ISSN 2464-4617 (print)
ISSN 2464-4625 (online)
Computer Science Research Notes - CSRN
http://www.wscg.eu
WSCG 2025 Proceedings

Improving Comparability of Temporal Evolution in 2D Embeddings of Ensemble Data

Simon Leistikow
University of Münster
Einsteinstraße 62
48149 Münster, Germany
simon.leistikow@uni-muenster.de

Vladimir Molchanov
University of Münster
Einsteinstraße 62
48149 Münster, Germany
molchano@uni-muenster.de

Lars Linsen
University of Münster
Einsteinstraße 62
48149 Münster, Germany
linsen@uni-muenster.de

ABSTRACT

Ensemble data refers to a set of simulations or measurements of the same phenomenon conducted under different settings such as varying control parameters of the simulation. When the data capture changes over time, the main analysis task is to compare the temporal evolution of the ensemble members.For a global analysis of the ensemble's temporal evolution, the data can be embedded in a 2D visual space, where each ensemble member is visualized by a 2D curve. For the embedding, the dissimilarities of the states of the ensemble members at different points in time shall be represented by the distances of the respective points in the 2D embedding. This goal is achieved by minimizing the stress functional of Multidimensional Scaling (MDS). However, direct visual comparison of the embedded curves can be complicated due to their non-trivial geometry. We propose an algorithm for the simplification of curve geometries in 2D MDS embeddings of ensemble data, where the target shape of a selected reference curve can be chosen by the user. We transform the original embedding in a controllable way consistent with the MDS stress functional. Thus, we improve the comparability of curves representing individual ensemble members, while minimizing the loss in accuracy of presented data dissimilarities. We perform several tests demonstrating the benefits of the proposed method, present our findings, and discuss numerical aspects of the algorithm.

Keywords

Ensemble data, Multidimensional Scaling

1 INTRODUCTION

Ensemble data refers to repeated measurements or simulations under varying initial conditions or parameterizations. A key task in ensemble data analysis is to explore how ensemble members' characteristics depend on the initial setup. Data for each member are often collected over time, with significant attributes recorded at each step. Thus, ensemble data are collections of evolving multidimensional datasets, each tied to specific initial conditions. Examples include:

- Simulations of physical phenomena such as weather, climate, blood flow, or medical treatments, performed under varying initial conditions or prediction models.
- Longitudinal cohort studies, for example, in epidemiology, healthcare, and demographics.

Ensemble data analysis involves detecting outliers, identifying qualitatively different regimes, and predicting critical parameter settings where ensemble behavior changes. These insights are useful for detecting simulation flaws, planning experiments, and predicting behavior for unexplored setups. Visualizing ensemble data is invaluable for these tasks, enabling comprehensive overviews and clear illustrations of findings.

Each time step of an ensemble member is represented by multidimensional data, which must be embedded in a lower-dimensional (usually 2D) space for visualization. The choice of embedding method depends on analysis goals, as different algorithms preserve different data characteristics. Multidimensional Scaling (MDS) [BG05] is a classical approach that represents similarities between multidimensional samples as distances in lower-dimensional space. It is ideal when the primary goal is to examine resemblance and disparity between objects.

Large ensemble datasets, with dense temporal sampling, can result in high stress values in MDS scatter-

plots, making visual analysis error-prone. However, temporal proximity is often more important than comparing states at distinct times. To address this, a temporal windowing approach can be applied to the stress function, emphasizing similarities between temporally close points. This modification improves the accuracy of visual analysis for temporally localized behaviors.

In MDS scatterplots, each observation is shown as a point, and points from consecutive time steps are connected (e.g., with line segments or spline curves) to indicate temporal progression. Each ensemble member is thus represented as a directed polyline or smooth curve, whose geometry reflects recorded temporal changes. Comparing ensemble members then becomes the task of comparing the positions and shapes of their 2D curves in the embedding.

These curves can exhibit complex geometries, such as flat segments, high-curvature regions, loops, or self-intersections, which hinder intuitive comparisons. In many cases, a distinguished reference ensemble member serves as a baseline for comparison, such as a median sample or ground-truth data. Simplifying the reference curve's geometry can make comparisons more intuitive. However, constraining the reference curve's shape conflicts with the classical MDS goal of stress minimization.

We propose an algorithmic approach to modify the reference curve's shape controllably. Users can rely on automatic simplification or sketch the desired geometry. The stress-minimization procedure then recomputes the projection while respecting the reference curve's constraints. This achieves an interactive trade-off between preserving temporal similarity and improving curve comparability. We demonstrate the method's effectiveness and discuss its numerical aspects through several experiments. The source code is available on GitHub [Lei25].

The remainder of the paper is structured as follows. First, we discuss relevant terminology and related methods in Section 2. In Section 3, we extend related methods to create a framework summarized in Section 3.4. Experiments with synthetic and real-world data are carried out and discussed in Section 4 where we also present. Finally, we conclude our paper in Section 5.

Our contributions can be summarized as follows.

1. We present an algorithmic simplification approach that allows to modify an ensemble member's projection in a 2D embedding obtained with Multidimensional Scaling and re-project the other members' projection with respect to the stress functional.
2. We evaluate our approach on synthetic data and apply it on the astroid impact ensembles data.
3. We provide an open-source implementation for reproducibility.

2 RELATED WORK

Ensemble Data. In physics and engineering, ensemble data are commonly used to study the sensitivity of simulations to varying control parameters. Over the years, numerous visual analysis techniques have been proposed, focusing on statistical descriptors [PWB+09b], uncertainty in weather forecasts [PWB+09a], and isocontour similarities in spatio-temporal data [FML16]. Kehrer and Hauser introduced the broader term *multi-faceted data* [KH13] to emphasize the complexity of modern data in terms of dimensionality, structure, format, and sources.

Dimensionality Reduction. Dimensionality reduction (DR) simplifies complex data by creating low-dimensional representations that preserve essential characteristics, making data easier to explore, understand, and visualize. DR is a key preprocessing step in modern data analysis, with the choice of method depending on application goals, data type, and size [VDMPVdH09].

DR techniques are broadly classified into linear and non-linear methods. Linear methods, such as Principal Component Analysis (PCA) [AW10] and Star Coordinates [ML19], apply linear transformations. Non-linear methods, such as t-Distributed Stochastic Neighbor Embedding (t-SNE) [vdMH08] and Uniform Manifold Approximation and Projection (UMAP) [MHSG18], preserve neighborhood relationships. Multidimensional Scaling (MDS) preserves pairwise similarities or dissimilarities and is computationally efficient compared to techniques such as Diffusion Maps [CL06] and Laplacian Eigenmaps [BN03]. Consequently, MDS is well-suited for analyzing similarities in multidimensional ensemble datasets.

MDS Algorithms. The stress functional measures how well distances in the MDS embedding represent original data dissimilarities, and MDS algorithms aim to minimize this stress. Minimization is typically iterative, with methods such as the Newton-Raphson approach [Bro87] and *Scaling by Majorizing a Complicated Function* (SMACOF) [BG05], the latter offering fast convergence. Borg et al. [BG05] provide a comprehensive overview of MDS algorithms, including variants for non-metric spaces (where dissimilarities are replaced by ranks) and Landmark MDS [LC09], which processes large datasets by embedding a representative subset and mapping remaining samples locally. Recent advances, such as Uncertainty-Aware MDS [HKW23] and Weighted MDS [MBR16], introduce flexibility by incorporating uncertainties or assigning weights to pairwise dissimilarities, which makes them particularly relevant to this work. We also carried out experiments

with Lagrange Multipliers [Ber82] and Spring Layouts [FR91], however, none of them could deliver compelling results in reasonable computing time. We still integrated them as drop-in replacement in the published source code. For the remainder of the paper, we focus on SMACOF as this is currently the only method suitable for maintaining an interactive approach.

Geometry Simplification. Our approach incorporates shape regularization of 2D curves. High-frequency features can be suppressed using smoothing techniques like the Savitzky-Golay (Savgol) filter [SG64], where the window size determines the level of smoothing. However, smoothing may alter the start and end points of curves. Alternatively, algorithms like Visvalingam-Whyatt [VW93] and Douglas-Peucker [DP73] approximate curve geometry. We adopt the Douglas-Peucker algorithm for its computational efficiency and high-quality results. Users can select the simplification strategy and fine-tune parameters (e.g., window size or epsilon) to achieve desired results.

3 IMPROVING COMPARABILITY OF TEMPORAL EVOLUTION IN MDS PLOTS

In this section, we introduce necessary notations and definitions used in SMACOF theory, formulate the proposed solution in terms of minimization of weighted stress functional and discuss various practical aspects of the algorithm.

3.1 Notations and Definitions

Let $\Delta = \{\delta_{ij}\}_{1\le i,j\le n}$ be a symmetric matrix, whose elements δ_{ij} are pairwise dissimilarities of observations in an ensemble dataset. The total number of observations in all ensemble members is n and we assume, that an ensemble member's number and temporal position of each observation can be uniquely deduced from its index. MDS aims to place a set of points $\mathbf{x}_i \in \mathbb{R}^k$, $1 \le i \le n$, so that their pairwise distances $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ match given dissimilarities δ_{ij}. Mathematically, the problem of locating n points $\mathbf{X} = \{\mathbf{x}_i\}$ can be expressed as minimization of following non-negative functional called *stress*

$$\sigma(\mathbf{X}) = \sum_{i<j} \omega_{ij} \left(\delta_{ij} - d_{ij}(\mathbf{X})\right)^2, \qquad (1)$$

where $\omega_{ij} \ge 0$ are some weights. If $\omega_{ij} = 1$, all observations and all pairwise distances between them are treated equally. In practice, however, there may exist a priority reflecting, e.g., uncertainty of measurements, importance of distinct observations or missing data. In our setting, observations belonging to a distinctive ensemble member that we refer to as reference member play a special role. Therefore, possibility to adjust ω_{ij} provides a great flexibility in adapting classical MDS approach to specific problems.

Minimization of stress (1) can be done by an iterative SMACOF approach. In every iteration, the algorithm majorizes the stress by a quadratic function and computes its critical point using the linear algebra calculus. Then, the algorithm results in a series of gradually refined configurations of $\mathbf{X}$ corresponding to non-increasing values of the stress. We refer the reader to [BG05] for detailed exposition of the method.

Iterations of the SMACOF algorithm stop, either after a maximal number of iterations was reached or when the series of configurations of $\mathbf{X}$ converged. Normalized stress

$$\sigma_1(\mathbf{X}) = \left(\frac{\sum_{i<j} \omega_{ij} \left(\delta_{ij} - d_{ij}(\mathbf{X})\right)^2}{\sum_{i<j} \omega_{ij} \delta_{ij}^2} \right)^{1/2} \qquad (2)$$

allows for assessing the quality of intermediate distributions $\mathbf{X}$ during iterations as well as of the final result. Note that normalized stress σ_1 differs from raw stress σ only by a constant factor and a monotone transformation. Therefore, both functionals share extreme points and minimization of one is equivalent to minimization of another. However, σ allows for an efficient minimization using the SMACOF algorithm, while σ_1 is independent of the absolute values of dissimilarities. A guideline introduced by Kruskal et al. [Kru64] suggests the following evaluations of the fit goodness depending on the normalized stress value: perfect for $\sigma_1 \le 0.025$, excellent for $0.025 < \sigma_1 \le 0.05$, good for $0.05 < \sigma_1 \le 0.1$, fair for $0.1 < \sigma_1 \le 0.2$, poor for $0.2 < \sigma_1$. Mair et al. [MBR16] criticized Kruskal's evaluation, however, formula for σ_1 in their paper differs from the standard definition.

When dimensionality of the embedding $k = 2$, solution $\mathbf{X}$ can be visualized as a set of points, so that each $\mathbf{x}_i$ is *projection* of i-th observation from the ensemble dataset. Projections of observations from the same ensemble member are usually connected by curves according to their temporal order. Throughout this paper, we use stress (1) to compute 2D MDS projection. When not specified differently, $\omega \equiv 1$. For evaluation of the projection quality, we apply normalized stress (2).

3.2 Simplification of the Reference Curve

MDS optimally transforms dissimilarities of multidimensional samples into 2D distances between respective projected points. Thus, closeness of the projected points can be interpreted as similarity of the corresponding observations. When projecting evolutionary data, points representing consequent observations are usually connected in the projected space, so that an emerging curve depicts the observed process. Then, various patterns of the curve geometry like cycles, high-curvature segments or vanishing tangential vectors can

be interpreted as periodic behavior, rapid changes or stationary points of the studied process.

Visualization of time-varying ensemble data results in a collection of 2D curves, each representing a single ensemble member or a set of observations. Comparing individual ensemble members between each other can be a tedious task due to complicated geometries and relative positions of the corresponding curves. To mitigate the problem, we propose to simplify the geometry of a distinguished curve representing what we call a reference ensemble member (or reference member, for short). When the shape of the reference curve becomes close to a straight line, a circle or an arbitrary smooth regular curve, other curves can be easier characterized to have similar or distinct shape. Therefore, visual clustering and outlier detection can be performed for all ensemble members shown in the projected space as curves.

Reasons for the selection of a particular reference ensemble member can be many-fold, e.g., it can be a ground truth data within a set of simulations, an ensemble member with the most reliable data, or just the most simple 2D curve selected by the user. Irrespective to why a particular ensemble member was chosen, our task is to transform the curve geometry towards the desired simplified shape and simultaneously preserve acceptable level of the stress function. I.e., we aim at an optimal trade-off between comparability of the ensemble members and accuracy of the modified MDS projection.

Without loss of generality, we assume that $\{\mathbf{x}_i\}_{1\le i\le l}$ are the MDS projections of the observations belonging to the reference member. Then, the curve passing through these points should be simplified. In our work, we consider two possible scenarios: First, the given geometry may be automatically simplified by a geometric algorithm or, second, a desired simple geometry can be manually sketched by the user in the projected space. In the former case, one may apply the Douglas-Peucker algorithm [DP73], which regularizes a given polyline while keeping its end points. In the latter case, the user interactively sketches a 2D curve, which ideally represents a smooth approximation to the reference curve in the MDS projection layout. Alternatively, the user may place a number of control points, which will be used to reconstruct a cubic-spline curve.

Let c stand for a curve either generated by the automatic approach or specified by the user. Our task is to force $\{\mathbf{x}_i\}_{1\le i\le l}$ to be placed along c, while preserving the pairwise distances as good as possible. Therefore, we run an optimization procedure, which minimizes functional

$$\sigma^{\text{ref}}(\mathbf{X}) = \sum_{1\le i,j\le l} \omega_{ij}^{\text{ref}} \left(\delta_{ij} - d_{ij}(\mathbf{X})\right)^2, \qquad (3)$$

for $\{\mathbf{x}_i\}_{1\le i\le l} \subset c$. For constraining positions of points along the reference curve, we use a continuous parametrization of c and Sequential Least-Squares Programming algorithm [Kra88]. During optimization, we ensure that temporal dimension remains consistent with the curve parametrization, i.e., points corresponding to later time moments get assigned higher values of the curve parameter. We also fix $\mathbf{x}_1$ to the start of c.

Here, we may use the flexibility of the weighted MDS method by utilizing ω_{ij}^{ref}. For example, instead of setting $\omega_{ij}^{\text{ref}} \equiv 1$, one may choose higher values of ω_{ij}^{ref} for smaller values of $|i-j|$, which would mean that similarities between close timesteps more strongly affect the placement of points along c. Thus, weighted MDS allows, in particular, for balancing local versus global similarities in the temporal dimension.

Quality of the fitting the reference points on c can be evaluated by means of normalized stress σ_1^{ref} constructed for σ^{ref} by analogy with Expression (2). Depending on the geometry of the specified c, minimization procedure can result in an unacceptable high value of σ_1^{ref}, meaning that the introduced distortion is too high. In such case, we report σ_1^{ref} to the user and let them decide, whether to continue with the current reference curve or repeat computation for new c. As soon as the placement of the reference points along the reference curve is accepted, we proceed with the next phase of the proposed algorithm.

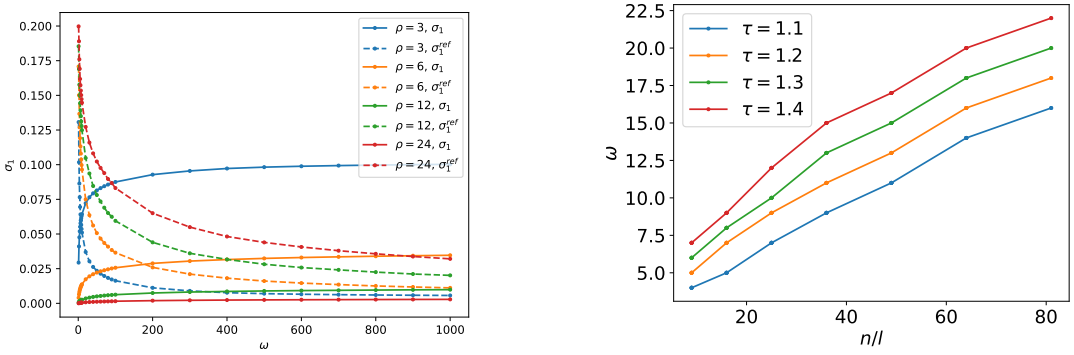
3.3 Data Reprojection and Optimal Weight

Let $\{\mathbf{q}_i\}_{1\le i\le l}$ be the solution to the minimization of stress (3). Our test is now to modify the weighted MDS stress function (1) so that the optimal configuration of the reference points would approach $\{\mathbf{q}_i\}$ while keeping the stress value low. We achieve this goal in two steps. First, we redefine distances between reference observations as follows:

$$\delta_{ij}^* = \begin{cases} \|\mathbf{q}_i - \mathbf{q}_j\| & \text{for } 1 \le i,j \le l, \\ \delta_{ij} & \text{otherwise,} \end{cases} \qquad (4)$$

i.e., we replace the actual dissimilarities between reference observations by the 2D distances between desired projections on the reference curve c. Second, we set

$$\omega_{ij}^* = \begin{cases} \omega & \text{for } 1 \le i,j \le l, \\ \omega_{ij} & \text{otherwise,} \end{cases}$$



(a) (b)

Figure 1: Numerical experiments for the evaluation of stress convergence and optimal weight. (a) Experiment showing the typical stress evolution for both σ_1 and σ_1^{ref} depending on ω. The target curve was $t(x) = 0$ in this case. (b) Experiment showing the optimal ω for a given stress tolerance τ. A proportional relationship can be observed between ω and the ratio n/l, where n is the number of time steps of the entire ensemble and l is the number of time steps of the reference curve.

where $\omega \gg \max \omega_{ij}$ is some constant. Putting all together, we look for a configuration of projected points minimizing following stress functional

$$\begin{aligned} s(\mathbf{X}) = & \sum_{i<j} \omega_{ij}^{*} \left(\delta_{ij}^{*} - d_{ij}(\mathbf{X})\right)^2 \qquad (5) \\ = & \ \omega \sum_{1 \le i < j \le l} \left(\|\mathbf{q}_i - \mathbf{q}_j\| - d_{ij}\right)^2 + \\ & \sum_{\substack{i<j \\ i>l \text{ or } j>l}} \omega_{ij} \left(\delta_{ij} - d_{ij}\right)^2 . \end{aligned}$$

An efficient solution to the presented minimization problem can be obtained by the SMACOF method. The role of factor ω is to weigh the constraint of the reference: For small ω, stress function s remains very close to stress σ given by Expression (1); the larger ω is, the more contribution positions of the reference points make to stress (5), thus, the more their placement resembles the relative positions of $\{\mathbf{q}_i\}$.

To find the optimal value of the weighting factor ω, we inspect the behavior of normalized stress σ_1 (2) and respective normalization σ_1^{ref} evaluated for the sequence of solutions as ω increases. While σ_1^{ref} monotonically decreases, σ_1 may increase due to growing distortion with respect to the original MDS projection. Therefore, placement of the reference points approaches their desired configuration along c at the cost of dropping accuracy as ω grows. Monitoring both stress functionals σ_1^{ref} and σ_1, the user determines such value of the weight, for which both errors lie within acceptable ranges, or restarts the procedure with newly defined reference shape c. The typical evolution of the stress functionals is depicted in Figure 1a. Although the weighting factor ω could be chosen directly by the user, it is difficult to determine its effect on the stress functionals. Instead, we rather propose to find an optimal ω for a tolerated error τ, i.e., a percentage of the stress of the initial MDS projection (σ_1^{init}), the choice of which we leave to the user, as it is specific to the data and task at hand. To find the optimal ω for a given τ, we search the $\omega \in \mathbb{R}^{\geq 1}$ for which the condition $s_\omega = \tau \cdot \sigma_1^{\text{init}}$ is satisfied. We achieve that by performing a binary search on the alternative bound integer domain $\{1, \ldots, \theta\}$ where θ is some high numerical value. In our experiments, $\theta = 10^3$ was sufficient. Depending on the data set, target curve, and the value of τ, the tolerance may never be reached, leading to an "infinite" ω. As consequence, the reference member will be deformed to the target curve while the rest of the ensemble remains mostly unchanged in the projection.

3.4 Algorithm

Summarizing the exposition of the last two subsections, the proposed algorithm can be described as follows:

1. In the 2D MDS projection (stress σ_1^{init}) of time-varying ensemble data, select a reference member.
2. Define a shape of the reference curve by running an automatic procedure or sketching the curve manually.
3. Compute optimal placement of the reference points along the reference curve by minimizing stress (3). Improve c if the goodness of fit is unsatisfactory.
4. Update dissimilarity matrix Δ according to Equation (4).
5. Compute a series of solutions minimizing stress (5) with varying weight ω using the SMACOF approach.

6. Determine an optimal solution, for which both quality measures σ_1 and σ_1^{ref} are acceptable. Here, the user may specify a tolerance τ (or acceptable error $\tau \cdot \sigma_1^{\text{init}}$), for which the optimal ω is computed. Alternatively, the user may specify a value for ω.

3.5 Implementation Notes

The simplest geometry of reference curve c is a straight line. In this particular case, minimization of functional (3) is equivalent to computing 1D MDS projection for the set of reference observations $\{\mathbf{x}_i\}_{1\leq i\leq l}$.

Usually, the MDS minimization routine is repeated with a number of randomized starting configurations in order to deal with the problem of local minima. We observed that the best solution (obtained with SMACOF) may have properties undesired for the interpretation of the temporal evolution of ensemble members. I.e., individual time steps may not follow a smooth curve. To avoid these artifacts, we compute the initial embedding using the method presented by Wickelmaier et al. [Wic03], which better preserves smooth curves in the embedding. As a trade-off, the stress may not be minimal. To find the optimal ω, however, we are required to start with minimal stress to guarantee a monotonously increasing sequence of stress values in order to apply the binary search. Further optimizing this initial embedding using SMACOF then minimizes the stress while better preserving smooth curves than when using random initializations. This optimized initial embedding can also be used to start iterations of the SMACOF algorithm in Step 5 of the proposed method. Though, several optimization runs are required to determine the optimal value for weight ω. The overall complexity of the presented method remains comparable to a single computation of MDS projection.

The proposed algorithm can be easily adapted to the situation when several ensemble members are selected as reference and, respectively, several distinct curves are specified. In this case, Expression (3) should be adapted accordingly. Detailed description is left beyond the scope of this paper.

4 RESULTS

In this section, we apply our approach to both an illustrative data set and a real world data set and report our findings. For a demonstration of how the algorithm could be implemented and how our results were generated, we refer to the accompanying video.

4.1 Illustrative Example

To understand the effect of our approach and especially how the simplification of the reference member affects the other ensemble members in the embedding, we define a set of three functions that contains

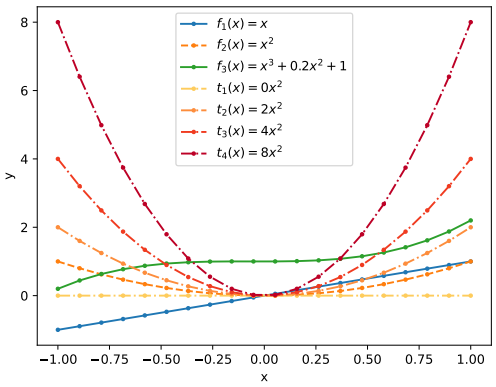


Figure 2: The illustrative dataset we use for our experiments. The dataset contains three 2D-curves, i.e., a linear function (f_1, blue), a quadratic function that serves as references (f_2, dashed orange) and a cubic function (f_3, green). Additionally, we depict the target curves (t_1, t_2, t_3, t_4) used for our experiments.

1. a linear function: $f_1(x) = x$,
2. a quadratic function: $f_2(x) = x^2$, and
3. a cubic function: $f_3(x) = x^3 + 0.2x + 1$.

From each function f_i, $i \in \{1,2,3\}$ we generate a 2D curve $c_i = \{p_j\}_{1\leq j\leq n}$ with n being the number of time steps, and $p_j = (d(j), f_i(d(j)))$ with $1 \leq j \leq n$ and $d(j) = 2 \cdot ((j-1)/(n-1)) - 1$, simply mapping the time step to the domain $[-1,1]$

The resulting curves are depicted in Figure 2. Here, we choose function f_2 to be our reference and represent it (here and in the following) as dashed line.

Optimal Weight In order to estimate the optimal weighting factor ω for a given stress tolerance τ, we consider (5) again in more detail. Let n be the number of time steps of the ensemble and l be the number of time steps of the reference run, then we define $\rho = n/l$ to be the ratio of all optimized distances and the ones that are affected by ω. We extended the illustrative data set described above by adding curves defined by functions of the form $f_i(x) = x^i + (-1)^i \cdot i$ in order to obtain certain ratios $\rho = n/l$. As target curve we chose $t(x) = 0$. We also define different levels of acceptable stress τ and find the optimal ω as described above. In Figure 1b, we plot the resulting ω over the ratio *rho* and observe a proportional relationship. The same experiment was conducted for another target shape, i.e., $t(x) = 8x^2$ the result of which is identical. We conclude that the target shape has no effect on the relationship of ω and ρ.

Effect of Number of Time Steps Our experiments show that the number of time steps per ensemble member has no impact on any of the considered stress functionals, given the ratio of number of time steps between

the members remains the same. A down-sampling of the temporal domain by a constant factor hence has no impact on the optimal weighting factor ω, as long as the number of time steps can still reasonably represent the structure of the data.

Effect of Number of Ensemble Members Again, consider a variation in the number of ensemble members or, in other words, a change of ratio ρ. Figure 1a shows the resulting σ_1 and σ_1^{ref} stress depending on the weighting factor ω for different ratios. We observe that in order to minimize σ_1^{ref} stress for higher values of ρ, a higher value has to be chosen for ω. At the same time, the σ_1 stress is preserved better. This meets our expectations as the number of distances that have to be preserved between the runs that are competing against the reference curve's distances in the optimization procedure is much higher.

Effect of Difference from the Original Curve To analyze the effect that an increasing stress has on the weight, we set our target curve to be a quadratic function of the form $f(x) = \alpha x^2$, depicted in Figure 2. We apply our method for α set to 0, 2, 4, and 8. Figure 1a shows the resulting stresses for $\alpha = 0$, i.e., $t(x) = 0$. For other values of α, i.e., an increased stress between original reference and target curve, the evolution is very similar, however the initial σ_1 stress is already much higher, increasing more rapidly when increasing when increasing α. Analogously, the σ_1^{ref} stress starts with higher value but decreases faster. The stress (or α) has, however, no effect on the optimal weighting factor ω.

Effect of Weight In summary, increasing the weighting factor ω leads to a decrease of σ_1^{ref} as weighted MDS preserves the respective distances more than the unweighted distances. At the same time, σ_1 is increased. Fig 1a shows that effect by example of our illustrative data set and the target curve that was set to $t(x) = 0$. We repeated the same experiment for all target curves described above (cf. Figure 2) and observed the same behavior. The resulting stresses for σ_1^{ref} and σ_1 were all strictly monotonously decreasing and increasing, respectively. To understand how the stresses realize in the actual embeddings, we observe two different target curves, $t(x) = 0$ and $t(x) = 8x^2$ for different ω, depicted in Figure 3.

Typical Stress Evolution The SMACOF method guarantees a strictly monotonously decreasing sequence of the optimized stress (cf. Equation (5)). The σ_1^{ref} stress is implicitly optimized as well and will end up minimized after convergence of SMACOF, however, may fluctuate during the process. The same is true for original MDS stress which, instead, ends up increased since already the replacement of original distances (cf. Equation (4)) introduces higher stress which is further increased in the optimization. For an illustration of the stress functionals during optimization, we refer to the accompanying video.

4.2 Deep Water Asteroid Impact Data Set

We also applied the approach on real-world data, i.e., the Deep Water Asteroid Impact Ensemble [GP17]. This data set contains 7 simulation runs initialized with different parameter configurations for the asteroid's size, incoming angle and height of the airburst, if any. The initial embedding is depicted in Figure 4a, where the start of the temporal evolution of the reference curve is depicted as square in the plot. We then applied the automatic simplification procedure by means of Douglas-Peuckert. The epsilon parameter was set to 0.05 and the resulting control points were used to fit a cubic spline, depicted as gray curve. The result of our approach, ω set to 50, is depicted in Figure 4b. The reference run is transformed to the target shape. Interestingly, the temporal evolution of the other ensemble members' curves remains mostly unchanged. The stress between the original and the target reference curve is low enough to maintain the global structure.

4.3 Complexity and Limitations

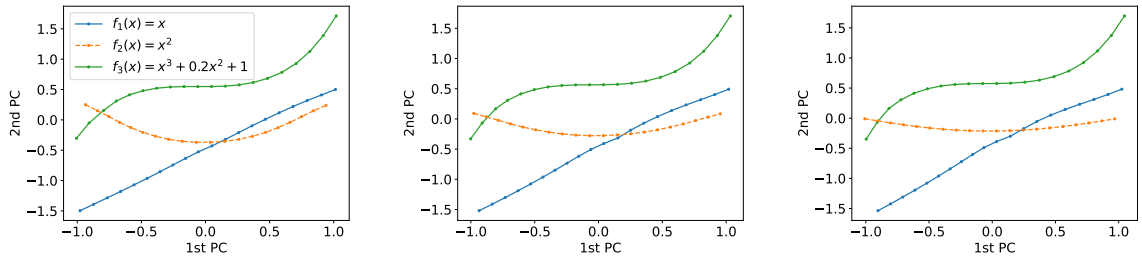
Table 1: Summary of Algorithmic Worst-Case Complexities. n and n_{ref} are the number of time steps of the ensemble and reference run, respectively. d is the dimensionality of the projection, k and k' are the number of iterations needed for the respective steps.

Algorithm	Worst-Case Complexity
Douglas-Peucker	$\mathcal{O}(n^2)$
SLSQP	$\mathcal{O}(n_{\text{ref}}^3 \times k)$
SMACOF (MDS)	$\mathcal{O}(n^2 d \times k')$

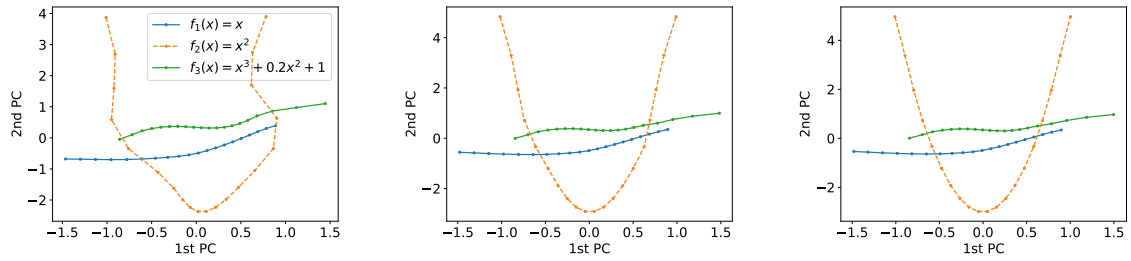
The algorithmic complexities of the individual steps can be seen in Table 1. When we assume that $n_{\text{ref}} \ll n$ $d \ll n$, $k \ll n$ and $k' \ll n$ hold, the total complexity can be summarized as $\mathcal{O}(n^2)$, however, the selection of algorithms for the steps was made in favor of interactivity of the implementation, where runtime is more relevant than complexity.

The runtime of the proposed algorithm depends on the input data, as well as the choices made by the user. Both finding the parameterization of the reference curve (cf. (3)) and minimizing the stress functional using SMACOF (cf. (5)) are iterative methods that stop after a certain amount of iterations is reached (we choose 100 throughout all experiments) or the delta of two consecutive stresses falls below a threshold (10^{-5} in our case). For that reason, it is not trivial to estimate a general runtime of our approach. However, we demonstrate the runtimes of the individual steps in the accompanying video.

Regarding the applicability of our approach, we want to discuss the following limitations. First, our approach

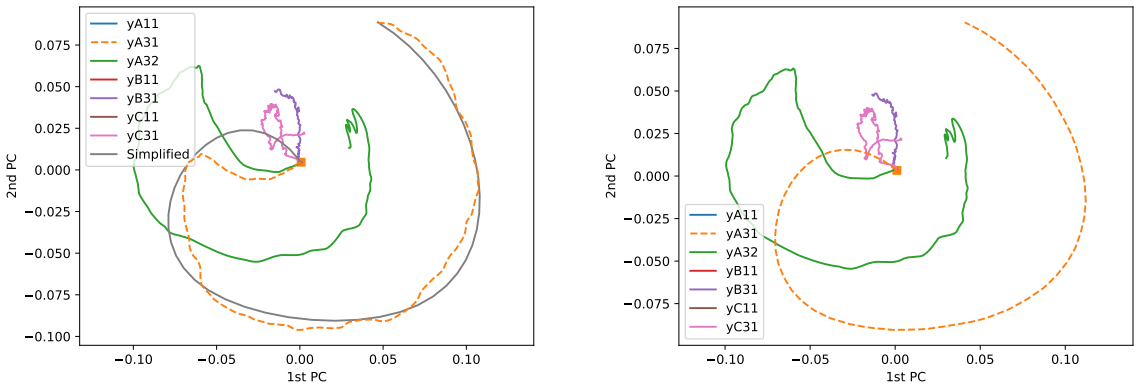


(a) $t(x) = 0, \omega = 5$ (b) $t(x) = 0, \omega = 50$ (c) $t(x) = 0, \omega = 500$



(d) $t(x) = 8x^2, \omega = 5$ (e) $t(x) = 8x^2, \omega = 50$ (f) $t(x) = 8x^2, \omega = 500$

Figure 3: Depiction of our approach applied on synthetic data and different target curves $t(x)$ (see Figure 2) for the reference run (dashed orange curve), and different weights ω. The number of time steps is 20. The horizontal and vertical axes represent the first and second principal component, respectively.



(a) (b)

Figure 4: Depiction of our approach applied to the Deep Water Asteroid Impact Ensemble [LHFL19]. (a) Classical 2D MDS and simplified target curve (gray) for the reference member (dashed orange curve), obtained by automatic geometry simplification. (b) Embedding achieved using our approach with ω set to 50. The horizontal and vertical axes represent the first and second principal component, respectively.

should only be used, if the intrinsic dimensionality of the data is two or larger. Otherwise, a 1D MDS embedding with time as horizontal axis should be preferred. When the simplified curve is quite different from the original, e.g., because of how the user sketched the curve, the start point may detach from the original position, even though we fix its position. One possible explanation could be, that the higher weighted intra-reference distances are pulling it away. Two special cases may render sketching a simplified curve difficult, that is (1) if an ensemble member has a periodic evolution and hence will be projected to a repeated cycle that overlaps with itself and (2) if an ensemble member happens to be static, i.e., will be projected to a single point. In these cases, high reference stress will be induced. Lastly, the simplified curve is drawn in the low-dimensional space from which we derive the distances by which we replace the original reference curve

distances. It may be more optimal to allow for geometric simplification in high-dimensional space, which we leave beyond the scope of this paper.

5 CONCLUSION AND FUTURE WORK

In this paper, we proposed an algorithmic approach for modifying the shape of a reference curve in 2D MDS embeddings with the goal of geometric simplification. In a controllable way, accuracy of the embedding is traded for comparability between members and the reference. The reference curve can be modified by the user using an automatic approach or by sketching a curve in the low-dimensional embedding. The approach was evaluated with synthetic data and applied on the deep water asteroid impact ensemble data set. This could demonstrate the geometric simplification of the reference run compared to the projection obtained by classical MDS. In the future, we want to extend the approach to multiple reference members. Additionally, we want to impose additional constraints such as applying higher weights to better preserve the temporal domain and analyze the induced trade-off between accuracy and comparability.

6 ACKNOWLEDGMENTS

This work was funded by Deutsche Forschungsgemeinschaft (DFG) grants LI 1530/28-1 – 468824876 and MO 3050/2-3 – 360330772.

REFERENCES

[AW10] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.

[Ber82] Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 1982.

[BG05] Ingwer Borg and Patrick J. F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer Science & Business Media, 2005.

[BN03] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[Bro87] Michael Browne. The young-householder algorithm and the least squares multidimensional scaling of squared distances. *Journal of Classification*, 4:175–190, 1987.

[CL06] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

[DP73] David H Douglas and Thomas K Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122, 1973.

[FML16] Alexey Fofonov, Vladimir Molchanov, and Lars Linsen. Visual analysis of multi-run spatio-temporal simulations using isocontour similarity for projected views. *IEEE Transactions on Visualization and Computer Graphics*, 22(8):2037–2050, August 2016.

[FR91] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991.

[GP17] Galen R. Gisler and John M. Patchett. Deep water impact ensemble data set. Technical Report LA-UR-17-21595, Los Alamos National Laboratory, 2017.

[HKW23] David Hägele, Tim Krake, and Daniel Weiskopf. Uncertainty-aware multidimensional scaling. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):23–32, 2023.

[KH13] Johannes Kehrer and Helwig Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19:495–513, 03 2013.

[Kra88] Dieter Kraft. A software package for sequential quadratic programming. Technical Report DFVLR-FB 88-28, Deutsche Forschungs- und Versuchsanstalt fÃ¼r Luft- und Raumfahrt (DFVLR), KÃ¶ln, Germany, 1988.

[Kru64] Joseph B Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[LC09] Seunghak Lee and Seungjin Choi. Landmark MDS ensemble. *Pattern Recognition*, 42(9):2045–2053, 2009.

[Lei25] Simon Leistikow. `https://github.com/sleistikow/simplified_embeddings`, 2025.

[LHFL19] Simon Leistikow, Karim Huesmann, Alexey Fofonov, and Lars Linsen. Aggregated ensemble views for deep water asteroid impact simulations. *IEEE Computer Graphics and Applications*, 40(1):72–81, 2019.

[MBR16] Patrick Mair, Ingwer Borg, and Thomas Rusch. Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate Behavioral Research*, 51(6):772–789, 2016.

[MHSG18] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018.

[ML19] Vladimir Molchanov and Lars Linsen. Shape-preserving star coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):449–458, 2019.

[PWB+09a] Kristin Potter, Andrew Wilson, Peer-Timo Bremer, Dean Williams, Charles Doutriaux, Valerio Pascucci, and Chris Johhson. Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. *Journal of Physics: Conference Series*, 180(1):012089, July 2009.

[PWB+09b] Kristin Potter, Andrew Wilson, Peer-Timo Bremer, Dean Williams, Charles Doutriaux, Valerio Pascucci, and Chris R. Johnson. Ensemble-Vis: A framework for the statistical visualization of ensemble data. In *2009 IEEE International Conference on Data Mining Workshops*, pages 233–240, 2009.

[SG64] Abraham Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, July 1964.

[vdMH08] Laurens van der Maaten and Geoffrey E. Hinton. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008.

[VDMPVdH09] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10:66–71, 2009.

[VW93] Maheswari Visvalingam and James D Whyatt. Line generalisation by repeated elimination of points. *The Cartographic Journal*, 30(1):46–51, 1993.

[Wic03] Florian Wickelmaier. An introduction to MDS. *Sound Quality Research Unit, Aalborg University, Denmark*, 46(5):1–26, 2003.