EuroVis Workshop on Visual Analytics (2023)
M. Angelini and M. El-Assady (Editors)

Multi-Ensemble Visual Analytics via Fuzzy Sets

N. Piccolotto, M. Bögl and S. Miksch

TU Wien

Abstract

Analysis of ensemble datasets, i.e., collections of complex elements such as geochemical maps, is widespread in science and industry. The elements' complexity arises from the data they capture, which are often multivariate or spatio-temporal. We speak of multi-ensemble datasets when the analysis pertains to multiple ensembles. While many visualization approaches were suggested for ensemble datasets, multi-ensemble datasets remain comparatively underexplored. Our years-long collaboration with statisticians and geochemists taught us that they frame many questions about multi-ensemble data as set operations. E.g., what are the most common members (intersection of ensembles), or what features exist in one member but not another (difference of members)? As classical crisp set relations cannot account for the elements' complexity, we propose to model multi-ensembles as fuzzy relations. We provide examples of fuzzy set-based queries on a multi-ensemble of geochemical maps and integrate this approach into an existing ensemble visualization pipeline. We evaluated two visualizations obtained by applying this pipeline with experts in geochemistry and statistics. The experts confirmed known information and got directions for further research, which is one Visual Analytics (VA) goal. Hence, our proposal is highly promising for an interactive VA approach.

CCS Concepts
• ***Human-centered computing*** → ***Visual analytics;***

1. Introduction

Many fields in science and industry use intricate algorithms and simulations to study real-world phenomena ranging from, e.g., diesel engines [MGJ*10] to weather [FKRW17]. The output of these computations is i) often complex, i.e., multivariate or spatio-temporal, and ii) usually dependent on the exact model configuration, such as parameter settings or initial conditions. Hence, a single run is inadequate to comprehend the phenomenon or its model. The collection of simulation outputs is known as an ensemble dataset [WHLS19]. Typical analytic tasks include getting an *overview* of the ensemble or *parameter analysis*. These tasks are impeded because each individual member may have all the intricate properties of a scientific dataset [KH13], but there are multiple slightly different copies of it. For multi-ensemble data, the number of ensembles quickly gets too large to visualize them directly, and it is not obvious how to achieve computations across many ensembles. Thus, we look for a principled and scalable approach allowing us to carry out necessary analytic tasks.

This paper draws from extensive discussions with our collaborating experts in statistics and geochemistry. They deal with multi-ensemble datasets (specifically, ensembles of time series or geographic maps) and have the same goals as suggested for ensemble visualization [WHLS19]. We realized that many answers they seek from the multi-ensemble datasets pertain to four multi-ensemble dimensions: *Ensemble*, *parameter setting*, *member*, and *feature*. Specifically, our collaborators are, e.g., interested in the most common members in all ensembles (*overview*) or parameter settings producing one member but not another (*parameter analysis*). In addition, these relations can be formulated as set relations when, for a suitable tuple of dimensions, one dimension acts as a set and another as an element. We can then answer the question about common members as an intersection relation of all ensembles (acting as a set) containing members (acting as elements). That is, assuming we have a way of doing that, and it accounts for the uncertainty arising from comparing complex objects. Set relations like those cannot be computed with classical crisp sets as there is too much uncertainty in real-world multi-ensemble data. We can rarely say for sure that a given member exhibiting complex properties of scientific data is in one ensemble but not in another. More likely, two members will be simultaneously different and similar in some but not other aspects. Therefore, in this paper, we propose using fuzzy set theory to model such relations. Fuzzy relations bring a formal basis to compute the uncertainty-aware set relations necessary to handle multi-ensemble data. Our contributions are:

- application of fuzzy set theory to multi-ensemble datasets (Section 4) yielding
- a multi-ensemble visualization pipeline based on fuzzy queries (Section 5); and
- evaluation of the resulting visualizations with domain experts (Section 6) using a real-world dataset.

DOI: 10.2312/eurova.20231092

2. Related Work

An ensemble constitutes a collection of objects that often exist in space, time, and a multivariate attribute space. Members of an ensemble are expected to be similar in some aspects, as ensemble data typically emerges from running a simulation with perturbed parameters. Several surveys cover approaches to ensemble data visualization [KH13, WHLS19, XLWD19]. Multi-ensembles are less explored, with a few examples being the following.

Köthur et al. [KWS*15] propose a Visual Analytics (VA) approach to compare two ensembles of time series. They compare all pairwise combinations of members with windowed cross-correlation at several offsets and lags, and visualize the aggregated results in a tilemap. This gives an overview of when in time the two ensembles show similarities, but it is not possible to relate them in any other ways. Cibulski et al. [CKS*17] deal with a simulation in engineering, where each simulation run yields a collection of surfaces. They propose a coordinated multiple views system and extensive data aggregation. Wang et al. [WLSL17] worked with a meteorological simulation. Analysts were interested in the influence of the grid size, in addition to convection parameters. Nested Parallel Coordinate Plots (NPCP) were suggested to support the analysis, but the ensembles shown in the NPCP are fixed and few. Finally, Piccolotto et al. [PBG*22] describe a VA approach for the analysis of up to ca. 20 ensembles of time series. In addition to visualizations supporting the pairwise comparison of ensembles, they suggest a set-aware clustering scheme to obtain an overview of the multi-ensemble.

Our four proposed dimensions may be viewed to form a 4D cube. Cubes were proposed both as concrete data structures [GCB*97, WFW*17, MLKS18] and conceptual models [BDA*17, LCC*20]. However, we do not frame our data model as such because the intuitive operations on 3D conceptual models vanish using a hypercube.

To summarize, recent work for visualization of multiple ensembles focuses on rather few ensembles, presumably because it eases visualization design. However, due to that fact, analysts are limited in the amount of data that can be processed and in the questions they can ask. With this paper, we aim to lay the required foundation for the visualization of many ensembles.

3. Data Definition and Data Model

We present a formal data definition for everything used in Section 6. The dataset we consider is the GEMAS geochemical survey [RBD*14]. A publicly available version of the dataset contains measurements of 18 elements associated to 2,108 locations across Europe. The “simulation model” we use is SBSS [NOFR15, MBN22, PBM*22]. For a high-level introduction to SBSS see [PBM*22], and for the statistical formulas and proofs, see [NOFR15, MBN22]. Like Principal Component Analysis for non-spatial data, SBSS obtains meaningful latent spatial dimensions from multivariate spatial data, such as the GEMAS dataset. Hence, one obtains a multi-ensemble dataset when varying SBSS parameters and a member is an SBSS latent dimension.

The four dimensions we consider are *parameter setting*, *ensemble*, *member* and *feature* (Figure 1). We define a *member* m_{ij} as

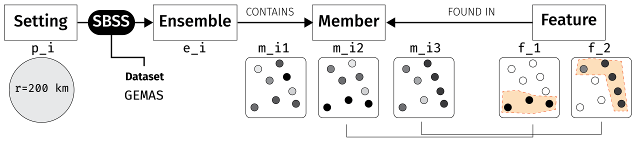


Figure 1: *Illustration of the data model (Section 3). SBSS produces an ensemble (e_i) of maps (m_{ij}) from a parameter setting (p_i) and a multivariate spatial dataset (GEMAS geochemical survey). Ensemble members are scalars (color intensity) at geographical locations (circles). Features are named spatial masks referring to a member.*

spatially distributed scalars, i.e., $m_{ij} \in \mathbb{R}^2 \times \mathbb{R}$, with $\mathbb{R}^2$ being the two-dimensional position. All members use identical locations. Locations may be on a grid, but we do not require it. An *ensemble* e_i is only a container for members, i.e., $e_i = \{m_{i1}, \ldots, m_{ik}\}$, with k being the amount of members in e_i. All ensembles have size k. One SBSS run with given *parameter* settings produces one ensemble. This paper considers only one SBSS parameter, which is a ball-shaped point neighborhood definition described by a circular radius. Hence, a parameter setting is a positive real number $p_i \in \mathbb{R}_+$. Finally, based on how domain experts analyze latent dimensions, we model a *feature* f as qualitative observations in a member, e.g., a pattern that could be related to mineral deposits. A feature is a partial ensemble member, i.e., the values of a member m_{ij} at locations inside a user-defined simple polygon s_f plus a text description t_f: $f = (t_f, m_{ij}[s_f])$.

4. Fuzzy Relations

The first key idea to our approach is fuzzy comparison, which is known as distance functions, dissimilarity metrics, or similarity measures. As the name implies, these functions quantify the similarity of two objects. Fuzzy comparison is crucial as all ensemble members are expected to be slightly different copies of each other, and perfect matches will be the exception. We assume appropriate similarity measures $sim(a,b) : X \times X \rightarrow [0,1]$ exist, where X is any of the four dimensions we consider and $sim(a,b) = 1$ iff $a = b$. The second key idea is to use these similarity measures to compute the association between instances of dimensions connected by an arrow in the data model (Figure 1). For example, a feature is highly associated with a member if it has similar values at the relevant locations. A parameter setting is highly associated with an ensemble if a similar parameter setting produced it. An ensemble is highly associated with a member if it is similar to a member in the ensemble. Fuzzy set theory, specifically fuzzy relations, is the formalism allowing us to combine and build upon these ideas.

Fuzzy Sets. Fuzzy sets extend classical crisp sets by relaxing the binary membership condition. It is replaced by a membership degree $\mu_A(x) \in [0,1]$ that encodes to which degree an element x belongs to set A. A fuzzy set is thus $A = \{(x, \mu_A(x)) \mid x \in X\}$. Fuzzy sets allow similar operations as crisp sets, such as intersection, union, or difference [Cha19]. In order to carry out these operations, the fuzzy sets taking part may first be written in a tabular form (Table 1). Fuzzy set relations are then computed column-wise, e.g., for a fuzzy intersection, a *min*() operator is applied per column and yields $\{(y_1, 0.1), (y_2, 0.2), (y_3, 0.7)\}$.

	y_1	y_2	y_3
x_1	0.1	0.5	0.9
x_2	1.0	0.2	0.7

Table 1: *Two fuzzy sets x_1 and x_2 written in tabular form.*

Fuzzy Relations. Fuzzy sets are defined on the same domain ($Y = \{y_1, y_2, y_3\}$ in Table 1). Fuzzy *relations* define the membership degree on the Cartesian product $X \times Y$ of two domains X and Y, i.e., $R = \{(x, y, \mu_R(x, y)) | (x, y) \in X \times Y\}$. They may also be written in tabular form (Table 2 with X being the *ensemble* dimension and Y being the *members*), which we call a relation matrix. The identical structure to the fuzzy set's tabular form is why we say a dimension can be viewed either as *set* or *element* in the relation: By defining e_i in Table 2 as the *sets*, we can compute, e.g., an intersection using column-wise *min*() operations just like described earlier and obtain the degree to which a member belongs to all ensembles. When we transpose Table 2, m_{ij} are the sets and e_i the elements.

R	m_{11}	m_{12}	m_{13}	m_{21}	m_{22}	m_{23}
e_1	1.0	1.0	1.0	0.1	1.0	0.6
e_2	1.0	0.0	0.4	1.0	1.0	1.0

Table 2: *Example relation matrix of* ensembles *and* members*.*

Building Relation Matrices. We build a relation matrix for all pair-wise combinations of dimensions in our data model. For four dimensions, we find six such relations. Three of those relations are characterized by a direct connection in the data model (Figure 1). Appropriate similarity metrics populate cells in these relation matrices. E.g., a cell with index i, j in EM (Ensemble × Member) holds the largest similarity between any member of e_i to m_j. In PE (Parameter × Ensemble), a cell holds the similarity between a parameter setting p_i and the setting p_j that produced e_j. In MF (Member × Feature), we compare the values at locations in m_i defined by the feature's spatial mask to the feature's reference values. Hence, we need one similarity measure for parameter settings and one for members. Our concrete choices are explained in Section 6.

Computing Indirect Relations. Talking about the association between dimensions without a direct connection in the data model is also natural. E.g., a parameter setting p_i can be highly associated with a feature if the feature is highly associated with a member, which is contained in the ensemble produced by p_i. Fuzzy composition allows a combination of two relations on different product sets with a shared domain, i.e., $R_1 = X \times Y, R_2 = Y \times Z$. There are a couple ways to carry out this composition [Cha19]; we propose the max-product composition. For example, the common dimension *member* is between ensembles and features. To compose EM and MF to the relation EF (Ensemble × Feature), degrees $\mu_{EM}(e_i, m)$ and $\mu_{MF}(m, f_j)$ for all m are multiplied, and the largest product is kept. This way, we can fill in the gaps (EF, PF, PM) using existing relation matrices (EM, MF, PE).

Handling Duplicates. For a sufficiently large multi-ensemble dataset, some members will be practically duplicates (e.g., m_{11} and m_{22} in Table 2). This fact influences possible downstream aggregations, e.g., weighted averages. Duplicates will have the same degrees in any fuzzy set computation. The duplicates' data will be over-weighted if we use the degrees as weights. To account for this, we propose to add another dimension *representatives*, which will be the centers of a partition-based clustering on the members (e.g., k-medoids [PJ09]). Clustering parameters (like k) can be defined by the analyst or automatically set with grid search and clustering quality metrics, such as a Silhouette index [LLX*10]. Fuzzy relations are defined for *representatives* in the same fashion as other dimensions described before. We assume that the clustering is of sufficient quality and captures the most important trends in the dataset. We expect that the partitioning can be interactively changed in a VA system. From now on, we will, for simplicity, say *members* but mean their *representatives*. This step is optional when no such aggregation is required later on.

5. Visualization Pipeline

The steps outlined in the previous section can be viewed as a visualization pipeline for multi-ensemble data, depicted in Figure 2. Steps connected by red lines mark our proposal. After formulating and computing a desired fuzzy set to analyze, we obtain instances of a dimension (e.g., *members*) together with the degree to which they belong to the fuzzy set. The result may again be viewed as a traditional ensemble dataset, with the difference that in addition to a categorical variable to which member the complex data belongs, we have a quantitative variable in the degree. While in itself not an advantage, it readily allows standard computations that may not be as straightforward with traditional ensemble data. For instance, one may compute aggregations such as weighted averages when using the degree as weight. Another option could be to treat the degree as another source of uncertainty to incorporate in data transformations if the VA pipeline is already aware of uncertainty [CCM09]. Once a fuzzy set is obtained, the pipeline suggested by Wang et al. [WHLS19] remains applicable. In their pipeline (black lines in Figure 2), an ensemble is optionally aggregated per member (using the degree, in our case), then a visualization step produces possibly multiple images, which are optionally composited into a single image. In the last step, one may again use the degree variable for visual encodings, e.g., as opacity.

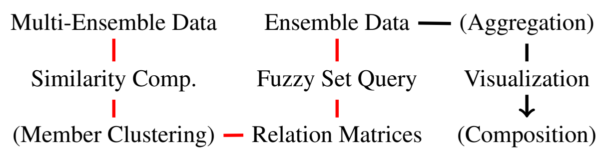


Figure 2: *Our proposed pipeline (red lines) combined with an existing ensemble visualization pipeline (black lines).*

6. Evaluation Using Real-World Geochemistry Data

In this section we apply the proposed visualization pipeline to the dataset and algorithm described in Section 3. Settings for the pa-

Proceedings published by Eurographics - The European Association for Computer Graphics.

rameter (ball radius) start from 50 km and increase up to 600 km in 50 km steps (12 values). The output of SBSS is an ensemble of maps, which show the latent dimensions. Latent dimensions in SBSS are defined only up to sign and order, i.e., they are not ordered and a pattern of high values can be equivalent to a pattern of low values. Hence, we select the absolute Spearman's rank correlation coefficient ρ as similarity measure for members. Values near 1 will point to high rank correlations, values near 0 to low correlations. The similarity between parameter settings is calculated as the ratio of ball radii. As the dataset is too big to just take a look at ($12 \times 17 = 204$ maps, the 17 instead of 18 variables arise from some necessary preprocessing of compositional data [Ait82, NOFR15]), we iterate over all possible values of the clustering parameter k and compute the Silhouette index [LLX*10]. We use this index as we need an unsupervised clustering quality metric due to the absence of ground truth data. We choose the value of k that maximizes this index, which is 18 (index value 0.573). Hence, we obtain 18 representatives of members in the multi-ensemble. For features we defined polygons (see Figure 3) that, from our discussion with geochemists, we know relate to differences in soil type or age [HMZ09]. Podzol is the typical soil type for boreal forests, found in Scandinavia. The soil in the Baltics is younger than in the Nordic countries and thus has different characteristics [RST*00]. The soil in eastern Spain is characterized by significant accumulations of calcium carbonate (Calcisol). The southern mainland part of Greece is a mix of several soil types, hence any features found there could indicate a latent process other than soil type. Each feature is characterized by consistently high or low values in the respective area (not depicted in Figure 3).

In the following two examples we apply the red steps of Figure 2 and inspect the degrees ourselves to verify first that this part works.

Which Members Show One Feature But Not Another? We pick something that we should be able to verify by visual inspection. We will look for members that contain the Scandinavia (S) and Greece (G) features, but not those in Spain (E) and the Baltics (B). The relation in question is MF (Member × Feature), and we query for $(S \cup G) \setminus (E \cup B)$. The best-matching member (degree 0.47) is shown in Figure 3: There are lots of very low values in Scandinavia and many very high values in Greece (good match to $S \cup G$), but in the Baltics there are neither high nor low values and in Spain the high values do not cover the whole feature polygon (poor but not terrible match to E, B). Hence, we can visually confirm the correctness and quality of the query result.

Which Parameter Settings Produce All Features? The relation in question is FP (Feature × Parameter) and we query for an intersection of features as *set*. Setting 1 (kernel radius 50 km) is least associated to this relation with a degree of 0.31, whereas settings 5–9 (radii 250–450 km) are most associated with a degree of around 0.58. This pattern of setting 1 at the bottom and settings 5–9 at the top persists also when we look at individual features, i.e., look at rows in FP. We can infer that maps obtained with the three smallest kernel radii 50–150 km do not show the defined features well. This seems reasonable, as too small kernels might not capture patterns over wide areas, but we discuss this further with an expert.

Next, we apply the complete pipeline and show resulting visu-

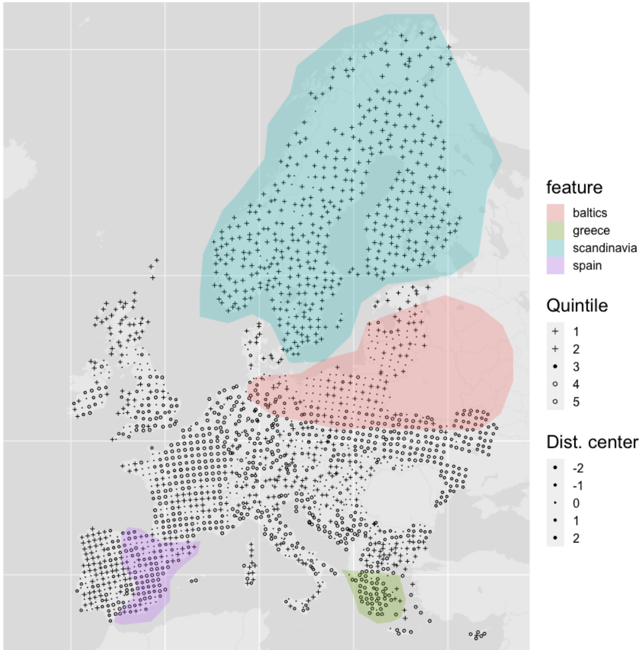
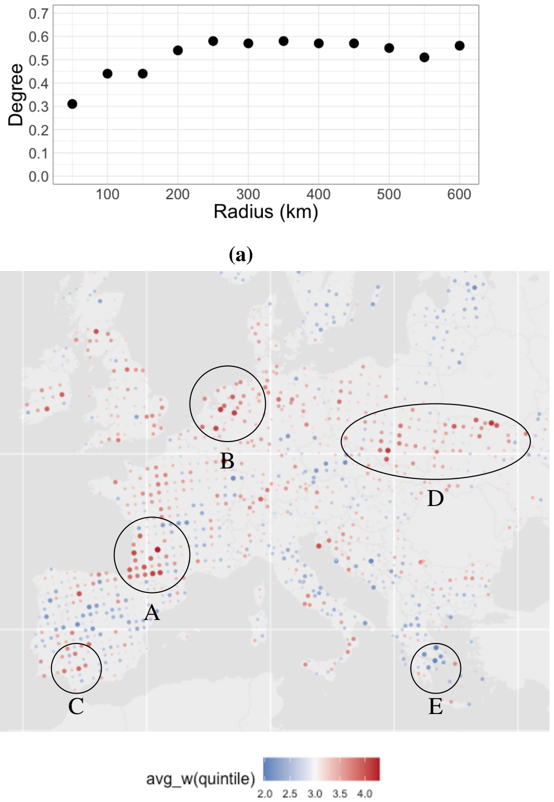


Figure 3: *An example latent dimension from SBSS (plotted following common practices in geochemistry [RFGD08]) with annotated features (polygons). Crosses mark low and circles mark high values, polygons the areas in which we want to find mostly high/low values. This latent dimension is a good match to the features in Greece and Scandinavia, but less to those in Spain and the Baltics.*

alizations to experts to verify the images' usefulness. The experts knew about the features we defined. Notably, we use one more relation matrix (ME), so the evaluation covers half of all matrices. We are convinced the remainder is also useful, but it is out of this paper's scope to present three more examples.

Figure 4a shows the latter of the previous two examples, i.e., the intersection of features (*set* dimension) defined in Scandinavia, Greece, Spain, and the Baltics (Figure 3) with regard to parameter settings (*element*). No *aggregation* is performed and we *visualize* the degree as points on a common axis, which are *composed* by superposition into a single visualization image. We chose a simple and rather abstract visualization approach to illustrate our point.

Figure 4b shows the intersection of ensembles (*set*) w.r.t. members (*element*), i.e., the most common members. The degrees were then used as weights to compute a weighted average of all ensemble members (*aggregation*). The map in Figure 4b shows this weighted average on a diverging color scale (*visualization*). As latent SBSS dimensions often show underlying physical processes [NOFR15], very high (red) and low (blue) values should highlight areas in Europe that are special with regard to various physical processes, such as climate, soil, land use or population density. And in fact it shows high values in south-west France (Figure 4b-A), which is an area that is distinct from its surroundings with regard to climate, soil, rocks, and land use, among others [HMZ09]. Similar arguments can be made for Norway or the Netherlands/Ruhr area (Figure 4b-B), which are also distinctive areas in many atlas maps.



(a)

(b)

Figure 4: *(a) Degree of association (y axis) of kernel radii (x axis) to features in Figure 3* (composition)*. (b) Cropped and annotated visualization* (aggregation) *showing the intersection of ensembles* (set) *w.r.t. members* (element)*. Size encodes distance from the central quintile. It suggests that, e.g, south-west France (A) is a common distinctive area. See Section 6 regarding other annotations.*

We showed Figure 4a to a statistics/SBSS expert along with a short introduction on how it should be read, that small kernels are less associated with all the features we defined (Figure 3). He speculated that small kernels do not capture enough points, and thus the estimation error in the SBSS procedure is too high. Indeed, when we investigated this, it became apparent that point neighborhoods of 50–150 km kernels are very small or empty. This changes from the 200 km kernel onwards. The "dip" at 550 km could be related to too large kernels, where the expert expects more noise to be captured, but this would require more research.

We then showed Figure 4b to an expert in geochemistry, who is very familiar with the dataset we used, noting that it shows a summary of the most common latent dimensions. Especially west Norway was interesting to our expert, as "[during the course of my research] this area repeatedly caught my eye as being special. Something is happening there." To investigate this, further research, e.g., getting more soil samples or conducting different experiments altogether, should be carried out. The high values in Spain (Figure 4b-C) and Ireland likely show mineral deposits, but would need to be compared to other maps to verify if the locations match. Our expert explained the patterns in south Poland and north Ukraine (Figure 4b-D) with glacial sediments, and in Greece (Figure 4b-E) with "the ophiolites with their high concentration of Nickel, Cobalt, Chromium, and Copper."

7. Discussion and Conclusion

Analysis of multi-ensembles is challenging as not only are ensemble members complex scientific datasets themselves, but there are now also potentially many ensembles. Existing visualization approaches focus on a small static collection of ensembles. As many analysis questions can be framed as set relations, we propose in this paper to model multi-ensembles as fuzzy relations. With those, set relations can be computed while simultaneously accounting for the complexity of ensemble members. We outlined our approach and illustrated it with a real-world example. Experts found both expected patterns in visualizations in Figure 4, that confirm known facts, and unexpected patterns, that spark further research. This is exactly the goal of VA [TC05]. Our approach thus seems very promising to be further explored in an interactive setting. We believe our proposal of fuzzy comparison to compute associations, then using fuzzy set theory to compute queries, can in principle easily adapted to other data types than those that our specific case required.

Some limitations apply. The choice of similarity measure is crucial. It should naturally map to the unit interval, which excludes some popular choices, e.g., Euclidean distance. Next, it must be "directly proportional" to the actual change in the similarity of two objects. Similarity measures that behave like sigmoids when plotted against the actual similarity change, or exhibit significant steps, are to be used cautiously. During fuzzy composition, two similarity measures may be multiplied, which we believe is fine when they show the properties mentioned earlier. Others may be more cautious than us and focus on dimensions that work on the same similarity measure (everything but *parameter setting* in our case). Regarding computational efforts, presented fuzzy set operations on fuzzy relations require simple column-wise operations, like $min()$, which are easily parallelized and scale well also to larger relations. More expensive are the pair-wise similarity computations, which pose a one-time $\mathcal{O}(N^2)$ effort for a static dataset.

Translating our approach to an interactive setting would induce interesting future work. Intuitive interactions, possibly based on direct manipulation, are necessary for analysts to build the desired fuzzy set relations. At the same time, one would need to ensure that the fuzzy calculations are transparent and explainable to human analysts. Revealing internals, like relation matrices, might be a start but insufficient. General visualization approaches exist for ensemble data, like dimension reduction [NA19], but few were suggested for fuzzy sets [PP10b, PP10a]. Combining and developing the respective ideas could yield powerful visualizations for large multi-ensemble datasets.

Acknowledgements This research was funded by the Austrian Science Fund (FWF) under grant P31881-N32.

References

[Ait82] AITCHISON J.: The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society 44*, 2 (1982), 139–177. 4

[BDA*17] BACH B., DRAGICEVIC P., ARCHAMBAULT D., HURTER C., CARPENDALE S.: A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes. *Computer Graphics Forum 36*, 6 (2017), 36–61. doi:10.1111/cgf.12804. 2

[CCM09] CORREA C. D., CHAN Y.-H., MA K.-L.: A framework for uncertainty-aware visual analytics. In *2009 IEEE Symposium on Visual Analytics Science and Technology* (Oct. 2009), pp. 51–58. doi:10.1109/VAST.2009.5332611. 3

[Cha19] CHAIRA T.: *Fuzzy Set and Its Extension: The Intuitionistic Fuzzy Set*, 1st ed. John Wiley & Sons, Inc., IEEE Xplore, Hoboken, New Jersey, [Piscataqay, New Jersey], 2019. doi:10.1002/9781119544203. 2, 3

[CKS*17] CIBULSKI L., KLARIN B., SOPOUCH M., PREIM B., THEISEL H., MATKOVIĆ K.: Super-Ensembler: Interactive visual analysis of data surface sets. In *Proceedings of the 33rd Spring Conference on Computer Graphics* (New York, NY, USA, 2017), ACM, pp. 1–9. doi:10.1145/3154353.3154362. 2

[FKRW17] 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 (Jan. 2017), 831–840. doi:10.1109/TVCG.2016.2598868. 1

[GCB*97] GRAY J., CHAUDHURI S., BOSWORTH A., LAYMAN A., REICHART D., VENKATRAO M., PELLOW F., PIRAHESH H.: Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. *Data Mining and Knowledge Discovery 1*, 1 (Mar. 1997), 29–53. doi:10.1023/A:1009726021843. 2

[HMZ09] HRNČIAROVÁ T., MACKOVČIN P., ZVARA I.: *Atlas Krajiny České Republiky*. Ministerstvo životního prostředí ČR, Výzkumný ústav Silva Taroucy pro krajinu a okrasné zahradnictví, Prague, 2009. 4

[KH13] KEHRER J., HAUSER H.: Visualization and Visual Analysis of Multifaceted Scientific Data: A Survey. *IEEE Transactions on Visualization and Computer Graphics 19*, 3 (Mar. 2013), 495–513. doi:10.1109/TVCG.2012.110. 1, 2

[KWS*15] KÖTHUR P., WITT C., SIPS M., MARWAN N., SCHINKEL S., DRANSCH D.: Visual Analytics for Correlation-Based Comparison of Time Series Ensembles. *Computer Graphics Forum 34*, 3 (June 2015), 411–420. doi:10.1111/cgf.12653. 2

[LCC*20] LI J., CHEN S., CHEN W., ANDRIENKO G., ANDRIENKO N.: Semantics-Space-Time Cube: A Conceptual Framework for Systematic Analysis of Texts in Space and Time. *IEEE Transactions on Visualization and Computer Graphics 26*, 4 (Apr. 2020), 1789–1806. doi:10.1109/TVCG.2018.2882449. 2

[LLX*10] LIU Y., LI Z., XIONG H., GAO X., WU J.: Understanding of Internal Clustering Validation Measures. In *2010 IEEE International Conference on Data Mining* (Sydney, Australia, Dec. 2010), IEEE, pp. 911–916. doi:10.1109/ICDM.2010.35. 3, 4

[MBN22] MUEHLMANN C., BACHOC F., NORDHAUSEN K.: Blind source separation for non-stationary random fields. *Spatial Statistics 47* (Mar. 2022), 100574. doi:10.1016/j.spasta.2021.100574. 2

[MGJ*10] MATKOVIĆ K., GRAČANIN D., JELOVIĆ M., AMMER A., LEŽ A., HAUSER H.: Interactive Visual Analysis of Multiple Simulation Runs Using the Simulation Model View: Understanding and Tuning of an Electronic Unit Injector. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov. 2010), 1449–1457. doi:10.1109/TVCG.2010.171. 1

[MLKS18] MIRANDA F., LINS L., KLOSOWSKI J. T., SILVA C. T.: TopKube: A Rank-Aware Data Cube for Real-Time Exploration of Spatiotemporal Data. *IEEE Transactions on Visualization and Computer Graphics 24*, 3 (Mar. 2018), 1394–1407. doi:10.1109/TVCG.2017.2671341. 2

[NA19] NONATO L. G., AUPETIT M.: Multidimensional Projection for Visual Analytics: Linking Techniques with Distortions, Tasks, and Layout Enrichment. *IEEE Transactions on Visualization and Computer Graphics 25*, 8 (Aug. 2019), 2650–2673. doi:10.1109/TVCG.2018.2846735. 5

[NOFR15] NORDHAUSEN K., OJA H., FILZMOSER P., REIMANN C.: Blind Source Separation for Spatial Compositional Data. *Mathematical Geosciences 47*, 7 (Oct. 2015), 753–770. doi:10.1007/s11004-014-9559-5. 2, 4

[PBG*22] PICCOLOTTO N., BÖGL M., GSCHWANDTNER T., MUEHLMANN C., NORDHAUSEN K., FILZMOSER P., MIKSCH S.: TBSSvis: Visual Analytics for Temporal Blind Source Separation. *Visual Informatics 6*, 4 (2022), 51–66. doi:10.1016/j.visinf.2022.10.002. 2

[PBM*22] PICCOLOTTO N., BÖGL M., MUEHLMANN C., NORDHAUSEN K., FILZMOSER P., MIKSCH S.: Visual Parameter Selection for Spatial Blind Source Separation. *Computer Graphics Forum 41*, 3 (2022), 157–168. doi:10.1111/cgf14530. 2

[PJ09] PARK H.-S., JUN C.-H.: A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications 36*, 2 (Mar. 2009), 3336–3341. doi:10.1016/j.eswa.2008.01.039. 3

[PP10a] PARK Y., PARK J.: Disk Diagram: An Interactive Visualization Technique of Fuzzy Set Operations for the Analysis of Fuzzy Data. *Information Visualization 9*, 3 (Sept. 2010), 220–232. doi:10.1057/ivs.2010.3. 5

[PP10b] PARK Y., PARK J.: Interactive visualization of fuzzy set operations. In *IS&T/SPIE Electronic Imaging* (San Jose, California, Jan. 2010), Park J., Hao M. C., Wong P. C., Chen C., (Eds.), p. 753002. doi:10.1117/12.840229. 5

[RBD*14] REIMANN C., BIRKE M., DEMETRIADES A., FILZMOSER P., O'CONNOR P.: *Chemistry of Europe's Agricultural Soils. Part A: Methodology and Interpretation of the GEMAS Data Set*, vol. Heft 102 of *Geologisches Jahrbuch*. Bundesanstalt für Geowissenschaften und Rohstoffe (BGR), Hannover, 2014. 2

[RFGD08] REIMANN C., FILZMOSER P., GARRETT R., DUTTER R.: *Statistical Data Analysis Explained*, first ed. John Wiley & Sons, Ltd, 2008. doi:10.1002/9780470987605. 4

[RST*00] REIMANN C., SIEWERS U., TARVAINEN T., BITYUKOVA L., ERIKSSON J., GILUCIS A., GREGORAUSKIENE V., LUKASHEV V., MATINIAN N. N., PASIECZNA A.: Baltic soil survey: Total concentrations of major and selected trace elements in arable soils from 10 countries around the Baltic Sea. *Science of The Total Environment 257*, 2 (Aug. 2000), 155–170. doi:10.1016/S0048-9697(00)00515-5. 4

[TC05] THOMAS J. J., COOK K. A.: *Illuminating the Path: An R&D Agenda for Visual Analytics*. National Visualization and Analytics Center, Department of Homeland Security, IEEE Computer Society, 2005. 5

[WFW*17] WANG Z., FERREIRA N., WEI Y., BHASKAR A. S., SCHEIDEGGER C.: Gaussian Cubes: Real-Time Modeling for Visual Exploration of Large Multidimensional Datasets. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (Jan. 2017), 681–690. doi:10.1109/TVCG.2016.2598694. 2

[WHLS19] 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 (Sept. 2019), 2853–2872. doi:10.1109/TVCG.2018.2853721. 1, 2, 3

[WLSL17] 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 (Jan. 2017), 81–90. doi:10.1109/TVCG.2016.2598830. 2

[XLWD19] XIE C., LI M., WANG H., DONG J.: A survey on visual analysis of ocean data. *Visual Informatics 3*, 3 (Sept. 2019), 113–128. doi:10.1016/j.visinf.2019.08.001. 2