Viewpoint-Based Quality for Analyzing and Exploring 3D Multidimensional Projections

Wouter Castelein[a], Zonglin Tian[b], Tamara Mchedlidze[c], and Alexandru Telea[d]
Department of Information and Computing Science, Utrecht University, Netherlands
wmcastelein@gmail.com, {z.tian, t.mtsentlintze, a.c.telea}@uu.nl

Keywords: multidimensional projections, visual quality metrics, perception, user studies

Abstract: While 2D projections are established tools for exploring high-dimensional data, the effectiveness of their 3D counterparts is still a matter of debate. In this work, we address this from a multifaceted quality perspective. We first propose a viewpoint-dependent definition of 3D projection quality and show how this captures the visual variability in 3D projections much better than aggregated, single-value, quality metrics. Next, we propose an interactive exploration tool for finding high-quality viewpoints for 3D projections. We use our tool in an user evaluation to gauge how our quality metric correlates with user-perceived quality for a cluster identification task. Our results show that our metric can predict well viewpoints deemed good by users and that our tool increases the users' preference for 3D projections as compared to classical 2D projections.

1 INTRODUCTION

Dimensionality reduction (DR), also called projection, is a popular technique for visualizing high-dimensional datasets by low-dimensional scatterplots. Tens of different DR techniques (Espadoto et al., 2019) have been designed to address the several requirements one has for this class of methods, such as computational scalability, ease of use, robustness to noise or small data changes, projecting additional points along those existing in an original dataset (out-of-sample ability), and visual quality.

Visual quality is a key requirement for DR methods. Globally put, a good projection scatterplot captures well the so-called *data structure* present in the original high-dimensional data in terms of point clusters, outliers, and correlations (Nonato and Aupetit, 2018; Espadoto et al., 2019; Lespinats and Aupetit, 2011). As such, high-quality projections are essential to allow users to reason about the data structure by exploring the visual structure of the scatterplot.

Projection techniques used for visualization purposes can typically create 2D or 3D scatterplots equally easily. For brevity, we call such scatterplots 2D and 3D projections respectively. In contrast to 2D projections, 3D projections have one extra dimension to project the data (thus, can in principle achieve higher quality). However, the user must choose a suitable viewpoint for analysis. Hence, to assess 3D projection quality, we cannot simply reuse the viewpoint-independent metrics used for 2D projections, but must also consider the viewpoint information.

Relatively few works studied 3D projections and mainly by comparing their ease of interpretation and use for selected tasks by means of user studies. In this paper, we aim to extend such insights by answering the following questions:

Q1: How can we **measure** the quality of 3D projections by means of quantitative metrics?

Q2: How do 3D projections **compare** with their 2D counterparts (generated on the same datasets by the same projection technique) from the perspective of these metrics?

Q3: How do our proposed quality metrics correlate with quality as **perceived** by actual users?

We answer these questions as follows. We measure 3D projection quality by a *function* (rather than a single value) that evaluates existing 2D quality metrics over a large set of 2D viewpoints of the 3D projection (Q1). Next, we quantitatively analyze 30 3D projections (five techniques run on six datasets) and find that most views of a 3D projection are of relatively high quality, with only a few poor views, and that these good views can have higher quality than a 2D projection made with the same technique for the same dataset (Q2). We propose an interactive tool for

[a] https://orcid.org/0000-0002-4964-4670
[b] https://orcid.org/0000-0001-5626-402X
[c] https://orcid.org/0000-0001-6249-3419
[d] https://orcid.org/0000-0003-0750-0502

exploring the viewpoint-based quality. We perform a user study to test which viewpoints users *perceive* to be good for a cluster separation task and if these have high quality values as *measured* by our metric (Q3). We find a correlation of perceived *vs* computed quality, which suggests that the latter can be used to predict the former. When showing users the computed quality as they search for good viewpoints, we found out that users tend to select even higher-quality views. Our study shows that users preferred in most cases the 3D projections (with their own selected viewpoints) to the static 2D projections, and that, in these cases, the computed quality of the selected views was at the high end of the spectrum of qualities reachable by viewpoints of a 3D projection, and comparable to the quality of corresponding 2D projections.

2 RELATED WORK

We start by listing some notations. Let $D = \{\mathbf{x}_i\}$ be a dataset of n-dimensional samples or points $\mathbf{x}_i \in \mathbb{R}^n$. A projection P maps D to $P(D) = \{\mathbf{y}_i\}$, where $\mathbf{y}_i \in \mathbb{R}^q$ is the projection of $\mathbf{x}_i$. Typically $q \ll n$, yielding 2D projections ($q = 2$) and 3D projections ($q = 3$) that are used to visualize D by depicting the respective scatterplots. We next use P_2 to denote a technique P that creates 2D projections ($q = 2$); P_3 for 3D projections ($q = 2$); and P when the dimension q is not important.

A quality metric is a function $M(D, P(D)) \rightarrow \mathbb{R}^+$ that tells how well the scatterplot $P(D)$ captures aspects of the dataset D. We next discuss metrics for 2D projections (Sec. 2.1) and 3D projections (Sec. 2.2).

2.1 Measuring 2D projection quality

Measuring 2D projection quality is a well-established field which can be split into three types of methods.

Quantitative metrics: M is computed by directly analyzing D and $P(D)$. Examples of M are listed below.

Trustworthiness T measures the fraction of points in D that are also close in $P(D)$ (Venna and Kaski, 2006). In T's definition (Tab. 1), $U_i^{(K)}$ are the K nearest neighbors of $\mathbf{y}_i$ which are not among the K nearest neighbors of $\mathbf{x}_i$; and $r(i, j)$ is the rank of $\mathbf{y}_j$ in the ordered-set of nearest neighbors of $\mathbf{y}_i$. High trustworthiness implies that visual patterns in $P(D)$ represent actual patterns in D, *i.e.*, the projection has few so-called false neighbors (Martins et al., 2014). Continuity C, a related metric, measures the fraction of points in $P(D)$ that are also close in D (Venna and Kaski, 2006). In C's definition (Tab. 1), $V_i^{(K)}$ is the set of points that are among the K nearest neighbors of $\mathbf{x}_i$ but not among the K nearest neighbors of $\mathbf{y}_i$; and $\hat{r}(i, j)$ is the rank of $\mathbf{y}_j$ in the ordered-set of nearest neighbors of $\mathbf{x}_i$. High continuity implies that data patterns in D are captured by $P(D)$, *i.e.*, the projection has few so-called missing neighbors (Martins et al., 2014).

Normalized stress N measures how well inter-point distances in $P(D)$ reflect the same inter-point distances in D, *i.e.*, how well users can retrieve distance information from the projection (Joia et al., 2011). Different distance metrics Δ^n for D, and Δ^q for $P(D)$ respectively can be used (Tab. 1), the most typical being the L_2 metric. Distance preservation can also be measured by the Shepard diagram (Joia et al., 2011), a scatterplot of the pairwise L_2 distances between all points in $P(D)$ *vs* the corresponding distances in D. Points close to the main diagonal show a good distance preservation. The diagram can be reduced to a single metric value by computing its Spearman rank correlation S, where $S = 1$ denotes a perfect (positive) correlation of distances in D and $P(D)$.

Many other quantitative metrics exist for 2D projection, *e.g.*, the neighborhood hit (NH) that captures how well $P(D)$ captures same-label clusters in D (Paulovich et al., 2008; Rauber et al., 2017; Aupetit, 2014); Distance Consistency (DSC) (Sips et al., 2009) or Class Consistency Measure (CCM) (Tatu et al., 2010; Sedlmair and Aupetit, 2015)), which tell how well $P(D)$ is separated into visually distinct, same-label, clusters. Additional visual separation metrics are given by (Albuquerque et al., 2011; Sedlmair et al., 2013; Motta et al., 2015). We do not use these metrics since they either need labeled data and/or do not measure how well $P(D)$ preserves aspects of D but rather how well humans separate $P(D)$ into visual clusters, which is a different task than ours.

Error views: In contrast to quantitative metrics which produce a single scalar value $M \in \mathbb{R}^+$, error views produce a *set* of values, typically one per projection point $\mathbf{y}_i$. These include the projection precision score (Schreck et al., 2010), which captures the aggregated difference between the distances of a point in $P(D)$ to its K nearest neighbors in D, respectively $P(D)$; stretching and compression (Aupetit, 2007; Lespinats and Aupetit, 2011), which measure the increase (stretching), respectively decrease (compression) of distances of a point to all other points in $P(D)$ *vs* the corresponding distances in D; and the average local error (Martins et al., 2014), which aggregates stretching and compression. Error views give fine-grained insight on where in a projection distance-preservation errors occur. Such views are mainly intended for human analysis, *i.e.*, they cannot be easily used to automatically compare many projection instances, which is our goal.

User studies: Both quantitative metrics and error

Table 1: Projection quality metrics used in this paper. All metrics range in $[0 = worst, 1 = best]$.

Metric	Definition (taken from (Espadoto et al., 2019))
Trustworthiness (T)	$1 - \frac{2}{NK(2N-3K-1)} \sum_{i=1}^{N} \sum_{j \in U_i^{(K)}} (r(i,j) - K)$
Continuity (C)	$1 - \frac{2}{NK(2N-3K-1)} \sum_{i=1}^{N} \sum_{j \in V_i^{(K)}} (\hat{r}(i,j) - K)$
Normalized stress (N)	$\frac{\sum_{ij} (\Delta^n(\mathbf{x}_i,\mathbf{x}_j) - \Delta^q(P(\mathbf{x}_i),P(\mathbf{x}_j)))^2}{\sum_{ij} \Delta^n(\mathbf{x}_i,\mathbf{x}_j)^2}$
Shepard goodness (S)	Spearman rank correlation of scatterplot $(\lVert \mathbf{x}_i - \mathbf{x}_j \rVert, \lVert P(\mathbf{x}_i) - P(\mathbf{x}_j) \rVert), 1 \le i \le N, i \neq j$

views do not directly gauge the *usefulness* of a projection. (Nonato and Aupetit, 2018) discuss this in detail and provide a taxonomy of tasks for projections and evaluation methods for these, mainly based on user studies. (Espadoto et al., 2019) acknowledge that task-based evaluations provide insight in the application *value* of projections but argue that evaluating metrics over large dataset collections and projection hyperparameter settings provides complementary insight in the projections' technical *quality* and is much easier to automate than user studies. Quantitative metrics are a first, necessary, step to assess projections, to be followed by more specific task-based user studies – an approach that we follow in our work.

2.2 Measuring 3D projection quality

All abovementioned 2D projection quality metrics can be easily computed for 3D projections. Recently, (Tian et al., 2021b) compared 29 projection techniques across 8 datasets using the T, C, and S metrics mentioned above, computed following their definitions in Tab. 1 for the 2D, respectively 3D, scatterplots $P(D)$. They found a small increase of quality metrics (on average, 3%) for the 3D *vs* the 2D projections. They also qualitatively (as opposed to our quantitative measurements) studied how users perceive projections and found that, for tasks involving explaining groups of close points by similar values of dimensions, 3D projections can show more insights than their 2D counterparts.

Simply transposing 2D quality metrics to 3D has a major issue. Even if such metrics score highly on a 3D projection, this does not mean that that projection *shows* data patterns to users well. Indeed, the metrics 'see' $P(D)$ in three dimensions; users see only 2D *views* of $P(D)$ from chosen viewpoints. Information encoded along the viewing direction is thus used by the metrics but is not seen by the user. This can *e.g.* artificially make the metrics indicate higher quality which the user actually does not see.

Apart from the above, and in line with the observations in (Nonato and Aupetit, 2018) for 2D projections, 3D projections can actually bring added value which cannot be easily captured by automatically-computed metrics. As such, 3D projections have been mainly assessed in the literature by user studies. Several such examples follow.

(Poco et al., 2011) compare 2D *vs* 3D projections and show that 3D scores better than 2D for the *NH* and *C* metrics. They refine this insight by a user study where 12 participants were asked to count clusters, order clusters by density, list all pairwise cluster overlaps, detect an object within a cluster, find the cluster closest to a given point (all operations involve *visual* clusters shown by $P(D)$). Users were better able to provide the correct answer for these tasks in 3D (74.4%) than in 2D (64.3%). Yet, the only statistically significant improvement was found for the last task. Also, users needed around 50% more time for these tasks in 3D. Overall the work suggests a slight improvement when using 3D, but it lacks certainty.

(Sedlmair et al., 2013) asked two experienced coders to rate how well classes of 75 labeled datasets were separable in a 2D projection, an interactive 3D projection, and a scatterplot matrix. They found that the 2D projection was often good enough to visualize separate classes and was also the fastest method to use. The interactive 3D projection scored better than the 2D one and the scatterplot matrix only for highly synthetic (abstract) datasets. A limitation of this study is that it involved only two users. In contrast, our evaluations for a similar task described in this paper involve 22 participants.

Several works compared 2D with 3D scatterplots and argued for the latter as better in capturing sample density variations (Sanftmann and Weiskopf, 2009; Sanftmann and Weiskopf, 2012) and having less information loss (Chan et al., 2014). However, it is important to note that interpreting 3D scatterplots whose axes directly encode data *dimensions* is very different from interpreting 3D projections where the three axes often have no meaning.

3 VIEWPOINT-DEPENDENT 3D PROJECTION QUALITY

A first conclusion from Sec. 2 is that (a) quantitative metrics are an useful, scalable, generic, and accepted first step for evaluating 2D projections but (b) we lack such metrics for the 3D case.

Metric design: We construct such 3D projection met-

rics based on the well-known and accepted 2D projection metrics (Sec. 2.1) as follows. Take a 3D projection $P_3(D)$ which is explored from multiple viewpoints using a virtual trackball metaphor. Let $\mathbf{p} \in \mathbb{R}^3$ be a viewing direction pointing to the center of $P_3(D)$. Let $Q(\mathbf{p}, P_3(D))$ be the *view* of $P_3(D)$ from direction $\mathbf{d}$, *i.e.*, the 2D scatterplot of the orthographic projection of $P_3(D)$ on a plane orthogonal to $\mathbf{p}$.

$Q(\mathbf{p}, P_3(D))$ is a 2D scatteplot, so we can measure its quality by directly applying all metrics in Tab. 1, or any other quality metric M for 2D projections, on it. Hence, we can describe the quality of $P_3(D)$ by a function $M(D, Q(\mathbf{p}, P_3(D))$ of the viewpoint $\mathbf{p}$. Note that we can ignore in-plane (around $\mathbf{p}$) rotations since these do not change the inter-point distances in the 2D scatterplot $Q(\mathbf{p}, P_3(D))$ that all metrics in Tab. 1 use.

To analyze M, we sample it over a set of viewpoints $V = \{\mathbf{p}_i | 1 \leq i \leq s\}$ uniformly distributed over a sphere using the spherical Fibonacci lattice algorithm (Gonzalez, 2010) with $s = 1000$. Other sampling methods can be readily used, *e.g.* (Camahort et al., 1998; Levoy, 2006). Sampling yields a dataset $\widetilde{M} = \{M(D, Q(\mathbf{p}, P_3(D)) | \mathbf{p} \in V\}$ which is our replacement of the scalar metric M to evaluate 3D projection quality.

As mentioned in Sec. 2.2, users do not see any information along the viewing direction $\mathbf{p}$. Hence, views Q have occluded points (along $\mathbf{p}$) which our metric $M(D, Q(\mathbf{p}, P_3(D))$ does not account for. We do not handle occlusions when computing $\widetilde{M}$ since all uses we know of quality metrics M for 2D projections in the literature have *exactly* the same accepted problem, albeit for a different reason, *i.e.* overdraw due to not-ideal T values.

Visual exploration tool: We implemented an interactive tool for exploring and comparing the 3D and 2D projections $P_3(D)$ and $P_2(D)$ and their computed quality metrics. We next describe this tool, which is key to our user evaluation (see next Sec. 5).

Figure 1 shows our tool's four views. Views (a) and (b) show the 3D projection $P_3(D)$, respectively the corresponding 2D projection $P_2(D)$, of a dataset D. Views (c,d) allow comparing P_2 and P_3 to decide which is better for the task at hand, as follows.

Quality distribution: View (c) renders $\widetilde{M}$ (for a user-chosen $M \in \{N, S, C, T\}$) over all directions V by color-coding points $\mathbf{p}$ on a sphere via an ordinal (red-yellow-green) colormap. For example, red points show viewing directions $\mathbf{p}$ from which $\widetilde{M}$ is low. The current viewpoint used in view (a) is always at the center of the sphere, see black cross in (b). Rotating either the 3D projection (a) or the sphere (b) updates the other view: Rotating the sphere allows users to find viewpoints of high quality $\widetilde{M}$ and see how the 3D projection looks from them. Rotating the 3D projec-

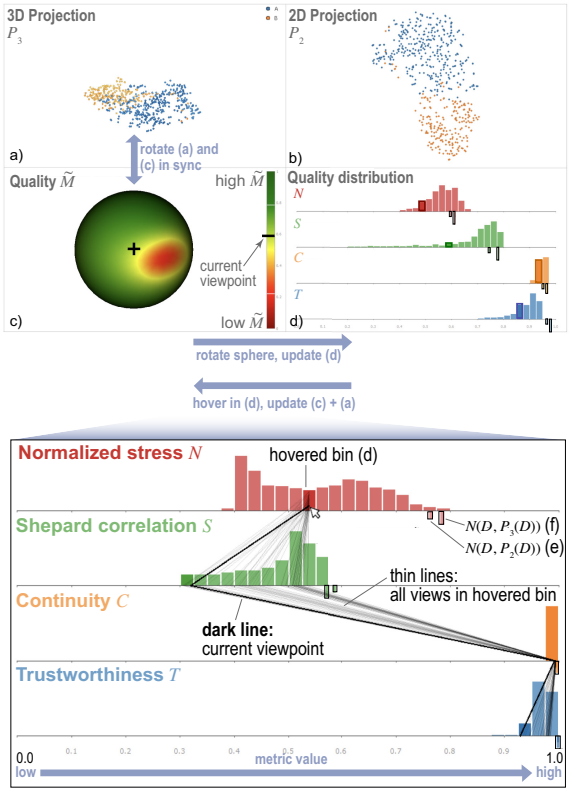

Figure 1: Tool for exploring 2D/3D projection quality.

tion allows users to find interesting patterns and see if they can trust them, *i.e.*, if $\widetilde{M}$ is high for those viewpoints. Our viewpoint exploration by sphere rotation is conceptually related to the mechanism in (Coimbra et al., 2016). However, the latter encodes *explanations* of the different viewpoints of a 3D projection, whereas we encode projection *quality*.

View (d) shows all the quality metrics N, S, C, and T for both $P_3(D)$ and $P_2(D)$ using one annotated histogram per metric, as follows (see also inset in Fig. 1 bottom for details). The histogram shows the number of views in V that have quality values $\widetilde{M}$ falling in a given bin; the range $[0, 1]$ of $\widetilde{M}$ is uniformly divided in 40 such bins. Hence, long bars indicate $\widetilde{M}$ values reached by many viewpoints; short bars indicate $\widetilde{M}$ values that only few viewpoints have. Histograms shifted to the right tell that the 3D projection has high quality from most viewpoints, as is the case for the C and T metrics in Fig. 1 (inset); histograms shifted to the left tell that the 3D projection has poor quality from most viewpoints, as is the case of the S metric in Fig. 1 (inset). Disagreement of the four histograms tells that it is hard to find views deemed good from the perspective of *all* four quality metrics.

Single-value metrics: A small, respective large, tick shown under the histogram tells the value of the quality metric for the 2D projection, *i.e.*, $M(D, P_2(D))$,

and respectively the value of M computed *directly* on the 3D projection, *i.e.*, $M(D, P_2(D))$. Seeing where the small tick falls within the histogram range tells how easy is to find viewpoints from which the 3D projection has a better quality than the single-view 2D projection. For example, Fig. 1 (inset) shows that the small-tick for N is very close to the right end of the respective histogram. There are only two, shallow, bars to the right of this tick. So, it is quite hard to find viewpoints in which the 3D projection has a higher N than the 2D projection. The large tick shows why computing a single value for M in 3D is not insightful. For example, in Fig. 1 (inset), the large tick for N indicates a very high quality, larger than almost all the per-viewpoint N values for the 3D projection and also larger than the N of the 2D projection. However, as we argued in Sec. 2.2, users do not inherently 'see' the 3D projection but only a 2D orthographic view thereof. As such, the quality value $M(D, P_3(D))$ indicated by the long tick can never be reached in practice.

Visiting viewpoints in quality order: We further link views (c) and (d) by interaction. When the user rotates the viewpoint sphere in (c), the bins of the four histograms in (d) in which the currently selected viewpoint (crosshair in (c)) falls are rendered in a darker hue. This helps the user to see all *four* quality metrics for the respective viewpoint. Conversely, when the user moves the mouse over a bar in (d), the sphere and 3D projection rotate to a viewpoint that has a quality value within the bar's interval. Moving the mouse inside the bar from bottom to top selects views with quality values increasing from the lower end to the higher end of the interval. This allows one to quickly scan, in increasing order, all 3D projection viewpoints with quality values in a given interval.

Using all four quality metrics: When hovering over a histogram bar, we also draw a Parallel Coordinates Plot (PCP) from the hovered bar to the other three histograms. If V_0 is the number of viewpoints in the hovered bar, the PCP will contain V_0 polylines (rendered half-transparent to reduce visual clutter), each showing the four quality values for the V_0 viewpoints. A thicker, more opaque, polyline shows the quality of the currently selected viewpoint. The PCP plot shows how, for a selected range of one quality metric (hovered bar), the quality of the other three metrics varies. For example, the PCP in Fig. 1 (inset) shows that all viewpoints with a N value around 0.53 (red hovered bar) have S values that cover almost the entire spectrum of S (since PCP lines fan out from the red bar to almost all green bars except the two rightmost ones), and very similar C and T values (since the lines fan in when reaching the orange and blue histograms respectively). Moving the mouse over the PCP plot selects the closest polyline and makes it the current viewpoint. This allows users to effectively explore the viewpoint space V using *all four* metric values to *e.g.* choose a viewpoint where one, or several, metrics have high values (if such a viewpoint exists).

4 QUANTITATIVE COMPARISON

We use the tool described in Sec. 3 to study how the viewpoint-dependent quality of 3D projections compares among several datasets and projection techniques and also how it compares with the quality of corresponding 2D projections, thereby answering Q1.

4.1 Datasets and techniques

We used 6 different real-world datasets and 5 projection techniques, so a total of 30 2D and 3D projection-pairs to explore. Datasets were selected from the benchmark in (Espadoto et al., 2019) and have varying numbers of samples and dimensions; have categorical, ordinal, or no labels; and come from different application areas (Tab. 2). Projection techniques were selected from the same benchmark and include global-*vs*-local, linear-*vs*-nonlinear, approaches, using both samples and sample-pair distances as inputs.

Table 2: Datasets and techniques used to compare 2D and 3D projections.

Dataset	Samples	Dims	Labels	Domain
AirQuality (Vito et al., 2008)	9357	13	-	physics
Concrete (Yeh, 2021)	1030	8	ordinal	chemistry
Reuters (Lewis and Shoemaker, 2021)	8432	1000	categories	text
Software (Meirelles et al., 2010)	6773	12	ordinal	software
Wine (Cortez et al., 2009)	6497	11	ordinal	chemistry
WBC (Dua and Graff, 2017)	569	30	categories	medicine

Technique	Linearity	Input	Locality
Autoencoders (AE) (Bank et al., 2020)	nonlinear	samples	global
MDS (Tenenbaum et al., 2000)	nonlinear	distances	global
PCA (Jolliffe, 2002)	linear	samples	global
t-SNE (van der Maaten and Hinton, 2008)	nonlinear	distances	local
UMAP (McInnes et al., 2018)	nonlinear	distances	local

For each technique-dataset combination (P, D), we computed the 2D and 3D projections $P_2(D)$ and $P_3(D)$ and next measured the single-value metrics ($M(D, P_2(D))$ and $M(D, P_3(D))$) and the viewpoint-dependent $\widetilde{M}$ for the four metrics in Tab. 1. We compute T and C with $K = 7$ neighbors as in (van der Maaten and Postma, 2009; Martins et al., 2015; Espadoto et al., 2019). Our results, and our tool's source code, are publicly available (The Authors, 2022).

4.2 Viewpoint-dependent metrics

Given our new way to evaluate quality as a viewpoint function $\widetilde{M}$ (Sec. 3), it makes sense to start exploring how $\widetilde{M}$ varies over all evaluated combinations of datasets, projection techniques, and quality metrics. Figure 2 shows a table, with a row per projection, ordered first by dataset and next by projection technique. Each row shows two snapshots of the quality sphere (as in Fig. 2c) for each quality metric, taken from two opposite viewpoints (chosen arbitrarily), so that one can see nearly the whole sphere. As we have four quality metrics, there is a total of 8 such sphere snapshots. Figure 2 further shows the four metric histograms (as in Fig. 1d).

Figure 2 leads us to the following insights.

Metric ranges: An immediate observation is that T and especially C have a (very) narrow range which is also close to 1 (maximal quality), *i.e.*, C and T have very high values *regardless* of the viewpoint. In contrast, S and N vary much more over all viewpoints V. We also see this in the C and T spheres which almost fully green, whereas the S and N spheres show much more color variation. It seems, thus, that C and T cannot really indicate good viewpoint quality since, according to them, nearly all viewpoints are good.

However, changes within the very small range of C and T could be just as significant as larger changes for the other, larger-range, metrics S and N. To test this, we visually compare viewpoints with highest, respectively lowest, T and C values, for two datasets and projection techniques (Fig. 3). We found similar results for all other projection-dataset combinations (see supplementary material). We see that viewpoints with maximum metrics show high point spread, thus allow understanding the projected data well; viewpoints with minimum metrics show far less structure (due to overlap of points). This is so even though the *ranges* of these metrics is quite small – C for AirQuality only differs by 0.02 between the best and worst values; and T for WBC only differs by 0.22 for WBC.

Metric distributions: A second finding from Fig. 2 is that we see no clear correlation of quality with *datasets* but rather with projection *techniques*. Given the above two observations, we recreate Fig. 3 by using the *actual* ranges of the metric histograms to yield Fig. 4. Also, we group projections by technique rather than dataset, to study how techniques affect quality metrics. We now cannot any longer compare the actual x positions of different metric histograms. However, we now can (a) see much better how views get distributed to metric values and thus interpret the *shapes* of these distributions; and (b) how quality metrics correlate with projection techniques.

Figure 4 tells us several insights. First, we see how important is to use viewpoint dependent metrics: Viewpoints are non-uniformly spread over the (wide or narrow) ranges of the metrics; and our previous analysis (Fig. 3 and related text) showed that small metric values can correspond to big visual differences. Hence, these small metric value changes *are* important predictors of visual quality.

Secondly, we see that, in most cases, the histograms for T, C, and S have their mass skewed to their right, *i.e.*, have their most and longest bars for the higher metric values, with only a few exceptions (N, Airquality MDS; S, Software AE). Hence, in most cases, users should not have a problem in finding high-quality-metric viewpoints in 3D projections, which partially counters the argument in previous papers that viewpoint selection is a problem for 3D projections (Poco et al., 2011). We further test whether such high-quality-metric viewpoints are indeed seen as high-quality by users themselves in Sec. 5.

Thirdly, if our four metrics inherently capture 'quality', their shapes should be similar (at least for the same dataset-technique combination). Figure 4 shows that this so in most cases for the T, C, and S metrics. In contrast, the N metric has quite different histogram shapes in most cases, tending to show a preference for lower quality values. This actually correlates with qualitative observations in earlier papers (Espadoto et al., 2019; Joia et al., 2011) that N is not a good way to assess the quality of multidimensional projections. We further explore how this correlates with the actual quality perception of users in Sec. 5.

Finally, let us interpret Fig. 4 from the perspective of projection *techniques*. We see that UMAP has more 'peaked' histograms, with mass shifted to the right, than the other four techniques – also seen in the amount of green in its sphere snapshots. Hence, if we want to use a 3D projection, UMAP generates many viewpoints of similar, consistent, quality, so picking a viewpoint with UMAP is easier than for the other techniques. Along this, Fig. 2 shows tha UMAP yields higher quality metrics than the other techniques from nearly all perspectives (datasets, metrics). Concluding, UMAP is the best technique to use for 3D projections from the perspective of our four quality metrics. Interestingly, Figs. 3 and 4 show that t-SNE does not yield higher quality values (spread over all viewpoints), nor a majority of views with consistent high quality values. This is in line with earlier findings (Tian et al., 2021b; Tian et al., 2021a) that showed that t-SNE generates 'organic', round, clusters which tend to fill in the projection space. In 3D, this small separation space between clusters means these clusters will *overlap* in most 2D views of the 3D projection, *i.e.*, poor values for the four quality met-

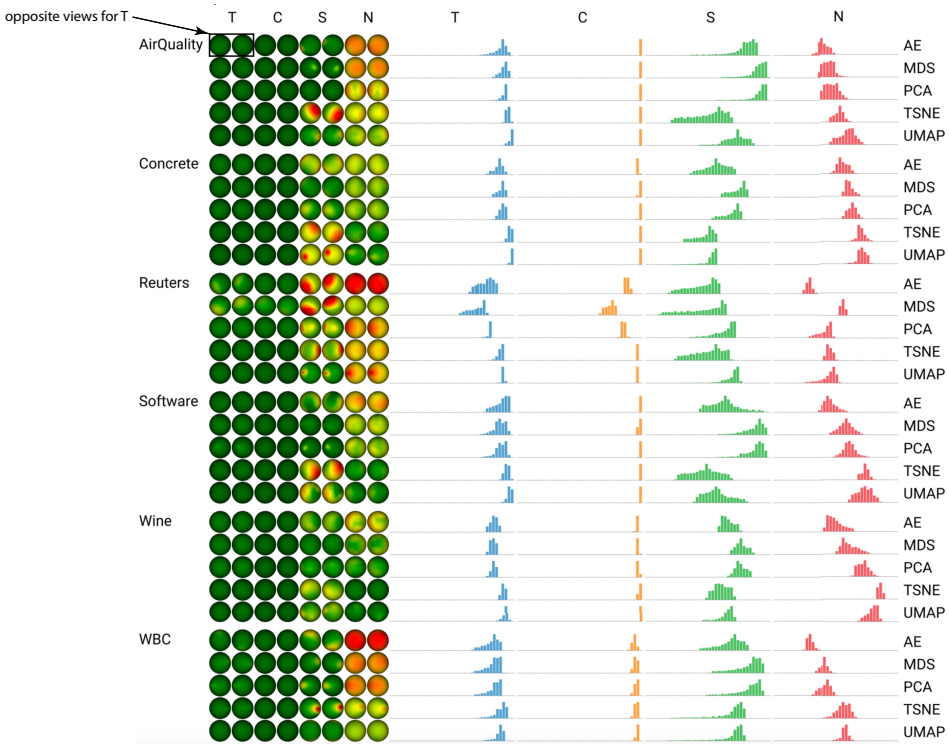

Figure 2: Exploration of viewpoint-dependent metrics (Sec. 4.2).

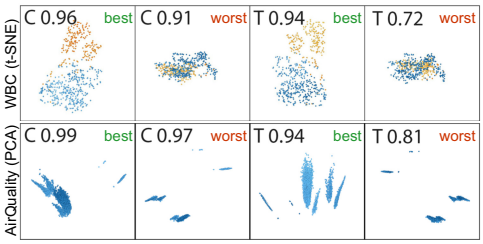

Figure 3: Comparison of best-*vs*-worst viewpoints of 3D projections from the perspective of T and C (Sec. 4.2).

rics computed for such views. Simply put: t-SNE may be best-quality for 2D projections (Espadoto et al., 2019) but not for 3D ones.

Comparing 2D and 3D projections: All above compare different *viewpoints* of 3D projections. Figure 5 next compares the quality of our 3D projections P_3 with their 2D counterparts P_2. The top table shows the single-value metrics for the 2D and 3D projections, *i.e.*, $M(D, P_2(D))$ and $M(D, P_3(D))$, averaged over all tested datasets and projection methods. Like (Tian et al., 2021b), we see that the 3D metrics (table second row) are slightly higher than the 2D ones (table first row). As argued earlier, this is not relevant, since users do not 'see' 3D projections but only 2D orthographic views thereof. The stacked barchart shows, for each dataset, projection technique, and metric, the fraction of viewpoints, of the total s that were computed, where 3D metrics exceeded the quality of the 2D projection. Note that, since we stack the bars of 5 projection techniques atop each other, 20% in the figure corresponds to all views V of a single technique-dataset pair.

Figure 5 gives several insights. For T, viewpoints of 3D projections outperform 2D projections only in a few cases. For all other metrics, many viewpoints do this: For N on *AirQuality*, over 50% of the 3D projection viewpoints have higher quality than the 2D projection. For all datasets and all metrics except T, we see multiple, differently colored, non-zero-height, stacked bars atop each other. So, many techniques create 3D projections having viewpoints that score better than their 2D counterparts. Hence, as our tool (Sec. 3) helps users in finding high-quality viewpoints, 3D projections can effectively provide higher-quality results than their 2D counterparts. We further analyze this in Sec. 5 from a user perspective.

On no *dataset* do all techniques score consistently better in 3D – that would be a dataset whose bar, in the plot, has five stacked fragments, each larger than 12.5% (since a 20% length bar indicates that *all* views of a technique-dataset pair score better in 3D than 2D). Yet, some *techniques* score consistently better in 3D for some metrics: For all but one of the AE projections (blue bars), almost all viewpoints (20%) have better N than the 2D projection – so, if we trust N, 3D AE projections are better. For PCA and t-SNE (green

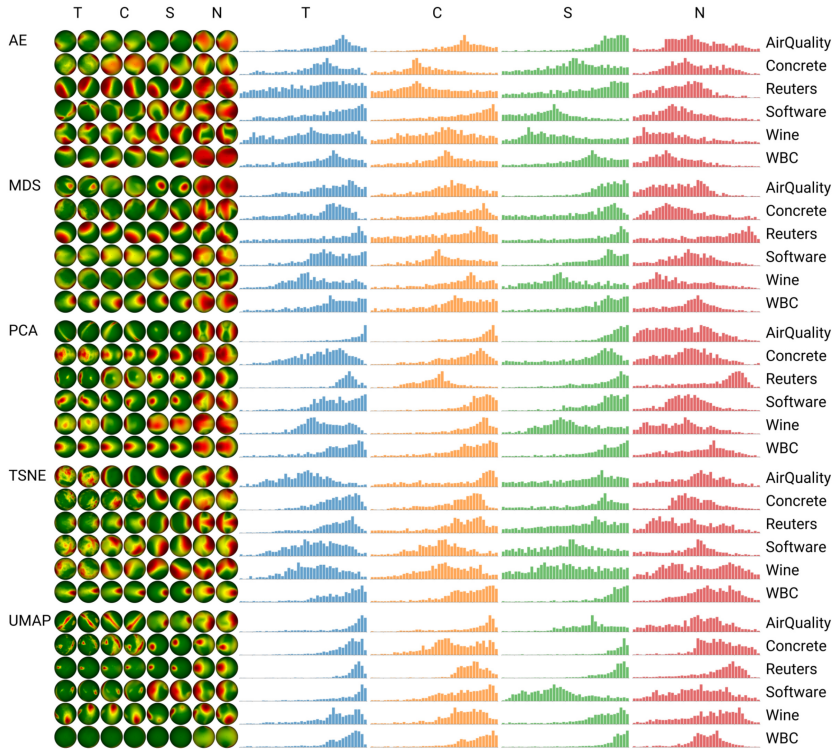

Figure 4: Refinement of Fig. 2 with metrics capped to their actual ranges and projections grouped per technique (Sec. 4.2).

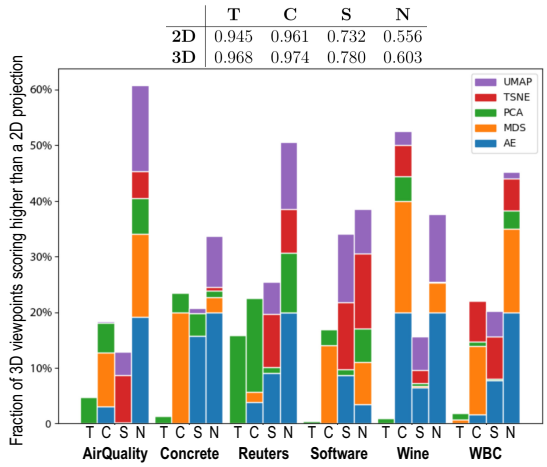

Figure 5: Comparison of quality metrics of 2D *vs* 3D projections over 6 datasets, 5 projection techniques (Sec. 4.2).

and red), we see far fewer viewpoints with better N than the 2D projection. Also, UMAP (purple bars) is better in 3D only in terms of S or N, but rarely for C and never for T. For MDS (orange bars), 3D viewpoints outperform 2D projections mostly in C.

Summarizing, we conclude that (a) 3D projections offer viewpoints with higher quality than corresponding 2D projections; (b) such viewpoints are not dominant for all studied quality metrics and datasets; (c) our exploration tool helps finding such viewpoints.

5 USER STUDY

Our quantitative evaluation shows that even small changes of quality metrics can strongly influence how a 3D projection looks from a certain viewpoint (Fig. 3 and related text). We further analyze the power of our proposed metrics for predicting good views of 3D projections by conducting a user study.

Projections and datasets: To make the study duration manageable (roughly 10-15 minutes), we picked a subset of the 30 (D, P) pairs used in our quantitative evaluation. The subset contains projections which (1) have discernible structure in terms of separated point-groups with similar coloring based on class labels; (2) finding a good viewpoint, showing strong visual cluster separation for the 3D projection, is not trivial. (3) the datasets have at least 1000 samples, so the space added by the third dimension (in 3D projections) has value. Our subset contains six pairs: (Wine, t-SNE); (Wine, PCA); (Concrete, t-SNE); (Reuters, AE); (Reuters, t-SNE); and (Software, t-SNE). Each pair contains the 3D projection and the corresponding 2D projection. Note that UMAP is not in this subset since UMAP projections tend to have a very high visual separation between clusters.

Study design: We aim to discover how users reason about the quality of views of a 3D projection in comparison to a 2D projection, and how their reasoning

correlates to our metric values and findings discussed in Sec. 4. The full study set-up, including training and tasks, is detailed in the supplementary material.

First, we explained our tool (Fig. 1) to users. We kept the explanation of quality metrics simple since deep understanding of the metric definitions (Tab. 1) was not needed for our study's tasks. Specifically, we told users that T and C measure the quality of neighborhood preservation; that N and S measure how distances in the projection reflect data distances; and that all metrics range between 0 (worst) and 1 (best).

We further explained users that they should search for 'good' viewpoints in 3D projections. We defined a good viewpoint as being one which showed the data as well-separated point groups that have similar colors (labels). In other words, users were *implicitly* tasked with finding views that have minimal overlap for different clusters and show most of the data structure in terms of class separation.

Usage of metrics: To study how metrics correlate with the users' choices of good viewpoints, we split the study into two parts. For the first three projections, further called the *blind* (B) set, users had to go through the first three (of the six) projection-pairs and select, for each pair, 3 different viewpoints of the 3D projection that they deemed good, without seeing the tool's metric widgets – that is, using only views (a) and (b) in Fig. 1. For the remaining three projections, further called the *guided* (G) set, we posed the same questions, but also showed the metric views (Fig. 1c,d) to the users. We explicitly stressed that one should use the metric values just as *suggestions* for finding interesting viewpoints to further search from, as these metrics do not measure class and cluster separation (which the task aims to maximize) but only local structure preservation. We randomized the order in which users saw the projections so that the projections in the B and G sets differed for each user.

For each viewpoint users picked, we also asked whether they preferred it to the 2D projection. Finally, we asked users to give their agreement on a 7-point Likert scale with the following statement: "A 3D projection, examined from various viewpoints, better displays data structure than a 2D projection." We explained the users that 'data structure' in this context means the ability to see reasonably-well separated clusters of points (which we know to exist in the studied datasets) and that, in general, a good projection scores high values of the quality metrics displayed in the visual exploration tool.

5.1 Study results

We invited around 50 people to our study. Twenty-two downloaded our tool and performed the study. At the end of the study, our tool saved the projections selected by the users as 'good' (a total of 66 per projection) and also the corresponding viewpoint-dependent quality metrics. These metrics were, as explained, not seen by users in the blind experiment, and respectively shown in the guided experiment. These data were anonymously sent by the participants back to us.

We next analyze these data to study how user preferences correlate with the computed quality metrics.

Do users prefer viewpoints with high metric values? Figure 6 shows the histograms of each metric and projection-pair in the evaluation set. Three box plots show the distributions of quality values in the actual histogram (H), for viewpoints in the blind set (B), and for viewpoints in the guided set (G). Comparing the histograms H and B, we see that, in almost all cases, users choose good viewpoints that have high values (for all metrics) even when *not* seeing any quality metrics. This is a first sign that quality metrics do correlate with what users see as good viewpoints.

If we compare the histograms H and G, we see that users 'zoomed in' and selected good viewpoints with, in most cases, even higher quality values than in the first case. This finding has to be interpreted with care. On the one hand, users could have been biased by the quality metrics displayed during the guided experiment. On the other hand, as explained earlier, we explicitly stressed that these metrics are only guidelines for finding interesting viewpoints and explicitly told users that, if they find other viewpoints as being better, they should ultimately go by their own preference. As such, the H-G comparison suggests us that quality metrics are useful predictors of users' preferences of good viewpoints. As such, showing the metric widgets during actual exploration of 3D projections can be useful since it helps users find high-quality viewpoints (G boxplots show clearly that users selected the high-end of the quality ranges) *and* users find high-quality viewpoints to be good (correlation of H and B boxplots). Yet, the strength of this correlation is not equal for all (dataset, projection) pairs.

Table 3 refines these insights by showing the p-values of a T-test (equal variance, one tail) for each projection, all four metrics. The test checks whether the average metrics for the B, G, and combined (B+G) conditions are significantly higher than the average metrics over all viewpoints V. We see that, for nearly all cases, this is so for the guided set G. For the combined set B+G, this is slightly less often so.

Do users prefer 3D or 2D projections? Figure 7 shows the percentage of 3D viewpoints that users preferred over the 2D projection of the same dataset for both the B and G conditions. Overall, in the G condition, users preferred the 3D projection over the 2D projection, and did so *more* than in the B condition.

Table 3: p values of t-testing whether the average metrics for user-selected viewpoints in the blind (B), guided (G), and both (B+G) sets are significantly higher than average values for all viewpoints V. Significant values ($p < 0.05$) are in bold.

	T			C			S			N		
	B	G	B+G	B	G	B+G	B	G	B+G	B	G	B+G
Wine t-SNE	**.029**	**.001**	**<.001**	**.01**	**<.001**	**<.001**	**.005**	**<.001**	**<.001**	.24	**.001**	**.003**
Wine PCA	**.015**	**<.001**	**<.001**	.121	**.041**	**.025**	**.003**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Concrete t-SNE	**.003**	**.035**	**.001**	**.004**	.111	**.004**	.159	.156	.08	.148	**.049**	**.029**
Reuters AE	.087	**<.001**	**<.001**	.232	**<.001**	**<.001**	.214	**<.001**	**<.001**	.448	**.025**	.071
Reuters t-SNE	.418	**.025**	.059	.595	**.029**	.111	.935	**.009**	.23	.679	.056	.197
Software t-SNE	**.04**	**<.001**	**<.001**	.088	**<.001**	**<.001**	.689	**<.001**	**.002**	.279	**<.001**	**<.001**

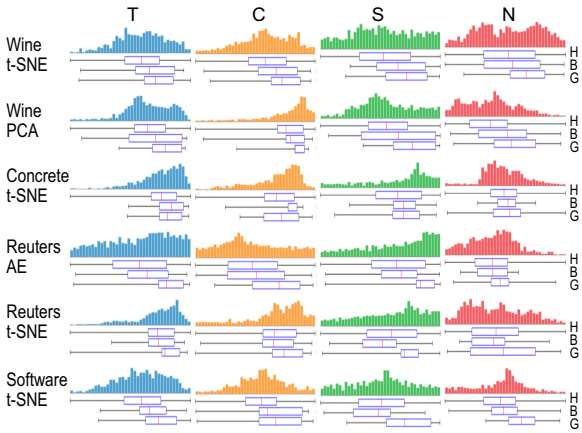

Figure 6: Distribution of metric values for all viewpoints (histograms and boxplots H), viewpoints in the blind set (boxplots B), and viewpoints in the guided set (boxplots G).

This, and the findings in Fig. 6 showing that users tend to pick high-quality views in the G condition, tell us that the metric widgets add to the user-perceived value of 3D projections. The 3D-*vs*-2D preference in the G condition was the strongest for the *Wine* dataset. Interestingly, for this dataset, we found the strongest correlation between metric values and user perceived quality (Tab. 3). For *Reuters* and *Software*, we see a much smaller 3D-*vs*-2D preference in the G condition in Fig. 7 and also little or no correlation between metric values and user-perceived quality (Tab. 3). This further reinforces our claim that, when metrics capture well user preference, displaying them only increases the perceived added-value of 3D projections. For *Software*, we see that 3D was preferred much less than 2D. Looking at Fig. 6, we see that this is the only case of the six (P, D) combinations in our study where *all* four quality metrics have a shallow tail to the right, *i.e.*, have only few 3D viewpoints in which any, let alone all, metrics is/are high. In other words, for such datasets where 3D quality metrics are more spread-out, it is hard to argue for the added-value of 3D projections *vs* their 2D counterparts. For *Reuters*, the obtained results also correlate with the fact that this is a much higher-dimensional dataset than all other studied ones (1000 *vs* a few tens of dimensions). This may indicate that the added-value of 3D projections potentially decreases for very high-dimensional datasets – a hypothesis we aim to explore in future work.

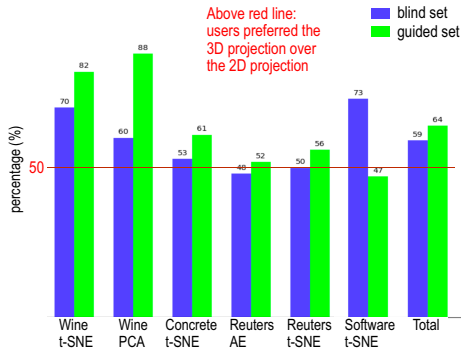

Figure 7: Percentage of cases where users preferred viewpoints of 3D projections *vs* 2D projections.

Finally, we consider the last question we asked our participants – whether, all taken into account, they preferred a 3D interactive projection to a static 2D projection for the task of assessing the data structure. *All* 22 participants responded with a value on the positive side of the scale (4 or higher), with an average of 5.94. This is additional evidence that, when aided by interaction, and by tools that help the selection of interesting viewpoints (like our quality metrics), 3D projections are an important alternative to be considered to classical, static, 2D projections.

6 DISCUSSION

Our results answers our questions Q1-Q3 as follows:

Q1: Simply *reusing* 2D projection quality metrics for 3D projections is misleading. These metrics will score higher values than their 2D counterparts but the respective 3D projections can appear massively different from different viewpoints. To address this, we need viewpoint-dependent quality metrics. Using such viewpoint-dependent quality metrics, as proposed in this paper, helps assessing the quality of 3D projections as these show significant variations between viewpoints of different projection techniques for different datasets. Such metrics are as simple and

fast to compute as their 2D counterparts.

Q2: Viewpoint-dependent quality metrics can reach *higher values* than their 2D counterparts, albeit for a small number of viewpoints. Such viewpoints can be easily found using the interactive visual metric-and-projection exploration tool we proposed. Hence, 3D projections can generate 2D images which are of higher quality than the static 2D projections typically used in DR, all other things equally considered, and generating such images (finding such viewpoints) is not hard if assisted by suitable interactive tooling.

Q3: Users' definition of "good viewpoints" (for the task of separating a 3D projection into distinct same-label clusters) correlates well with high values of our viewpoint-based quality metrics. This correlation is little influenced by the projection technique but more so by the dataset being explored. Separately, enabling visual exploration of the quality metrics increases the users' preference for a 3D projection *vs* a 2D one for performing the same task. Summarizing the above, using our quality metrics during the visual exploration helps using 3D projections in multiple ways.

Limitations: Computing quality metrics (Tab. 1) is linear in the number of viewpoints s and dataset points N. For $s = 1000$ and our studied datasets (N in the thousands), this takes a few minutes. This is not an issue for our study goal as we can precompute all the metric values for all tested datasets prior to the actual study. Using our metric-based exploration tool (Fig. 1) at interactive rates on *unseen* datasets would require faster metric computation – which can be trivially implemented by *e.g.* GPU parallelization.

Our findings are restricted to the sample of 6 datasets and 22 users who evaluated only 6 of the 30 dataset-projection combinations. It is possible that the correlation of user preference with viewpoint-dependent metrics, thus the predictive power of the latter for choosing good viewpoints, is different for datasets with other traits (intrinsic dimensionality, sparsity, and cluster structure). Using more datasets to study this correlation can bring valuable insights and be used to improve our visual tool to recommend good viewpoints as a function on these traits. Also, using more quantitative tasks to gauge how users select suitable 3D viewpoints (and measuring the time needed for this) is an important direction for future work. Similarly, our findings are restricted to the five projection techniques we studied. However, as (Espadoto et al., 2019), average quality metrics evaluated on a total of 45 projection techniques show quite similar values. As such, we believe that our findings – and the added-value of our proposed visual tool for choosing good viewpoints for 3D projections – will hold for most, if not all, such techniques.

7 CONCLUSIONS

We have presented a novel approach to measuring the quality of 3D multidimensional projections and using this quality to drive projection exploration. We defined (and measured) quality as a viewpoint-dependent distribution based on accepted quality metrics for 2D projections. We showed that viewpoint-dependent metrics capture the visual variability in 3D projections much better than aggregated, single-value, quality metrics. We further proposed a visual interactive tool for finding high-quality viewpoints. Finally, we conducted an user experiment showing that our proposed viewpoint-dependent quality metrics correlate well with user-perceived good viewpoints and, also, that our viewpoint-exploration tool increases the preference of users for 3D projections as compared to classical, static, 2D projections.

We next aim to extend our evaluation with more datasets, tasks, and projections find even more accurately when, and how much, 3D projections can bring added value atop their 2D counterparts. In particular, we aim to study LAMP (Joia et al., 2011) which scales well with the number of points and dimensions. We also aim to extend our evaluation to involve more sophisticated visualizations of high-dimensional data using pre-conditioned 3D projections that display data using density estimation rather than raw scatterplots such as the Viz3D system (Artero and de Oliveira, 2004). Since Viz3D takes several measure to reduce limitations of raw 3D projections (and scatterplots), using our viewpoint-selection assistance tools may further increase the added-value of 3D projection visualizations as opposed to classical 2D ones.

REFERENCES

Albuquerque, G., Eisemann, M., and Magnor, M. (2011). Perception-based visual quality measures. In *Proc. IEEE VAST*, pages 11–18.

Artero, A. O. and de Oliveira, M. C. F. (2004). Viz3D: Effective exploratory visualization of large multidimensional data sets. In *Proc. SIBGRAPI*.

Aupetit, M. (2007). Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 10(7–9):1304–1330.

Aupetit, M. (2014). Sanity check for class-coloring-based evaluation of dimension reduction techniques. In *Proc. BELIV*, pages 134–141. ACM.

Bank, D., Koenigstein, N., and Giryes, R. (2020). Autoencoders. arXiv:2003.05991 [cs.LG].

Camahort, E., Lerios, A., and Fussell, D. (1998). Uniformly sampled light fields. In *Proc. EGSR*, pages 117–130.

Chan, Y., Correa, C., and Ma, K. L. (2014). Regression

cube: a technique for multidimensional visual exploration and interactive pattern finding. *ACM TIS*, 4(1).

Coimbra, D., Martins, R., Neves, T., Telea, A., and Paulovich, F. (2016). Explaining three-dimensional dimensionality reduction plots. *Inf Vis*, 15(2):154–172.

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553. https://archive.ics.uci.edu/ml/datasets/wine+quality.

Dua, D. and Graff, C. (2017). Wisconsin breast cancer dataset. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic).

Espadoto, M., Martins, R., Kerren, A., Hirata, N., and Telea, A. (2019). Toward a quantitative survey of dimension reduction techniques. *IEEE TVCG*, 27(3):2153–2173.

Gonzalez, A. (2010). Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, 42(1):49–64.

Joia, P., Coimbra, D., Cuminato, J. A., Paulovich, F. V., and Nonato, L. G. (2011). Local affine multidimensional projection. *IEEE TVCG*, 17(12):2563–2571.

Jolliffe, I. (2002). *Principal Component Analysis*. Springer.

Lespinats, S. and Aupetit, M. (2011). CheckViz: Sanity check and topological clues for linear and nonlinear mappings. *CGF*, 30(1):113–125.

Levoy, M. (2006). Light fields and computational imaging. *Computer*, 39(8):46–55.

Lewis, D. and Shoemaker, P. (2021). Reuters dataset. https://keras.io/api/datasets/reuters.

Martins, R., Coimbra, D., Minghim, R., and Telea, A. C. (2014). Visual analysis of dimensionality reduction quality for parameterized projections. *Computers & Graphics*, 41:26–42.

Martins, R., Minghim, R., and Telea, A. C. (2015). Explaining neighborhood preservation for multidimensional projections. In *Proc. CGVC*, pages 121–128.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426v2 [stat.ML].

Meirelles, P., Santos, C., Miranda, J., Kon, F., Terceiro, A., and Chavez, C. (2010). A study of the relationships between source code metrics and attractiveness in free software projects. In *Proc. SBES*, pages 11–20.

Motta, R., Minghim, R., Lopes, A., and Oliveira, M. (2015). Graph-based measures to assist user assessment of multidimensional projections. *Neurocomputing*, 150:583–598.

Nonato, L. and Aupetit, M. (2018). Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. *IEEE TVCG*.

Paulovich, F. V., Nonato, L. G., Minghim, R., and Levkowitz, H. (2008). Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *IEEE TVCG*, 14(3):564–575.

Poco, J., Etemadpour, R., Paulovich, F. V., Long, T., Rosenthal, P., Oliveira, M. C. F., Linsen, L., and Minghim, R. (2011). A framework for exploring multidimensional data with 3D projections. *CGF*, 30(3):1111–1120.

Rauber, P. E., Falcão, A. X., and Telea, A. C. (2017). Projections as visual aids for classification system design. *Inf Vis*, 17(4):282–305.

Sanftmann, H. and Weiskopf, D. (2009). Illuminated 3D scatterplots. *CGF*, 28(3):642–651.

Sanftmann, H. and Weiskopf, D. (2012). 3D scatterplot navigation. *IEEE TVCG*, 18(11):1969–1978.

Schreck, T., von Landesberger, T., and Bremm, S. (2010). Techniques for precision-based visual analysis of projected data. *Inf Vis*, 9(3):181–193.

Sedlmair, M. and Aupetit, M. (2015). Data-driven evaluation of visual quality measures. *CGF*, 34(3):545–559.

Sedlmair, M., Munzner, T., and Tory, M. (2013). Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE TVCG*, pages 2634–2643.

Sips, M., Neubert, B., Lewis, J., and Hanrahan, P. (2009). Selecting good views of high-dimensional data using class consistency. *CGF*, 28(3):831–838.

Tatu, A., Bak, P., Bertini, E., Keim, D., and Schneidewind, J. (2010). Visual quality metrics and human perception: An initial study on 2D projections of large multidimensional data. In *Proc. AVI*, pages 49–56. ACM.

Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323.

The Authors (2022). Viewpoint-based comparison of 2D and 3D projections – datasets, software, and results. https://github.com/WouterCastelein/Proj3D_views.

Tian, Z., Zhai, X., van Driel, D., van Steenpaal, G., Espadoto, M., and Telea, A. (2021a). Using multiple attribute-based explanations of multidimensional projections to explore high-dimensional data. *Computers & Graphics*, 98(C):93–104.

Tian, Z., Zhai, X., van Steenpaal, G., Yu, L., Dimara, E., Espadoto, M., and Telea, A. (2021b). Quantitative and qualitative comparison of 2D and 3D projection techniques for high-dimensional data. *Information*, 12(6).

van der Maaten, L. and Hinton, G. E. (2008). Visualizing data using t-sne. *JMLR*, 9:2579–2605.

van der Maaten, L. and Postma, E. (2009). Dimensionality reduction: A comparative review. Technical report, Tilburg Univ., Netherlands. Tech. rep. TiCC 2009-005.

Venna, J. and Kaski, S. (2006). Visualizing gene interaction graphs with local multidimensional scaling. In *Proc. ESANN*, pages 557–562.

Vito, S., Massera, E., Piga, M., Martinotto, L., and Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors & Actuators B*, 129(2):750–757. https://archive.ics.uci.edu/ml/datasets/Air+Quality.

Yeh, I.-C. (2021). Concrete compressive strength dataset. https://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength.