Parallel Coordinates-based Visual Analytics for Materials Property

Diwas Bhattarai, Jian Zhang and Bijaya B. Karki
School of Electrical Engineering and Computer Science, Louisiana State University, Baton Rouge, LA 70803, U.S.A.

Keywords: Parallel Coordinates, Multivariate Visual Analytics, Materials Property, Viscosity Data.

Abstract: Because of major advances in experimental and computational techniques, materials data are abundant even for specific classes of materials such as magma-forming silicate melts. A given material property M can be posed as a complex multivariate data problem. The relevant variables or dimensions are the values of the property itself, the factors which influence the property (pressure P, temperature T, multicomponent composition X), and meta data information I. Here we present an innovative visual analytics system for the melt viscosity (η), which can be represented by M (η, P, T, X_1, X_2, …, I_1, I_2, …). Our system consists of a viscosity data store along with a web-based visualization support. In particular, we enrich the parallel coordinates plot with non-standard features, such as derived axes/sub-axes, dimension merging, binary scaling, and nested plot. It offers many insights of relevance to underlying physics, data modeling, and guiding future experiments/computations. Other material properties such as density can be incorporated as new attributes and corresponding new axes in the plot. Our aim is to collect all published data on various melt properties and develop a framework supporting database, visualization and modelling functions.

1 INTRODUCTION

Data on materials properties are common in many fields of science and engineering. Researchers usually have to go through the data collection, pre-processing, exploration, and finally modeling phases to better understand a given physical property. Advanced techniques to deal with massive amounts of materials data have been gaining interest in recent years. However, the multivariate nature of these data introduces further challenges. They involve variables of different types which require different representations. For instance, the value of material property under consideration itself may be a scalar, vector or tensor quantity. The parameter space in which the property is defined includes variables such as pressure, temperature, and composition. Other information such as methodology (experiment or computation), publication (year, authors, source), model-predicted values, uncertainties/errors, etc. can be useful in the analysis. The actual data values along with metadata can be examined for the completeness, trends, correlations, and modeling. An important class of materials belongs to magma-forming silicate melts (Kono and Sanloup, 2018), and here we take the melt viscosity as a use case for materials property (e.g., Hui and Zhang 2007; Karki et al., 2013).

In this paper, we present the parallel coordinates plot (PCP) enriched with several features to visualize the viscosity data (Figure 1). PCP is one of the widely used multivariate data visualization techniques (Inselberg 2009) to grasp an overall view of data and to reveal the relationships, clusters, etc. Many recent works have been focused towards either expanding feature set of plot components, reducing visual clutter or making correlations more visible (Heinrich and Weiskopf 2013; Johansson and Forsell, 2016). However, user centric analyses based on PCP are still rare. Our study presents a user specific enhancement of PCP for visual analytics on the viscosity data. It is implemented as part of a web-based framework for managing, exploring and analyzing the data.

Figure 2 shows the components of the proposed framework, where PCP serves as the foundation for the visualization platform. In the framework, data are stored and managed in the database. The modelling platform allows us to build materials property models based on standard forms optimized with finely selected data. The user interface for modeling and visualization is implemented in the client-side web browser. As part of an analytics framework, our PCP visual system has several unique features compared to other stand-alone tools: 1) Our system connects to a database and loads data directly from the database.

Bhattarai, D., Zhang, J. and Karki, B.
Parallel Coordinates-based Visual Analytics for Materials Property.
DOI: 10.5220/0007375400830095
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 83-95
ISBN: 978-989-758-354-4
Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

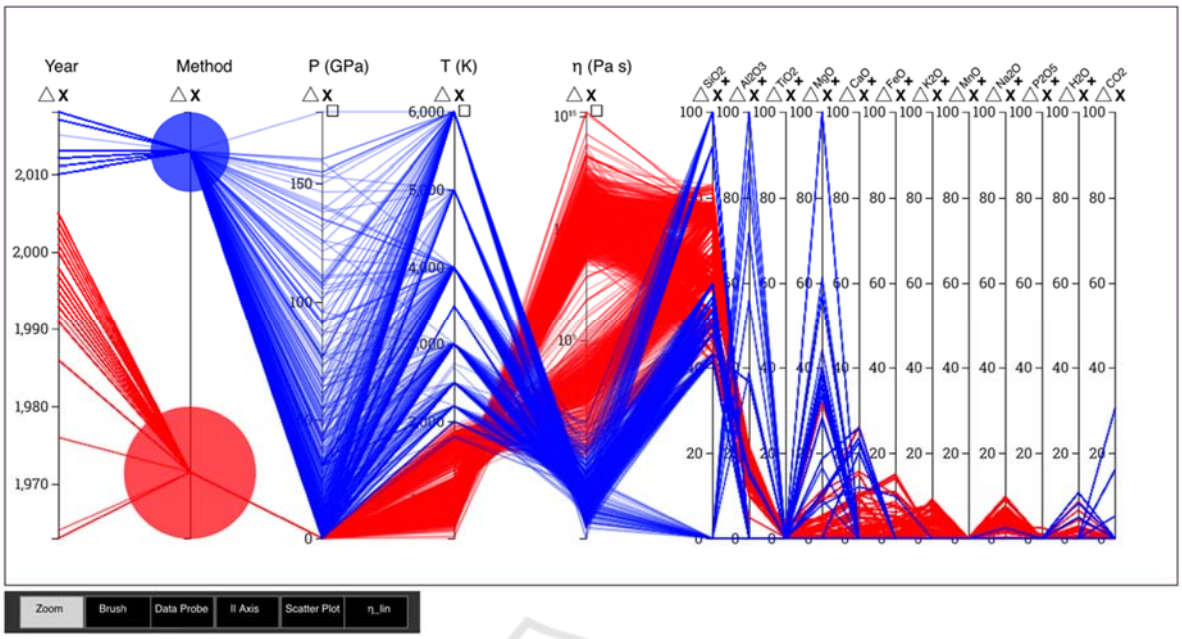

Figure 1: Parallel coordinates plot of melt viscosity data with respect to 17 dimensions. Computational and experimental data are displayed in blue and red, respectively. Numerical axes include viscosity η, pressure P, temperature T, etc., and categorical axis includes "Method". Linear and logarithmic scales are used. The value range 0 – 100 wt% is used for all component axes. Triangle button inverts the axis, cross button removes selected dimension, and square button performs binary scaling. Control panel at the bottom contains additional exploration tools such as *Zoom*, *|| Axis* (derived axis), *Data Probe*, and others.

2) It also connects to a modeling platform. Insight from visual analysis can be directly used in model construction and resulting models can be visualized for evaluation. 3) Finally, the whole framework is web-based. Users can conduct analysis using our system anywhere without the need for installation. The framework is expected to help gain insight (qualitative) to the given materials property data and explore quantitative details for further analysis.

Web Interface		
Database	**Visualization Platform**	**Modelling Platform**
Multi-criteria search and selection	Visual analytics	Model builder and evaluator

Figure 2: Layers of the proposed framework.

2 RELATED WORK

Visualization is used for gaining insights onto materials data, particularly when they have direct relevance to 3D space and time. Examples include atomic configurations (crystal structures), charge distributions and bonding, and many microscopic phenomena such as molecular diffusion, crack propagation, fluid dynamics (e.g., Bohara and Karki, 2014). In contrast, the data representing the bulk (macroscopic) properties such as density, elasticity, viscosity, conductivity, etc. usually do not require sophisticated visualization because the number of data values generated by a single (complete) study tends to be small. Over the years, these materials data have piled up so we want to collect and analyze them. This has been the situation with silicate melts whose physical properties are highly sought after. In particular, the viscosity is perhaps the most important property governing all magmatic processes including melt transport, magma mixing, and volcanic eruptions (Abe 1997; Solomatov 2007; Zhang et al. 2007; Adjaoud et al. 2011).

We aim to provide a database and web visualization platform to facilitate data exploration and modeling. Previous works on the viscosity data collections were confined to either published materials or simple databases with some model calculations (Hui and Zhang 2007; Giordano et al., 2008). The data are mostly found in Excel or CSV format and are not centrally located. Experiments have been helpful to generate the viscosity data, however, they are usually confined in the low

temperature-low pressure regime (Shaw 1972; Urbain et al. 1982; Wang et al. 2014). Computational techniques have started to generate data over experimentally inaccessible conditions (e.g., Adjaoud et al. 2011; Karki et al. 2013; Ghosh and Karki 2017). A combination of experimental and computational data appears to be a promising avenue to full understanding of the viscous behavior of magmatic melts. Exploring these data due to the multivariate nature can be realized only with visualization.

Parallel coordinates plot (PCP) is widely used in the visualization of multivariate data (Inselberg 2009; Heinrich and Weiskopf 2013; Johansson and Forsell 2016). PCP treats all variables/dimensions on equal footing as vertical axes and renders all data points in the parallel axis layout. By displaying full information on the same display, it helps us judge correlations, detect outliers, and identify clusters (Inselberg 1997). The primary technique to display a multidimensional data row (a data item) in PCP is by using a set of polylines connecting successive axes. Using continuous curves to map data values instead of polylines which are non-differentiable at the axis intersections can enable over-plotted line segment detection and cluster visualization with curve bundling (Graham et al., 2003; Zhou, 2008). Binning techniques for data density estimation allow to understand the distributions of data values on multiple variables (Geng et al., 2011; Nguyen 2018).

A quick glance at a relatively dense PCP may look over crowded or even intimidating for untrained users. Interaction is crucial for the effectiveness of the PCP-based visual analytics (Siirtola and Raiha, 2006). Techniques such as brushing (Fisherkeller, 1988; Roberts et al., 2015) and pinching (Inselberg, 2009) allow a selection of a subset of data. Axis-aligned brush in particular gives the user ability to filter data rows along a selected axis (Turkay et al., 2011). The parallel coordinates matrix (Heinrich et al. 2012) plots all permutations of dimensional positions which is comparable to a scatter plot matrix (Zhou and Weiskopf, 2018). Dimensional re-ordering can be applied with appropriate metrics to prioritize the location of each dimensional axis (Lu et al., 2016; Peltonen and Lin, 2017). Unimportant or less important dimensions can be removed from the main PCP view and placed in another view for context (Riehmann et al., 2012; Kaur and Karki, 2018).

When multiple datasets are involved, a single view is expected to support both intra- and inter-dataset analyses. This can be achieved with the nested PCP (Wang et al., 2016). It embeds multiple nested axes pairs, one for each dataset, between the corresponding primary axes pair. The nested axes can have different scaling factors than the primary axes to allow better view of the data lines in each dataset plot. The nested PCP can provide in-depth visualization by breaking down a dataset into multiple subsets corresponding to different clusters or categories or arbitrary groups in the data (Kaur and Karki, 2018).

It has been common to bring visualization to the web platform (e.g., Jourdain et al., 2011; Donato et al., 2018). The convenience of working with an application without having to download or install anything locally is one of the reasons for the web browsers' popularity for visualization application. Further, web technology is greatly platform agnostic, ensuring greater accessibility with minimal user effort. Graphical components can be managed using Scalable Vector Graphics (SVG) or Canvas with web scripting programming language Javascript. Chart components can be constructed using just Javascript or through a popular data driven Javascript library D3 (Bostock et al., 2011). Chart components developed with D3 can be used for highly-interactive data-driven visualizations (Donato et al., 2018).

3 DATABASE

Our enhanced parallel coordinates system is part of an analytic framework that consists of three pillars: the database, visualization, and modelling platforms (Figure 2). Data are internally managed in a relational database management system. Below we describe the data itself and methods of data compilation.

3.1 Data Format

A given material property such as the melt viscosity which is considered in this study can be viewed as a multivariate entity:

$$M \rightarrow M(\eta, P, T, X_1, X_2, \ldots, I_1, I_2, \ldots) \qquad (1)$$

In this multidimensional representation, the value of the property itself is considered as one of the variables (attributes). It is a scalar quantity for the melt viscosity (η). The property can also be a multi-valued quantity, for instance, diffusion coefficients (defined per atomic species) or elastic stiffness tensor. The parameter space in which the property is defined/determined involves factors such as pressure P, temperature T, and composition X. The compositional factor is itself multi-component (X_1, X_2, …), which represent molar fractions (or weight percentages) of over ten oxides in the case of molten silicates. Additional information of the data such as

methodology, publication details, research group, and comments may also provide valuable insight during analysis. The meta data information can be included as additional attributes (I_1, I_2, …). Researchers are also interested in building models using property values with respect to the parameter space and assess the uncertainties. To compare the predicted results with the actual data, we can use one or more derived variables. Thus compiled full information is visually mapped for meaningful analysis and modeling of the materials property (e.g., melt viscosity) in the question. Any other properties can be represented in the same format by simply adding their data values as additional attributes to *M*.

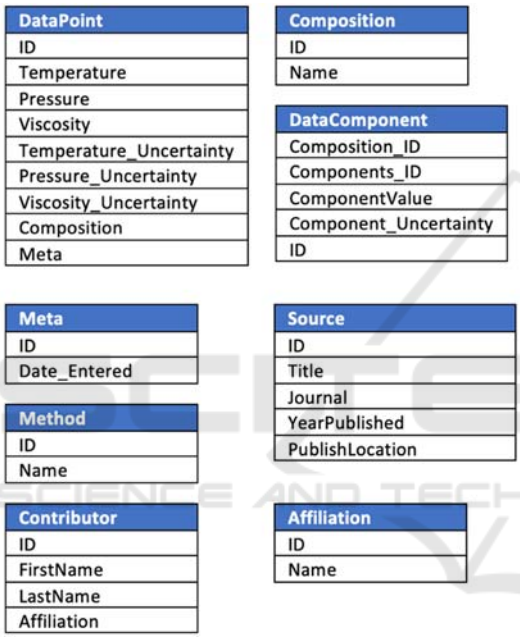

Figure 3: Database schema showing central data tables.

The viscosity data are collected from various published sources. Data format across all datasets usually varies and needs to be transformed into a standard format before storage and analysis. At the heart of it, the dataset is a list of viscosity values collected over different pressure, temperature, and composition ranges. Previous works on viscosity database have focused on only essential components for analysis and model building. Here we are also including metadata and any other relevant information (uncertainty, model values) along with the actual viscosity values. Similarly, rather than storing data in a general text file format, we organize a table structure for efficient data storage and retrieval (Figure 3). Data values are stored as real numbers in respective standard units whereas each composition component value is stored as a molar mass/weight percentage (wt%). Therefore, summing up all components for a single data point always yields a 100%. Detailed information for each data point can be extracted by joining the appropriate tables (Figure 3). In the web-application, users can filter results not only based on the actual data variables (viscosity value, *P*, *T* or *X*), but also by the metadata.

3.2 Data Compilation

We have compiled melt viscosity data from both experimental and computational sources. The measured viscosity-temperature data at ambient pressure exist for melts of several compositions (Hui and Zhang 2007). Our database incorporates these data previously used for developing predictive models. Steady but important advances in experimental measurements have brought new data at elevated pressures. We have yet to compile these high-pressure data, which mostly have appeared in recent years. The third category consists of calculated viscosity results. Our database includes the data from first-principles molecular dynamics simulations, which are generally considered to be highly accurate (Karki et al., 2013; Ghosh and Karki, 2017). There are much more viscosity data generated by other computational methods yet to be gathered. We anticipate to eventually have several thousand records in the melt viscosity database.

4 PARALLEL COORDINATES

To visually analyze the silicate melt viscosity data, we choose to adopt parallel coordinates plot as this technique can address various visualization challenges related to large multivariate dataset. In essence, PCP maps all data items with respect to all dimensions on a single display. The data polylines go across the display space sequentially through each dimensional axis. Since visual clutter is inevitable, a standard PCP with just vertical axes and polylines is bound to be inadequate. We consider various standard and non-standard PCP features, falling broadly into two categories. First, the interaction with the data variables/dimensions is explored in detail with the derived axes, axis merging, and bi-scaling. Axis can also provide space to display additional information and overlays such as categorical bubbles or histograms. Second, we explore ways to interact with the data itself through polylines. One or more polylines can be selected out of the entire data with appropriate color mapping and alpha blending.

4.1 Standard PCP Features

A standard PCP contains data polylines going through a consecutive list of parallel vertical axes for a data record (Figure 1). These lines intersect each axis at the corresponding scaled dimensional values. The axes thus divide a 2D drawing surface with respect to a k-dimensional data space onto $k-1$ sub-surfaces. Axes scaling is done with respect to maximum and minimum data values per dimension. Axes act as a main container for visual elements such as labels, markers, and overlays. These visual cues allow user to observe data space and read specific values. A linear (uniform) scale is used for most of the numerical data values. Categorical values are represented either by transforming each category value to a point or bubble at the mapped axis location.

Plotting a large multi-dimensional dataset on a standard PCP may occlude some data. Data selection techniques, such as axis-aligned brushing, probing, and pinching are used to filter the data. Brushing is useful along single axis as well as when combined with brushes from multiple axes. Logical relations between dimensions such as "and" or "or" are used to construct higher order brushes (Turkay et al., 2011). Similarly, pinching can be used to make data selection from the $k-1$ data sub-surfaces itself. Polylines may also be given discrete color map to differentiate between categories or continuous color map to represent numerical axis values. Relationship between any two axes may emerge in the form of positive or negative correlation in any of the sub-surfaces between axes (i.e., between adjacent axes). This may require examining different axis layouts. Axis reordering and flipping techniques are used to overcome the correlation identification problem. Our PCP system is also augmented with data table and pair-wise scatter plot.

4.2 Non-standard PCP Features

On initial load, our PCP system orders the axes by placing the most significant dimensions around the material property value dimension, which is viscosity in this study (Figure 1). It places pressure and temperature axes on left side of the viscosity axis and place the composition axes (silica component followed by other components) on the right. The metadata axes are placed further away. Other axes orderings can also be explored with drag and drop user interaction on any axis. Proper axis scaling is crucial for true representation of data. While each categorical value is given uniform space across the height of its axis, the numerical values are dealt differently. In the standard PCP, the lowest domain values appear at the bottom while the highest ones at the top of the axis. This is true for all dimensions except the components. All of the components, regardless of their domain extents, are plotted from 0 to 100 to effectively show their proportions for a given data point. Both linear and non-linear scaling have been used for data representation. Since viscosity spans a large range, data patterns become difficult to observe as the data size over the full range increases. A logarithmic scale is used for viscosity.

Axes are usually laid out uniformly across the display space with axis spacing $\Delta x = X_D/(k-1)$, where X_D represents the display width. This spacing layout provides equal significance to all dimensions. In the context of viscosity data, the users might be interested in exploring the relationships between the parameter space and viscosity values much more than the relationships between composition components. Therefore, to provide a larger space for interesting dimensions while also keeping the context of the overall data, our system uses variable axial spacing technique to separate the PCP into *data* and *component* regions (Figure 1). The *data* region can occupy one half of the display space and contains most important axes such as viscosity, temperature, pressure, method, and year. The *component* region is given the rest of the space with some padding. This region packs many (about a dozen) component axes.

4.2.1 Derived Axes/Sub-axes

Different data regimes for a dimension can be explored in large screen estate by augmenting a derived axis/sub-axis next to a primary axis. We can display the viscosity values using a linear scale on the derived axis alongside the logarithmic primary viscosity axis. Another scenario is that user may want to focus on a sub-range of a variable. For example, the derived sub-axis maps filtered domain values (say, 1200 K < T < 1600 K) to range $0 \leq r \leq y'$ using the same scale type as that of its primary axis (Figure 4). The data lines not falling in the chosen interval simply ignore the derived axis/sub-axis and continue along their paths to the next primary axis. This provides context of overall data along with detail at a specific range. The minimum length of the derived sub-axis is kept at half the length of its primary axis. However, if the chosen axes pair points contain more than 50% of the total data, the length is then made proportional to the number of points falling in the chosen dimensional domain range. Hence the length of the derived axis y':

$$y' = \begin{cases} n'(Y_D - p), & n' > 0.5 \\ 0.5(Y_D - p), & n' \leq 0.5 \end{cases} \quad (2)$$

where n' is the ratio of the number of filtered data points to the total number of data points, Y_D is the vertical extent of the display, and p refers to vertical padding value to accommodate dimension label and other controls on top of the axis. The horizontal position of all the axes are re-calculated taking the new axis into account. The derived axis is constructed such that a line through the middle point of both the primary and the derived axis is orthogonal to both the axes. This means translating the derived axis by $0.5(Y_D - p - y')$ in the vertical direction. Its position and height can be adjusted interactively. More derived axes are considered in the following sections.

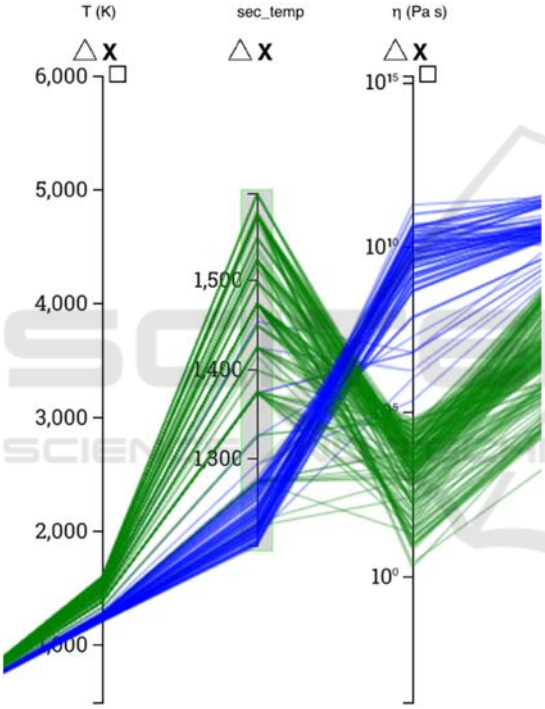

Figure 4: Derived temperature secondary axis showing a temperature range 1200-1600 K. Green represents viscosity values less than 10^4 Pa s and blue represents the rest.

4.2.2 Dimension Merging

Often times, domain experts are interested in looking at various binary joins or ternary systems though the database covers multicomponent composition. For instance, MgO-SiO_2 join is the considered to be the most important binary. The components CaO, FeO and MnO can be treated on the same footing as MgO because these oxides play the role of structure modifiers and are highly mobile. They can be combined together and viewed as one compositional variable. On other hand, SiO_2, Al_2O_3 and TiO_2 components together form silicate polyhedral network and are mostly immobile. These oxides can be treated collectively as one compositional variable. In PCP, any two components can be merged together.

Merging components results in the insertion of a new derived axis which acts as an independent dimension itself. The data polylines are re-rendered by incorporating the new axis (shown later in Figure 8). Since composition component is stored as a percentage, any two components can be directly summed together. Further, since a merged axis behaves as any other primary axes, it can itself be merged with other components. No component can be added twice so that the total molar mass of the composition for each data row is always at a 100%.

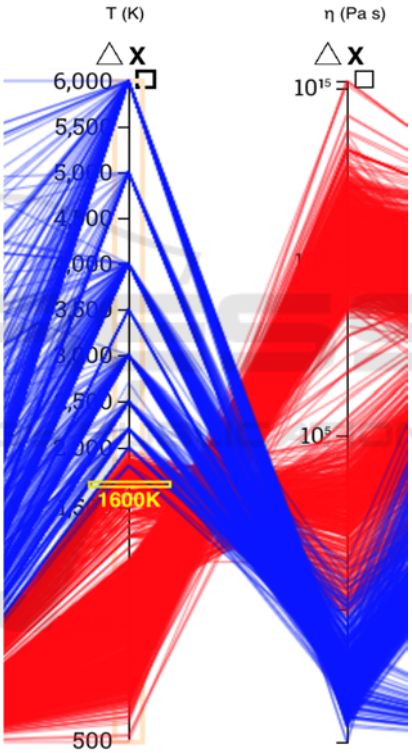

Figure 5: Binary scaling applied to the temperature axis with a cutoff point at 1600 K. Axis is scaled proportionately with respect to the number of data lines falling under and over the specified cut off point.

4.2.3 Binary Scaling

Binary scaling contains two different scales on an axis divided at a user chosen domain cut-off value. Let n_1 be the number of data points whose values are equal or less than the cutoff and n_2 be the rest of the data points. The sub-lengths are assigned for the two scales along the complete axis as follows:

$$l_1 = \frac{Y_D - p}{(n_2/n_1) + 1}; \quad l_2 = Y_D - p - l_1 \qquad (3)$$

We can now create two scales of dimension specific scale type with different domain and ranges: $0 \leq r_1 \leq l_1$ and $l_1 < r_2 \leq (Y_D - p)$. This scheme can be further extended to implement more than two scales on the same axis. Figure 5 shows a bi-scaled axis such that the lower range 500 to 1600 K is stretched while the upper part is compressed.

4.2.4 Categorical Bubbles

In PCP, it is desirable to treat numerical and categorical axes differently. Categorical variables consist of distinct categories which are difficult to directly map onto an axis. Therefore, a transformation to a metric scale must be done such that each category gets uniform space in the axis. One such data transformation can be done by overlaying circles or bubbles of varying radius on axes as category markers (Tuor et al., 2018). The radius of each bubble can be utilized to show different data properties. For instance, we can map the radius (r) with the frequency (f) of the data points falling under a category:

$$r = \sqrt{\frac{f}{f_{max}}} \; \frac{\Delta x}{2} \qquad (4)$$

Here, f_{max} is the number of data rows for category with maximum frequency. We utilize categorical bubbles to differentiate between experimental and calculated data points. Bubbles can also be used for user interactions such as to hide and show data based on mouse click (Figure 6). Categorical bubbles can also be split into smaller bubbles with respect to data values from another axis. For instance, both computational and experimental category bubbles can be split with respect to different viscosity data regimes (shown later in Figure 8). The split bubbles are stacked on top of each other such that the sum of their radii equal to the radius of the merged bubble. The radius of each split bubble is proportional to the frequency of data row falling under both the original and split categories.

4.2.5 Nested PCP

A recent study has shown that using nested PCP to visualize model parameter correlation between different datasets is more useful than superimposed or juxtaposed PCP representations (Wang et al., 2016). As we show, this technique can be further extended to analyze two or more subsets (groups) of the data corresponding to different intervals on selected numerical axis (for instance, low pressure versus high

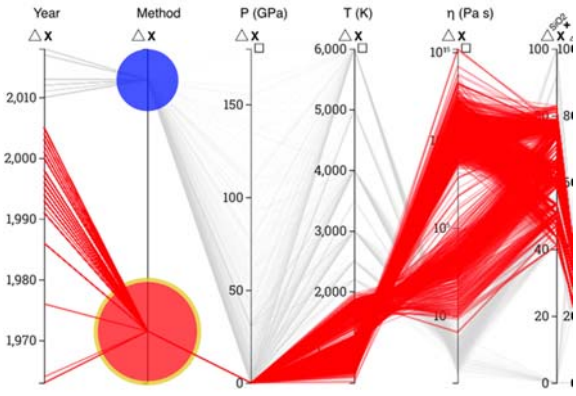

Figure 6: Experimental data selection by clicking categorical bubble (lower) shown in red while the calculated data are shown in background (gray polylines).

pressure regime) or different categorical values (for instance, experimental versus computational). Nested PCP resides symmetrically about the mid line between two adjacent axes under consideration (Wang et al., 2016; Kaur and Karki, 2018).

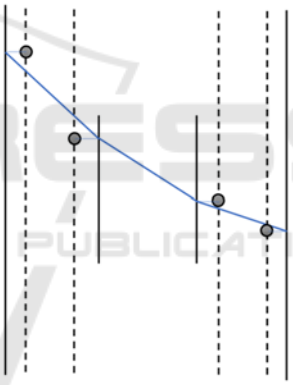

Figure 7: Nested PCP along with four control points for Bezier curves between nested and primary axes. The polylines connecting between the primary and nested axes are replaced by two curves.

The vertical space between two primary axes is divided into Y/n_s uniform regions where n_s is the number of nested categories and $Y = Y_D - p$. The categories are sorted by the mean location of each polyline on either one or both of the primary axes. Each nested axes is then constructed symmetrically from the middle of its category region where the end points of the j^{th} nested axis (j = 1, 2, 3,…, n_s counting from the bottom) are first constructed using $(j - 0.5)Y/n_s \pm \Delta d$, where Δd can vary between $0.2Y/n_s$ to $0.4Y/n_s$. A translation is then applied between 0 to $\pm 0.1Y/n_s$ depending on the location of the maximum pixel value of the polylines on the selected primary axis. The horizontal spacing of the

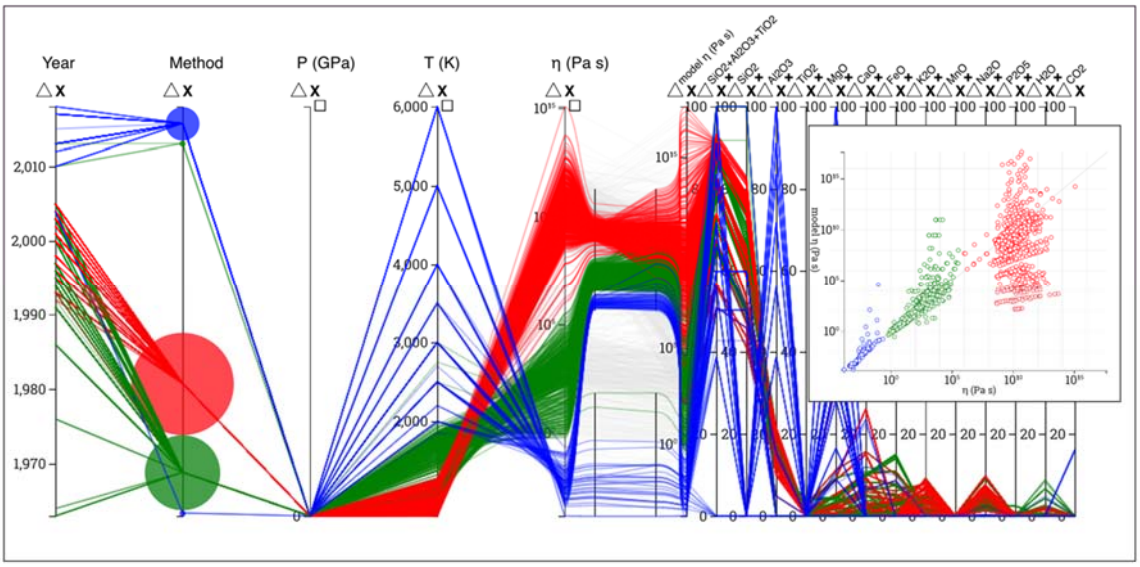

Figure 8: The original and model η values plotted with discrete colormap to highlight three clusters in the viscosity data at zero pressure (also shown in the scatter plot). The color follows a positively correlated pattern in the model result axis as well. The lower nested plot shows the pure MgO-SiO_2 system. The upper nested plot includes all compositions by using a merged SiO_2+Al_2O_3+TiO_2 axis to obtain model values. The computation categorical bubble is split into two while the experiment category bubble is split into three bubbles. Components are in wt%.

i^{th} nested axis h_i is given by $(i - 0.5)\Delta x \pm \Delta h$ where i = 1, 2, …, k-1 and Δh can vary between 0.1Δx and 0.4Δx. This spacing scheme allows no overlap between the axes horizontally or vertically. A single polyline previously going between two primary axes are now replaced with two curves and a line. We use cubic Bezier curve with control points' x location at$(i - 1)\Delta x \pm \alpha\Delta n_x$ from primary axes where Δn_x is the distance between primary and nearest nested axis. Similarly, $(i - 0.5)\Delta x \pm \Delta h \pm \alpha\Delta n_x$ from nested axes where α can vary between 0.1 to 0.3 (Figure 7). The y values of these control points are made the same as the y values of the data line in primary and nested axes, respectively. The curves join primary axis with the nested axis and a straight line is drawn between the two nested axes for each data point. The nested axes display the full data extent range based on the minimum and maximum for domain extent of their subset (that is, a local scale is applied).

We use the nested PCP to compare the actual data and model-predicted values using the primary η axis and the derived η_{model} axis (Figure 8). We consider two scenarios of the zero pressure η-T-X model (Karki et al., 2013). In one case, we apply the model to pure MgO-SiO_2 binary system, where X represent the molar fraction of silica. In the other case, we apply the model to the multicomponent system by taking X as the sum of SiO_2, Al_2O_3 and TiO_2 fractions (i.e., using the merged component variable). There are two corresponding nested plots (Figure 8). For model result comparison it is desirable to use the same domain range for both axes in each nested plot category. Similarly, we use the nested plot to show the anomalous behavior of silicate rich composition at 3000K at two different pressure ranges (see section 6). Here we use the local domain extent for two nested plots corresponding to viscosity and pressure.

5 IMPLEMENTATION

The overarching goal of this work is to develop a web-based data analysis and modelling framework (Figure 2) for physical properties of materials by using silicate melt viscosity as an example. The users from geoscience field and, in general, diverse materials science communities may be interested in exploring these data. This work also describes a non-standard PCP implementation as a part of a web-based data analysis platform. As such, the web-application was developed using client server model. This application can be accessed using any standard web browser.

Server: The server was built using popular server-side python web-framework Django (Django) along with MySQL (MySQL) for database management. The client requests are addressed through a web API.

Client: The client application was built using client-side web-framework AngularJS (AngularJS) along with d3 for charting. D3 is a JavaScript library for manipulating documents based on data. D3 can be used to draw chart components using any HTML component such as SVG or Canvas. D3 also has a declarative syntax with Publish-Subscribe pattern interface which seamlessly works with highly interactive charts. However, due to large data size and computationally intensive task of rendering, performance is a crucial part of the application. Therefore, the application was designed with performance in mind by carefully choosing between Canvas and SVG elements for any PCP component. The client also contains a chart control panel to activate and deactivate PCP controls (derived axis, brushing, probing and others).

Other material properties can be incorporated as new attributes and corresponding new axes in the plot as long as they are defined in the same parameter (*P-T-X*) space. For instance, to visualize the melt density ρ, the density value and uncertainty entries are made in the "DataPoint" table (Figure 9). The density data at zero pressure are estimated using the density model for the multi-component melt system (Lange and

DataPoint
ID
Temperature
Pressure
Viscosity
Density
Temperature_Uncertainty
Pressure_Uncertainty
Viscosity_Uncertainty
Density_Uncertainty
Composition
Meta

Figure 9: Updated data table.

Carmichael, 1987; Lange et al., 1997). The density (ρ) axis appears in the PCP (Figure 10). The data for the computational category extend over the entire axis whereas those for the experimental category are confined in the upper half. Density takes smaller values at higher temperatures. The data lines between the density and viscosity axes tend to show positive correlation – the higher density, the higher viscosity.

6 VISUAL ANALYTICS OF VISCOSITY

In this section, we present the details of our visual data analysis of the silicate melt viscosity database. Viewing the viscosity axis using a logarithmic scale, we realize that the viscosity values span the large range of orders of magnitudes: 10^{-4} to 10^{15} Pa s. (Hui and Zhang, 2007) We also notice that viscosity show a bi-modal or tri-modal distribution (Figures 1 and 10). The outliers in the high viscosity region are from experimental sources and cover a narrow temperature range < 1000 K and zero pressure. On the other hand, outliers in the low viscosity region are mainly from computational sources at high temperatures (4000 – 6000 K) in the low-pressure regime (0 – 20 GPa).

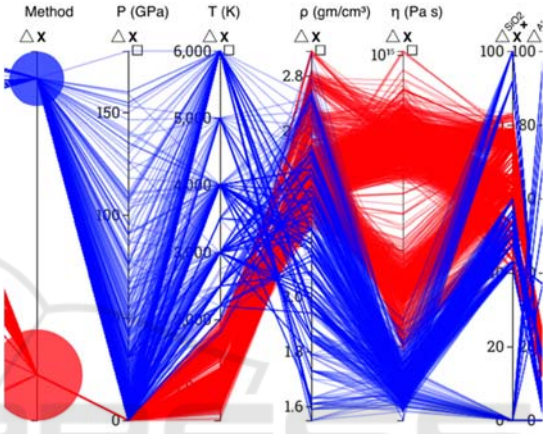

Figure 10: PCP showing two melt properties: viscosity (η) and density (ρ).

The polylines going across all axes are colored either red or blue to represent the methodology categories (experimental or computational). On the "Method" axis we see that the red circle is much larger than the blue circle (Figure 1 and 6), which means that the majority data are the measured values. Looking closely, we can also find that almost all experimental data are at ambient pressure (0 GPa) and low temperatures (< 2000 K). In Figure 6, two groups can be seen in the viscosity axis for the experimental data. One group is characterized by super-high viscosity and low temperature, and the other is characterized by high viscosity and sub-low temperature (representing two category bubbles in Figure 8). The experimental data in our database come from publications as old as 1965. The metadata axis "Year" shows a steady increase in the experimental data since the 1990s (Figure 1). The computational data, on the other hand, are relatively new and are available only from 2010 onwards. The calculated data cover much wider range of temperature (2000 – 6000 K) and pressure (0 to over 150 GPa), but the majority are found to be in the low-pressure regime. While we have not included the

experimental data available in recent years, it seems that the broad ranges of temperature and pressure data for silicate melts were previously unattainable from just experimental sources.

The composition of magma includes several oxides along with volatiles H_2O and CO_2. The computational results offer full range of MgO, SiO_2, and Al_2O_3 contents (0 to 100 wt% for each). Experimentally studied fall in the silica range 40 to 80 wt%. Pure silica or silica-rich melts tend to be highly viscous as suggested by calculated data in the low pressure-low temperature regime. The SiO_2 as network former makes the melt highly polymerized and highly viscous. However, MgO as structure modifier lowers the melt viscosity. Small amounts of volatiles can cause significant changes in melt viscosity. These dependencies can be observed by brushing along the merged dimension axis with controlled temperature and pressure ranges. The PCP method is considered to be highly effective in visually judging correlations between dimensions which are mapped to the adjacent axes. Negative correlation between the viscosity and temperature has manifested as data lines crossing each other (Figure 5). This can be further enhanced by constraining pressure and composition. Selecting the computational category at zero pressure (blue data lines), we find that the silica and MgO end members show the strongest and weakest negative correlation, respectively. If we select experimental data by clicking the red circle, we can see that the viscosity varies more than six orders of magnitude for a relatively small change in temperature, showing negative correlation with each other (Figure 6).

The viscosity-inverse temperature relationship appears to hold at all pressures and for all compositions. PCP has successfully captured this fundamental nature. Viscosity depends on pressure in a complicated way, however. Brushing the pressure axis in the range of 60 – 80 GPa we can see that several calculations were performed at 3000, 4000, and 6000 K (Figure 11). However, the viscosity value does not change as much when translating the brush towards higher pressure regime of 120 – 180 GPa. Here we can see that there are several data points at 4000 and 6000 K. The viscosity value is still found to be in the same lower region. These steps hint that viscosity of silicate melts changes much more at low temperatures and pressures, but not so much in high *P-T* regime. Interestingly, silica-rich melts display an anomalous behavior at 3000 K in that the viscosity first decreases and then increases as pressure increases. This can be observed in the two nested plots for the viscosity and pressure axes

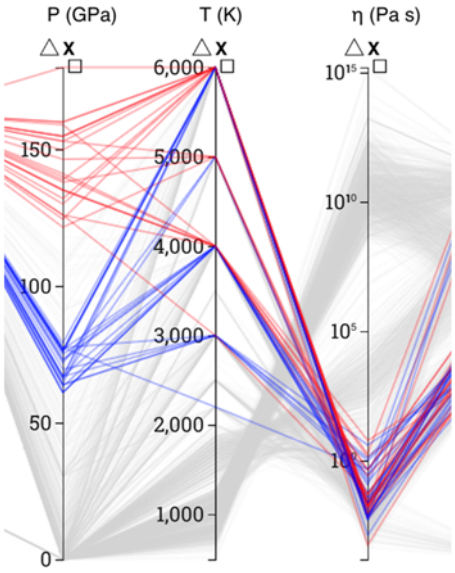

Figure 11: Two pressure regimes 60-80 GPa (blue) and above 120 GPa (red) are highlighted. Even with wide high-pressure regime, the viscosity remains almost in the same region as relatively low pressure.

corresponding to low- and high-pressure regimes (Figure 12). The data lines cross each other at low pressures (negative correlation) whereas they run parallel at high pressures (positive correlation).

There exist many models for viscosity-temperature relationships (e.g., Shaw, 1972; Mauro et al., 2009; Hui and Zhang, 2007). New models can be developed using more data points available in our database. For illustration, we consider the Arrhenian model previously developed for the MgO-SiO_2 binary at zero pressure (Karki et al., 2013):

$$\ln \eta_{\text{model}}(T, X) = (A_0 + A_1 X^4) + \frac{E_{A_0} + E_{A_1} X^4}{RT} \quad (5)$$

where the pre-exponential factor and the activation energy vary as the 4th power of the molar SiO_2 fraction X, and R is the ideal gas constant. We make two types of model assessment and display the results using the nested plots with respect to the η and η_{model} axes. For the MgO-SiO_2 binary, the model works well as shown by nearly parallel horizontal data lines between the two axes (the lower nested plot in Figure 8). To evaluate the model for the whole dataset, we take X as the sum of the molar fractions of SiO_2, Al_2O_3 and TiO_2. Many data lines follow the horizontal trend and also the model data points show the cluster patterns that are found on the viscosity axis (the upper nested plot in Figure 8). These signify the dominant role of the silica in controlling the melt viscosity. However, there are several exceptions (e.g.,

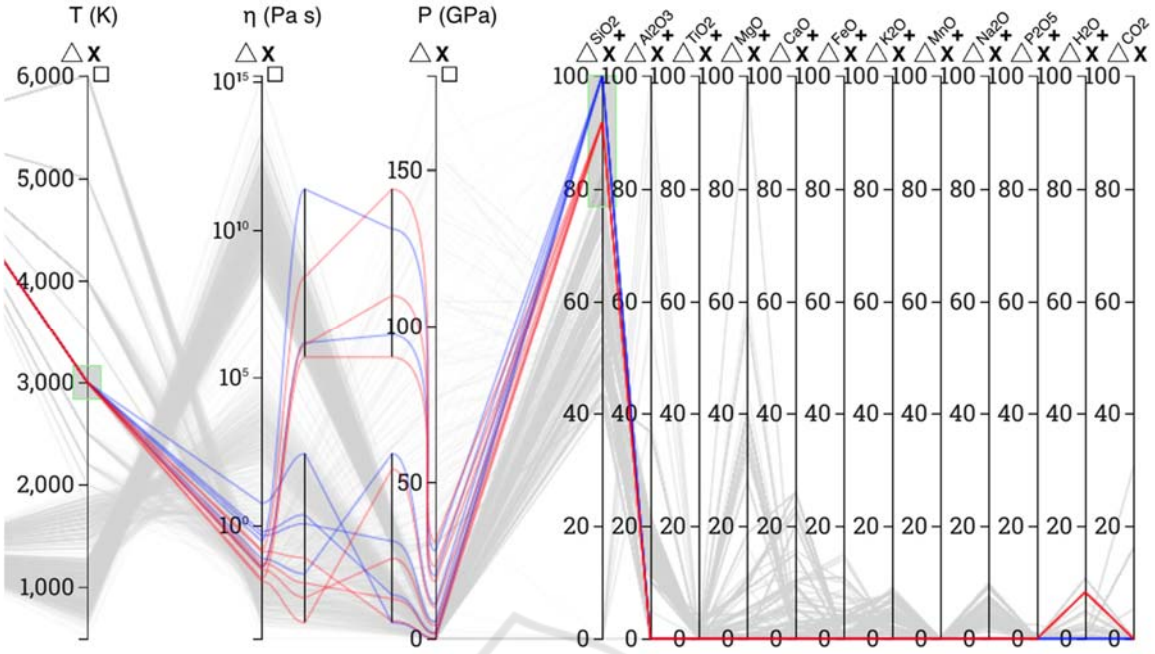

Figure 12: Data selection of silica-rich compounds at 3000 K showing anomalous behavior at low pressures below 15 GPa (lower nested plot) and normal behavior at higher pressures (upper nested plot). Blue and red represent pure and hydrous silica liquids, respectively. The oxide components are given in wt%.

polylines skewness, line crossing), which on further exploration are associated with other compositional factors (e.g., Al_2O_3 content of the melt). This means that the binary model is not sufficient, and a general multicomponent viscosity model is needed.

7 CONCLUSIONS

Advances in the computational and experimental techniques are producing ever larger amounts of data on materials properties. These data collected from various sources are used for gaining physical insight, identifying missing values, exploring systematics, and data modeling. Melt viscosity is one of the most important properties that govern production, transport, and eruption of silicate magmas. Visual exploration of available viscosity data is crucial in developing a generalized model that works across a large range of conditions of geological relevance. We have presented a data store along with a web-application for visual analytics of multivariate viscosity data. We explore parallel coordinates plot enriched with various non-standard features such as derived axes/sub-axes, dimension merging, binary scaling, nested plots. The data analysis has provided several insights about viscosity values themselves and their relationships with pressure, temperature, and composition. We plan to further develop the system into a robust platform where verified users can upload, apply standard and custom models, and analyze their own data along with the ones already present in the database for rapid model development and evaluation workflow. We also anticipate to incorporate other material properties such as density and diffusivities into our framework.

ACKNOWLEDGEMENTS

This work was funded by National Science Foundation (EAR-1426530). The authors thank Y. Zhang (University of Michigan) for useful feedback.

REFERENCES

Abe, Y. (1997). Thermal and chemical evolution of the terrestrial magma ocean. *Physics of Earth and Planetary Interiors*, 100:27-39.

Adjaoud, O., Steinle-Neumann, G., and Jahn, S. (2011). Transport properties of Mg_2SiO_4 liquid at high pressure: Physical state of a magma ocean. *Earth and Planetary Science Letters*, 312:463-470.

AngularJS, https://angularjs.org

Bohara, B and Karki, B. B. (2014). Clutter reduction in visualization of particle (atom) trajectories with

position merging, *International Conference on Computer Graphics Theory and Applications (GRAPP 2014)*, pages 265-271

Bostock, M., Ogievetsky, V., and Heer, J. (2011). D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics* (Proc. Infovis).

Django, https://www.djangoproject.com

Donato, D. V., Patrignani, M., and Squarcella, C. (2018). Exploring flow metrics in dense geographical networks. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018)*, pages 52-61.

Fisherkeller, M. A., Friedman, J. H., and Tukay, J. W. (1988). An interactive multidimensional data display and analysis system. In *Dynamic Graphics for Statistics*, pages 111-120.

Geng, Z., Peng, R., Laramee, S., Roberts, J. C., and R. Walker, R. (2011). Angular Histograms: Frequency-Based Visualizations for Large, High Dimensional Data. *IEEE Transactions on Visualization and Computer Graphics*, 17:2572-2580.

Ghosh, D. B. and Karki, B. B. (2017). Transport properties of carbonated silicate melt at high pressure. *Science Advances* 3:e1701840.

Giordano, D., Russell, J. K., Dingwell, D. B. (2008). Viscosity of magmatic liquids: A model. *Earth and Planetary Science Letters*, 27:123-134.

Graham, M., and Kennedy, J. (2003). Using curves to enhance parallel coordinate visualization. *International Conference on Information Visualization (IV 2003)*, pages 10–16.

Heinrich, J. and Weiskopf, D. (2013). State of the art of parallel coordinates. *Eurographics*, pages 95–116.

Heinrich, J., Stasko, J., and Weiskopf, D. (2012). The parallel coordinates matrix. *Eurographics Conference on Visualization*, pages 37–41.

Hui, H. and Zhang, Y. (2007). Toward a general viscosity equation for natural anhydrous and hydrous silicate melts. *Geochimica et Cosmochimica Acta*, 71:403-416.

Inselberg, A. (1997). Multidimensional detective. *IEEE Symposium on Information Visualization (INFOVIS 1997)*, pages 100-107.

Inselberg, A (2009). *Parallel coordinates: visual multidimensional geometry and its application*. Springer, New York.

Johansson, J. and Forsell C. (2016). Evaluation of parallel coordinates: Overview, categorization and guidelines for future research. *IEEE Transactions on Visualization and Computer Graphics*, 22:579-588.

Jourdain, S., Ayachit, U., and Geveci, B. (2011). ParaViewWeb, a web framework for 3D visualization and data processing. *International Journal of Computer Information Systems and Industrial Management Applications*, 3:870-877.

Karki, B. B., Zhang, J. and Stixrude, L. (2013). First-principles viscosity and derived models for MgO-SiO_2 melt system at high temperatures. *Geophysical Research Letters*, 40:94-99.

Kaur, G. and Karki, B. B. (2018). Bifocal parallel coordinates plot for multivariate data visualization. *Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (*VISIGRAPP 2018)*, pages 176-183.

Kono Y. and Sanloup, C. (2018). *Magmas Under Pressure Advances in High-Pressure Experiments on Structure and Properties of Melts*, Elsevier Inc.

Lange, R. A. and Carmichael, I. S. E. (1987). Densities of Na_2O-K_2O-CaO-MgO-FeO-Fe_2O_3-Al_2O_3-TiO_2-SiO_2 liquids: New measurements and derived partial molar properties. *Geochimica et Cosmochimica. Acta* 51:2931-2946.

Lange, R. A. (1997). A revised model for the density and thermal expansivity of K_2O-Na_2O-CaO-MgO-Al_2O_3-SiO_2 liquids from 700 and 1900 K: extension to crustal magmatic temperatures. *Contributions to Mineralogy and Petrology*, 130:1-11.

Lu, L. F., Huang, M. L., and Zhang, J. (2016). Two axes re-ordering methods in parallel coordinates plots. *Journal of Visual Languages and Computing*, 33:3–12.

Mauro, J. C., Yue, Y., Ellison, A. J., Gupta, P. K., and Allan, D. C. (2009). Viscosity of glass-forming liquids. *Proceedings of National Academy of Sciences* 106:19780–19784.

MySQL, http://www.mysql.com

Nguyen, H. and Rosen, P. (2018). DSPCP: A data scalable approach for identifying relationships in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 24:1301-1315.

Peltonen, J. and Lin, Z. (2017). Parallel coordinates plot for neighbour retrieval. *Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (*VISIGRAPP 2017)*, pages 40-51.

Roberts, R., Laramee, R. S., Smith, G. A., Brookes, P. and D'Cruze, T. (2018). Smart brushing for parallel coordinates, *IEEE Transactions on Visualization and Computer Graphics. Pages 1-1*.

Riehmann, P., Opolka, J., and Froehlich, B. (2012). The product explorer: Decision making with ease. *International Working Conference on Advanced Visual Interfaces (AVI'12)*, pages 423-432.

Shaw, H. R. (1972). Viscosities of magmatic silicate liquids: an empirical method of prediction. *American Journal of Science*, 272:870–893.

Siirtola, H. and Raiha, K. (2006). Interacting with parallel coordinates. *Interacting with Computers*, 18:1278-1309.

Solomatov, V. S. (2007). In Evolution of the Earth, D. Stevenson, Ed., vol. 9 of Treatise on Geophysics, G. Schubert, Ed. Elsevier, Amsterdam, p. 91.

Tuor, R., Evéquoz, F., and Lalanne, D. (2018). Parallel bubbles: categorical data visualization in parallel coordinates. *In Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications* (*VISIGRAPP 2018)*, pages 299-306.

Turkay, C., Filzmoser, P., and Hauser, H. (2011). Brushing dimensions: a dual visual analysis model for high dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 17:2591-2599.

Urbain, G., Bottinga, Y., Richet, P. (1982). Viscosity of liquid silica, silicates, and alumino-silicates. *Geochimica et Cosmochimica. Acta* 46:1061-1072.

Wang, Y., Sakamaki, T., Skinner, L.B., Jing, Z., Yu, T., Kono, Y., Park, C., Shen, G., Rivers, M. L., and Sutton, S. R. (2014). Atomistic insight into viscosity and density of silicate melts under pressure. *Nature Communications*, 5: 3241.

Wang, J., Liu, X., Shen, H. W., and Lin, G. (2016). Multiresolution climate ensemble parameter analysis with nested parallel coordinates plots. IEEE *Transactions on Visualization and Computer Graphics*, 23:81-90.

Zhang, Y., Xu, Z., Zhu, M., and Wang, H. (2007). Silicate melt properties and volcanic eruptions. *Reviews of Geophysics*, 45: RG4004.

Zhou, H., Yuan, X., Qu, H., Cui, W., and Chen, B. (2008). Visual clustering in parallel coordinates. *Computer Graphics Forum* 27:1047–1054.

Zhou, L. and Weiskopf, D. (2018). Indexed-points parallel coordinates visualization of multivariate correlations. *IEEE Transactions on Visualization and Computer Graphics*, 24: 1997-2010.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS