informs RQ1 and RQ2. Finally, for RQ3, we discussed visualizations with a microclimate simulation expert using an appropriate dataset. In this section we use two-letter shortcuts for people: Just letters indicate authors (e.g., NP) and a trailing number refers to participants (e.g., ME1).

Procedure.   All sessions started with a 30 minutes introduction where we explained our problem context and the visualizations independently from the available datasets in the prototype. The slides are

Component	Mean			Std.dev
Insight	6.26	1 ————•— 7		1.06
Confidence	5.11	1 ———•—— 7		2.03
Essence	5.32	1 ———•—— 7		1.45
Time	6.08	1 ————•— 7		0.91
Total	**5.83**	1 ———•+— 7		**1.50**

Table 1: *Results of the ICE-T evaluation with visualization experts. Responses were on a 7-point Likert scale. A total mean greater than five (small bar) is considered a success.*

Fig. 6: *Leaf visualizations for SBSS regionalization (R) and kernel (K) parameter, and output (W).*

available in the supplemental material. After the introduction, visualization experts continued with the questionnaire. The other experts used the prototype on a dataset and parameter settings they were familiar with. A semi-structured interview followed for all participants.

6.1 Visualization Experts (RQ1, RQ2)

We evaluated our visualization design heuristically with visualization experts according to the ICE-T method [65]. While a good design does not imply that the visualizations are effective, we think the inverse most likely holds (bad design → ineffective). Our chosen method is a good compromise between insights gained and the time requested from participants. We asked five participants (four Ph.D. students and one post-doc) from various universities to join our evaluation. We mostly met them over Zoom, and the sessions took around one hour each. According to ICE-T guidelines, five people are sufficient. Participants were free to use the prototype with various datasets on their own computers. They could always return to the visualization while filling out the ICE-T questionnaire. ICE-T responses are on a 7-point Likert scale. We asked them to share their thought process to understand their critique better.

Table 1 holds the results of these questionnaires, split by ICE-T component. The complete responses are available as supplemental material. Wall et al. [65] state that a visualization design is successful when the mean score exceeds five, which we clearly achieved with an overall mean of 5.83. Our visualization's worst-scoring component (mean 5.11) is Confidence, which is also the one with the highest standard deviation. While participants agreed that we use "meaningful and accurate visual encodings" (question Q18 in the ICE-T questionnaire) and "avoid misleading representations" (Q19), they mostly disagreed that our visualization "promotes understanding beyond individual data cases" (Q20) or highlights data quality issues (Q21). It would take some effort to detect duplicate or invalid data cases in our visualization, but that was a conscious design choice. The second-worst component is Essence, which also has the second-highest standard deviation, indicating disagreement between participants. In fact, the two most contested questions here were whether the visualization "facilitates generalizations and extrapolations" (Q16) or "helps understand how variables relate in order to accomplish different analytic tasks" (Q17). Low ratings in the former were, e.g., because the Discrepancy Dendrogram assesses individual clusters but does not indicate differences between elements. This issue could be tackled in the future by specially crafted comparison visualizations. In the latter question, some participants focused on the "*different* analytic tasks" and argued that our visualization does not fulfill this criterion due to its singular focus.

On the other hand, participants rated the Insight and Time components very well. Two questions of the former seemed somewhat controversial, as they are associated with higher standard deviations (1.79 and 1.64). One participant *somewhat disagreed* that the visualization "facilitates perceiving relationships in the data" (Q2). Their reasoning was as follows. We show data cases as leaves in the Discrepancy Dendrogram and also in a gallery to the side. However, all data cases are separate visualizations, so it would be akin to showing individual bars instead of a histogram. However, they also realized that this was not a goal of our visualization design. The other contested question was whether the visualization "helps identify unusual or unexpected, yet valid, data characteristics" (Q5). One participant *somewhat*
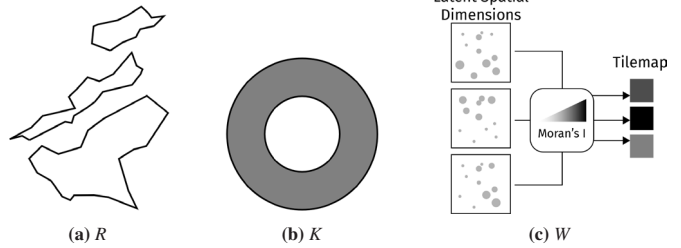
disagreed, mentioning that data cases with unusual or unexpected features would be hard to spot if the distance metrics would not consider these. We do not see this as an issue because the chosen dissimilarity metrics might as well measure local differences.

6.2 SBSS (RQ1, RQ2)

Two experts (SE1 and SE2) in statistics and SBSS, who were not part of the design process, used our visualizations on familiar datasets. They were recruited from the authors' professional network as they were required to have knowledge of SBSS. They both hold a Ph.D. in statistics and published on spatial data analysis. Sessions took around 2 hours. We guided them in the process as much as necessary, e.g., formulated possible analysis goals and answered any questions they had. After that, we continued with a semi-structured interview, inquiring about their confidence in findings, possible insights, and how these relate to prior expectations.

Datasets and Parameter Settings. The experts used two spatial datasets. SE2 worked on the *Colorado* dataset, which is a geochemical survey of 960 locations and 27 variables in Colorado, USA. Both SE2 and NP contributed parameter settings to investigate, as was agreed upon prior to the interview. SE2 provided an R script to obtain regionalizations (10 slices along four directions) and kernels (0–200 km radii). NP added regionalizations obtained in a prior study [51]. SE1, on the other hand, worked on the meteorological *Veneto* dataset, which consists of 72 locations and 7 variables in Veneto, Italy. Parameter settings were obtained in a pilot session by SE1 and NP together using an existing prototype [51]. We computed outputs for a full factorial of selected regionalizations and kernels for both datasets. In total, 42 settings were available for the *Veneto* and 48 for the *Colorado* dataset.

Dissimilarity Measures. We chose appropriate functions together with our collaborators. For the unmixing matrix W, we use the MD-Index [27], a specialized comparison tool for unmixing matrices. For two kernels (K), we compute the difference of their so-called Spatial Kernel Matrix [45]. We compare two regionalizations (R) by counting location pairs for which the region assignment is not identical.

Leaf Visualizations. We used three visualizations to represent R, K, and W (Fig. 6). For R, we showed as multiple polygons representing the concave hull of regions (Fig. 6a). In tooltips, these were integrated into interactive Leaflet maps. For K, we showed concentric circles representing the ring size (Fig. 6b), also overlaying them to the spatial context with Leaflet in tooltips (Fig. 5-E). We visualized W as a tilemap where each tile represented one latent dimension (Fig. 6c). Tiles were colored in a univariate continuous gray color map showing Moran's I [44], a measure for spatial autocorrelation. High values of that measure point to large-scale spatial patterns, which analysts might find easier to interpret. Tiles were ordered as the SBSS algorithm returned respective dimensions. Tooltips of tiles showed static plots of latent dimensions overlayed on OpenStreetMap.

SE1. NP guided SE1 to focus on SA because other than SE2, SE1 initially focused more on the spatial relationship between regionalizations (R) and locations in the dataset. Regarding SA, SE1 was interested in the influence of the kernel (K) parameter on the output.
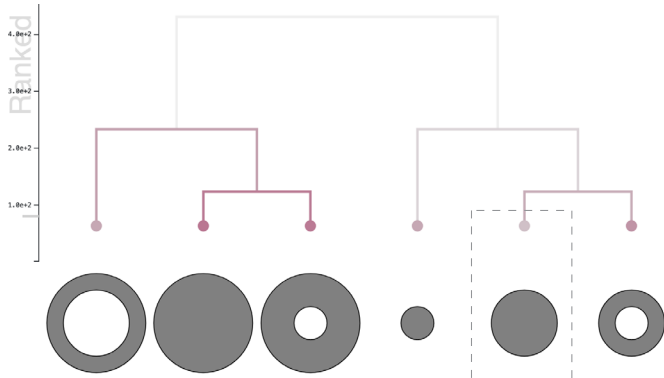
Fig. 7: ●KW *Discrepancy Dendrogram (average linkage with some clusters collapsed, cropped) for the* Veneto *dataset. The dashed box marks the most stable kernel setting identified by SE1.*
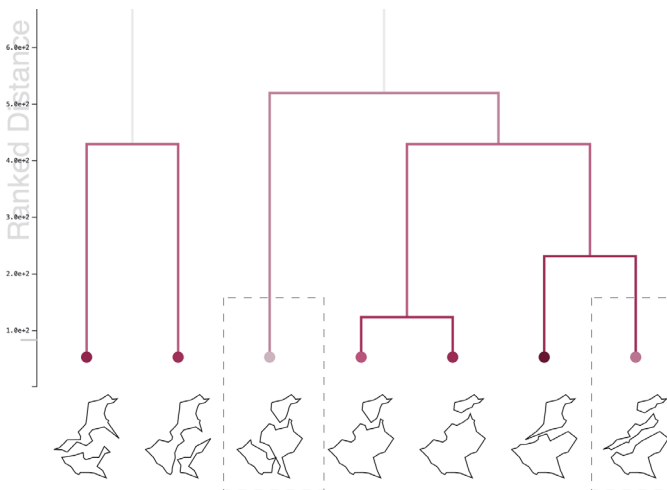

Fig. 8: ●RW *Discrepancy Dendrogram (complete linkage with some clusters collapsed, cropped) for the* Veneto *dataset. Dashed boxes mark the most stable regionalization settings identified by SE1.*

Fig. 9: ●WR *Discrepancy Dendrogram for the* Colorado *dataset. Blue lines mark data cases with variation in W despite similar R. Closer inspection revealed that the presence of a hole of at least 100 km size in associated K settings distinguishes these cases (red arrow).*

NP pointed SE1 to a ●KW dendrogram configuration and explained that the red color points to sensitive parameter settings. Almost all K clusters were colored red. As they were wider in W, it indicated that the other parameter (R) exerts more influence on the output than K. SE1 switched to average linkage to account for any outliers that may skew the complete linkage criterion. Using this view (Fig. 7), they found that K with a radius 0–60 km was the least red compared to others. Hence, this setting was most stable regarding the choice of R, with K=0–30 km a close second. SE1 explained that most locations in the dataset are within 75 km, so a kernel up to 60 km will likely capture most of the spatial dependency structure. SE1 also observed kernels up to 90 km radius (three big circles on the left in Fig. 7) generally showing wider clusters in W than the smaller kernels due to their stronger red color. SE1 concluded that two levels of spatial variability exist in the dataset.

Next, a ●RW configuration of the Discrepancy Dendrogram, was investigated (Fig. 8). Here, clusters of the 3-partitions chosen by altitude and precipitation were the most stable, meaning they were more independent of the choice of K than other partitions. This fact was initially surprising to SE1. However, SE1 reconciled it such that the two partitions are similar in that they both separate Veneto's mountainous and flat region. However, another separation in the plane seemed necessary. 2-partitions with just the mountain-flat separation were linked to wider clusters in the output, thus more sensitive to the choice of K.

In the interview, SE1 voiced many positive sentiments. They found the visualization "not difficult" to understand, and the construction of the Discrepancy Dendrogram was logical and easy to follow. SE1 liked
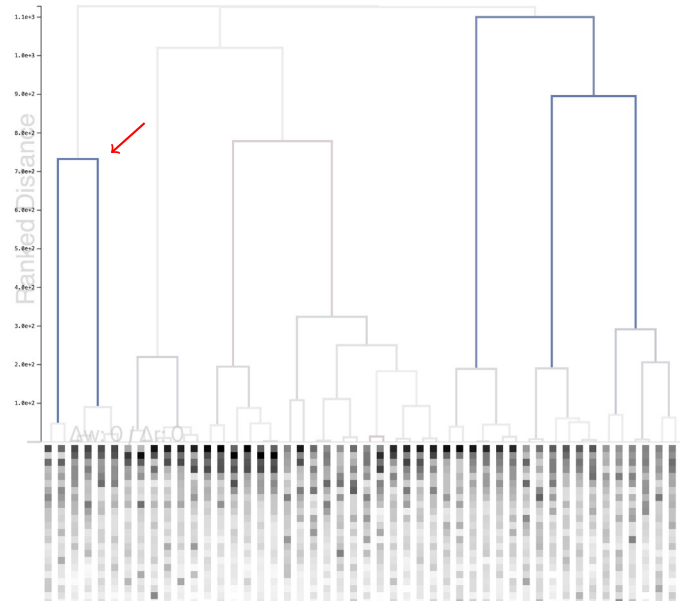
the interactive maps and that "you can analyze the data by looking at different aspects in different ways." "Half of the work is made [with this tool]," so analysis time is saved compared to the "classical methods." In sum, SE1 found our visualizations "help evaluate the parameters" and identified an interesting parameter subspace to consider for future analysis: Smaller K in higher resolutions, as 0–60 km kernels were found to be most stable. SE1 could see our visualizations working for people who are "not completely expert [sic]" in SA. Based on these sentiments, we think RQ1 and RQ2 can be answered positively.

SE1 thought that the Discrepancy Dendrogram is not very easy to interpret but also attributed this to lack of familiarity with our approach and visualizations. Other than SE2, SE1 did not confirm or challenge expectations about parameter importance/sensitivity, as they find it necessary to compare multiple datasets before concluding anything. In the same spirit, SE1 remarked that a proper data analysis pipeline uses multiple complementing methods, prohibiting sweeping conclusions using our visualizations alone.

SE2. First, SE2 focused on a ●WK Discrepancy Dendrogram. SE2 observed many red lines and asked if it was correct to conclude that those outputs are less sensitive to kernel (K) choice, which it was. SE2 was then interested in regionalizations (R) and switched to ●WR. There, SE2 observed a very salient pattern (Fig. 9): Most of the dendrogram was gray, indicating that cluster diameters match well between W and R. Thus, R is an important parameter for the *Colorado* dataset. A few clusters showed blue highlights, indicating clusters of sensitive R parameter settings. SE2 looked at one of the clusters (red arrow in Fig. 9), saw the same R combined with various K, and considered the local dendrogram shape. SE2 concluded that two groups of W exist for this R setting (10 horizontal slices): One using very "un-local" kernels (K) with a 100 km hole and another group containing the dataset's remaining K settings. Hence, the choice of K matters a lot for this particular R setting. Other salient blue patterns were visible on the dendrogram's right side but not investigated by SE2. SE2 then returned to the ●WK configuration, but set the leaves to show R and investigated how these parameter settings were distributed in the dendrogram. They observed mostly neat clusters (by SBSS output W) of 6 data cases and identical R in each cluster, which was another hint that R is the more important parameter.

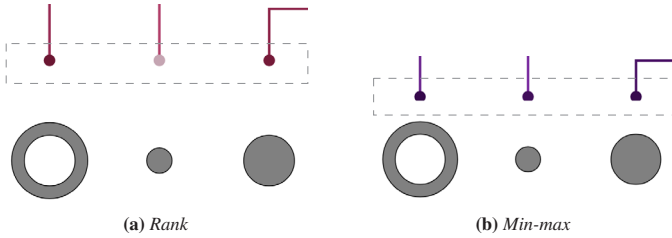NP suggested looking at a parameter-focused dendrogram, after

(a) *Rank*   **(b)** *Min-max*

Fig. 10: *Rank and min-max distance normalization highlight different relations. Note the enclosed circles' color. With ranked distance (a), the small kernel is shown as most stable of the three (lighter color). Min-max distances (b) show that the absolute difference is low (saturated colors). Our visualizations offer more precise visual encodings in addition to color hue for such comparisons (see, e.g., Section 5.1).*

which SE2 changed it to ●KW. Here SE2 suggested that one K setting (0–50 km radius) is much more stable than the others due to its lighter color and wrongly concluded that R choice matters less for that. While their first assessment (more stable than others) was correct, the second part did not consider the magnitude of the cluster diameter difference in W. If SE2 would have used min-max normalized distances Discrepancy Dendrogram (Fig. 10), they would have seen that also for that K, the cluster diameter difference in W was very high in absolute terms. Finally, SE2 also considered ●RW to investigate the stability of R. Here, a completely different picture than for ●KW emerged: The lines touching dendrogram leaves were gray instead of red, thus suggesting that less variation happens within R settings than between them. This image was consistent with ●WR, underlining the importance of the regionalization (R) parameter even more.

In the interview afterward, SE2 offered mainly positive comments. The visualization is "very intuitive to use", and it "speeds up analysis because one can see all parameter combinations at once." It "does exactly what it's supposed to do" because "the color pointed [them] to [data cases] where there was something going on" and, therefore, SE2 is "very confident in results obtained with this tool, I don't doubt it." They mentioned that "many observations would not have been possible without this [visualization]" and that, therefore, it can be a qualitative complementary to the quantitative methods they use in their research. E.g., as a more systematic replacement for the trial & error they do now. SE2 confirmed their suspicion that R is the more important parameter using our visualizations. Again, we think these sentiments strongly support both RQ1 and RQ2.

SE2 also mentioned that the Discrepancy Dendrogram was not particularly easy to understand. I.e., the syntax, so to say, was clear (red, gray, and blue pointing in the direction of wider clusters), but translating that into actionable steps in parameter analysis was difficult. SE2 expects that this effect will get smaller with more familiarity with the visualizations. Finally, SE2 admitted they mostly looked at the tilemap in the leaves to judge W similarity. Since tiles contain a summary (Moran's I) of actual maps, the visualization may be misleading. A possible remedy could be a glyph design incorporating a derived feature and map similarity.

6.3   Microclimate Simulations (RQ3)

To demonstrate that the approach used in our prototype is transferable to other problem contexts (the goal of design studies [60]), we applied it to microclimate simulation results [64]. Such simulation models predict meteorological variables (e.g., air temperature or humidity) in a very small area, typically for a single street or a building. Microclimate simulations are critical nowadays as the climate crisis pressures cities and real estate developers to adapt to changing climate conditions. Usually, stakeholders, like city planners and architects, use existing simulation models and do not develop them themselves. Hence, parameter space analysis so far was mostly done by studying derived features (e.g., maximum temperature) with respect to grid size, often carried out with visual inspection, and computing relations (e.g., correlation)

between individual variables. Analysts have certain expectations about parameter relations. These come partially from known model limitations (e.g., the model does not perform well in extreme conditions) and partially from the modeled physical reality (e.g., the humidity of cold vs. hot air or wind chill effects).

In the conducted session, two authors (NP, JS) of this paper met with a microclimate simulation expert ME1, who has a Ph.D. in civil engineering and was recruited from the authors' professional network. JS controlled the prototype and suggested findings that ME1 assessed, while NP took notes.

Dataset and Parameter Settings.   The experts' use case was to analyze the climatic conditions around a potential building (available as a 3D model) in several cities, seasons and meteorological conditions (called a *scenario*) to find the best location. The tested cities were Vienna, Helsinki, and Gothenburg in various seasons. ME1 computed a dataset containing 12 parameter settings and respective outputs. The low number of data cases follows the simulation model's computational demands as a single run takes several minutes to a couple of hours. The four outputs were wind speed (O_W), temperature on the surface (O_S) and in the air (O_A), and humidity (O_Q) at 6 am after a simulated interval of 24 h. The output values are spatially distributed on a grid. Parameters of the model were air temperature (P_A) and humidity (P_Q) as time series over 24 h, and wind speed and direction (P_W). We agreed to use Euclidean distance to measure similarity.

Leaf Visualizations.   Three visualizations were used to show the model's parameters and outputs, both as leaf and tooltip visualizations. For the spatially distributed outputs (O_W, O_S, O_A, O_Q), we used heatmaps with univariate color scales of varying hue. Time series (P_A, P_Q) were shown as line charts. Wind speed and direction were shown as arrows, with speed as length and direction as rotation.

ME1.   In the beginning, we asked ME1 about the most important output in the dataset, which ME1 answered to be the surface temperature (O_S). The goal was to identify a scenario where O_S is both low and stable so as to not be a threat to the human circulatory system. At the same time, general parameter-output relations were of interest. To achieve these tasks, JS set up a Discrepancy Dendrogram with O_S as primary distance and cycled through parameters as alternative distance. We started with a ●$O_S P_W$ configuration, i.e., compared surface temperature output to the wind (direction and speed) parameter. The dendrogram showed many red lines, indicating wider clusters in P_W and thus generally no strong association between P_W and O_S. JS changed from ranked to min-max distances to see if the pattern persists when the magnitude is considered, which it did. This relation was expected for ME1. We also observed an O_S outlier with temperatures up to 36 °C, which seemed unexpected (red arrow in Fig. 11). ME1 recalled that "the simulation model in question aims to capture extreme conditions in summer, like overheating, and there is really the question of how it performs in other conditions and different climates." ME1 concluded that the outlier might be a failure case of the model. Later analysis showed that the presumed model failure was related to extreme temperatures in the P_A parameter. However, it became clear that wind alone "does not really make a difference" when it comes to surface temperature.

JS then switched to other parameters. Air temperature (P_A) was strongly correlated, as expected (Fig. 12). A similar picture emerged for humidity (P_Q), except for a group of three scenarios (Fig. 11-A) that arrived at similar O_S with significantly varying P_Q settings. ME1 noted that to determine the actual impact of P_Q here, one has to account for the different seasons and cities. This observation was noted as something to investigate later, as, at the time, season and city were not displayed in the prototype. JS then proceeded to compare other outputs with parameters. Our visualizations showed, and ME1 confirmed, the known relationship between humidity and air temperature. The next interesting observation came from the connection between wind and temperature. Wind parameter (P_W) and output (O_W) were not strongly correlated, and air temperature (P_A) was identified as another relevant factor (red arrow in Fig. 12a). Regarding how temperature could influence wind, ME1 mentioned horizontal and vertical mixing effects but that those would be smaller than the wind-to-temperature
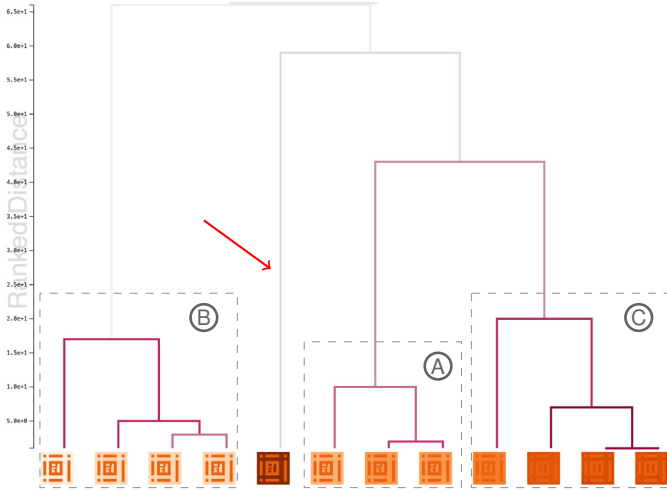
Fig. 11: ●$O_S P_W$ *Discrepancy Dendrogram used for microclimate simulations. Red lines indicate wider clusters in P_W and thus little influence of that parameter on O_S. The red arrow marks a data case suspected to be a model failure. Data cases enclosed by A were also investigated with a* ●$O_S P_Q$ *configuration. Data cases enclosed by A–C were considered for the final location choice.*
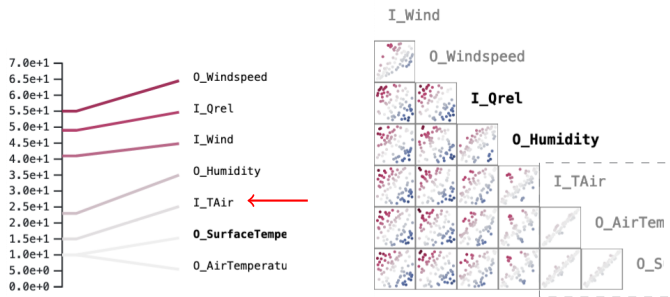
(a) *Subset Sensitivity View (Section 5.3) of data cases in Fig. 11-A.*

(b) *Shepard Matrix (Section 5.4) of microclimate dataset.*

Fig. 12: *Subset Sensitivity View (a) of cluster in Fig. 11 shows that air temperature P_A is be the driving parameter for surface temperature O_S, as expected. Shepard diagrams of air-related parameters/outputs in Shepard Matrix (b) show that this relation holds for all data cases.*

effects. ME1 speculated that some correlations might come from the used 3D grid slices being on pedestrian level (1.8 m) while surface temperature is only valid for the slice at 0 m.

Asked about disadvantages or improvements, ME1 mentioned not picking a winning location for their use case because the city and season were missing in the visualization. NP checked together with JS later. Of all the surface temperature (O_S) clusters (A–C in Fig. 11), O_S was least sensitive to air temperature (P_A) in the three scenarios enclosed by A. They belonged to Helsinki (2/3) and Vienna (1). Thus, Helsinki could be identified as the most suitable choice due to the more constant surface temperature. This choice is also consistent with the latest report of the Intergovernmental Panel on Climate Change [23], which predicts more stable mean temperature for Northern than Central Europe.

To summarize, we could apply our visualizations in a domain they were not originally designed for in the following way. We could find a suitable location for the building, which was the main goal for ME1, thus solving this domain's SA task. ME1 could reconcile visualization images with domain knowledge and find interesting relations to investigate in the future, like the humidity parameter's impact. We see this session as evidence to support RQ3, that our visualizations can be transferred to other contexts.

7 LIMITATIONS

As we rely on cluster diameters, the particular choice of nested partitions will greatly influence our sensitivity index, visualization image and, ultimately, the analysis outcome. The partitions are in turn influenced by the dataset, dissimilarity measure, clustering algorithm, and its parameters. We took care to select reasonable defaults, but they may not work for every situation. While it may be a demanding task, truthful clusterings can be obtained (cf. Section 2.3) and the particular groupings could be modifiable by the analyst. Another consequence of relying on relative cluster diameter differences for SA is that the sensitivity index likely changes when new data is considered, thus the visualization image may be unstable with regard to additions to the underlying dataset. While that may seem like a big constraint, we argue that the same is true for visual SA of multivariate parameters: If they are sampled too coarsely or in too narrow intervals, then the analysis outcome may change a lot when the previously excluded parameter space is considered.

Our approach (Section 4.4) roughly corresponds to a one-at-a-time sensitivity index, i.e., a local method. Saltelli et al. [56] argue that local methods are only appropriate when the model under investigation is demonstrably linear. We did not confirm whether SBSS (Section 6.2) or the microclimate simulations (Section 6.3) are linear models. However, we do not see this as an issue for two reasons. First, local indices in the SA literature make precise quantitative statements for the whole parameter. As we defined our index only for subsets of data cases, it does not do that. Second, we developed the index for visual guidance in an interactive visualization. As all relevant data cases are visible in detail at any time, the analyst may consider much more context and existing domain knowledge than they would when interpreting only a single number, as demonstrated in Section 6.

8 DISCUSSION AND CONCLUSION

Based on requirements and observations in the context of SBSS, we developed a data type agnostic approach to visual SA. It only requires dissimilarity measures and thus works for complex parameters and outputs alike. The core innovation is measuring variation in parameter settings and outputs by cluster diameters. SA then becomes possible by looking at the difference of the same cluster's diameter in parameter and output space. Evaluation participants expressed high confidence in our visualizations. Future work may improve this paper's proposal by accounting for noise or simultaneously supporting multiple parameters.

The Discrepancy Dendrogram and supporting visualizations (Section 5) were also received very well by evaluation participants, especially considering the task complexity and short training time (around 30 minutes). The construction of the Discrepancy Dendrogram was logical for all participants, and the prototype provided sufficient interactions and levels of detail. The successful heuristic evaluation (Section 6.1) further supports this evidence. SBSS and microclimate simulation experts could confirm suspected or expected parameter-output relations with our visualizations, while mentioning the need to familiarize themselves more with our approach. E.g., the regionalization parameter R is more important for SBSS than the kernel configuration K (suspected by SE2), or that surface temperature mainly depends on air temperature (expected by ME1). Further, they could make high-level decisions (building location, ME1), find new relevant parameter subspaces (smaller kernels, SE1), or just obtain interesting observations (kernels with holes, SE1 and SE2). Considering the utility of the Discrepancy Dendrogram it will also be interesting to apply our approach to other visualization idioms, e.g., to DR scatterplots (Section 2.3).

We noted, e.g., during introductory explanations, that some participants found it mentally demanding to reason simultaneously about 1) groups of elements instead of single elements and 2) two distances within a group of elements. This issue is, to some extent, inherent to the problem we want to solve. On the other hand, we think rephrasing SA or finding visual representations so that analysts can reason about single elements instead of groups has much simplification potential. Achieving this would allow even more powerful SA visualizations potentially applicable to many contexts (Section 6.3).

ACKNOWLEDGMENTS

REFERENCES

[1] A. M. Abuzuraiq and H. Erhan. The Many Faces of Similarity. In *RE: Anthropocene, Proc. CAADRIA*, vol. 1, pp. 485–494. Assiciation for Computer-Aided Architectural Design Research in Asia, Hong Kong, 2020. doi: 10.52842/conf.caadria.2020.1.485 2

[2] F. Bachoc, M. G. Genton, K. Nordhausen, A. Ruiz-Gazen, and J. Virta. Spatial Blind Source Separation. *Biometrika*, 107(3):627–646, 2020. doi: 10.1093/biomet/asz079 1

[3] Z. Bar-Joseph, D. K. Gifford, and T. S. Jaakkola. Fast Optimal Leaf Ordering for Hierarchical Clustering. *Bioinformatics*, 17(Suppl 1):S22–S29, 2001. doi: 10.1093/bioinformatics/17.suppl_1.S22 4

[4] M. Beham, W. Herzner, M. E. Gröller, and J. Kehrer. Cupid: Cluster-Based Exploration of Geometry Generators with Parallel Coordinates and Radial Trees. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1693–1702, 2014. doi: 10.1109/tvcg.2014.2346626 2

[5] S. Bergner, M. Sedlmair, T. Moller, S. N. Abdolyousefi, and A. Saad. ParaGlide: Interactive Parameter Space Partitioning for Computer Simulations. *IEEE Transactions on Visualization and Computer Graphics*, 19(9):1499–1512, 2013. doi: 10.1109/tvcg.2013.61 2

[6] R. Blanch, R. Dautriche, and G. Bisson. Dendrogramix: A Hybrid Tree-Matrix Visualization Technique to Support Interactive Exploration of Dendrograms. In *Proc. PacificVis*, pp. 31–38. IEEE, New York, 2015. doi: 10.1109/pacificvis.2015.7156353 3

[7] E. Borgonovo and E. Plischke. Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, 248(3):869–887, 2016. doi: 10.1016/j.ejor.2015.06.032 2

[8] S. Bruckner and T. Möller. Result-Driven Exploration of Simulation Parameter Spaces for Visual Effects Design. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1468–1476, 2010. doi: 10.1109/tvcg.2010.190 2

[9] D. G. Cacuci and M. Ionescu-Bujor. A Comparative Review of Sensitivity and Uncertainty Analysis of Large-Scale Systems—II: Statistical Methods. *Nuclear Science and Engineering*, 147(3):204–217, 2004. doi: 10.13182/04-54cr 2

[10] R. J. G. B. Campello, D. Moulavi, and J. Sander. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*, vol. 7819 of *Lecture Notes in Artificial Intelligence*, pp. 160–172. Springer, Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-37456-2_14 3

[11] N. Cao, D. Gotz, J. Sun, and H. Qu. DICON: Interactive Visual Analysis of Multidimensional Clusters. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2581–2590, 2011. doi: 10.1109/tvcg.2011.188 3

[12] M. Cavallo and Ç. Demiralp. Clustrophile 2: Guided Visual Clustering Analysis. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):267–276, 2019. doi: 10.1109/tvcg.2018.2864477 3

[13] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, H. Schulz, M. Streit, and C. Tominski. Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, 2017. doi: 10.1109/tvcg.2016.2598468 3

[14] L. Cibulski, T. May, J. Schmidt, and J. Kohlhammer. COMPO*SED: Composite Parallel Coordinates for Co-Dependent Multi-Attribute Choices. *IEEE Transactions on Visualization and Computer Graphics*, 2023. in press. doi: 10.1109/tvcg.2022.3180899 2

[15] L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer. PAVED: Pareto Front Visualization for Engineering Design. *Computer Graphics Forum*, 39(3):405–416, 2020. doi: 10.1111/cgf.13990 2

[16] J. de Leeuw. Shepard Diagram. In B. S. Everitt and D. C. Howell, eds., *Encyclopedia of Statistics in Behavioral Science*, p. bsa612. Wiley, Chichester, 2005. doi: 10.1002/0470013192.bsa612 5

[17] R. Desai, F. Anderson, J. Matejka, S. Coros, J. McCann, G. Fitzmaurice, and T. Grossman. Geppetto: Enabling Semantic Design of Expressive Robot Behaviors. In *Proc. CHI*, pp. 369:1–369:14. ACM, New York, 2019. doi: 10.1145/3290605.3300599 2

[18] H. Doraiswamy, N. Ferreira, M. Lage, H. Vo, L. Wilson, H. Werner, M. Park, and C. Silva. Topology-Based Catalogue Exploration Framework for Identifying View-Enhanced Tower Designs. *ACM Transactions on Graphics*, 34(6):230:1–230:13, 2015. doi: 10.1145/2816795.2818134 2

[19] C. Eichner, H. Schumann, and C. Tominski. Making Parameter Dependencies of Time-Series Segmentation Visually Understandable. *Computer Graphics Forum*, 39(1):607–622, 2020. doi: 10.1111/cgf.13894 2, 5

[20] T. Galili. Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering. *Bioinformatics*, 31(22):3718–3720, 2015. doi: 10.1093/bioinformatics/btv428 3

[21] T. Günther and H. Theisel. Inertial Steady 2D Vector Field Topology. *Computer Graphics Forum*, 35(2):455–466, 2016. doi: 10.1111/cgf.12846 2

[22] Z. Guo, M. O. Ward, E. A. Rundensteiner, and C. Ruiz. Pointwise Local Pattern Exploration for Sensitivity Analysis. In *Proc. VAST*, pp. 131–140. IEEE, New York, 2011. doi: 10.1109/vast.2011.6102450 2

[23] J. M. Gutiérrez, R. G. Jones, G. T. Narisma, L. M. Alves, M. Amjad, I. V. Gorodetskaya, M. Grose, N. A. B. Klutse, S. Krakovska, J. Li, D. Martínez-Castro, L. O. Mearns, S. H. Mernild, T. Ngo-Duc, B. van den Hurk, and J.-H. Yoon. Atlas. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, eds., *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1927–2058. Cambridge University Press, Cambridge, New York, 2021. doi: 10.1017/9781009157896.021 9

[24] D. M. Hamby. A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models. *Environmental Monitoring and Assessment*, 32(2):135–154, 1994. doi: 10.1007/bf00547132 2

[25] S. Hazarika, H. Li, K.-C. Wang, H.-W. Shen, and C.-S. Chou. NNVA: Neural Network Assisted Visual Analysis of Yeast Cell Polarization Simulation. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):34–44, 2020. doi: 10.1109/tvcg.2019.2934591 2

[26] W. He, J. Wang, H. Guo, K.-C. Wang, H.-W. Shen, M. Raj, Y. S. G. Nashed, and T. Peterka. InSituNet: Deep Image Synthesis for Parameter Space Exploration of Ensemble Simulations. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):23–33, 2020. doi: 10.1109/tvcg.2019.2934312 2

[27] P. Ilmonen, K. Nordhausen, H. Oja, and E. Ollila. A New Performance Index for ICA: Properties, Computation and Asymptotic Analysis. In *Latent Variable Analysis and Signal Separation*, Lecture Notes in Computer Science, pp. 229–236. Springer, Berlin, Heidelberg, 2010. doi: 10.1007/978-3-642-15995-4_29 6

[28] M. Ionescu-Bujor and D. G. Cacuci. A Comparative Review of Sensitivity and Uncertainty Analysis of Large-Scale Systems—I: Deterministic Methods. *Nuclear Science and Engineering*, 147(3):189–203, 2004. doi: 10.13182/nse03-105cr 2

[29] B. Iooss and P. Lemaître. A Review on Global Sensitivity Analysis Methods. In *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, vol. 59 of *Operations Research/Computer Science Interfaces Series*, pp. 101–122. Springer, Boston, 2015. doi: 10.1007/978-1-4899-7547-8_5 2

[30] T. Jankun-Kelly and Kwan-Liu Ma. A Spreadsheet Interface for Visualization Exploration. In *Proc. VIS (Cat. No. 00CH37145)*, pp. 69–76, 2000. doi: 10.1109/visual.2000.885678 2

[31] H. Jeon, H.-K. Ko, J. Jo, Y. Kim, and J. Seo. Measuring and Explaining the Inter-Cluster Reliability of Multidimensional Projections. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):551–561, 2022. doi: 10.1109/tvcg.2021.3114833 3

[32] J. Knittel, A. Lalama, S. Koch, and T. Ertl. Visual Neural Decomposition to Explain Multivariate Data Sets. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1374–1384, 2021. doi: 10.1109/tvcg.2020.3030420 2

[33] Y. Koyama, D. Sakamoto, and T. Igarashi. Crowd-Powered Parameter Analysis for Visual Design Exploration. In *Proc. UIST*, pp. 65–74. ACM, New York, 2014. doi: 10.1145/2642918.2647386 2

[34] B. C. Kwon, B. Eysenbach, J. Verma, K. Ng, C. De Filippi, W. F. Stewart, and A. Perer. Clustervision: Visual Supervision of Unsupervised Clustering. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):142–151, 2018. doi: 10.1109/tvcg.2017.2745085 3

[35] A. Ligmann-Zielinska. Spatially-Explicit Sensitivity Analysis of an Agent-Based Model of Land Use Change. *International Journal of Geographical Information Science*, 27(9):1764–1781, 2013. doi: 10.1080/13658816.2013.782613 2

[36] L. Lilburne and S. Tarantola. Sensitivity Analysis of Spatial Models. *International Journal of Geographical Information Science*, 23(2):151–168, 2009. doi: 10.1080/13658810802094995 2

[37] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu. Understanding of Internal Clustering Validation Measures. In *Proc. ICDM*, pp. 911–916. IEEE, New York, 2010. doi: 10.1109/icdm.2010.35 4

[38] M. Luboschik, M. Röhlig, A. T. Bittig, N. Andrienko, H. Schumann, and C. Tominski. Feature-Driven Visual Analytics of Chaotic Parameter-Dependent Movement. *Computer Graphics Forum*, 34(3):421–430, 2015. doi: 10.1111/cgf.12654 2

[39] M. Luboschik, S. Rybacki, F. Haack, and H.-J. Schulz. Supporting the integrated visual analysis of input parameters and simulation trajectories. *Computers & Graphics*, 39:37–47, 2014. doi: 10.1016/j.cag.2013.09.004 2, 5

[40] S. L'Yi, B. Ko, D. Shin, Y.-J. Cho, J. Lee, B. Kim, and J. Seo. XCluSim: A Visual Analytics Tool for Interactively Comparing Multiple Clustering Results of Bioinformatics Data. *BMC Bioinformatics*, 16(11):S5:1–S5:15, 2015. doi: 10.1186/1471-2105-16-s11-s5 3

[41] J. Marks, P. Beardsley, B. Andalman, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, J. Seims, and S. Shieber. Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation. In *Proc. SIGGRAPH*, pp. 389–400. ACM, New York, 1997. doi: 10.1145/258734.258887 2

[42] A. Marrel, B. Iooss, M. Jullien, B. Laurent, and E. Volkova. Global Sensitivity Analysis for Models with Spatially Dependent Outputs. *Environmetrics*, 22(3):383–397, 2011. doi: 10.1002/env.1071 2

[43] K. Matković, H. Abraham, M. Jelović, and H. Hauser. Quantitative Externalization of Visual Data Analysis Results Using Local Regression Models. In *Machine Learning and Knowledge Extraction*, vol. 10410 of *Lecture Notes in Computer Science*, pp. 199–218. Springer, Berlin, Heidelberg, 2017. doi: 10.1007/978-3-319-66808-6_14 2

[44] P. A. P. Moran. Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2):17–23, 1950. doi: 10.2307/2332142 6

[45] C. Muehlmann, F. Bachoc, and K. Nordhausen. Blind Source Separation for Non-Stationary Random Fields. *Spatial Statistics*, 47:100574:1–100574:24, 2022. doi: 10.1016/j.spasta.2021.100574 1, 2, 6

[46] K. Nordhausen, H. Oja, P. Filzmoser, and C. Reimann. Blind Source Separation for Spatial Compositional Data. *Mathematical Geosciences*, 47(7):753–770, 2015. doi: 10.1007/s11004-014-9559-5 1, 2

[47] D. Orban, D. F. Keefe, A. Biswas, J. Ahrens, and D. Rogers. Drag and Track: A Direct Manipulation Interface for Contextualizing Data Instances within a Continuous Parameter Space. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):256–266, 2019. doi: 10.1109/tvcg.2018.2865051 2

[48] S. Pajer, M. Streit, T. Torsney-Weir, F. Spechtenhauser, T. Möller, and H. Piringer. WeightLifter: Visual Weight Space Exploration for Multi-Criteria Decision Making. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):611–620, 2017. doi: 10.1109/tvcg.2016.2598589 2

[49] N. Piccolotto, M. Bögl, T. Gschwandtner, C. Muehlmann, K. Nordhausen, P. Filzmoser, and S. Miksch. TBSSvis: Visual Analytics for Temporal Blind Source Separation. *Visual Informatics*, 6(4):51–66, 2022. doi: 10.1016/j.visinf.2022.10.002 2, 3, 4

[50] N. Piccolotto, M. Bögl, and S. Miksch. Visual Parameter Space Exploration in Time and Space. *Computer Graphics Forum*, 2023. in press. doi: 10.1111/cgf.14785 2, 3

[51] N. Piccolotto, M. Bögl, C. Muehlmann, K. Nordhausen, P. Filzmoser, and S. Miksch. Visual Parameter Selection for Spatial Blind Source Separation. *Computer Graphics Forum*, 41(3):157–168, 2022. doi: 10.1111/cgf14530 1, 2, 3, 6

[52] J. Raimbault, C. Cottineau, M. L. Texier, F. L. Néchet, and R. Reuillon. Space Matters: Extending Sensitivity Analysis to Initial Spatial Conditions in Geosimulation Models. *Journal of Artificial Societies and Social Simulation*, 22(4):10:1–10:22, 2019. doi: 10.18564/jasss.4136 2

[53] F. J. Rohlf. 12 Single-link Clustering Algorithms. In P. R. Krishnaiah and L. N. Kanal, eds., *Classification Pattern Recognition and Reduction of Dimensionality*, vol. 2 of *Handbook of Statistics*, pp. 267–284. Elsevier, 1982. doi: 10.1016/s0169-7161(82)02015-x 3

[54] I. B. Rojo, M. Gross, and T. Günther. Visualizing the Phase Space of Heterogeneous Inertial Particles in 2D Flows. *Computer Graphics Forum*, 37(3):289–300, 2018. doi: 10.1111/cgf.13420 2

[55] A. Saltelli. Sensitivity Analysis for Importance Assessment. *Risk Analysis*, 22(3):579–590, 2002. doi: 10.1111/0272-4332.00040 2

[56] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, and Q. Wu. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, 114:29–39, 2019. doi: 10.1016/j.envsoft.2019.01.012 2, 9

[57] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis: The Primer*. Wiley, Chichester, 2007. doi: 10.1002/9780470725184 2

[58] A. Schulz, J. Xu, B. Zhu, C. Zheng, E. Grinspun, and W. Matusik. Interactive Design Space Exploration and Optimization for CAD Models. *ACM Transactions on Graphics*, 36(4):157:1–157:14, 2017. doi: 10.1145/3072959.3073688 2

[59] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller. Visual Parameter Space Analysis: A Conceptual Framework. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2161–2170, 2014. doi: 10.1109/tvcg.2014.2346321 1, 2

[60] M. Sedlmair, M. Meyer, and T. Munzner. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012. doi: 10.1109/tvcg.2012.213 8

[61] J. Seo and B. Shneiderman. Interactively Exploring Hierarchical Clustering Results [Gene Identification]. *Computer*, 35(7):80–86, 2002. doi: 10.1109/mc.2002.1016905 3

[62] I. M. Sobol'. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe modelirovanie*, 2(1):112–118, 1990. 2

[63] C. A. Steed, D. M. Ricciuto, G. Shipman, B. Smith, P. E. Thornton, D. Wang, X. Shi, and D. N. Williams. Big Data Visual Analytics for Exploratory Earth System Simulation Analysis. *Computers & Geosciences*, 61:71–82, 2013. doi: 10.1016/j.cageo.2013.07.025 2

[64] M. Vuckovic, J. Schmidt, T. Ortner, and D. Cornel. Combining 2D and 3D Visualization with Visual Analytics in the Environmental Domain. *Information*, 13(1):7:1–7:21, 2022. doi: 10.3390/info13010007 2, 8

[65] E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A Heuristic Approach to Value-Driven Evaluation of Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2019. doi: 10.1109/tvcg.2018.2865146 6

[66] J. Wang, X. Liu, H.-W. Shen, and G. Lin. Multi-Resolution Climate Ensemble Parameter Analysis with Nested Parallel Coordinates Plots. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):81–90, 2017. doi: 10.1109/tvcg.2016.2598830 2

[67] M. Wattenberg, F. Viégas, and I. Johnson. How to Use t-SNE Effectively. *Distill*, 2016. doi: 10.23915/distill.00002 3

[68] J. Weissenböck, A. Amirkhanov, E. Gröller, J. Kastner, and C. Heinzl. PorosityAnalyzer: Visual analysis and Evaluation of Segmentation Pipelines to Determine the Porosity in Fiber-Reinforced Polymers. In *Proc. VAST*, pp. 101–110. IEEE, New York, 2016. doi: 10.1109/vast.2016.7883516 2

[69] W. Willett, J. Heer, and M. Agrawala. Scented Widgets: Improving Navigation Cues with Embedded Visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007. doi: 10.1109/tvcg.2007.70589 2

[70] J. Xia, Y. Zhang, J. Song, Y. Chen, Y. Wang, and S. Liu. Revisiting Dimensionality Reduction Techniques for Visual Cluster Analysis: An Empirical Study. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):529–539, 2022. doi: 10.1109/tvcg.2021.3114694 3

[71] D. Xu and Y. Tian. A Comprehensive Survey of Clustering Algorithms. *Annals of Data Science*, 2(2):165–193, 2015. doi: 10.1007/s40745-015-0040-1 3

[72] H. Yang, R. Ballester-Ripoll, and R. Pajarola. SenVis: Interactive Tensor-based Sensitivity Visualization. *Computer Graphics Forum*, 40(3):275–286, 2021. doi: 10.1111/cgf.14306 2