MLVis International Workshop on Machine Learning in Visualisation for Big Data (2021)
D. Archambault, J. Peltonen, and I. Nabney (Editors)

Revealing Multimodality in Ensemble Weather Prediction

N. Galmiche[1], H. Hauser [1], T. Spengler[2], C. Spensberger[2], M. Brun[3] and N. Blaser[1]

[1] Department of Informatics and Center for Data Science, CEDAS, University of Bergen, Norway
[2] Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway
[3] Department of Mathematics, University of Bergen, Norway

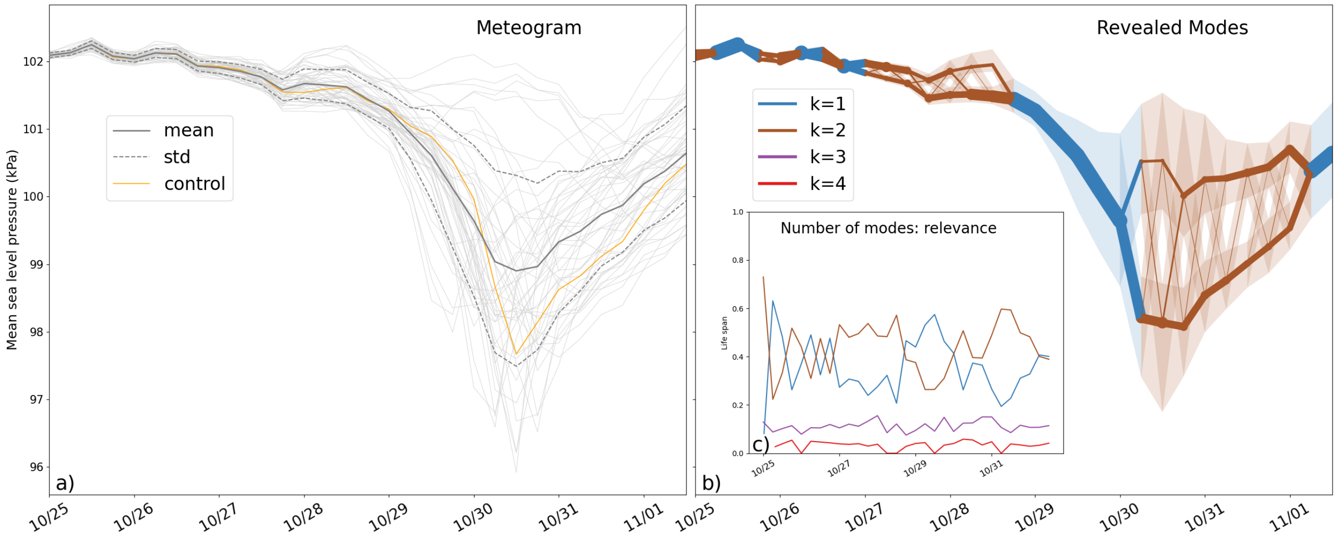

Figure 1: *a)* Meteogram for mean sea level pressure forecast near New York City before the passage of hurricane Sandy, initialised 25 October 2012, 00 UTC. The *control* line represents the control member, i.e., the unperturbed simulation run. Mean and standard deviation are commonly used to assess the ensemble forecast and its uncertainty. *b) Most relevant components* view displaying automatically revealed modes by our method. *c) Life span* plot, indicating the relevance of assuming 1, 2, 3, or 4 modes at each instant according to our algorithm.

Abstract

Ensemble methods are widely used to simulate complex non-linear systems and to estimate forecast uncertainty. However, visualizing and analyzing ensemble data is challenging, in particular when multimodality arises, i.e., distinct likely outcomes. We propose a graph-based approach that explores multimodality in univariate ensemble data from weather prediction. Our solution utilizes clustering and a novel concept of life span associated with each cluster. We applied our method to historical predictions of extreme weather events and illustrate that our method aids the understanding of the respective ensemble forecasts.

CCS Concepts

• ***Applied computing*** → *Earth and atmospheric sciences;* • ***Human-centered computing*** → *Visual analytics;* • ***Computing methodologies*** → *Unsupervised learning;*

1. Introduction

Scientific disciplines and modern applications dealing with complex non-linear systems commonly resort to stochastic ensemble methods to provide an overview of physically possible outcomes, where the ensemble spread originates from slightly varied initial conditions as well as model imperfections. While ensemble methods have become a well-established tool to estimate the uncertainty of predictions, their analysis, interpretation, and visualization still provide major challenges [WHLS18]. Given that the computed simulations, each referred to as a member of the ensemble, contain

DOI: 10.2312/mlvis.20211073

a wealth of information, it is non-trivial to informatively summarize the collective nature of the ensemble. This is especially true in the case of distinct likely outcomes, i.e., multimodality.

The prevalent method to summarise an ensemble prediction is to report the mean as the most likely outcome and the standard deviation as an assessment of uncertainty. This implicitly assumes a unimodal Gaussian distribution and therefore discards crucial information in the case of multimodality. Exploring multimodality in weather forecasting is of great socio-economic importance, as it can provide a more nuanced assessment of the forecast. In particular, it allows domain experts and users to identify potentially threatening weather outcomes that might not be discernible with a classical unimodal approach [PDRHW05].

In this paper, we focus on univariate meteograms, i.e., ensembles of N univariate time-series of common length T. Such ensembles are computationally expensive and typically only around 50 members are simulated. Meteograms are a common visualisation in the domain to provide an overview of the possible evolution of a given variable of interest at a fixed geographic point (e.g., Fig.1). In this work, we focus on exploring and analyzing the possible situation of multimodality in such ensemble data; in particular, we address the following tasks:

T1 Estimate the number of modes at each time step
T2 Derive summary statistics for each mode at each time step
T3 Determine when distinct modes appear and/or disappear
T4 Determine all possible connections between consecutive modes

To support these tasks, we provide a new solution based on a graph resulting from clustering and the concept of life span. Our threefold contribution is to: **a)** introduce the concept of life span for each mode, a proxy of its relevance; **b)** provide a visualization tool revealing all possible clusters according to their relevance in one image; **c)** suggest an interpretation of the distribution.

2. Related work

Visualizing ensembles is challenging as it needs to cope with the additional *member* dimension [WHLS18]. Potter et al., with EnsembleVis, and Nocke et al., with SimEnvVis, proposed interactive visualization solutions to help experts with understanding ensemble simulations [PWB*09, NFB07]. Multimodality is not explicitly treated in these frameworks, although non-unimodal behavior can be deduced using *Plume Charts*. Parameter space analysis is one objective of ensemble visualization and both interactive and automated methods exist [SMG*15, WLSL16]. Although associated challenges and solutions are ensemble-related, our task is closer to what Wang et al. [WHLS18] defined as *ensemble trend analysis* and *comparison*. Here, current challenges include to reveal different trends using as few assumptions as possible [FKRW16, OBJ15], faithfully representing these trends [FKRW16], and illustrating the reasons for selecting these trends [JKW16].

Clustering is a broad topic and an active area of research that has applications in numerous domains. Javed et al. recently carried out a benchmark study specifically on time series clustering using 112 different time series datasets and the most popular clustering methods [JLR20]. Typical approaches for time series clustering are distance-based and most of them use either the Euclidean distance or dynamical time warping (DTW) [HHB15, WLSL16], among others such as shape-based distances [PG15]. Using distance measures that match the requirements of the application domain is of particular importance [LGZY16]. In their study, Javed et al. [JLR20] assumed that the number of clusters was known. Several techniques exist to estimate this number k of clusters in a dataset, such as the elbow and silhouette methods [KR09, SKRR15] or the gap statistic [TWH01]. Similarly, methods exist that find the number of clusters automatically such as DBSCAN [EKSX96] or kNN [CH67]. However, the reasons that lead to a final k value might depend on other hyper-parameters in complex ways.

3. Revealing Multimodality

In what follows, we first explain the design of our solution, before then detailing the new algorithm and visualization specifics.

3.1. Design and Overview

Based on the tasks **T1**–**T4** and the state of the art as outlined above, we arrived at the following design rationales:

DR1: The ensemble distribution can have any shape. No a priori information is available concerning the underlying distribution of the ensemble. In particular, the number of modes and type of probability distribution of each mode are unknown.

DR2: All ensemble members matter. Prediction ensembles, as studied here, are typically small ($N \approx 50$) and the notion of outliers must be considered cautiously, as each ensemble member can describe the actual outcome best-possibly, no matter how different it is from the other members. Furthermore, clear communication is essential for weather forecasting, as it concerns not only meteorologists but also users whose decisions depend on the prediction. As a consequence, while simplicity and readability should be favoured, it remains of great importance to not ignore singularities in the ensemble distribution, especially if they describe extreme scenarios.

DR3: All scales matter. In the atmosphere, large and small scales co-exist in space and time, describing distinct phenomena (diurnal cycle, seasonality, etc.). To ensure an accurate representation, one scale should not be pre-selected at the expense of another. Consequently, thresholds of any kind are to be avoided, where possible.

DR4: Not a black box. After discussing with meteorologists, we agreed that one *correct* number of modes k does not exist. Thus, the process of determining k together with its uncertainty and the consequences of choosing a particular value of k over another should remain transparent. The underlying motivation is that only experts in the application domain can make an informed decision regarding the most meaningful number of clusters in ambiguous cases.

Based on these rationales, we arrived at the following design of our solution: study the members at each time step t separately, then try all the possible values of k for each t, and summarize the obtained outcome for each t and each k simultaneously in one image. This simultaneous visualization remains comprehensible using the concept of *life span* that emphasizes modes that *live longer*.

3.2. Algorithm

Given the ensemble denoted as *members*, a clustering *model* with parameter number k of clusters (e.g., k-means) is devised and a monotonic *score* function that returns a real-valued quality measure, given a set of clusters (e.g., cluster inertia). The algorithm completes 3 main tasks (see Algorithm 1) that result in a graph. The first task is motivated by **DR1**: at each time step t, apply *model* to the N data-points for all possible values of $k \in \{1, \ldots, N\}$. For each t and k, *model* returns a list of clusters such that each member belongs to exactly one of the k clusters. Then, for each time t, each k is associated with its *normalized score* $s(k,t)$, and a *life span* $l(k,t)$, defined as $l(k,t) = s(k,t) - s(k+1,t)$. Thirdly, the corresponding vertices and edges are created in the graph. Each vertex v_i stands for a cluster and stores the associated members, t_i and k_i (time t and value k at creation), its *normalized score* $s_i = s(k_i,t_i)$ and *life span* $l_i = l(k_i,t_i)$. Edges connect vertices between t and $t+1$, provided that vertex v_j corresponds to vertex v_i in terms of the score, i.e., $[s_i, s_i+l_i[\cap [s_j, s_j+l_j[\neq \emptyset$. The life span of the resulting edge is defined as the length of this intersection. The members of the edge are members belonging to both v_i and v_j (see Figure 2).

The normalization process follows **DR3**, by adapting the method to time-related changes of score scales due to error growth with increasing lead time in the forecast.

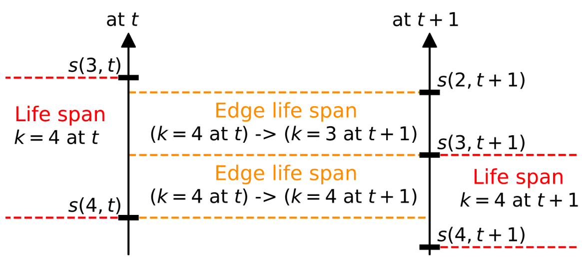

Figure 2: Schematic diagram of normalized scores for selected k at times t and $t+1$, showing scores and life spans of nodes and edges.

Algorithm 1: Graph construction

Input: *members* ($T \times N$ array), a *score* function, a clustering *model*
Output: a graph summarizing the clustering outcome for each k and t

```
for t = 0 .. (T-1) do
    X = members[t];
    for k = 1 .. N do
        Fit model(k) to X, store clusters in clusters[t, k];
        scores[t, k] = score(clusters[t, k]) ;
    for k = 1 .. N do
        s(t, k) = (scores[t,k] - min_k(scores[t,k])) / (max_k(scores[t,k]) - min_k(scores[t,k])) ;
        l(t, k) = s(k+1, t) - s(k, t);
Create a vertex for each cluster in clusters[t, k] ;
Create an edge v_i -> v_j if at least one member is in vertex v_i
    at t and in v_j at t+1, and if v_j corresponds to v_i in terms of
    their score (see also Fig. 2)
```

3.3. Visualization

We provide two main visualization solutions: the *entire graph* and the *most relevant components* view. In both views, the x-axis represents time and y the considered physical value. Vertices are placed at their corresponding time step and their y-value is defined as the mean of the members it represents. The shaded area around a component (vertex or edge) corresponds to the standard deviation of its members. Although each member represents a similarly likely outcome, the proportion of members in each mode retains essential information. Consequently, the more members in a component the thicker the component. Furthermore, to distinguish between the different assumptions for k (that is to say, how many clusters were assumed when a specific cluster was created), we identify each choice k with a color (colors from ColorBrewer [HB03]).

The *entire graph* view shows the output of the clustering model for all values of k. To ensure readability without discarding crucial information, the opacity of the components is based on their life span, thus revealing only the most relevant components. As the occurrence of multiple values of k is meaningful, their corresponding output is concurrently displayed. This allows an effortless exploration using no assumptions nor thresholds, thus following **DR2**. This visualization leads to an unusual combination of a categorical color scale, associated with ordered data ($k = 1, 2, ..N$), so that the colors remain distinguishable even across largely varying opacities.

To guide the exploration, we provide two additional views: a) a *life span* plot showing the life span of each k assumption and each t; b) a *suggestion bar* displaying the most long-lived clustering k for each time step. This provides quantitative information to the user while facilitating the decision-making process as motivated by **DR4**. The user can either predetermine the number of modes, or let the algorithm decide.

The *most relevant components* view shows only one value of k at each time step, which is thought to be the most relevant (either chosen automatically as in the suggestion bar or manually re-adjusted by the user). As there is no concurrent display on this plot, the opacity of the components can be set to 1. This view contains less information than the *entire graph* and favours readability.

Both views can be seen as the final product of our graph-based method. While some users appreciate the information-rich and transparent nature of the *entire graph* (with the *suggestion bar* and *life span* plot), others prefer the more classical visual summary provided by the *most relevant components* view.

4. Weather forecasting case studies

The needs and challenges of ensemble analysis and visualization meet in weather prediction due to the diversity of its audience, their manifold objectives and the complex nature of the atmospheric system. Further, this application is prone to multimodality due to the combination of chaos and stability in the atmosphere. We limit our focus here on extreme weather events, as they demand a detailed assessment of uncertainty and risk, where the users have a particular interest in potentially multimodal outcomes that might be attached to significant socio-economic risks. For simplicity and consistency, we used k-means as clustering model and inertia as score for all case studies. More options are available in our implementation.

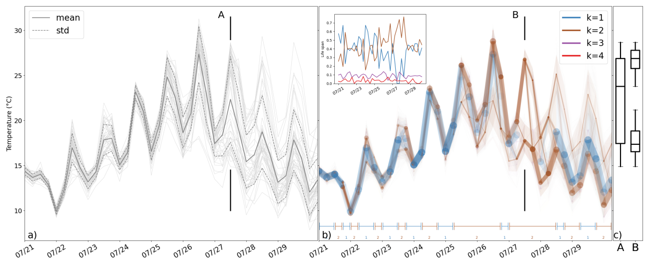

Figure 3: *a)* Meteogram for 2m temperature forecast near Bergen, Norway, initialised 21 July 2019 00 UTC during the 2019 heatwave. *b) Entire graph* view, showing all assumptions about the number of clusters with the *suggestion bar* and the *life span* plot. Starting 27 July, the bimodality is so clear that the unimodal assumption (blue line) almost disappears. *c)* Boxplots representing interpretations A and B for 27 July at noon.

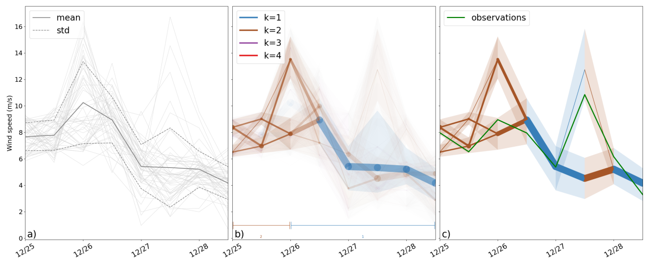

Figure 4: *a)* Meteogram for forecast of wind speed near Paris, initialised 25 Dec. 1999 00 UTC before the passage of the storm Lothar. *b) Entire graph* view, suggesting that unimodality is the most relevant interpretation for 27 Dec., while revealing a threatening peak in wind speed if bimodality were to be assumed. *c) Most relevant components* view, where the user followed the method suggestions except for one time step, where $k = 2$ was specified instead of $k = 1$, emphasizing the potential risk. These observations support **DR2** & **DR3**: a sudden variation predicted by few members.

As first case, we consider the temperature meteogram for Bergen, Norway, during the July 2019 heatwave (Fig. 3). This synoptic condition is linked to a weather situation known as Scandinavian blocking, imposing significant forecasting challenges [PH03]. A large diurnal temperature range is evident, featuring an alternation between uni- and bi-modality, revealing that the nighttime temperature forecast is more certain than its daytime counterpart. From July 27, a clear bi-modality emerges with a misleading “mean and standard deviation” interpretation: "22.3±4.7°C, i.e., a [17.6, 27.0] range" that suggests great uncertainty. Our method implies a 51% (respectively 49%) probability of being inside [16.4, 19.6] (resp. [25.2, 27.2]), which yields a more accurate description considering the distribution of all members. A similar analysis of the ensuing day shows that 20% of the members are actually higher than the upper bound (22.8°C) of the predicted range, with some members reaching up to 29.3°C. While our method is capable of revealing this situation appropriately, a standard, unimodal interpretation would have missed to provide a meaningful explanation.

As a second case, we show the meteogram for New York City, encompassing hurricane Sandy in October 2012, causing disastrous damage as it was about to undergo extratropical transition [EWA*17]. Forecasting such transitions is a well-known challenge and a regular source of multimodality in ensemble prediction: “Will a given tropical cyclone transition to become extratropical or not?”. The bimodality for Sandy illustrates that mode detection does not simply rely on the spread, as the spread is wider on 10/30 than on 10/28, yet 10/30 is considered unimodal while 10/28 is detected as bimodal (Fig. 1, page 1). Our method is able to clearly and accurately represent the creation and fusion of modes.

As a third case, the European winter storm Lothar caused severe socio-economic consequences for large parts of central Europe and was relatively poorly forecasted [PDRHW05]. This case (Fig. 4) demonstrates the importance of transparency in the interpretation of the member distribution. In the meteogram, we clearly see that almost all members agree on the outcome for Dec. 27 and that only a few predict severe wind speed. While considering the distribution as an asymmetric mode with a long tail would be appropriate from a statistical viewpoint, the potential hazard associated with the long tail represents a significant risk. Thus, from the meteorological viewpoint, it might be preferred to distinguish between 2 scenarios: one relatively harmless, composed of the majority of the members, and a second, less likely, but rather severe scenario depicted by few members. Although our method suggests that there is only one mode, the *entire graph* view allows experts to observe that considering 2 modes could be a pertinent choice, revealing a dangerous situation that would be ignored if assuming unimodality. We provide experts with the ability to tune the most relevant number of modes, choosing for example two modes for 27 Dec. The *most relevant components* view will then be automatically updated so that 2 modes are considered that day (Fig. 4.c).

5. Conclusion and future work

As result of a long-term interdisciplinary collaboration between meteorology and data science, we introduce a new, graph-based solution, based on clustering and the concept of life span, that successfully reveals multimodality in ensemble weather prediction, together with its imminent uncertainty. We demonstrate the performance of our method in the context of three case studies, all featuring extreme weather events for which a detailed risk assessment is imperative. Our method integrates a carefully designed computational approach (graph construction based on clustering) with a visualization solution that enables the user to bring in the necessary level of expert judgement, e.g., when considering unusual ensemble members.

Our main future objective is to adapt the algorithm and visualization to multivariate and spatio-temporal ensembles. Furthermore, even though instant-wise mode evolution is a feature of our method, we aim at allowing the user to use a time window in order to emphasize general patterns. In the meantime, we plan to provide new interactive tools to enhance the graph exploration while proposing fully automated suggestions.

Acknowledgments. We thank the climate and energy transition strategy of the University of Bergen (UiB) for funding this research and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the ensemble data. Parts of this work have been carried out in CEDAS, i.e., UiB’s Center for Data Science.

References

[CH67] COVER T., HART P.: Nearest neighbor pattern classification. *IEEE transactions on information theory 13*, 1 (1967), 21–27. 2

[EKSX96] ESTER M., KRIEGEL H.-P., SANDER J., XU X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (1996), KDD'96, AAAI Press, p. 226–231. 2

[EWA*17] EVANS C., WOOD K. M., ABERSON S. D., ARCHAMBAULT H. M., MILRAD S. M., BOSART L. F., CORBOSIERO K. L., DAVIS C. A., DIAS PINTO J. R., DOYLE J., ET AL.: The extratropical transition of tropical cyclones. part i: Cyclone evolution and direct impacts. *Monthly Weather Review 145*, 11 (2017), 4317–4344. 4

[FKRW16] FERSTL F., KANZLER M., RAUTENHAUS M., WESTERMANN R.: Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE transactions on visualization and computer graphics 23*, 1 (2016), 831–840. 2

[HB03] HARROWER M., BREWER C. A.: Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal 40*, 1 (2003), 27–37. 3

[HHB15] HAO L., HEALEY C. G., BASS S. A.: Effective visualization of temporal ensembles. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2015), 787–796. 2

[JKW16] JAREMA M., KEHRER J., WESTERMANN R.: Comparative visual analysis of transport variability in flow ensembles. *J. WSCG 24*, 1 (2016), 25–34. URL: http://wscg.zcu.cz/WSCG2016/!_2016_Journal_WSCG-No-1.pdf. 2

[JLR20] JAVED A., LEE B. S., RIZZO D. M.: A benchmark study on time series clustering. *Machine Learning with Applications 1* (2020), 100001. 2

[KR09] KAUFMAN L., ROUSSEEUW P. J.: *Finding groups in data: an introduction to cluster analysis*, vol. 344. John Wiley & Sons, 2009. 2

[LGZY16] LIU R., GUO H., ZHANG J., YUAN X.: Comparative visualization of vector field ensembles based on longest common subsequence. In *2016 IEEE Pacific Visualization Symposium (PacificVis)* (2016), IEEE, pp. 96–103. 2

[NFB07] NOCKE T., FLECHSIG M., BOHM U.: Visual exploration and evaluation of climate-related simulation data. In *2007 Winter Simulation Conference* (2007), IEEE, pp. 703–711. 2

[OBJ15] OBERMAIER H., BENSEMA K., JOY K. I.: Visual trends analysis in time-varying ensembles. *IEEE transactions on visualization and computer graphics 22*, 10 (2015), 2331–2342. 2

[PDRHW05] PALMER T., DOBLAS-REYES F., HAGEDORN R., WEISHEIMER A.: Probabilistic prediction of climate using multi-model ensembles: from basics to applications. *Philosophical Transactions of the Royal Society B: Biological Sciences 360*, 1463 (2005), 1991–1998. 2, 4

[PG15] PAPARRIZOS J., GRAVANO L.: k-shape: Efficient and accurate clustering of time series. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (2015), pp. 1855–1870. 2

[PH03] PELLY J. L., HOSKINS B. J.: How well does the ecmwf ensemble prediction system predict blocking? *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 129*, 590 (2003), 1683–1702. 4

[PWB*09] POTTER K., WILSON A., BREMER P.-T., WILLIAMS D., DOUTRIAUX C., PASCUCCI V., JOHNSON C. R.: Ensemble-vis: A framework for the statistical visualization of ensemble data. In *2009 IEEE International Conference on Data Mining Workshops* (2009), IEEE, pp. 233–240. 2

[SKRR15] SUBBALAKSHMI C., KRISHNA G. R., RAO S. K. M., RAO P. V.: A method to find optimum number of clusters based on fuzzy silhouette on dynamic data set. *Procedia Computer Science 46* (2015), 346–353. 2

[SMG*15] SPLECHTNA R., MATKOVIĆ K., GRAČANIN D., JELOVIĆ M., HAUSER H.: Interactive visual steering of hierarchical simulation ensembles. In *2015 IEEE Conference on Visual Analytics Science and Technology (VAST)* (2015), IEEE, pp. 89–96. 2

[TWH01] TIBSHIRANI R., WALTHER G., HASTIE T.: Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*, 2 (2001), 411–423. 2

[WHLS18] WANG J., HAZARIKA S., LI C., SHEN H.-W.: Visualization and visual analysis of ensemble data: A survey. *IEEE transactions on visualization and computer graphics 25*, 9 (2018), 2853–2872. 1, 2

[WLSL16] WANG J., LIU X., SHEN H.-W., LIN G.: Multi-resolution climate ensemble parameter analysis with nested parallel coordinates plots. *IEEE transactions on visualization and computer graphics 23*, 1 (2016), 81–90. 2