Uncertainty in Interactive WMDS Visualizations

Lata Kodali * John Wenskovitch Nathan Wycoff Leanna House Chris North

Virginia Tech

ABSTRACT

Visualizations are useful when learning from high-dimensional data. However, visualizations can be misleading when they do not incorporate measures of uncertainty; e.g., uncertainty from the data or the dimension reduction algorithm used to create the visual display. In our work, we extend a framework called Bayesian Visual Analytics (BaVA) on a dimension reduction algorithm, Weighted Multidimensional Scaling (WMDS), to incorporate uncertainty as analysts explore data visually. BaVA-WMDS visualizations are interactive, and possible interactions include manipulating variable weights and/or the coordinates of the two-dimensional projection. Uncertainty exists in these visualizations on the variable weights, the user interactions, and the fit of WMDS. We quantify these uncertainties using Bayesian models exploring randomness in both coordinates and weights in a method we call Interactive Probabilistic WMDS (IP-WMDS). Specifically, we use posterior estimates to assess fit of WMDS, the range of motion of coordinates, as well as variability in weights. Visually, we display such uncertainty in the form of color and ellipses, and practically, these uncertainties reflect trust in fitting a dimension reduction algorithm. Our results show that these displays of uncertainty highlight different aspects of the visualization, which can help inform analysts.

Index Terms: Human-centered computing—Visualization techniques—Multidimensional Scaling—Uncertainty

1 INTRODUCTION

In addition to parallel coordinate plots and heatmaps, dimension reduction methods are often used to visualize high-dimensional data [22]. The main advantage of dimension reduction is that when high-dimensional data are projected into low-dimensional spaces (e.g., two- or three-dimensional spaces for visualization), analysts may rely on personal spatial reasoning skills to identify data structures. For example, analysts may visually and quickly identify clusters, outliers, and/or other data patterns in projections that are not readily apparent when datasets are considered in their raw forms [31]. However, unlike parallel coordinate plots and some heat maps, low-dimensional data projections are, by definition, summaries or *approximations* of high-dimensional data, and naturally include model-lack-of-fit error or uncertainty. Thus, all findings from low-dimensional projections of high-dimensional data should be tempered by measures of the uncertainty; e.g., low measures could reinforce findings, whereas high measures should reduce confidence in findings. Yet, in practice, visual projections rarely communicate any measurements of uncertainty. With focus on multidimensional scaling (MDS) [15] - a common dimension reduction method - this paper presents a statistically sound and easy-to-interpret approach for quantifying and visualizing uncertainty in projections of high-dimensional data.

In particular, we consider both global and local (within-projection) measures of uncertainty in data projections. For example, MDS works by minimizing a stress function to estimate low-dimensional coordinates of high-dimensional data. Though unitless and without clear points of reference (which we address in this paper), the final stress of an MDS projection is effectively a global measure of projection lack-of-fit. However, there can be cases in which a majority of observations in a dataset is represented well in a projection, while a subset is represented poorly. The well-positioned observations reduces the global MDS stress enough for algorithm convergence, but the poorly projected portion of the data remains. Without local measures of uncertainty, analysts cannot determine which observations to trust in MDS visualizations, and thus must choose to either trust all observations equally well or avoid MDS altogether [6]. Notably, this choice presents itself when considering all dimension reduction methods, not just MDS [5, 19, 24].

Visualizations that incorporate uncertainty are truer representations of the data than those without [3]. Some work has been performed in the area of visualizing uncertainty and rely on techniques such as background color [11, 25], Voronoi tessellations [1, 18], glyphs [32], interactive probing [29], and error ellipsoids [33]. In contrast to these techniques, we create a Bayesian method for both computing and displaying global and local uncertainty in dimension-reduced projections, particularly those created by Weighted Multidimensional Scaling (WMDS) [15].

Specifically, we make the following contributions in this work:

- We develop both global and local normalized measures of probabilistic uncertainty across WMDS visualizations to enable direct comparisons between projections. Our global measure is a variant of stress called cn-stress.
- We provide two methods for visualizing the uncertainty of individual observation coordinates as measured by observation contribution to cn-stress and/or posterior estimates of coordinate variance (uncertainty region).
- We provide a method for quantifying the uncertainty in model parameters as influenced by user interactions. This includes:
  - Bayesian models for projecting and interacting with WMDS.
  - A Monte Carlo experiment to assess the distribution of interactions.

The remainder of the paper is as follows. Section 2 discusses the framework that underlies our method. Section 3 describes the Bayesian models used in Interactive Probabilistic WMDS (IP-WMDS), as well as our measures of global and local uncertainty. Section 4 demonstrates our method on two example datasets. We discuss future work and limitations of this method in Section 5, and we also include two appendices to further expand on our statistical models.

2 BACKGROUND

Our method to visualizing uncertainty is motivated by a framework called Bayesian Visual Analytics (BaVA) [12] for developing interactive visualizations. Though BaVA formally has technical underpinnings, over time the term has come to refer generally to the process of parameterizing feedback from analysts to include within analytic methods for visualization e.g., [4, 8, 9, 11, 26, 27, 30]. The main advantage of BaVA is that it provides an approach for transforming static analytic methods to dynamic methods that may

*e-mail: latak215@vt.edu

respond to analysts' needs, discoveries, and judgments. Using analytic methods enhanced with BaVA, visual explorations of data may parallel analysts' sense making processes of data.

The challenge with BaVA, or any visual analytic method for multidimensional data, is that visualizations of high-dimensional datasets are inevitably inaccurate. The data *represent* a larger population and visualizations are, by definition, *summaries*; visualizations in three or fewer dimensions cannot display high-dimensional data perfectly. Yet, analysts interact with data visualizations in BaVA, as if they are accurate represents of the truth; measures of inaccuracy or uncertainty in visualizations are not conveyed visually to analysts.

The degree of uncertainty depends heavily on the visual analytic method and methods for interaction. For this paper, we focus on the analytic method, Weighted Multidimensional Scaling (WMDS) [15]. In previous work, BaVA has been applied to WMDS (BaVA-WMDS) via two primary forms of interactions [12, 13, 17, 26]. In this section, we highlight those forms of interaction and briefly explain WMDS.

2.1 Interactions

Interactions with projection visualization systems fall into two categories (described by Self et al. [26]): Parametric Interaction (PI) and Observation-Level Interaction (OLI). Briefly, PI makes use of standard user interface controls such as sliders, text boxes, and spinboxes to allow analysts to directly set the value of a parameter. For example, analysts may adjust eigenvalues in covariance matrices for principal component analyses [14], change K (the number of clusters) in a K-means algorithm [20], or increase or decrease the importance of variables in WMDS [26]. The advantage of PI is that there is no ambiguity in what analysts want to change to create new visualizations. Disadvantages of PI are that analysts must know what the parameters mean to make specifications that make sense and the dimensionality of the parameter space must be manageable enough for analysts to assess easily.

In contrast, OLI permits analysts to stay in the data space (i.e., not enter the parameter space) to commit interactions. Analysts make adjustments to representations of observations *in visualizations* to indirectly influence model parameters and create new visualizations. For example, to change the number of clusters found in data, analysts may drag clusters of observations together or spread observations apart [30]; or, to change the direction of visual projections, analysts may adjust the location of observations in visualizations. After analysts interact with observations, underlying algorithms and/or probabilistic models of OLI interpret the interactions in a parametric form to then apply to the entire dataset. Crucially, analysts need not learn the algorithms or models to use OLI.

Tools that enable analysts to explore data with PI and OLI offer tremendous flexibility. Andromeda [26, 27] is a recent example that offers both when applying WMDS to visualize data. In the next section, we explain WMDS in detail and how PI and OLI works for exploring high-dimensional data.

2.2 Weighted Multidimensional Scaling (WMDS)

Broadly, WMDS offers an approach for projecting data linearly to a reduced-dimensional space. In this paper, the reduced space has two dimensions for visualization purposes. The direction in which data are projected by WMDS depends on weights assigned to variables in the data; variables with high weights have more influence on the direction than those will low weights. Similarly, variables with high weights have more influence on the interpretation of WMDS projections. In WMDS projections, relative proximity among observations implies relative similarity of the same observations in the high-dimensional space, particularly in the variables with high weights.

Formally, consider a dataset $X_{n\times p} = (x_1, x_2, \ldots, x_n)'$ (containing n observations and p variables) with associated weights $\omega = (w_1, w_2, \ldots, w_p)$ such that $\sum_{k=1}^{p} w_k = 1$. The goal is to find reduced- or low-dimensional coordinates $R_{n\times 2} = (r_1, r_2, \ldots, r_n)'$ that preserve the relationships in the high-dimensional space [15]. When relationships are measured by distance, the preservation is accomplished by minimizing a stress function

$$\min_{r_1,r_2,\ldots,r_n} \text{stress} = \min_{r_1,r_2,\ldots,r_n} \sum_{1\le i<j\le n} (d_{ij}^{r} - d_{ij}^{\omega})^2, \qquad (1)$$

where, for observations i and j, d_{ij}^{r} and d_{ij}^{ω} represent a distance metric in the low-dimensional space and a weighted distance metric in the high-dimensional space, respectively. Typically, the low-dimensional metric is Euclidean distance, and in this paper, the high-dimensional metric is weighted Euclidean distance, $d_{ij}^{\omega} = \sqrt{\sum_{k=1}^{p} w_k (x_{ik} - x_{jk})^2}$.

Common techniques to minimizing the stress function are gradient descent and Scaling by MAjorizing a COmplicated Function (SMACOF) [2, 7]. Gradient descent aims to find a direction in which the objective function is locally minimized, while SMACOF attempts to majorize the objective function and subsequently optimize the majorized function. Both techniques lead to a local optimum and use a greedy approach in finding this optimum. In this work, we use the WMDS solution found via the SMACOF approach.

In the context of BaVA, WMDS is extended to have both *forward* and *backward* versions. When analysts perform PI (i.e., directly increase or decrease a weight), those changes are reflected in ω^*. The high-dimensional distances ($d_{ij}^{\omega^*}$) are updated, and new low-dimensional coordinates (R^*) are found by minimizing Equation 1. This scheme is referred to as **Forward WMDS**; i.e., typical WMDS is renamed as Forward WMDS.

Alternatively, when analysts perform OLI (i.e., manipulate observations), a two-step process occurs. First, using updated coordinates R^*, weights ω^* are determined by optimizing the stress function for the weights (not coordinates):

$$\min_{w_1^*,w_2^*,\ldots,w_p^*} \text{stress} = \min_{w_1^*,w_2^*,\ldots,w_p^*} \sum_{1\le i<j\le n} (d_{ij}^{r} - d_{ij}^{\omega^*})^2. \qquad (2)$$

Because updating the location of some observations does not automatically create a new layout for all data, the second step involves updating all observation locations using the weights ω^* found from Equation 2. Thus, to project new low-dimensional coordinates, we solve Equation 1 using ω^* to finally get R^{**}. This two step procedure for updating the observation locations is called **Backward WMDS**.

Regardless of implementing Forward WMDS or Backward WMDS, solutions for R^* and ω^* will have inaccuracies. As mentioned earlier, inaccuracies may result from many sources, including complexities in datasets and limitations of optimization algorithms. The question is, "How inaccurate?" How uncertain are representations of data, relative to the actual data? In the next section, we detail our approach for quantifying uncertainty in visualizations and conveying the uncertainty visually to analysts.

3 Methods to Quantify Uncertainty

For each type of interaction, PI and OLI, we use Bayesian versions of both Forward WMDS and Backward WMDS [23] to quantify local and global uncertainty in projections. A local measure of uncertainty concerns a single interaction and reflects the estimated variance of the projected coordinates and/or weights. A global measure of uncertainty theoretically considers all possible interactions and reflects the trustworthiness of all projected coordinates (thus, pairwise relationships) in a visualization relative to other projections.

To work toward presenting local and global uncertainty visually in Section 3.5, we start by explaining Bayesian Forward and Backward WMDS (Sections 3.1 and 3.2). In these models, sources of variance for parameters (e.g., low-dimensional coordinates and weights) are

Table 1: Summary of Notation used in Sections 2 and 3.

Notation	Description
n	number of observations ($i = 1,2,\ldots,n$)
p	number of variables or dimensions ($k = 1,2,\ldots,p$)
X	high-dimensional data of size $n \times p$ ($X_{n\times p}$)
w_k	individual weight for dimension k.
ω	variable weight vector of size p such that $\sum_{k=1}^{p} w_k = 1$
d_{ij}^{ω}	high-dimensional pairwise distance
r_i	two-dimensional coordinate for observation i, (r_{i1}, r_{i2})
R	low-dimensional coordinates (here 2 dimensions, $R_{n\times 2}$)
d_{ij}^{r}	low-dimensional pairwise distance
c-stress	calibrated stress (see Eq. 5)
cn-stress	calibrated normalized stress (see Eq. 6)
cn-stress$_i$	raw contribution from observation (obs) i (see Eq. 7)
cn-stress$_i^*$	percentage of contribution from obs i (see Eq. 8)
S^{ω}	Monte Carlo sample of variable weights ω
S^{R}	Monte Carlo sample of coordinates R
$\tilde{\omega}$	matrix of MCMC iterations of weights ω
$\tilde{R}$	matrix of MCMC iterations of Procrustes transformed coordinates R
MM	range of top 10 values of $w_k \in \tilde{\omega}$ for all k
SD	standard deviation of coordinate $r_i \in \tilde{R}$ for all i

defined explicitly. Then, we derive a calibrated, goodness of fit metric for individual projections in Section 3.3 and estimate the distribution of this metric across sampled projections in Section 3.4. Because the magnitude of the metric is data dependent, learning the distribution of the metric is necessary for interpretation. We summarize all the notations used in previous sections and throughout Sections 3.1–3.5 in Table 1.

3.1 Forward Interactive Probabilistic WMDS (FIP-WMDS): Local Uncertainty in Coordinates

As originally described by Oh and Raftery [23], Forward Interactive Probabilistic WMDS (FIP-WMDS) models high-dimensional pairwise distances d_{ij}^{ω} between observations $(i,j)\ \forall\ i,j$ conditional on low-dimensional distances d_{ij}^{r} and variance σ^2. Because distances are nonnegative, we consider each d_{ij}^{ω} to be governed by a truncated Normal distribution, and assign prior distributions as follows:

$$\begin{aligned} d_{ij}^{\omega} \mid d_{ij}^{r}, \sigma^2 &\sim N(d_{ij}^{r}, \sigma^2)\cdot \mathbb{1}_{(d_{ij}^{\omega}>0)}\cdot \\ \sigma^2 &\sim IG(a,b) \\ r_i &\sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \Lambda = \begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}\right), \text{ for } i = 1,2,\ldots,n \\ \lambda_q &\sim IG(c_q, d_q), \text{ for } q = 1,2. \end{aligned} \tag{3}$$

Note that low-dimensional pairwise distances are not directly modeled a priori, but derived from estimates of low-dimensional coordinates r_i ($i \in \{1,..,n\}$) with the diagonal covariance Λ. Also, priors for $R = (r_1, r_2, \ldots, r_n)^T$ and σ^2 are approximately conjugate and the hyperprior on λ_q is conjugate. Details for the remaining hyperparameters, a, b, c_q, and d_q where $q = 1,2$ can be found in Appendix A. Using Markov Chain Monte Carlo (MCMC), we estimate $m = \binom{n}{2} = \frac{n(n-1)}{2}$ pairwise distances $d_{ij}^{r}, R, \lambda_1, \lambda_2$, and σ^2 a posteriori [23].

Visualizations of data X result from posterior estimates of R. However, because we model d_{ij}^{ω} conditional on d_{ij}^{r}, not R, MCMC samples of R have inconsistent rotations, translations, reflections, and scale. Thus, before we estimate R, we standardize the MCMC samples of R using a Procrustes transformation [7, 21]. Specifically, we post process each MCMC draw for R using a Procrustes transformation relative the SMACOF WMDS solution. In turn, estimate low-dimensional coordinates R via the posterior mode of the joint posterior $f(R, \lambda_1, \lambda_2, \sigma^2 \mid X, \omega)$. For the remaining parameters, we estimate σ^2, λ_1, and λ_2 using their posterior means.

To understand the movement of coordinates in T iterations of the MCMC or the uncertainty of the Procrustes transformed coordinates $\tilde{R}$, we can consider one of two measures. One is a variance for each dimension of R, λ_1 and λ_2. Estimates of these parameters, e.g., $\hat{\lambda_q} = E[\lambda_q | X, \omega]$ directly reflect variance of coordinates. Note that $\hat{\lambda}_1$ and $\hat{\lambda}_2$ will be the same for each coordinate i ($i \in \{1,...,n\}$). Second, we consider estimates for the standard deviation (sd) of coordinates $r_i = (r_{i1}, r_{i2})$ for i, that is the $\hat{s}_{i1} = sd(\{r_{i1}^1, r_{i1}^2, \ldots, r_{i1}^T\})$ and $\hat{s}_{i2} = sd(\{r_{i2}^1, r_{i2}^2, \ldots, r_{i2}^T\})$. We denote $SD_{n\times 2} = \{(\hat{s}_{i1}, \hat{s}_{i2})\}_{i=1}^{n}$.

3.2 Backward Interactive Probabilistic WMDS (BIP-WMDS): Local Uncertainty in Weights

Backward Interactive Probabilistic WMDS (BIP-WMDS) executes in response to an observation-level interaction, and models low-dimensional pairwise distances conditional on ω and σ^2. The goal of BIP-WMDS is to estimate weights a posteriori, as well as to understand the variability of these weights. To do so, we have the following hierarchical model for m low-dimensional pairwise distances:

$$\begin{aligned} d_{ij}^{r} \mid \omega, \sigma^2 &\sim N(d_{ij}^{\omega}, \sigma^2)\cdot \mathbb{1}_{(d_{ij}^{r}>0)}\cdot \\ \omega &\sim \text{Dirichlet}(a = (a_1, a_2, \ldots, a_p)) \\ \sigma^2 &\sim IG(b,c) \end{aligned} \tag{4}$$

We chose the prior for σ^2 to capitalize on conjugacy, and we chose the Dirichlet prior for ω to honor the constraint that the weights should sum to 1 ($\sum_k = 1^p \omega_k = 1$). Details concerning the posterior distribution for ω and σ^2, the Gibbs sampler, convergence diagnostics, and initialization are provied in Appendix A.

The posterior estimation for this model presents similar challenges to those presented by FIP-WMDS in Section 3.1. For variable weights, ω, the posterior mean should approximate $\{1/p\}_{k=1}^{p}$ by construct and relying on the individual Maximum A Posterior (MAP) estimators for each ω_k in ω will contradict the constraint that $\sum_{k=1}^{p} w_k = 1$. Thus, our posterior estimate for ω are taken from the maximum draw of the joint posterior $f(\omega, \sigma^2 \mid X, R)$.

To understand the variability of weights, we consider $\tilde{\omega} = \{\omega^1, \omega^2, \ldots, \omega^T\}$, which represents the MCMC draws of ω. We summarize the optimal values of the weights using the top 10 draws of the joint posterior to understand the range of w_k for $k = 1,2,\ldots,p$. A range is an interval of the minimum and maximum value, and we use MM to reflect this meaning. Specifically, we denote $MM_{k\times 2} = \{\text{range}(\hat{w}_k^1, \hat{w}_k^2, \ldots, \hat{w}_k^{10})\}_{k=1}^{p}$ for the range of these top 10 values for each w_k.

3.3 Global Uncertainty in Projections

Together, we refer to the above models as Interactive Probabilistic WMDS (IP-WMDS). IP-WMDS describes and quantifies uncertainty within one projection, i.e., provided one user interaction (e.g., specification of ω^* or R^*) Now, we expand IP-WMDS to allow us to consider the entire space of interactions by carefully defining measures to compare different visualizations.

To motivate the expansion, consider two visualizations of data X with dimension 20×5 in Figure 1. Using FIP-WMDS, the left visualization relies on weights $\omega_1 = (0.26, 0.43, 0.21, 0.03, 0.07)$ and the right visualization relies on $\omega_2 = (0.19, 0.47, 0.07, 0.01, 0.26)$. While relative locations appear similar between these two displays, WMDS projections represent similarity with proximity, and as a result the conclusions that an analyst may draw can differ. For example, observations 19 and 3 are relatively closer (more similar) in the left plot than they are in the right plot. In attempt to compare and choose one visualization, we might assess stress. That is, when we estimate R as defined in Section 3.1 and solve for the stress using

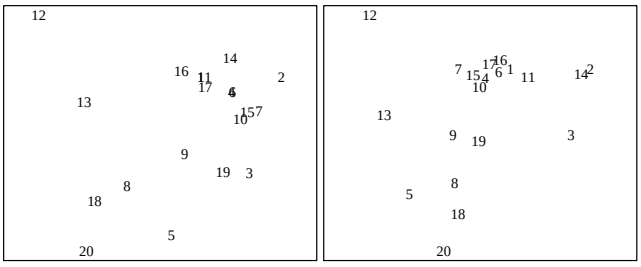


Figure 1: Example of comparing two different visualizations created by different interactions of the data. In these visualizations, the proximity of two observations implies similarity between those observations. Thus, conclusions about similarity (or dissimilarity) of observations are different between the two visuals.

Equation 1, the left and right plots have stresses 77.5 and 140.1, respectively. This would suggest that the left plot is considerably better than the right plot, however we must consider scale of the distance measures embedded in the stress calculations. The scale of high-dimensional distances will change with changes in weights. Thus comparisons of stress as defined in Equation 1 do not apply across interactions.

One idea for calibrating the stress comes from Kruskal [15]. He considered a "goodness of fit" measure that reflects the ease of summarizing high-dimensional with low-dimensional versions. We apply a similar idea and calculate c-stress, where c-stress calibrates the original stress by the sum of squared high-dimensional distances:

$$\text{c-stress} = \frac{\text{stress}}{\sum_{1\le i<j\le n}(d_{ij}^{\omega})^2} = \frac{\sum_{1\le i<j\le n}(d_{ij}^{r}-d_{ij}^{\omega})^2}{\sum_{1\le i<j\le n}(d_{ij}^{\omega})^2}. \quad (5)$$

The c-stress values for the left and right visualizations in Figure 1 are 0.10 and 0.18 respectively. However, while $\sum_{1\le i<j\le n} d_{ij}^{\omega_1} \neq \sum_{1\le i<j\le n} d_{ij}^{\omega_2}$, it can be shown that $\sum_{1\le i<j\le n}(d_{ij}^{\omega_1})^2 = \sum_{1\le i<j\le n}(d_{ij}^{\omega_2})^2 = n(n-1)$. Thus, the calculations of c-stress for Figure 1 are still on different scales, and we construct yet another measure.

Our construction, denoted cn-stress, penalizes differences in proportional distances, rather than absolute distances. Specifically, we use scalars $a = \sum_{1\le i<j\le n} d_{ij}^{r}$ and $b = \sum_{1\le i<j\le n} d_{ij}^{\omega}$ and normalize c-stress such that

$$\text{cn-stress} = \frac{\sum_{1\le i<j\le n}\left(d_{ij}^{r}/a - d_{ij}^{\omega}/b\right)^2}{\sum_{1\le i<j\le n}\left(d_{ij}^{\omega}/b\right)^2}. \quad (6)$$

Now, we are calibrating stress by differing constants in that $\sum_{1\le i<j\le n}(d_{ij}^{\omega_1}/b_1)^2 \neq \sum_{1\le i<j\le n}(d_{ij}^{\omega_2}/b_2)^2$ because $b_1 \neq b_2$. This means that cn-stress has been scaled to a common unit that can be compared across projections with different weights. In Figure 1, visualizations both have approximately cn-stress=0.016. This means that the visualizations in Figure 1 have comparable *global* uncertainty in the directions they were projected; i.e., given they specified weight vector.

We delve deeper into the visualization by observing the fit of WMDS on each coordinate of the display. We consider the concept of "stress-per-point," which essentially aims to identify outliers when fitting MDS [7, 21]. We consider an observation i's contribution to stress, denoted cn-stress$_i$, to be

$$\text{cn-stress}_i = \frac{\sum_{j=1;j\neq i}^{n}(d_{ij}^{r}/a - d_{ij}^{\omega}/b)^2}{\sum_{i=1}^{n}\sum_{j=1}^{n}(d_{ij}^{\omega}/b)^2}. \quad (7)$$

We note cn-stress$_i$ will appear rather arbitrary in the same sense as raw stress. To compare cn-stress$_i$ across different visualizations, the contribution we monitor per observation i is the following percentage:

$$\text{cn-stress}_i^{*} = \frac{\text{cn-stress}_i}{2\cdot\text{cn-stress}} = \frac{\sum_{j=1;j\neq i}^{n}(d_{ij}^{r}/a - d_{ij}^{\omega}/b)^2}{\sum_{i=1}^{n}\sum_{j=1}^{n}(d_{ij}^{r}/a - d_{ij}^{\omega}/b)^2}. \quad (8)$$

Higher proportions indicate that those coordinates are more difficult to find, and we note $\sum_{i=1}^{n}\text{cn-stress}_i^{*} = 1$. We call cn-stress$_i^{*}$ as the observation contribution to stress. This contribution can be a proxy for trust of that observation location in the visualization and subsequently pairwise relationships to the other coordinates.

3.4 Distributions of Uncertainty

To interpret measures of uncertainty, it is helpful to have points of references or ranges in which measures are probabilistically reasonable. We offer relative comparisons of uncertainty by estimating distributions of uncertainty, for a given dataset.

We take a Monte Carlo approach to explore the space of IP-WMDS solutions, and all related measures of uncertainty from each. That is, Models 3 and 4 can be applied to the space of interactions (parametric and observation-level) through a representative sample of interactions. Based on the sample, we estimate distributions of uncertainty.

We find our sample of interactions via Monte Carlo for $j = 1, 2, \ldots, 1000$ in the following manner:

1. Sample $\omega_j \sim \text{Dirichlet}(\vec{1}_p)$.
2. Use ω_j to find the SMACOF WMDS solution of R_j.

From this process, we have a representative sample of parametric and observation-level interactions to explore, denoted $S^{\omega} = \{\omega_1, \omega_2, \ldots, \omega_{1000}\}$ and $S^{R} = \{R_1, R_2, \ldots, R_{1000}\}$.

The first global measure we discuss comes from cn-stress (introduced in Section 3.3), which measures the fit of WMDS. The uncertainty will come from the distribution of cn-stress, $\pi(\text{cn-stress})$, i.e., cn-stress across all visualizations. For a given interaction, we can compare the cn-stress value from this visualization to the distribution. We can also further look to this measure by observation.

Another global measure is cn-stress$_i^{*}$ (introduced in Section 3.3), which measures the difficulty in fitting WMDS for coordinate i. When observing the distribution of cn-stress$_i^{*}$, $\pi(\text{cn-stress}_i^{*})$, we learn the difficulty level of fitting that observation across visualizations. The interpretation here is that if observation contribution to stress is low, then analysts can trust the location and pairwise relationships of that observation. In contrast, if the contribution to stress is high, analysts should take caution in trusting the location of that coordinate.

Next, we discuss local measures of uncertainty in the context of our modeling schemes for FIP-WMDS and BIP-WMDS using our samples of interactions, S^{ω} and S^{R}. This implementation is summarized in Table 2. Local measures here indicate uncertainty within a particular visualization, i.e., uncertainty from S_j^{ω} or S_j^{R}. For FIP-WMDS, we use the standard deviation of each coordinate i, *SD*, to understand the uncertainty region. For BIP-WMDS, we use the range of optimal weights for each weight k, *MM*, to understand the variability of the weights.

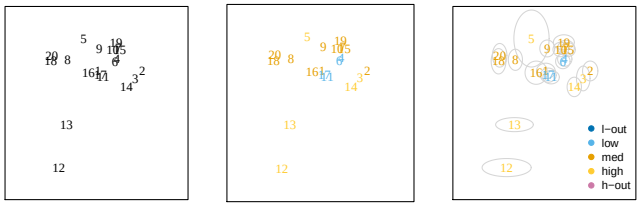
3.5 Visualizing Uncertainty

In our BaVA-WMDS projections, we can visualize the effect of cn-stress$_i^{*}$ from the distribution $\pi(\text{cn-stress}_i^{*})$ as well as the standard deviation of observation locations SD_i for each observation i. To better understand extremity in the movement of coordinates (uncertainty region), we plot 3 times the standard deviation value as an ellipse around each observation location.

Table 2: Summary of IP-WMDS Implementation.

	FIP-WMDS	BIP-WMDS
step 1	Run with inputs $S_j^{\omega} = \omega_j$ and X.	Run with inputs $S_j^{R} = R_j$ and X.
step 2	Store cn-stress, cn-stress$_i^*$ $\forall\, i$, and $\hat{\sigma}^2$.	
step 3	Store $\hat{R}_j$, SD_j, $\hat{\lambda}_1$, and $\hat{\lambda}_2$	Store $\hat{\omega}_j$ and MM_j.

To display the effect of cn-stress$_i^*$, we use colors and color observations. Based on simulations, we find the empirical distribution of π(cn-stress$_i^*$) to be right-skewed (since it is strictly nonnegative). We then apply a log transformation, and have $\pi(\log(\text{cn-stress}_i^*))$ which is approximately normally distributed. We consider the following quantiles: 0.0015, 0.25, 0.75, and 0.9985. If for observation i, the log(cn-stress$_i^*$) is low (below 0.25), we use cool colors to depict this observation is not difficult to fit with WMDS. As the log(cn-stress$_i^*$) value increases, we use warm colors to depict the difficulty of fit with pink representing extremely difficult (above 0.9985). We show an example of using color and ellipses in Figure 2 using a color blind palette[1]. In the legend of Figure 2 (c), the color scheme is for low outliers (abbreviated as l-out) is below 0.0015, low is between 0.0015 and 0.25, medium (abbreviated as med) is between 0.25 and 0.75, high is between 0.75 and 0.9985, while for high outliers (abbreviated as h-out) is above 0.9985. Lastly, we note the coordinates plotted are from FIP-WMDS since posterior estimates of coordinates from FIP-WMDS will closely match that of the SMACOF WMDS solution used in BIP-WMDS.



(a) (b) (c)

Figure 2: Visualizing Uncertainty from Coordinate Range and Trust of Observation Location. In (a), we show a projection of 20 observations without uncertainty information. In (b), we encode difficulty of fit with color. In (c), we provide a dual encoding of difficulty of fit using both ellipses and color. In the legend, l-out means a low outlier and h-out means a high outlier.

4 Results

We apply our methods on two example datasets concerning animals and characteristics describing these animals. Both example datasets were created from a larger dataset containing 49 observations and 65 variables, slightly reduced from the original dataset [16] to remove highly-correlated and sparse dimensions. Each of the variables represents a characteristic of the animals normalized on a scale of 0–100. For example, under the variable Furriness, bears will have a high value in this variable while snakes would be low. For demonstrative purposes, we randomly sampled 20 observations for our visualizations and abbreviate them as shown in Table 3 below.

4.1 Example 1 with $n = 20$ and $p = 5$

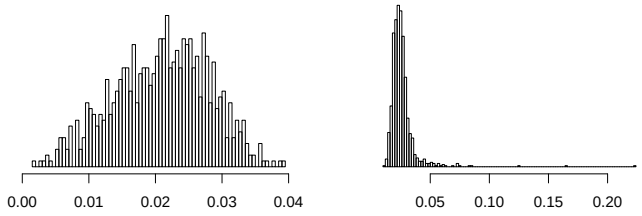
In this example, we randomly sample 5 attributes: Cave, Fish, Quadrupedal, Lean, and Bush. We show results for uncertainty measures from the scheme discussed in Section 3.4 by sampling 1000 interactions for both weights and observation locations. In Figure 3, we plot an empirical distribution of π(cn-stress). The average cn-stress from FIP-WMDS is 0.02 with a range of [0, 0.04]

[1] http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/

Table 3: Animals and abbreviations for examples.

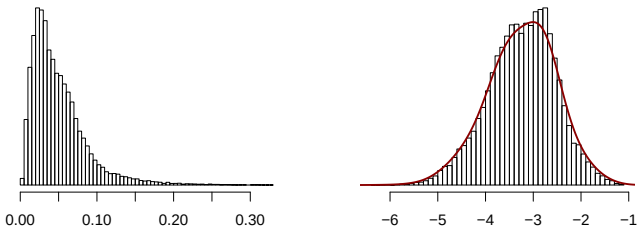
Animal	Abb.	Animal	Abb.
Mouse	Mu	Rabbit	Rb
Zebra	Ze	Bat	Ba
Leopard	Le	Grizzly Bear	Gr
Hamster	Ha	Rhinoceros	Rh
Polar Bear	Po	German Shepherd	Ge
Collie	Co	Spider Monkey	Sp
Dalmatian	Da	Mole	Ml
Beaver	Be	Dolphin	Do
Raccoon	Rc	Siamese Cat	Si
Moose	Ms	Walrus	Wa

while from BIP-WMDS is 0.03 with a range of [0.01, 0.22]. At first glance, they appear very different, however, most mass of the distribution for π(cn-stress) of BIP-WMDS is within the same range of FIP-WMDS. This suggests fitting weights can be more difficult, and we have a few outliers (difficult to fit interactions) for BIP-WMDS.



(a) (b)

Figure 3: Plots of π(cn-stress) across interactions (a) from FIP-WMDS and (b) from BIP-WMDS. The distribution of difficulty of fitting WMDS appears similar in both (a) and (b) with the bulk of values below 0.05. The right-tail of (b) shows for a few interactions, weights are difficult to find in BIP-WMDS.



(a) (b)

Figure 4: Plots of (a) π(cn-stress$_i^*$) and (b) $\pi(\log(\text{cn-stress}_i^*))$ for all observations i from IP-WMDS. Note that the log transformation is approximately normally distributed.

Figure 4 plots the empirical distribution π(cn-stress$_i^*$) *for all* i in both FIP-WMDS and BIP-WMDS, as well as its log transformation. We see that the log transformation is approximately normally distributed.We also observe the empirical distribution for an *individual* observation i (rather than all i). In Figure 5, we overlay the density curve from the overall distribution (in Figure 4) for 3 individual animals, Mouse, Zebra, and Leopard. We glean that the Mouse is relatively easier to fit across visualizations than the Leopard.

In Figure 6, we show how adding layers of uncertainty informs analysts about visualizations as described in Section 3.5. We choose two visualizations to display in Figure 6 based on the best and worst directions (out of our sample of 1000 interactions). "Best" would be visualizations with the most blue coordinates (Figure 6 (a)-(c)) and "worst" would be visualizations with the least blue coordinates (Figure 6 (d)-(f)). To help interpret these visualizations, a list of animal abbreviations is in Table 3 followed by weight estimates with

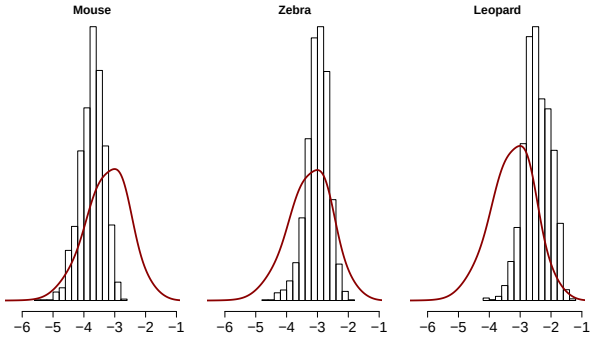


Figure 5: Plots of $\pi(\log(\text{cn-stress}_i^*))$ for only observation i. Note that the Mouse is relatively easier to fit across visualizations than the Leopard.

uncertainty for the variables in Table 4. From Table 4, we see that variables Quadrupedal and Bush are weighted highly in Figures 6 (a)-(c), whereas Figures 6 (d)-(f), assign most weight to only Fish and Bush.

Notice that the observation Bat appears as an outlier in Figure 6(a). However is it really an outlier within the high-dimensional space? Without measures of uncertainty, we would not know. In Figure 6(b), this observation is colored pink. This means that the observation Bat contributes unusually more than others to the global uncertainty of the projection. That is, this observations is not represented well in the visualization, relative to the rest. Also, we see from Figure 6(c) that the coordinates within the visualization could vary dramatically. That is, probabilistically, the coordinate for Bat could move slightly toward the remaining observations - thus suggesting it is slightly less of an outlier than appears in Figure (a). To summarize, the observation Bat does seem different from the remaining observations, but not as different as might be implied by an WMDS without uncertainty.

Though Figures 6(a)-(c) have more blue observations than Figures 6(d)-(f), the global stress of Figures 6(d)-(f) (cn-stress=0.012) is less than 6(a)-(c) (cn-stress=0.027). Thus, it becomes a choice for the user to decide which projection to trust more for making insights. If a user is looking to make insights across all observations, the user might prefer graphs 6(d)-(f). Whereas, if a user is interested in specific observations, e.g., German Shepard (Ge) and Raccoon (Rc), the user might choose projections 6(a)-(c), as they are colored in blue.

Table 4: True weights ω_j (True), posterior estimate of weights, $\hat{\omega}_j$ (Max Post), and range of top 10 weights, MM, from BIP-WMDS for Figure 6. Quadrupedal is abbreviated to Quad.

Fig 6(a)-(c)				Fig 6(d)-(f)		
True	Max Post	Range		True	Max Post	Range
0.16	0.16	[0.12, 0.18]	Cave	0.04	0.03	[0.02, 0.06]
0.14	0.15	[0.09, 0.17]	Fish	0.48	0.39	[0.37, 0.41]
0.32	0.33	[0.32, 0.40]	Quad	0.05	0.11	[0.11, 0.16]
0.11	0.06	[0.03, 0.08]	Lean	0.08	0.11	[0.06, 0.12]
0.27	0.30	[0.26, 0.34]	Bush	0.35	0.35	[0.30, 0.35]

4.2 Example 2 with $n = 20$ and $p = 30$

To see how our method applies when the number of attributes in a dataset (p) exceeds the number of observations (n) ($p > n$), we randomly sample 30 attributes from the animals dataset and show them in Table 5.

Again, we show results for uncertainty measures from the scheme discussed in Section 3.4 by sampling 1000 interactions for both weights and observation locations. In Figure 7, we plot an empirical

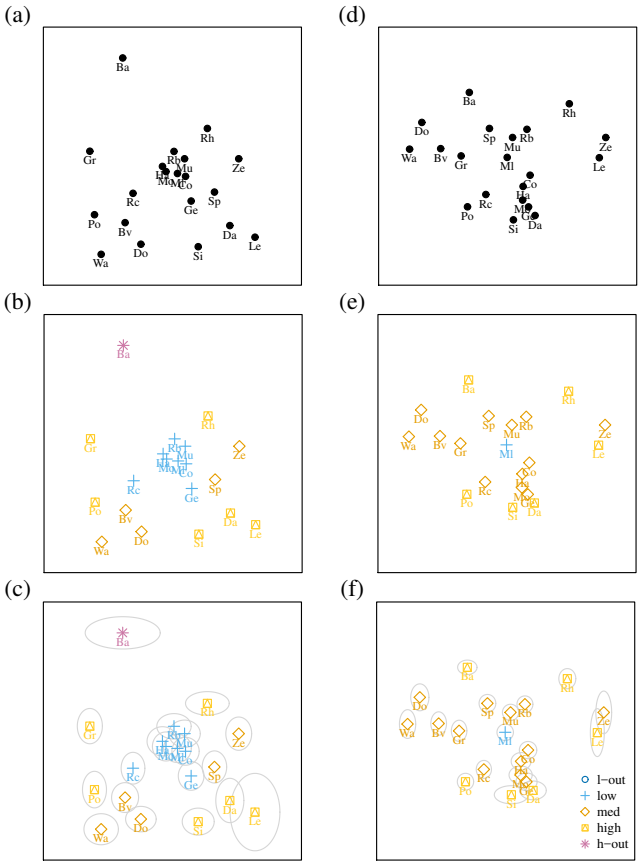


Figure 6: Visualizing Uncertainty Regions and Trust of Coordinates from the Best (Most Blue Observations) and Worst (Least Blue Observations) Directions of Example 1. Plots (a) and (d) have no visual of uncertainty making comparisons of each display difficult, plots (b) and (e) are uncertainty related to trust in observations locations and relationships, and plots (c) and (f) are uncertainty in both. For example, the interaction in (a) helps analysts understand the Racoon's (Rc) relationship to Polar Bear (Pb) more so than does the interaction in (b). See Table 3 for all the animal abbreviations. In the legend, l-out means a low outlier and h-out means a high outlier. For ease of viewing, symbols associated with the colors are utilized.

distribution of $\pi(\text{cn-stress})$. The average cn-stress from FIP-WMDS is 0.07 with a range of [0.05, 0.09], and BIP-WMDS is 0.08 with a range of [0.06, 0.14]. When increasing the number of attributes, we find the distributions of $\pi(\text{cn-stress})$ are very similar in FIP-WMDS and BIP-WMDS. There are fewer extremes in the tail of the distribution for BIP-WMDS as well.

Figure 8 shows plots of the empirical distribution $\pi(\text{cn-stress}_i^*)$ and $\pi(\log(\text{cn-stress}_i^*))$ for all i followed by Figure 9 of the same individual observations from Figure 5. In Figure 8, with the increase in variables, we essentially have one group of observations that is relatively easier to fit than the other group. This makes our log transformation appear as a mixture of two normal distributions describing those groups. In Figure 9, we draw a similar conclusion as in Example 1 that the Mouse is relatively easier to fit with WMDS than the Leopard and Zebra. Perhaps Mouse is in the relatively easy-to-fit group while Zebra and Leopard are in the other group.

Lastly, we show the best and worst visualizations for the $p = 30$ case in Figure 10. We provide the entire table of weights for each display in Appendix B. In Figure 10 (a), we see a direction with the most blue observations also has less extremes in the uncertainty region than does the display with the least blue. This is the opposite scenario of Figure 6, and shows that the uncertainty region and

Table 5: 30 Randomly Sampled Attributes for Example 2.

Cave (Ca)	Oldworld (Ow)	Hooves (Hv)	Spots (Sp)
Fish (Fi)	Hibernate (Hi)	Smelly (Sm)	Size (Sz)
Quadrupedal (Qu)	Coastal (Co)	Water (Wa)	Plains (Pl)
Lean (Ln)	Bulbous (Bb)	Hops (Hp)	Tree (Tr)
Bush (Bh)	Patches (Pa)	Stalker (St)	Horns (Hn)
Toughskin (Ts)	Inactive (Ia)	Grazer (Gr)	Timid (Ti)
Chewteeth (Ct)	Agility (Ag)	Walks (Wk)	Tail (Tl)
Buckteeth (Bt)	Nocturnal (Nt)		

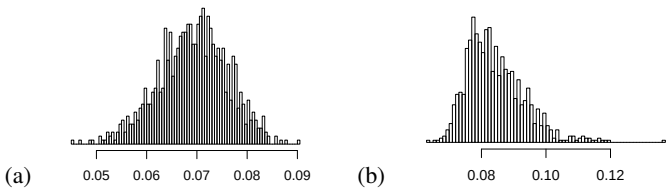


Figure 7: Plots of $\pi(\text{cn-stress})$ across interactions from (a) FIP-WMDS and (b) BIP-WMDS for $p = 30$. The distribution of difficulty of fitting WMDS appears more similar than when $p = 5$. In (b), there are less outliers for BIP-WMDS.

observations contributions to model fit of WMDS do not have a monotonic relationship. Thus, incorporating both types of uncertainty into the visualization is needed to understand the effects of fitting WMDS and guides the explorations for the analyst.

5 Discussion and Conclusion

In this work, we developed IP-WMDS, a probabilistic adaptation of WMDS to quantify and visualize measures of uncertainty in data projections. Uncertainty in projections may result from nuances in data, choices in interaction, and WMDS directly. To account for differences in these sources, we extend IP-WMDS to forward and backward versions, termed FIP-WMDS and BIP-WMDS, respectively. FIP-WMDS applies when assessing WMDS uncertainty in response to Parametric Interactions (PIs), whereas BIP-WMDS applies in response to Observation-Level Interactions (OLIs). To establish points of reference for measures of uncertainty, we estimate distributions of uncertainty from PI and OLI using Monte Carlo.

By incorporating uncertainty into visualizations, we provide analysts a means to make discoveries from data effectively and responsibly. That is, how analysts interact with data or make discoveries from data may be influenced by the knowing ranges of possible low-dimensional coordinates or intervals of optimal weights on variables, rather than just point estimates of each. The next step is to conduct user studies to assess the influence uncertainty measures have on analysts' explorations of data. To conduct these studies, we plan to incorporate uncertainty quantification into the BaVA tool Andromeda [27]. However, our current implementation of IP-WMDS, based on using 1000 interactions for both FIP-WMDS and

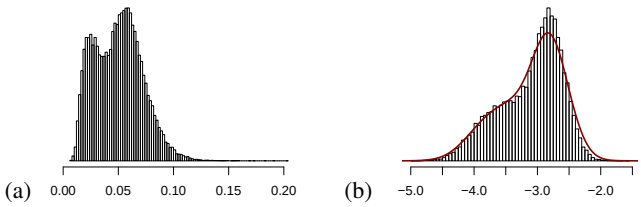


Figure 8: Plots of (a) $\pi(\text{cn-stress}_i^*)$ and (b) $\pi(\log(\text{cn-stress}_i^*))$ for all observations i from IP-WMDS when $p = 30$. Note that the log transformation is approximately a normal mixture of 2 components. This suggests we have a group that is easier to fit than another.

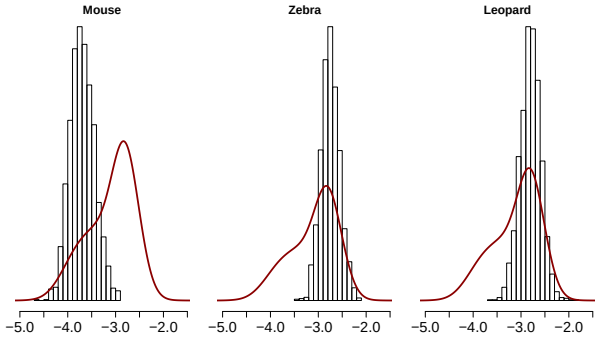


Figure 9: Plots of $\pi(\log(\text{cn-stress}_i^*))$ for only observation i. Even when including more attributes ($p = 30$), the Mouse is relatively easier to fit across visualizations than Leopard as in Figure 5.

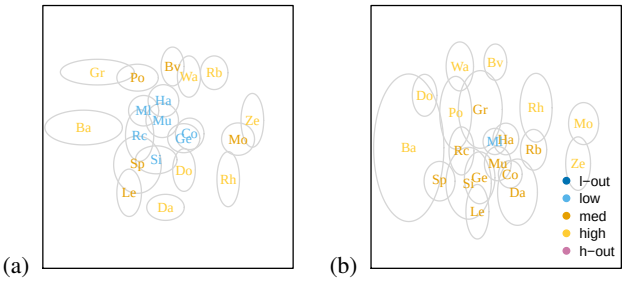


Figure 10: Plots of (a) Best and (b) Worst Directions for $p = 30$. Note that uncertainty regions are relatively smaller in (a) than in (b) which is opposite from Figure 6 (e) and (f) reflecting importance of displaying both measures of uncertainty. See Table 3 for all the animal abbreviations. In the legend, l-out means a low outlier and h-out means a high outlier.

BIP-WMDS is computationally intensive, and, as the number of observations in a dataset increases, so does computational time. To improve computational time, one idea is to approximate IP-WMDS to avoid the need for MCMC, or, at minimum, speed up the MCMC. Another idea is to combine FIP-WMDS and BIP-WMDS into one modeling effort, where both coordinates and weights unknown.

A Details of IP-WMDS

A.1 FIP-WMDS posterior distribution

We write the joint posterior for this model:

$$f(R,\lambda_1,\lambda_2,\sigma^2 \mid X,\omega) \propto (\sigma^2)^{-(m/2+a+1)} \prod_{q=1}^{2} \lambda_q^{-(n/2+c_q+1)} \cdot \exp\left\{ -\frac{\text{stress}}{2\sigma^2} - \sum_{1\le i<j\le n} \log\Phi\left(\frac{d_{ij}^r}{\sigma}\right) - \frac{b}{\sigma^2} - \frac{1}{2}\sum_{i=1}^{n} r_i^T \Lambda^{-1} r_i - \sum_{q=1}^{2} \frac{d_q}{\lambda_q} \right\}.$$

We note stress is defined as in Equations 1 and 2.

A.2 Gibbs sampler for FIP-WMDS

To draw samples from this posterior distribution, we slightly adjust he Gibbs sampler approach in Oh and Raftery [23]. We use a truncated normal proposal for σ^2 rather than a random walk. Parameters are updated as follows in the Gibbs sampler using T iterations. See Oh and Raftery [23] for more details.

A.3 FIP-WMDS Settings for Example 1 and 2

For both examples, we run our Gibbs sampler for $T = 50,000$ iterations with a burn-in of 10,000. Hyperparameter settings for σ^2 are $a = b = 0.05$ and Λ are $g_1 = g_2 = h_1 = h_2 = 0.01$. These settings

Algorithm 1 Gibbs sampler for FIP-WMDS.

Input: X, ω, and d_{ij}^{ω} (m high-dimensional pairwise distances).

Output: $\hat{R}, \hat{\lambda}_1, \hat{\lambda}_2$, and $\hat{\sigma}^2$.

1: **Initialize** $R^{(0)}, \lambda_1^{(0)}, \lambda_2^{(0)}, \sigma^{2(0)}$.
2: **For** t **in** $1:T$ {
3: Update $R^{(t)} \sim f(R_{n\times 2}^{(t)} \mid X_{n\times p}, \omega, \lambda_1^{(t-1)}, \lambda_2^{(t-1)}, \sigma^{2(t-1)})$ using a Metropolis-Hastings (MH) step and proposal $N(R^{(t-1)}, c_1\sigma^{2(t-1)}/(n-1))$ for constant c_1.
4: Update $\sigma^{2(t)} \sim f(\sigma^{2(t)} \mid X_{n\times p}, \omega, R_{n\times 2}^{(t)}, \lambda_1^{(t-1)}, \lambda_2^{(t-1)})$ using a MH step and truncated normal proposal.
5: Update $\lambda_q^{(t)} \sim f(\lambda_q^{(t)} \mid X_{n\times p}, \omega, R_{n\times 2}^{(t)}, \sigma^{2(t)}) \equiv IG\left(c_q + n/2, d_q + \frac{1}{2}\sum_{i=1}^{n} r_{iq}^{2(t)}\right)$ for $q = 1, 2$.
6: } **END**

form weakly-informative priors, and sensitivity analyses on hyperparameter values in (0,1] shows no sensitivity in the setting these values.

We have tuning constants for our MCMC to assure reasonable acceptance in our MH steps. They are for the random walk proposal of coordinates R and the truncated normal proposal of σ^2. For R, the step size is $c_R\frac{\sigma^2}{(n-1)}$, and for Example 1, $c_R \approx 13.5$ across different variable weight vectors while in Example 2, $c_R \approx 15$. For σ^2, the tuning parameter is a constant c_s, and for Example 1, $c_s = 28$ while for Example 2, $c_2 = 30$.

For initialization of the Gibbs sampler, Oh and Raftery [23] suggest using estimates (in their case classical MDS) from SMACOF WMDS to help reduce number of iterations T. This is especially important when number of observations n is large. In our case with $n = 20$, we used a random initialization of our coordinates and values of 4 and 5 for σ^2 and both λ_1 and λ_2.

Convergence of our MCMC is determined by the Geweke diagnostic, our acceptance rates in our MH steps, and trace plots [10]. For R and σ^2, we aim for acceptance rates within [0.20, 0.40] with R on average a rate of 0.30 and σ^2 on average a rate of 0.27.

A.4 BIP-WMDS posterior distribution

We write the joint posterior distribution for this model:

$$f(\omega, \sigma^2 \mid X, R) \propto (\sigma^2)^{-(m/2+b+1)} \left(\prod_{k=1}^{p} w_k^{a_k-1}\right) \exp\left\{-\frac{(\text{stress}+2c)}{2\sigma^2} - \sum_{1\le i<j\le n} \log\Phi\left(\frac{d_{ij}^{\omega}}{\sigma}\right)\right\}.$$

A.5 Gibbs sampler for BIP-WMDS

Parameters are updated in the following Gibbs sampler using T iterations. We use the same parameter setting in our Dirichlet proposal for $\omega^{(t)}$ as Sewell and Chen [28]. Such a proposal allows for small changes in the weight for a large constant v.

A.6 BIP-WMDS Settings for Example 1 and 2

For both examples, we run our Gibbs sampler for $T = 100{,}000$ iterations with a burn-in of 10,000. Hyperparameter settings for σ^2 are $a = b = 0.05$. This is still a weakly-informative prior and there is no sensitivity in setting values in $(0, 1]$.

We have tuning constants for our MCMC to assure reasonable acceptance rates in our MH steps. They are for the random walk proposal of weights ω and the truncated normal proposal of σ^2. For ω, the constant $v = 100$ and $v = 2050$ for Example 1 and 2 respectively. For σ^2, the tuning parameter is a constant c_s, and

Algorithm 2 Gibbs sampler for BIP-WMDS.

Input: X, R, and d_{ij}^{r} (m low-dimensional pairwise distances).

Output: $\hat{\omega}$ and $\hat{\sigma}^2$.

1: **Initialize** $\omega^{(0)}, \sigma^{2(0)}$.
2:
3: **For** t **in** $1:T$ {
4: Update $\omega^{(t)} \sim f(\omega^{(t)} \mid X, R, \sigma^{2(t-1)})$ using a MH step and proposal of Dirchlet$(v \cdot \omega^{(t-1)})$.
5: Update $\sigma^{2(t)} \sim f(\sigma^{2(t)} \mid X, R, \omega^{(t)})$ using a MH step and truncated normal proposal.
6: } **END**

for Example 1, $c_s = 28$ while for Example 2, $c_2 = 35$. Lastly, we initialize the variable weights using $\{1/p\}_{k=1}^{p}$ and a value of 4 for σ^2.

Lastly, for convergence, we use the same diagnostics mentioned in FIP-WMDS. Again, for ω and σ^2, we aim for acceptance rates within [0.20, 0.40] with ω on average a rate of 0.21 and σ^2 on average a rate of 0.26.

B Variability of Weights in Example 2

Table 6: True weights ω_j, posterior estimate of weights, $\hat{\omega}$, and range of top 10, $r\omega$, from BIP-WMDS.

True	Max Post	Range		True	Max Post	Range
0.054	0.038	[0.027, 0.06]	Ca	0.010	0.034	[0.03, 0.034]
0.020	0.028	[0.018, 0.03]	Fi	0.098	0.037	[0.031, 0.044]
0.078	0.011	[0.011, 0.069]	Qu	0.029	0.028	[0.02, 0.028]
0.146	0.076	[0.026, 0.076]	Ln	0.083	0.037	[0.021, 0.047]
0.052	0.023	[0.018, 0.052]	Bh	0.011	0.026	[0.022, 0.029]
0.005	0.026	[0.025, 0.034]	Bt	0.005	0.018	[0.018, 0.026]
0.000	0.015	[0.015, 0.03]	Tl	0.018	0.021	[0.014, 0.028]
0.006	0.027	[0.009, 0.031]	Ti	0.181	0.067	[0.059, 0.085]
0.051	0.028	[0.026, 0.053]	Pl	0.021	0.022	[0.022, 0.045]
0.012	0.016	[0.015, 0.035]	Pa	0.039	0.017	[0.017, 0.031]
0.097	0.044	[0.022, 0.066]	Hv	0.014	0.037	[0.023, 0.038]
0.025	0.035	[0.018, 0.039]	Hi	0.043	0.052	[0.013, 0.052]
0.025	0.015	[0.015, 0.027]	Wa	0.007	0.034	[0.03, 0.04]
0.065	0.038	[0.024, 0.056]	Hp	0.003	0.021	[0.021, 0.033]
0.008	0.047	[0.029, 0.049]	St	0.003	0.035	[0.019, 0.035]
0.004	0.056	[0.017, 0.056]	Ow	0.028	0.028	[0.016, 0.037]
0.003	0.013	[0.013, 0.044]	Sz	0.023	0.017	[0.017, 0.039]
0.004	0.036	[0.023, 0.041]	Bb	0.035	0.048	[0.048, 0.078]
0.082	0.049	[0.024, 0.049]	Sm	0.008	0.038	[0.038, 0.065]
0.019	0.042	[0.03, 0.044]	Hn	0.024	0.029	[0.02, 0.029]
0.009	0.043	[0.029, 0.048]	Ag	0.018	0.054	[0.022, 0.054]
0.008	0.023	[0.023, 0.048]	Wk	0.024	0.036	[0.021, 0.036]
0.054	0.056	[0.01, 0.057]	Ct	0.002	0.040	[0.025, 0.04]
0.028	0.022	[0.008, 0.028]	Ts	0.025	0.019	[0.019, 0.043]
0.052	0.037	[0.034, 0.058]	Sp	0.035	0.045	[0.037, 0.051]
0.061	0.040	[0.031, 0.057]	Tr	0.021	0.017	[0.01, 0.036]
0.019	0.027	[0.017, 0.035]	Ia	0.076	0.029	[0.028, 0.05]
0.005	0.034	[0.022, 0.057]	Gr	0.033	0.041	[0.026, 0.041]
0.006	0.022	[0.013, 0.028]	Co	0.044	0.036	[0.036, 0.054]
0.002	0.033	[0.016, 0.044]	Nt	0.038	0.038	[0.025, 0.051]

Funding

This work is supported by the National Science Foundation under Grants 1447416 and 1141096.

References

[1] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7):1304–1330, 2007.

Advances in Computational Intelligence and Learning. doi: 10.1016/j.neucom.2006.11.018

[2] M. Avriel. *Nonlinear programming: analysis and methods*. Prentice-Hall, 1976.

[3] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization*, pp. 3–27. Springer, 2014.

[4] L. Bradel, C. North, L. House, and S. Leman. Multi-model semantic interaction for text analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 163–172, Oct 2014. doi: 10.1109/VAST.2014.7042492

[5] M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pp. 1–8. ACM, New York, NY, USA, 2014. doi: 10.1145/2669557.2669559

[6] J. Chuang, D. Ramage, C. Manning, and J. Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 443–452. ACM, New York, NY, USA, 2012. doi: 10.1145/2207676.2207738

[7] J. De Leeuw and P. Mair. Multidimensional scaling using majorization: Smacof in r. *Journal of Statistical Software*, 2011.

[8] A. Endert, P. Fiaux, and C. North. Semantic interaction for sensemaking: inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, 2012.

[9] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 473–482. ACM, 2012.

[10] J. Geweke. Interpretation and inference in mixture models: Simple mcmc works. *Computational Statistics & Data Analysis*, 51(7):3529–3550, 2007.

[11] C. Han, L. House, and S. C. Leman. Expert-guided generative topographical modeling with visual to parametric interaction. *PloS one*, 11(2):e0129122, 2016.

[12] L. House, S. Leman, and C. Han. Bayesian visual analytics: Bava. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 8(1):1–13, 2015.

[13] X. Hu, L. Bradel, D. Maiti, L. House, C. North, and S. Leman. Semantics of directly manipulating spatializations. *IEEE Transactions on Visualization and Computer Graphics*, 19:2052 – 2059, 12/2013 2013. doi: 10.1109/TVCG.2013.188

[14] D. H. Jeong, C. Ziemkiewicz, B. Fisher, W. Ribarsky, and R. Chang. ipca: An interactive system for pca-based visual analytics. *Computer Graphics Forum*, 28(3):767–774, 2009. doi: 10.1111/j.1467-8659.2009.01475.x

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

[16] C. H. Lampert, H. Nickisch, S. Harmeling, and J. Weidmann. Animals with attributes: A dataset for attribute based classification, 2009.

[17] S. C. Leman, L. House, D. Maiti, A. Endert, and C. North. Visual to parametric interaction (v2pi). *PloS one*, 8(3):e50474, 2013.

[18] S. Lespinats and M. Aupetit. Checkviz: Sanity check and topological clues for linear and non-linear mappings. *Computer Graphics Forum*, 30(1):113–125, 2011.

[19] J. Lewis, L. Van der Maaten, and V. de Sa. A behavioral investigation of dimensionality reduction. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 34, pp. 671–676, 2012.

[20] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, Mar 1982. doi: 10.1109/TIT.1982.1056489

[21] P. Mair, I. Borg, and T. Rusch. Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate Behavioral Research*, 51:1–18, 11 2016.

[22] T. Munzner. *Visualization analysis and design*. CRC press, 2014.

[23] M.-S. Oh and A. E. Raftery. Bayesian multidimensional scaling and choice of dimension. *Journal of the American Statistical Association*, 96(455):1031–1044, 2001. doi: 10.1198/016214501753208690

[24] M. Sedlmair, M. Brehmer, S. Ingram, and T. Munzner. Dimensionality reduction in the wild: Gaps and guidance. *Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC, Canada, Tech. Rep. TR-2012-03*, 2012.

[25] C. Seifert, V. Sabol, and W. Kienreich. Stress maps: analysing local phenomena in dimensionality reduction based visualisations. In *Proceedings of the 1st European Symposium on Visual Analytics Science and Technology (EuroVAST10)*, vol. 1, 2010.

[26] J. Z. Self, X. Hu, L. House, S. Leman, and C. North. Designing usable interactive visual analytics tools for dimension reduction. In *CHI 2016 Workshop on Human-Centered Machine Learning (HCML)*, p. 7, 05/2016 2016.

[27] J. Z. Self, R. K. Vinayagam, J. T. Fry, and C. North. Bridging the gap between user intention and model parameters for human-in-the-loop data analytics. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, HILDA '16, pp. 3:1–3:6. ACM, New York, NY, USA, 2016. doi: 10.1145/2939502.2939505

[28] D. K. Sewell and Y. Chen. Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657, 2015.

[29] J. Stahnke, M. Drk, B. Mller, and A. Thom. Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):629–638, Jan 2016. doi: 10.1109/TVCG.2015.2467717

[30] J. Wenskovitch and C. North. Observation-level interaction with clustering and dimension reduction algorithms. In *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics*, HILDA'17, pp. 14:1–14:6. ACM, New York, NY, USA, 2017. doi: 10.1145/3077257.3077259

[31] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Proceedings of Visualization 1995 Conference*, pp. 51–58, Oct 1995. doi: 10.1109/INFVIS.1995.528686

[32] C. M. Wittenbrink, A. T. Pang, and S. K. Lodha. Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics*, 2(3):266–279, Sep 1996. doi: 10.1109/2945.537309

[33] Y. Wu, G. X. Yuan, and K. L. Ma. Visualizing flow of uncertainty through analytical processes. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2526–2535, Dec 2012. doi: 10.1109/TVCG.2012.285