

ARTICLE OPEN

Data analytics and parallel-coordinate materials property charts

Jeffrey M. Rickman¹

It is often advantageous to display material properties relationships in the form of charts that highlight important correlations and thereby enhance our understanding of materials behavior and facilitate materials selection. Unfortunately, in many cases, these correlations are highly multidimensional in nature, and one typically employs low-dimensional cross-sections of the property space to convey some aspects of these relationships. To overcome some of these difficulties, in this work we employ methods of data analytics in conjunction with a visualization strategy, known as parallel coordinates, to represent better multidimensional materials data and to extract useful relationships among properties. We illustrate the utility of this approach by the construction and systematic analysis of multidimensional materials properties charts for metallic and ceramic systems. These charts simplify the description of high-dimensional geometry, enable dimensional reduction and the identification of significant property correlations and underline distinctions among different materials classes.

npj Computational Materials (2018)4:5 ; doi:10.1038/s41524-017-0061-8

INTRODUCTION

The mapping of material properties, pioneered by Ashby and coworkers, leads to the creation of charts that condense a large quantity of information into a useful representation that has the virtue of revealing important property correlations and that facilitates materials selection and design.^{1,2} These (usually two-dimensional) charts form the basis of an optimized, systematic methodology for materials selection based on materials informatics. In particular, this program involves the identification of technical design requirements and the associated materials indices used for materials selection to meet these requirements.³ The utility of material property charts is evidenced by their centrality in the materials selection process and, more specifically, their ubiquity in engineering design courses.

The correlations between materials properties have been expressed in terms of generic rules-of-thumb and also couched more formally in terms of constraints on dimensionless property groups.⁴ In many cases, however, materials property correlations or design specifications are intrinsically highly multidimensional, and one therefore naturally seeks a representation that conveys the underlying complexity of these relationships. For example, mechanical properties such as stiffness and strength, as described by the Young's modulus and the ultimate tensile strength, respectively, may depend on the magnitude of interatomic interactions and the density, as described by the melting temperature and the mass density, respectively. This situation may be further complicated by the presence of a confounding variable, such as a particular materials property, that can lead one to ascribe spurious associations between other properties.⁵ In addition, for a given implementation, materials selection is often subject to practical, multidimensional constraints (e.g., on specimen dimensions, mechanical properties, cost, etc.) that are dictated, for example, by a fabricated component's function.

The visualization and analysis of complex, high-dimensional materials data presents several challenges. First, information on many axes must be conveyed in a digestible, graphical format permitting one to identify relationships among variables. (In some cases, it is possible to apply dimensional reduction techniques to find an approximate representation of the data.⁶) As the data may reflect different physical properties spanning a wide range of values, a proper normalization is required to make meaningful comparisons. Moreover, it is necessary to define metrics that quantify trends and provide a means to interpret the data. For these reasons, as noted by Bryden et al., despite the implementation of visualization tools for multidimensional data analysis in fields such as computer science⁷ and genetics,⁸ the exploitation of such tools to interpret materials data is still in its infancy.⁹

One especially useful tool for multidimensional visualization is parallel coordinates. In a parallel-coordinate representation, as developed by Inselberg,^{10–12} a set of d parallel axes replaces the conventional d -dimensional orthogonal Cartesian axes such that each point has a unique representation. In particular, a point given by the coordinates $(c_1, c_2, c_3, \dots, c_d)$ in a d -dimensional Cartesian system is represented in parallel coordinates by locating each coordinate on its respective (parallel) axis and then joining these points to form a polyline. As an example, consider the point $(0, 3, 1)$ in Cartesian coordinates. Its representation in parallel coordinates is given in Fig. 1a. More generally, it has been shown that mathematical objects may be readily mapped from Cartesian to parallel coordinates, and that the latter coordinate system is well-suited to multivariate statistical analysis.¹² Another example of an important duality is presented in Fig. 1b, which show that lines in a Cartesian coordinate system map into (polyline intersection) points in parallel coordinates.

In this work, we illustrate the utility of combining the methods of data analytics with parallel coordinates to represent and interrogate high-dimensional materials property data. While

¹Department of Materials Science and Engineering and Department of Physics, Lehigh University, Bethlehem, PA 18015, USA
Correspondence: Jeffrey M. Rickman (jmr6@lehigh.edu)

Received: 23 September 2017 Revised: 14 December 2017 Accepted: 27 December 2017
Published online: 29 January 2018

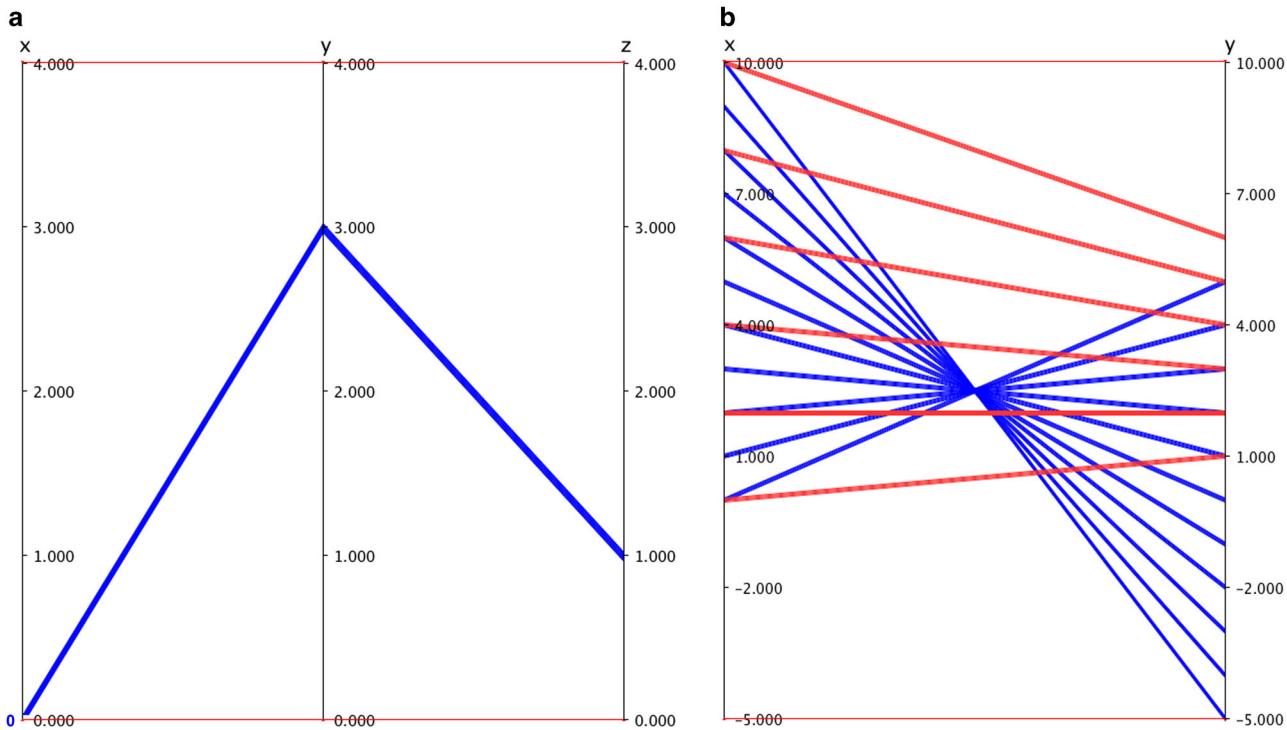


Fig. 1 **a** The point $(0, 3, 1)$ (in Cartesian coordinates) represented in parallel coordinates. Note that the intercepts on the axes are joined by line segments to make a polyline. **b** An illustration of the point-line duality. A line (shown in blue) with a negative slope m , namely $y = 5 - x$, in Cartesian coordinates maps into a point in parallel coordinates, the point lying at the intersection of polyline segments. A line (shown in red) with a positive slope m , namely $y = (x/2) + 1$, also maps into an (extrapolated) intersection point, now to the right of the rightmost axis

Table 1. A property table summarizing the thermomechanical properties considered here, along with their reference values for nickel

Properties	Symbol	Ref. Value
Hardness (Mohs)	H	4.0
Young's Modulus (GPa)	E	200
Tensile Strength (MPa)	UTS	317
Density (g/cc)	r	8.91
Melting Temperature ($^{\circ}$ C)	T_m	1455
Thermal Expansion Coefficient (1/K)	CTE	1.34×10^{-5}
Thermal Conductivity ($W/(m \cdot K)$)	TC	90.0

parallel-coordinate plots have been employed in other science and engineering contexts, including Internet attack detection, process monitoring and analysis of climate data, their application to material science problems has been quite limited.^{13–15} There are, however, several recent noteworthy papers in this area, including the recent work by Bryden et al.⁹ on visualization, Gorai et al.¹⁶ on the development of tools for thermoelectric materials design and Kamath et al.¹⁷ on the analysis of additively-manufactured parts. Our focus here is on the construction and systematic analysis of multidimensional materials properties charts for metallic and ceramic systems using the parallel-coordinate framework. For clarity and to illustrate the main concepts, we have focused here on these two materials classes; the same approach outlined above can, of course, be applied to other materials classes (e.g., polymers, electronic materials) with the same benefits. As will be seen below, when combined with data analytics, this approach constitutes a powerful tool for identifying important property relationships that can guide materials selection Table 1.

RESULTS

Parallel coordinates and correlations

To illustrate the construction and analysis of multidimensional materials property charts, we consider two distinct materials classes, namely elemental metals and (mostly) technical ceramics, each comprising 25 systems. The thermomechanical properties that will be examined are listed in the table below. Also shown in the table are the corresponding values for a reference system, chosen for convenience to be nickel, that provides a useful basis for comparison. These reference values (denoted with the subscript '0') will be used to normalize property values to create dimensionless variables that will be denoted with a prime (e.g., $H' = H/H_0$).

As a first illustration of a multidimensional property chart, Fig. 2 displays a parallel-coordinate chart showing the normalized values of the $p = 7$ properties listed in the table for the $n = 25$ elemental metals. It should be noted that the order of the parallel axes was chosen here to highlight important pairwise correlations. While there is inherent clutter in this presentation (an issue to be addressed below), some information can be gleaned upon inspection. For example, given the behavior illustrated in Fig. 2, it is evident that there is some degree of positive correlation between the pairs E' and T'_m , H' and T'_m and H' and UTS' .

It is of interest to explore these correlations in more detail. The dependence of the stiffness and hardness on the melting temperature follows intuitively from the fact that the melting temperature is a measure of bond strength. One can ask whether, given their association with T'_m , E' and H' are correlated and whether this correlation is direct. Several authors have explored this relationship using indentation measurements finding that in many, but not all cases, E' and H' are positively correlated.^{18,19} From the metals data shown in Fig. 2 one finds that the correlation coefficient between E and H $\rho_{EH} = 0.78$. Performing a hypothesis

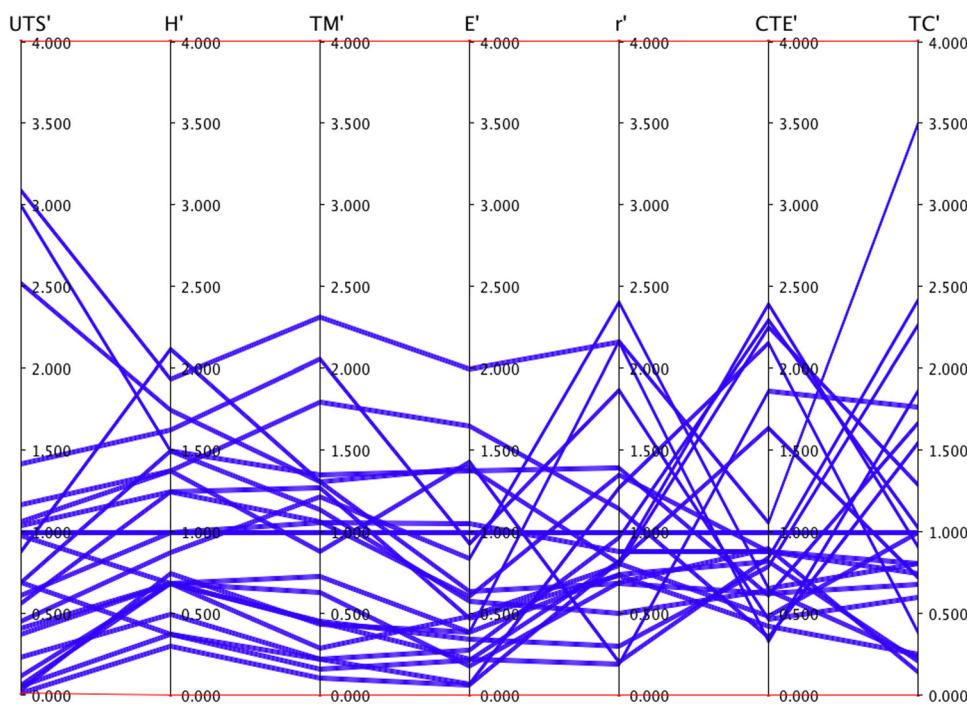


Fig. 2 A parallel-coordinate chart that displays the values of the normalized thermomechanical properties for 25 the elemental metals listed in the property table

test, one can conclude that since $\rho_{EH} > \rho_{crit}$ (where $\rho_{crit} = 0.396$ is the critical value) for $n - 2$ degrees of freedom, one can reject the null hypothesis that E and H are uncorrelated at the 5% level of significance.²⁰ Thus, there is some significance to this property association. To determine the degree of correlation between stiffness and hardness after the influence of the melting temperature is eliminated, one calculates the partial correlation coefficient²¹

$$\rho_{EH-T_m} = \frac{\rho_{EH} - \rho_{ET_m}\rho_{T_mH}}{\sqrt{1 - \rho_{ET_m}^2}\sqrt{1 - \rho_{T_mH}^2}}, \quad (1)$$

where the subscripts denote the properties. One finds that $\rho_{EH-T_m} = 0.29$. Assessing the significance of this coefficient with a hypothesis test, one finds a critical value of 0.488 at the 5% level of significance,²⁰ implying that there may, in fact, be no direct correlation between E and H when controlling for other factors, such as the melting temperature.

Property trends and cluster validation

We now make a comparison of properties across materials classes. Figure 3a displays a parallel-coordinate chart showing the normalized values of the $p = 7$ properties listed in the table for the $n = 25$ elemental metals and the $n = 25$ (mostly) technical ceramics. Given the clutter in this figure, it is useful to construct a somewhat simpler representation to highlight important property trends. For each class of materials, one convenient and robust measure of centrality, especially in higher dimensions, is the geometric median, \tilde{x} . In this context, the geometric median of the property points is defined such that

$$\tilde{x} = \operatorname{argmin}_{i=1}^p d(x, x_i), \quad (2)$$

where $d(x, x_i)$ denotes the Euclidean distance between points x and x_i .²² (We note that the use of other metrics here is also possible.) These median values are shown in Fig. 3b. Several trends are evident upon inspection of these charts. As a class, the selected

ceramics have lower densities, lower coefficients of thermal expansion and higher melting temperatures than the selected metals, as expected. Given that melting temperature is a measure of bond strength, it is not surprising that these ceramics are, in general, harder than the metals.

With this division into materials classes, one can determine the degree to which these classes are distinct. From a physical point of view, the nature of the interatomic bonding in these two classes is, of course, different and this difference leads to distinct properties. From a data analytics perspective, one can ask whether this expected difference in properties leads to a sensible division of the data into clusters. To make this determination, one can employ clustering metrics known as validation measures that are used in data analytics to measure the quality of a cluster analysis. Two such metrics are particularly useful in this context, namely the Thornton separability,²³ τ , and the Dunn^{24,25} Δ , indices, respectively. To define these metrics, suppose that one wishes to compare two point clusters (e.g., corresponding to metals and ceramics), C_m and C_c , that comprise a distinct partitioning, C , of a set of points. Then, if there are N total points, x_i is the i -th point and $x_{i'}$ is its nearest neighbor

$$\Delta = \frac{\min_{x_i \in C_m, x_j \in C_c} d(x_i, x_j)}{\max_{x_s, x_t \in C_k : k \in C} d(x_s, x_t)}, \quad (3)$$

and

$$\tau = \frac{\sum_{i=1}^N (f(x_i) + f(x_{i'}) + 1) \bmod 2}{N}, \quad (4)$$

where the indicator function $f(x_i) = 0$ if $x_i \in C_m$ and 1 otherwise. The Dunn index is then the ratio of the minimum intercluster distance to the maximum intraclass distance, with larger values signifying greater separability, while the Thornton index is the fraction of points whose nearest neighbor is in the same (materials) class.²⁶ Evidently, $0 \leq \tau \leq 1$, with values closer to 1 indicating well-separated clusters, those closer to 0.5 indicating little to no clustering and those close to 0 indicating a strong

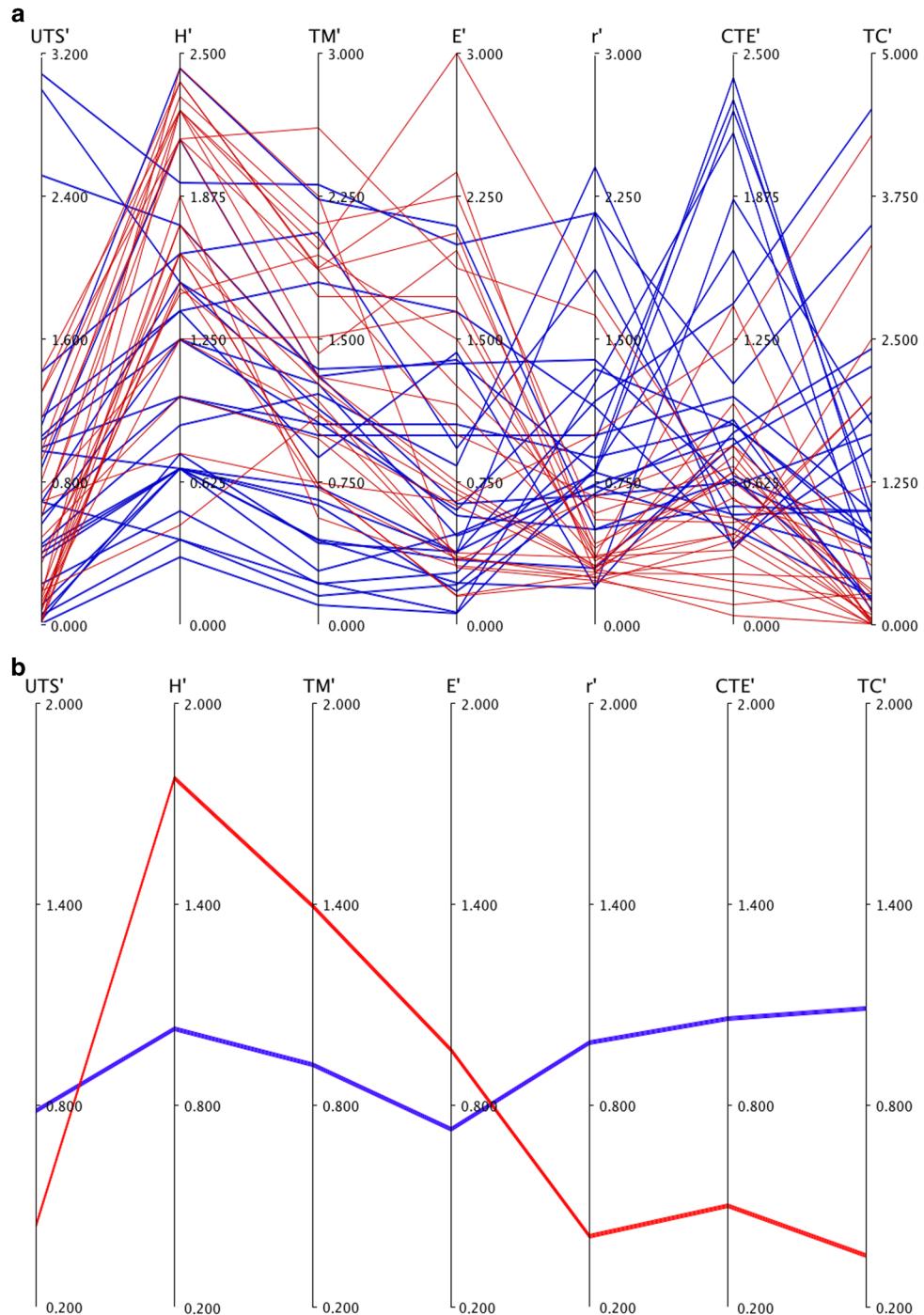


Fig. 3 **a** A parallel-coordinate chart showing the normalized values of the p properties listed in the table for the $n = 25$ elemental metals (blue) and the $n = 25$ (mostly) technical ceramics (red). **b** The geometric median of properties for both the metals and ceramics. This analysis removes much of the clutter seen in **a**

inverse correlation between classes (e.g., an alternating grid of points).

For this data one finds that $\Delta = 0.10$ and $\tau = 0.82$, respectively. The larger value of τ indicates that points corresponding to the same type of material have a relatively high degree of association. However, the relatively small value of Δ implies that the material clusters may be anisotropically dispersed such that clusters are in proximity in some direction in the multidimensional space. Thus, there is a moderate degree of separation between these two materials classes, suggesting some distinction between these two

types of materials. It is of interest to compare these results with those obtained by a data-mining clustering algorithm, namely k-means clustering.^{27,28} This unsupervised learning methodology assigns points to clusters such that (property) space is divided into Voronoi cells. When applied to partition the data here into 2 clusters, the resulting clusters differ somewhat from those based on materials class. To highlight these differences, one can calculate the purity index, π ($0 \leq \pi \leq 1$), for the k-means clusters, a measure of the fraction of data points that have been correctly classified by a clustering scheme.²⁹ In this case, $\pi = 0.66$, indicating

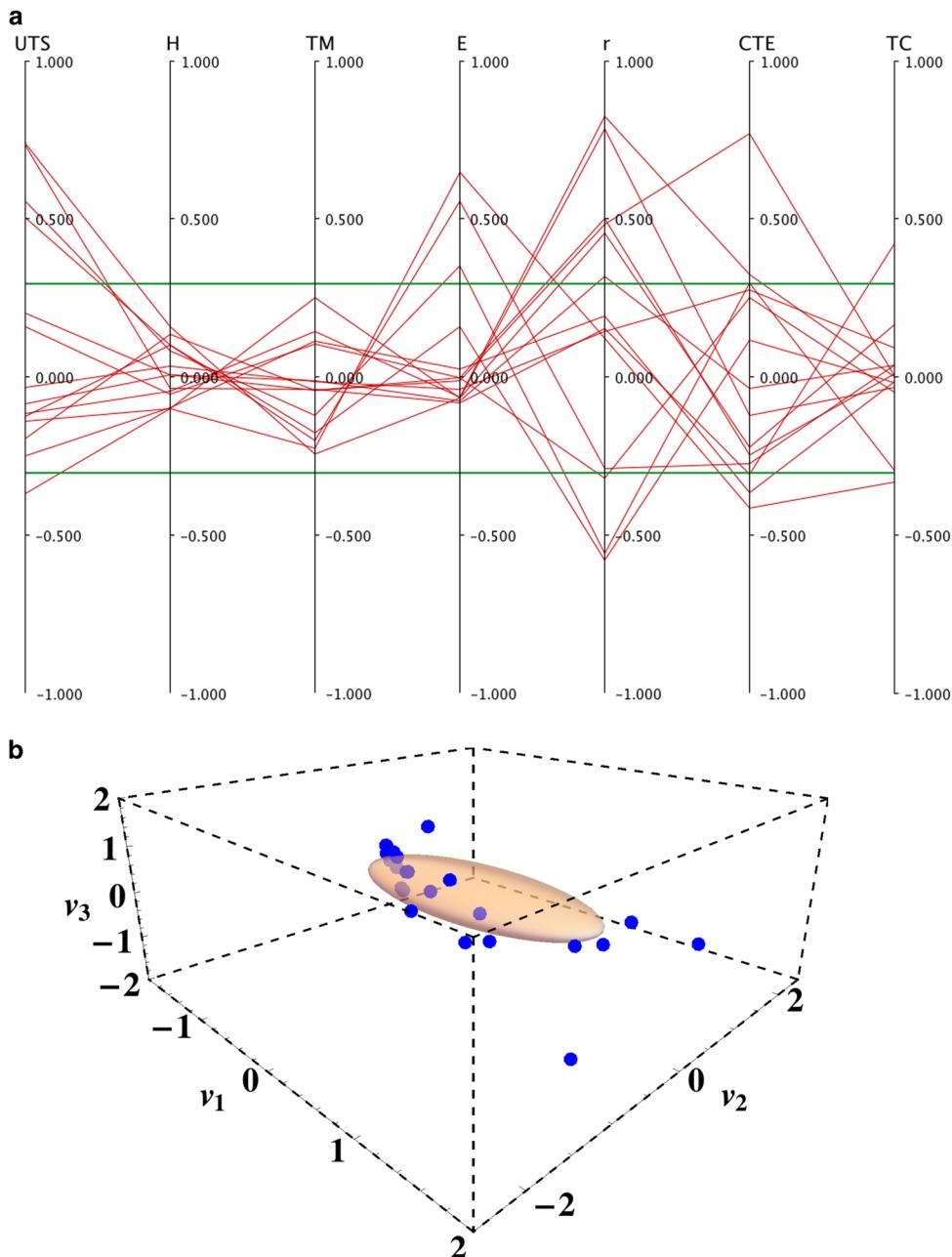


Fig. 4 **a** A parallel-coordinate plot of the fractional difference between the original (from Fig. 3) and the reconstructed ceramic property data from the PCA (red). For comparison, horizontal lines (green) show fractional differences of ± 0.25 . **b** The property data for ceramics plotted in the $k=3$ subspace obtained from the PCA. The ellipsoid with semi-major axes obtained from the PCA eigenvalues is also shown

that the k-means classification reflects some of the underlying physics. Evidently, the proximity of the property class clusters inhibits a clear spatial partitioning based on material properties.

Dimensional reduction

Parallel coordinates are also useful in displaying the impact of dimensional reduction strategies. With these strategies one seeks an approximate representation of the data in fewer dimensions.⁶ In this case, we apply a principal-component analysis (PCA)^{30,31} to effect this reduction. In brief, if we form $n \times 1$ column vectors $\vec{x}_j (j = 1, 2, \dots, p)$ of observations of the p properties and assemble these column vectors to form an $n \times p$ data matrix, then we can readily construct the corresponding $p \times p$ covariance matrix Σ having an eigenvalue spectrum $\{\lambda_i\} (i = 1, 2, \dots, p)$ and

whose eigenvectors, $\vec{v}_i (i = 1, 2, \dots, p)$, form an orthonormal basis.³² In a PCA one projects the \vec{x}_j onto a subspace of dimension $k < p$ by constructing a projection operator from the k eigenvectors of Σ having the largest corresponding eigenvalues.

The number of principal components required for a satisfactory representation of the data can be determined by using the rule of thumb that a component j is retained if $\lambda_j > \sum_{i=1}^p \lambda_i / p$.³² While in this case a two-dimensional representation is satisfactory, one extra dimension beyond what is necessary was added to more completely explain the data. Retaining one extra principal component, a PCA was applied to the ceramic property information shown in Fig. 3 to reconstruct the data in the relevant subspace with $k=3$. These three components account for 87% of the total variance. In Fig. 4a a parallel-coordinate plot shows the fractional difference between the original and the reconstructed

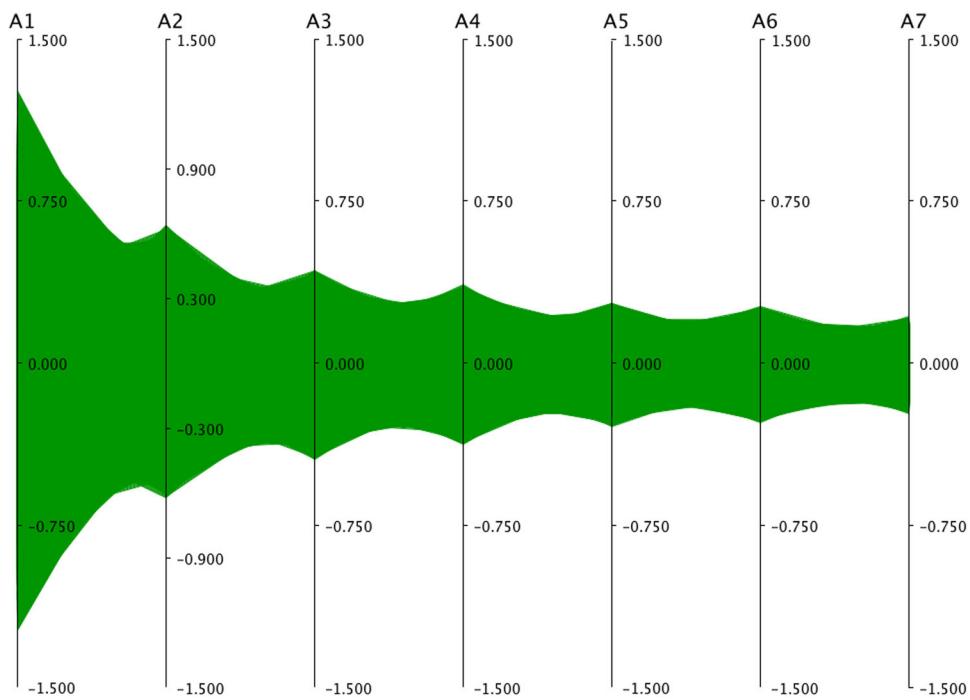


Fig. 5 The property envelope for the ceramic class in the principal axis frame in parallel coordinates

data from the PCA. In this format it is straightforward to determine the range of validity of the reconstruction. For this purpose horizontal lines show fractional differences of ± 0.25 . Figure 4b shows the data plotted in the $k=3$ subspace determined by the PCA. Also shown is the ellipsoid that captures much of the behavior and is given by $(v_1/\lambda_1)^2 + (v_2/\lambda_2)^2 + (v_3/\lambda_3)^2$.

Visualization of class envelopes

One may also wish to visualize the outer envelopes of materials classes to determine their relative hypervolumes in the materials property space. Parallel coordinates provides a powerful framework for this purpose. One approach to the construction of these envelopes is to determine the convex hulls associated with the respective collection of material points. Alternatively, one can use the results of a PCA to obtain an approximate, hyperellipsoidal envelope surrounding the points. More specifically, employing this latter option, one takes the envelope at position \vec{u} to be the quadratic form $g(\vec{u}) = (\vec{u} - \vec{\mu})^T \Sigma^{-1} (\vec{u} - \vec{\mu})$, where $\vec{\mu}$ is the vector of mean values for properties and Σ^{-1} is the (inverse of the) aforementioned covariance matrix. The associated d -dimensional hypervolume

$$V = \frac{2\pi^{p/2}}{p\Gamma[p/2]} |\Sigma|^{1/2}, \quad (5)$$

where Γ denotes the gamma function³³ and $|\cdot|$ denotes a determinant.

Figure 5 shows the property envelope for the ceramic class in the principal axis frame. The overall shape of the envelope can be rationalized by noting that there is a mapping duality between an ellipse in d -dimensional Cartesian coordinates and a hyperbola in parallel coordinates.³⁴ The doubly serrated shape of the envelope is a succession of hyperbolas that highlights the unequal lengths of the principal axes. A similar envelope can be constructed for the metals class. Upon comparing their corresponding volumes using Eq. (5), one finds that $V_{\text{metal}}/V_{\text{ceram}} = 1.95$, indicating that the metals class occupies a larger fraction of the multidimensional property space.

Materials selection

Finally, we examine the application of parallel-coordinate charts to facilitate materials selection. The selection process typically involves the imposition of constraints among properties. Given the aforementioned duality relations between multidimensional Cartesian coordinates and parallel coordinates, one can describe these constraints in the latter representation. For example, for the points on the line $y_i(x) = mx + b_i$, the corresponding polylines in parallel coordinates intersect at the point $(1/(1-m_i), b_i/(1-m_i))$, assuming a unit spacing between parallel axes.¹¹ Thus, for points constrained to be between the lines y_1 and y_2 (with say $m_2 > m_1 > 0$), the corresponding polylines pass through a constraint trapezoid with bases $|b_2 - b_1|/(1-m_i)$ ($i = 1, 2$) and height $(m_2 - m_1)/[(1-m_1)(1-m_2)]$ in parallel coordinates.

To illustrate the application of this approach, consider a materials selection problem in which one wishes to identify a metal for a relatively light structure that is both stiff and strong and that meets cost specifications. For example, one may seek a metal having a (normalized) specific stiffness $E'/\rho' > 1.4$, a (normalized) specific strength $UTS'/\rho' > 1.0$ and a normalized cost $\text{Cost}' < 2.0$. For these requirements, the corresponding constraint trapezoids collapse to line segments along the abscissa. Thus, for the stiffness relation, one looks for a polyline segment with a positive slope that, upon extrapolation, intercepts the abscissa between $1/(1-1.4) = -2.5$ and 0. Similarly, the strength relation leads to a polyline segment with a positive slope that intercepts the abscissa between 0 and 1.0. Figure 6 shows a truncated parallel coordinates plot focusing on three properties with a few candidate materials that satisfy at least some of these constraints. The white segments (i.e., the collapsed trapezoids) on the plot highlight regions that satisfy the constraints. In this plot beryllium is shown in red, molybdenum and chromium in green and titanium and aluminum in blue. Also, shown are extrapolated polyline segments for molybdenum and chromium (dashed green) and beryllium (dashed red) that are seen to intercept the abscissa between -2.5 and 0. By contrast, the polyline segments for titanium and aluminum (dashed blue) do not satisfy this constraint. It should also be noted that only beryllium satisfies the strength constraint, as indicated by the extrapolated polyline

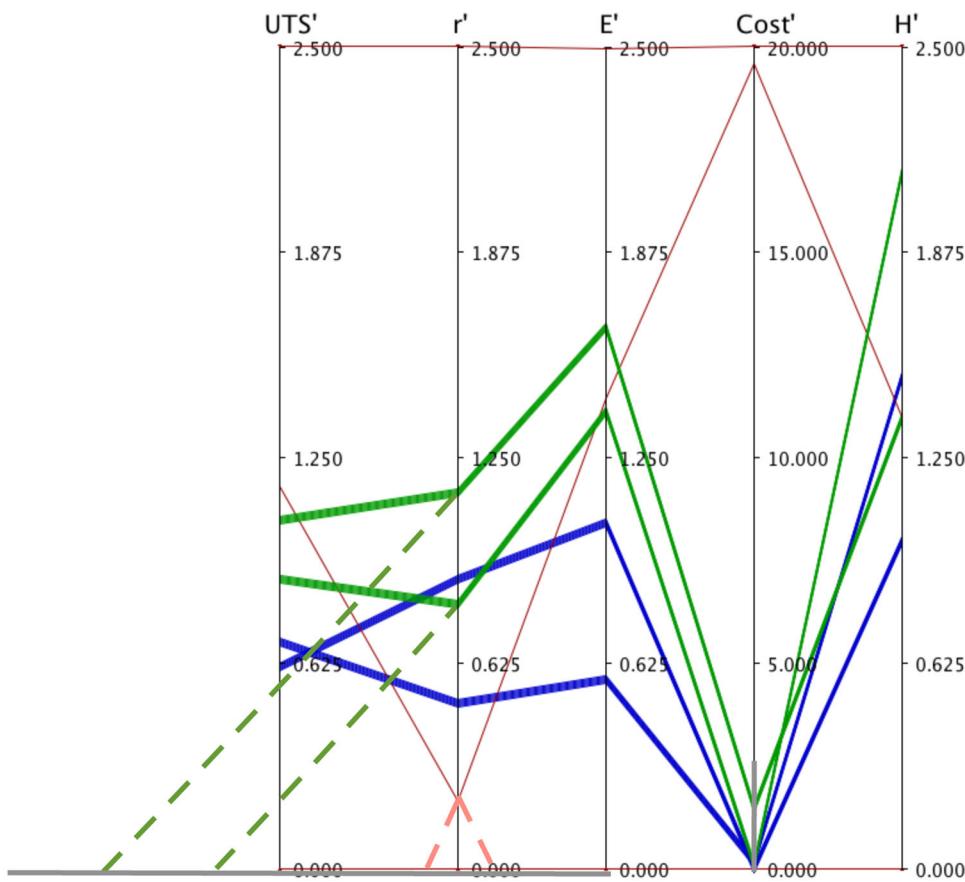


Fig. 6 A truncated parallel coordinates plot focusing on three properties, namely the (normalized) stiffness, strength and cost, with a few candidate materials that satisfy at least some of the stated constraints. Beryllium is shown in red, molybdenum, and chromium in green and titanium and aluminum in blue. The gray regions identify constraints for the polyline segments. The extrapolated polyline segments (dashed lines) show which materials satisfy the design constraints

(dashed red). Finally, the selected metals, with the exception of beryllium, satisfy the cost constraint, and therefore no materials satisfies all three constraints.

DISCUSSION

In this work, we illustrated the utility of combining the methods of data analytics with a parallel coordinates representation to construct and interpret multidimensional materials property charts. This construction, along with associated materials analytics, permits the identification of important property correlations, quantifies the role of property clustering, highlights the efficacy of dimensional reduction strategies, provides a framework for the visualization of materials class envelopes and facilitates materials selection by displaying multidimensional property constraints. Given these capabilities, this approach constitutes a powerful tool for exploring complex property interrelationships that can guide materials selection.

This paper illustrates the utility of combining higher-dimensional data visualization with data analytics to quantify useful relationships among materials properties. For reference, a summary of the important aspects of this analysis is given below.

- The use of cluster validation to identify cluster centrality (via the geometric median) and remove clutter, the bane of parallel-coordinate plots (Fig. 3a, b). This analysis facilitates a discussion of property trends in metallic versus ceramic systems.
- The use of the Thornton and Dunn metrics to characterize the quality of clusters. As discussed in the subsection entitled

Property Trends and Cluster Validation, these metrics allow one to quantify the degree to which there are separable clusters based on materials properties. In this way, one can quantify, for example, the extent to which metals and ceramics differ in terms of their attributes.

- The use of the purity index to quantify the correctness of property-point clustering. In this context, as noted above, the value of the index indicates that the proximity of the property class clusters inhibits a clear spatial partitioning based on material properties. Thus, the purity index complements the Thornton and Dunn metrics and gives one a quantitative measure of the degree to which materials classes are distinct.
- The visualization of property envelopes in a high-dimensional space (Fig. 5) exploits the geometric duality between ellipsoids in Cartesian space and hyperboloids in parallel-coordinate space.¹¹ In this case, parallel coordinates permit one to infer important directions (i.e., principal axes) of the hyperellipsoid that defines the property envelope in Cartesian space and thereby extend the Ashby property envelope to higher dimensions.
- The use of parallel coordinates in combination with principal component analysis (PCA) enables one to quantify readily the degree of dimensional reduction achieved with PCA (Fig. 4a, b). In this way, one can determine whether a description of a materials class by a small number of linear property combinations is accurate.

While for the purposes of illustration and clarity we have focused here on two materials classes, the same approach outlined above can, of course, be applied to other materials

classes (e.g., polymers, electronic materials) with the same benefits. It is especially well-suited to situations in which there are many coupled property variables and/or constraints, and one wishes to quantify systematically these relationships while navigating through a high-dimensional property space. Generalizations of this approach are also possible. For example, one may extend the representation and associated analysis to spatially heterogeneous systems with attendant property variations via position-dependent parallel coordinates. These extensions and other applications of this methodology are currently under investigation.

METHODS

We employ here data analytics in conjunction with parallel coordinates, a technique for visualizing and interpreting multidimensional data. The data analytics methods used in our analysis include: principal component analysis (PCA),³² the Thornton separability²³ and the Dunn clustering indices^{24,25} and centrality measures.²² The parallel-coordinate representation¹¹ of the data enables one to identify trends and material property envelopes in high-dimensional systems.

Data availability

The author will make available, upon request, the data used in the applications described in this work. It is understood that the data provided will not be for commercial use.

ACKNOWLEDGEMENTS

The author wishes to thank Lehigh University for support with funds from a presidential initiative, namely the Nano/Human Interface. This initiative aims to create the capability for data-driven discovery in which humans interact immersively with machines to interpret large, multidimensional data sets. This work was supported with funds from a Lehigh University presidential initiative, namely the Nano/Human Interface.

AUTHOR CONTRIBUTIONS

J.M.R. is the primary author (and guarantor) of this work. He collected the data and performed the analysis.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Computational Materials* website (<https://doi.org/10.1038/s41524-017-0061-8>).

Competing interests: The author declares no competing financial interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Ashby, M., Shercliff, H. & Cebon, D. *Materials: Engineering, Science, Processing and Design*, 3rd edn (Elsevier, 2014).
2. Ashby, M. F. *Materials Selection in Mechanical Design*, 5th edn (Elsevier, 2017).
3. Ashby, M., Cope, E. & Cebon, D. Materials Selection for Engineering Design. In *Informatics for Materials Science and Engineering* (ed. Rajan, K.) (Elsevier, 2013).
4. Ashby, M. Checks and estimates for material properties. I. ranges and simple correlations. *Proc. R. Soc. Lond. A* **454**, 1301–1321 (1998).
5. Coolican, H. *Research Methods and Statistics in Psychology*, 6th edn (Psychology Press, 2014).
6. Lee, J. A. & Verleysen, M. *Nonlinear Dimensionality Reduction* (Springer, 2007).
7. Keim, D. A. Information visualization and visual data mining. *IEEE Trans. Vis. Comput. Graph.* **7**, 100–1007 (2002).
8. Schroeder, M. P., Gonzalez-Perez, A. & Lopez-Bigas, N. Visualizing multi-dimensional cancer genomics data. *Genome Med.* **5**, 9 (2013).
9. Bryden, A., Rajan, K. & LeSar, R. Visualization in materials research: rendering strategies of large data sets. In *Informatics for Materials Science and Engineering* (ed. Rajan, K.) (Elsevier, 2013).
10. Inselberg, A. The plane with parallel coordinates. *Vis. Comput.* **1**, 6991 (1985).
11. Inselberg, A. *Parallel Coordinates: VISUAL Multidimensional Geometry and its Applications* (Springer, 2009).
12. Wegman, E. J. Hyperdimensional data analysis using parallel coordinates. *J. Am. Stat. Assoc.* **85**, 664–675 (1990).
13. Choi, H., Lee, H. & Kim, H. Fast detection and visualization of network attacks on parallel coordinates. *Comput. Secur.* **28**, 276–288 (2009).
14. Wang, X., Medasani, S., Marhoon, F. & Albaazzaz, H. Multidimensional visualization of principal component scores for process historical data analysis. *Ind. Eng. Chem. Res.* **43**, 7036–7048 (2004).
15. Steed, C. A., Fitzpatrick, P. J., Swan, J. E. & Jankun-Kelly, T. J. Tropical cyclone trend analysis using enhanced parallel coordinates and statistical analysis. *Cart. Geogr. Inform. Sci.* **36**, 251–265 (2009).
16. Goria, P. et al. TE design lab: a virtual laboratory for thermoelectric materials design. *Comp. Mat. Sci.* **112**, 368–376 (2016).
17. Kamath, C., El-dasher, B., Gallegos, G. F., King, W. E. & Sisto, A. Density of additively-manufactured, 316L SS parts using laser powder-bed fusion at powers up to 400 W. *Int. J. Adv. Manuf. Technol.* **74**, 65–78 (2014).
18. Prasad, S. L. A., Mayuram, M. M. & Krishnamurthy, R. Response of plasma-sprayed alumina-titania composites to static indentation processes. *Mater. Lett.* **41**, 234–240 (1999).
19. Bao, Y. W., Wang, W. & Zhou, Y. C. Investigation of the relationship between elastic modulus and hardness based on depth-sensing indentation measurements. *Acta Mater.* **52**, 5397–5404 (2004).
20. Crow, E. L., Davis, F. A. & Maxfield, M. W. *Statistics Manual: With Examples Taken from Ordnance Development* (Dover, 1960).
21. Sharma, J. K. *Business Statistics*, 2nd edn (Dorling Kindersley, 2007).
22. Fletcher, P. T., Venkatasubramanian, S. & Joshi, S. The geometric median on Riemannian manifolds with application to robust atlas estimation. *NeuroImage* **45**, S143–S152 (2009).
23. Thornton, C. Separability is a learner's best friend. In *Proceedings of the 4th Neural Computation and Psychology Workshop: Connectionist Representations* (ed. Bullinaria, J. A. et al.) (Springer-Verlag, 1997).
24. Brock, G., Pihur, V., Datta, S. & Datta, S. clValid, an R package for cluster validation. *J. Stat. Softw.* **25**, 4 (2008).
25. Bible, J., Datta, S. & Datta, S. Cluster analysis: finding groups in data. In *Informatics for Materials Science and Engineering* (ed. Rajan, K.) (Elsevier, 2013).
26. Greene, J. R. Feature subset selection using Thornton's separability index and its applicability to a number of sparse proximity-based classifiers. In *Proceedings of the Pattern Recognition Association of South Africa* (2001).
27. King, R. S. *Cluster Analysis and Data Analysis: An Introduction* (Mercury Learning and Information, 2015).
28. Jain, A. K. Data clustering: 50 years beyond k-means. *Pattern Recogn. Lett.* **31**, 651–666 (2010).
29. Manning, C. D., Raghavan, P. & Schütze, H. *Introduction to Information Retrieval* (Cambridge University Press, 2008).
30. Jackson, J. E. *A User's Guide to Principal Components* (Wiley, 2003).
31. Samudrala S., Rajan K. & Ganapathysubramanian B. Data dimensionality reduction in materials science. In *Informatics for Materials Science and Engineering* (ed. Rajan, K.) (Elsevier, 2013).
32. Jobson, J. D. *Applied Multivariate Data Analysis*, Vol. 2 (Springer-Verlag, 1992).
33. Arfken, G. B. & Weber, H. J. *Mathematical Methods for Physicists*, 6th edn (Elsevier, 2005).
34. Heinrich, J. & Weiskopf, D. State of the art of parallel coordinates, *Eurographics 2013* (Eurographics Assoc., 2013).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018