Published in final edited form as:

Proc SPIE. 2012 ; 8294: 82940T. doi:10.1117/12.907486.

Increasing the perceptual salience of relationships in parallel coordinate plots

Jonathan M. Harter^a, Xunlei Wu^a, Oluwafemi S. Alabi^a, Madhura Phadke^b, Lifford Pinto^b, Daniel Dougherty^c, Hannah Petersen^d, Steffen Bass^d, and Russell M. Taylor II^a

Jonathan M. Harter: jmharter@cs.unc.edu; Xunlei Wu: xunlei@cs.unc.edu; Daniel Dougherty: doughe57@msu.edu; Hannah Petersen: hp52@phy.duke.edu; Steffen Bass: bass@phy.duke.edu; Russell M. Taylor: taylorr@cs.unc.edu

^aComputer Science, University of North Carolina, Chapel Hill, NC, USA

^bComputer Science, North Carolina State University, Raleigh, NC, USA

^cStatistics, Michigan State University, Lansing, MI, USA

^dPhysics, Duke University, Durham, NC, USA

Abstract

We present three extensions to parallel coordinates that increase the perceptual salience of relationships between axes in multivariate data sets: (1) luminance modulation maintains the ability to preattentively detect patterns in the presence of overplotting, (2) adding a one- vs.-all variable display highlights relationships between one variable and all others, and (3) adding a scatter plot within the parallel-coordinates display preattentively highlights clusters and spatial layouts without strongly interfering with the parallel-coordinates display. These techniques can be combined with one another and with existing extensions to parallel coordinates, and two of them generalize beyond cases with known-important axes. We applied these techniques to two real-world data sets (relativistic heavy-ion collision hydrodynamics and weather observations with statistical principal component analysis) as well as the popular car data set. We present relationships discovered in the data sets using these methods.

Keywords

Parallel coordinates; scatter plot; multivariate visualization

1. INTRODUCTION

Our team is collaborating with groups of researchers studying statistics, high-energy physics, meteorology, medical imaging, microbiology, and biomaterial science at Michigan State, Duke, and UNC as well as Sandia National Laboratories. Each of these groups has high-dimensional multivariate data sets (in some cases coming from ensemble simulation runs), from which they seek to determine relationships among the many variables. Scientists are also interested in parameter sensitivities, i.e. the influence of parameter perturbation on simulation outputs. We showed a conventional parallel-coordinates (PC) technique to our collaborators to examine their data. Despite the fact that they can look at the high degree multivariate data in 2D and trace individual data samples, they found these plots to be difficult to work with in the presence of larger numbers of data points, larger numbers of variables, and more varied relationships. We developed three extensions to traditional PC to increase the perceptual salience of multivariate relationships.

- **Luminance-Modulated Parallel Coordinates (LMPC)** apply a random or variable-value-dependent luminance modulation to mitigate line cluttering and to re-introduce structure into overplotted PC plots,

- **One-vs.-All Parallel Coordinates (OVAPC)** generalize the scatter-plot matrix and display the relationship between a single variable in the data set and all other variables, and
- **Scatter Plots Overlay (SPO)** combine the pairwise cluster-detection and distribution-shape analysis benefits of scatter plots with the multi-variable display capabilities of parallel coordinates.

In this paper, these three techniques are shown to complement each other; to be compatible with existing PC variations (e.g. spline-based display, brushing or clustering algorithms, interactive axis reordering, *etc.*); and to increase the preattentive salience of relationships between variables of interest in real-world data sets.

The rest of this paper is organized as follows: Sections 2 through 4 present each technique along with related work. Section 5 discusses the trade-offs associated with using these methods in side-by-side and overlaid views. Section 6 demonstrates the application of these techniques in various combinations to three case studies and describes the real-world relationship insights gained using them.

2. LUMINANCE-MODULATED PARALLEL COORDINATES

Luminance-modulated parallel coordinates (LMPC) is a new technique that combines the ability of vector-field display techniques such as line-integral convolution (LIC) to preattentively display dense orientation fields with a PC display of multivariate relationships. It extends the useful number of lines that can be visualized using PC beyond the point at which overplotting causes regions of the display to be densely cluttered. Figure 1 shows this technique applied to a collaborator's weather observation data set.

2.1 Related Work

Parallel Coordinates—Among the most successful tools for multidimensional multivariate visualization is the parallel-coordinates technique introduced by Inselberg.¹ In this method, each variable of a multivariate dataset receives its own vertical axis. These axes are placed one after the other, separated by a fixed distance. Lines are then drawn from one axis to the adjacent axis connecting values that belong to the same tuple. Each line represents one data point in a high-dimensional space. Figure 3(a) shows an example for a small subset of data points from one of our collaborator's data sets.

Density-based methods have been developed to remove overplotting by summing brightness contributions from all lines in the image and then normalizing the luminance channel of each line. These methods produce continuous density estimates^{2,3} highlight trends and clusters in the data, and reduce the brightness of sparse regions. The resulting plots produced by these methods range from standard PC when the data size is small to a summary display with a large amount of lines. LMPC does not show relative line density in overplotted regions, opting instead to leave the entire covered region (and all outliers) visible while reintroducing local line orientation details.

Clustering methods have been developed that remove overplotting by displaying only lines that represent clusters of values.⁴ Novotný and Hauser extend this by also identifying outliers through histogram analysis.⁵ Zhou *et al.* use a clustering algorithm and polyline splatting to selectively and iteratively enhance or reduce the value of previously-drawn lines.⁶ This de-emphasizes and eventually removes lines from the plot, effectively thinning it to show only lines that indicate clusters. By selecting different pair of variables, they emphasize different 2D subspaces. LMPC achieves similar cluster-detection results without removing any lines from the plot. It splats all lines as independent entities (maintaining

dense coverage of the image) and thus preserves the display of local and global line orientations.

Vector-Field Display—LMPC was also inspired by the spot noise⁷ and line-integral convolution⁸ techniques that were developed for the dense display of vector-field data. Rather than using local splats or convolution to produce its texture, LMPC splats an entire polyline mapped from a data sample by adjusting its luminance with either a random number or a user selected attribute value.

When used together with hue for brushing, LMPC behaves in a manner quite similar to the color weaving LIC display technique described in.⁹ It takes advantage of the same preattentive separation of individual colored strands and large-scale mixing to enable both identification of large-scale patterns and determination of small-scale features.

Proximity Based Shading—LMPC is not the first extension to PC to vary the luminance of lines. Johansson *et al.*¹⁰ applied high-precision textures to PC line clusters in conjunction with transfer functions in order to highlight cluster characteristics. Steed *et al.*¹¹ extended their work with a method called aerial perspective shading in which the saturation and luminance of lines are varied depending on where the lines intersect an axis of interest. Lines that intersect a specified region of that axis are displayed boldly, while others fade into the plot background. In this way, luminance and saturation fulfill a role similar to that of hue in line brushing. LMPC extends the use of luminance variation to show relationships (when mapped to the values on an axis) and to maintain detail in regions of overplotting (when mapped to random variation).

2.2 Approach

Random Luminance Modulated Parallel Coordinates—In its most basic form, LMPC multiplies the luminance of a PC polyline with a factor determined by a uniformly distributed random variable between 0.5 and 1. Each line remains fully opaque. The modulation produces a plot similar to that of spot noise, which produces everywhere a local linear brightness gradient along the line. Before overplotting occurs, random LMPC (RLMPC) looks quite similar to a regular PC, since the intensity variation is not high enough to impede the perception of the individual lines against a dark background. In an overplotted region, RLMPC maintains the distinction of each polyline from its neighbors whose luminance values are randomly modulated, thus preserves the capability of each polyline to indicate its orientation.

The choice of modulating the luminance of each polyline has another benefit over density-based methods beyond improved shape perception. Because the minimal modulating factor is 0.5, no polyline is removed. This choice also leaves hue free for its traditional use in brushing of clusters within parallel coordinates.

Attribute Luminance Modulated Parallel Coordinates—In attribute LMPC (ALMPC), the luminance of a PC poly-line is modulated by a user-selected variable's value, again within the range [0.5,1]. In addition to mitigating overplotting and clutter, ALMPC produces brightness gradients along each variable axis that is correlated with the selected variable, which can stimulate the perception of relationships between the selected variable and all other variables. In Figure 1(b), the luminance variable is h_2O_{vap} . The overplotting seen in Figure 1(a) disappears when ALMPC is added in (b), and line orientations are more apparent. By looking at the luminance distribution on axis T_{air} , we can conclude h_2O_{vap} and T_{air} are negatively correlated. This is consistent with tracing polylines between the two axes. Within the same figure, it is evident that spd_{wind} is also negatively correlated with h_2O_{vap} .

bat_{volt} is positively correlated with h_2O_{vap} ; large h_2O_{vap} values cluster between 60% to 85% of pca_2 ; and pca_3 is independent with h_2O_{vap} . This benefit is similar to the One- vs.-All Plots techniques discussed in section 3.

Animated random resampling—Once a region has become heavily overplotted, the most-recently-plotted set of lines will obscure previously-plotted sets of lines. As in the case of spot noise, LMPC continues to produce an everywhere-dense estimate of local orientation, but the estimate will be based only on the most-recently-plotted lines. If the lines are plotted in a random order, this should produce an unbiased sampling. To avoid any remaining effects due to rendering order bias, we suggest continuously updating the display, cycling through all data values. This will provide the viewer with a continuous resampling of the data, indicating by its dynamic changes any distribution differences caused by sampling.

3. ONE-VS.-ALL PLOTS

To compare one variable to another in PC, the viewer traces the line passing from the first variable's axis to the target variable's axis. This multi-stage tracing precludes the rapid discovery of relationships between two variables which are mapped to distant axes. As the number of axes between them and the amount of data increase, it becomes more challenging to discover their relationship.

Because of their familiarity with the data, our collaborators often have an intuition for which axes are most likely to be important. Also, our statistics collaborators' analyses on data sets (principal-component analysis, eigen analysis) often reveal which variable axes explain most of the variance or which are most strongly correlated with other data sets. Each of these produces a small set of likely candidate axes that scientists use to compare to all other axes to look for important relationships.

One- vs.-all PC (OVAPC) is a new extension to PC that adds another row of plots below a standard PC, where the right-side of each block in the row displays the same variable, plotting it against all other variables (left-hand side of each block). This augments the ability to trace across the entire graph (PC's traditional strength) with the ability to preattentively compare one data set against all others. Section 6 describes how this augmented display increases preattentive salience of relationships beyond the capability of PC or scatter plots alone.

3.1 Related Work

Axis reordering—Animated and interactive axis re-ordering in PC has been used to make axes that may be related adjacent to one another.¹² This enables the viewer to compare pairs of axes while retaining the ability to trace values to other axes. To perform rapid comparison of a single potentially-dominant variable against all other axes requires a number of axis reorderings that is linear in the number of variables. Each axis re-arrangement alters two or more patterns such that relationships seen in a previous ordering may be lost. Thus, the detection of a relationship between a dominant variable and two others requires a number of rearrangements on the order of the number of axes squared (to place the two variables to either side of the dominant variable). The *Case Studies* section describes a case where the comparison of a dominant variable with two other variables resulted in the discovery of relationships in the data. Preattentive detection of similar relationships between the dominant variable and three or more variables is not possible using axis reordering alone.

Scatter plot matrix—Viau *et. al.*¹³ showed the benefit of using parallel coordinates in conjunction with the scatter-plot matrix.¹⁴ The OVAPC plot is a generalization of the

scatter-plot matrix applied to PC. In this case, only a subset (perhaps only one) of the values in the data set is plotted against all other variables. Removal of the redundant mapping, axis selection, and label placement and coloring are optimized for PC display to provide rapid readability of the resulting sub-matrix. The main benefit over plot-of-plots is the reduction of the number of plots, enabling the plots to be drawn at a much larger size.

3.2 Approach

For the specialized case in which one variable's relationship to all others is important, OVAPC extends parallel coordinates to enable one-*vs.*-all variable comparisons. For a selected variable p_{ova} , a 2-axis PC subplot $\Pi_j(\mathbf{X})$ is constructed between p_j and p_{ova} . This construction is then repeated $n - 1$ times, excluding p_{ova} itself (because $\Pi_{ova}(\mathbf{X})$ is trivially linear). For conventionality, the fixed variable is always mapped to the right axis. Although it is not required, to conserve space as well as to maintain consistent appearance with traditional PC, these subplots are concatenated in the same order as the traditional PC arrangement. The j^{th} axis of OVAPC becomes a *dual axis* representing both a_{ova} for the left PC subplot, $\Pi_{j-1}(\mathbf{X})$, and the subsequent a_j with respect to the right PC subplot, $\Pi_j(\mathbf{X})$. The comparison between a traditional parallel coordinate graph and an OVAPC plot over the same variable space is displayed in Figure 2.

If there is more than one potentially-dominant variable, then more than one row of OVAPC displays can be added. In the limit, this becomes a PC matrix. We have found it more useful for the viewer to be able to interactively select which variable is mapped to the OVAPC display. Combining these approaches enables the viewer to add as many new rows as seems potentially useful, mapping each to their variable of choice.

4. SCATTER PLOTS OVERLAY (SPO)

Compared to PC, scatter plots (SP) can only display relationships between two variables at a time. However, SP have been shown in controlled studies to enable viewers to reliably distinguish twice as many different correlation levels with less bias compared to PC.¹⁵ Because they display 2D projections of higher-dimensional shapes, they also directly display the silhouette of 3D spatial data sets. They also enable rapid detection of nonlinear regression between two data sets by drawing points along a curve in space.¹⁶ Because of these benefits and because points can both be preattentively distinguished from lines and take up only a small fraction of the display space of lines, we find that augmenting a PC display by superimposing a matrix of SP can improve the overall perceptual salience of relationships in the same-sized display region.

4.1 Related Work

Scatter plots—The scatter-plot matrix displays a number of adjacent 2D scatter plots, each of which shows a different 2D projection of a high-dimensional data set, enabling the pairwise study of relationships among the variables.¹⁴ SPO directly combines a subset of the full scatter-plot matrix with PC, augmenting PC's multivariate tracing with the preattentive 2D clustering and linear-relationship-determination capabilities of scatter plots.

Combination Display—Wong *et al.* and others have developed combination displays that show the same data set in SP and PC form within different display windows.¹⁷ Yuan *et al.* extended this and developed a method using multidimensional scaling and animation to smoothly transition from a PC display to the combined display of parallel coordinates and scatter plots in the same graph.¹⁶ Their technique converts two or more dimensions of parallel coordinate plots into scattered points intersecting with the parallel coordinate lines. In the case of two dimensions, these plots are similar to the plots used in SPO but differ due

to an additional transform to map them to PC lines. Whereas Yuan combined SP and PC into an integrated whole, SPO relies on the preattentive visual separation of points and lines and simply plots the points over the same region as the PC lines using a different color. This enables an undistorted view of both the SP and PC.

4.2 Approach

The implementation is straightforward. A SP subplot $\Sigma_j(\mathbf{X})$ represents the relationship between variables p_j and p_{ova} . The vertical axis a_j mapped to variable p_j and the horizontal axis b_{ova} is mapped to p_{ova} . Points are then plotted in the same space that the PC subplot $\Pi_j(\mathbf{X})$ occupies. In this way, SP and PC are independent of each other and the same data can be analyzed separately.

SPO can be used with either traditional PC or with our OVAPC approach. Figure 3 displays three one- vs.-all plot configurations over the same data. OVASP and OVAPC complement each other in that they each reveal correlations that may be masked by the other. For instance, Figure 3(b) displays a clear, positively-sloped diagonal for v_y vs. y , denoting a predominantly positive correlation, and an ellipse for x vs. y . It is more difficult to see either of these correlations in the cross-hatched line patterns of the same plots in 3(a). On the contrary, Figure 3(a) displays a cluster of lines toward the center of $dsig_y$ branching to y , which is less apparent from the same plot in Figure 3(b). This indicates the value of showing both representations.

OVASP can be used in place of, or in addition to, traditional parallel coordinates and OVAPC. Each SP subplot $\Sigma_j(\mathbf{X})$ contains axes $a_{j,0}$ and $a'_{j,1}$ which correspond to variables p_j and p_{ova} respectively. This is a similar convention to the axes used in OVAPC. However, unlike OVAPC, axis $a'_{j,1}$ is perpendicular to $a_{j,0}$, as is conventional for a traditional SP. For consistency, $a'_{j,1}$ is connected to the bottom of $a_{j,0}$ as opposed to the top. For n variables, n SP are constructed ($n + 1$ if p_{ova} vs. p_{ova} is included). Since each OVASP plot $\sigma_j(X)$ shares a common axis $a_{j,0}$ with each OVAPC plot $\lambda_j(X)$, they are able to occupy the same plot. For each point $\mathbf{x}_i = (x_{i,0}, x_{i,1}, \dots, x_{i,m}) \in X$, values $x_{i,j}$ for $0 \leq j \leq m - 1$ are plotted in subplot σ_j . The bottom plot in Figure 3 displays a OVASP instance. The white points are overlaid on the red OVAPC plot, although they can exist independently. Figure 3 shows both an OVAPC plot and the same plot with an overlaid OVASP plot.

Because a SP uses a single point to display each data value whereas PC displays a curve between two axes, SP always has much lower visual density for the same number of data values compared to PC. At higher data densities, a random subset of the points can be displayed using SPO to produce an unbiased display of relationships. To reduce the impact of small-sample effects, the selection of points can be animated over time to continually resample the data values.¹⁸

5. COMBINING TECHNIQUES

These three techniques are compatible with one another and with other common PC extensions. Depending on the available space and information sought by the viewer, different combinations of the above techniques are optimal. The following are suggested configurations that our collaborators found to be most useful. Because of the large number of variable axes in these data sets, screen space was at a premium – just showing a PC plot takes up half of a page demonstrated by Figure 5. Thus, it is important to be able to combine the described techniques, rather than displaying each in a separate window.

5.1 Parallel Coordinates and OVA Plots

A basic configuration includes rendering a traditional PC and the corresponding OVAPC side by side, seen in Figure 2. Although each OVAPC subplot is independent of the others, we concatenate all of them by merging the right axis of $\Pi_j(\mathbf{X})$ with the left axis of $\Pi_{j+1}(\mathbf{X})$. The orders of vertical axes between OVAPC and PC should be the same. This not only saves screen space, but aligns axis configuration between OVAPC and PC. Such alignment is important to correlate the two different multivariate data representations. With this configuration, the entire PC plot is available for analysis, with the OVAPC plot providing an added level of detail for any one variable. Our implementation provides the user full control over both variable ordering and the selection of the OVA variable at any time. Similarly, the locations of each OVASP subplot's vertical axis is the same with that of PC. Both the OVASP subplot and the PC subplot maintain the same aspect ratio.

5.2 Parallel Coordinates and SPO

In some cases, visualization space is extremely limited. If there is just enough room for a PC plot, an OVASP plot can be overlaid on this plot. In this way, the user is able to analyze the PC information while at the same time having access to OVA correlations. This configuration takes no more screen space than PC itself.

5.3 Other Common Techniques

Each of the proposed techniques is compatible with using color to distinguish different clusters, with various straight-line or curved-line display techniques, and with brushing on the underlying PC display. They are also directly compatible with interactive axis reordering and with animated methods that display different subsets of the data over time. Because OVAPC subplots are disjoint from each other, axis reordering does not affect the relationship pattern, displayed in the graph.

6. CASE STUDIES

We present case studies on these techniques as they were applied to three real-world data sets. The first two data sets are from our collaborators, and the third is the well-known cars data set. We describe interesting relationships among multiple variables that our collaborators were able to detect in their data using our new techniques.

6.1 Relativistic Heavy-Ion Collision Data

Figure 2 shows OVA plots (OVAPC and OVASP) and SPO techniques on relativistic heavy ion collision simulation data¹⁹ with 7 variables. Our PC extensions provided several advantages over traditional PC when looking for multiple-variable relationships within this figure that enabled insights into the data. In Figure 2(a), it is difficult to correlate $n(1)$ and T . In Figure 2(b), a viewer can observe the individual relationships between $n(1)$ and T , *without changing axis order*, by examining their individual relationships with $mu(1)$, the OVA variable. The second subplot in Figure 2(b) suggests $n(1)$ and $mu(1)$ are almost identical. The last subplot in Figure 2(b) suggests T and $mu(1)$ are negatively correlated. Thus $n(1)$ and T are negatively correlated as well.

Comparing simulations—Several different simulation implementations have been used to generate datasets for this particular relativistic heavy ion collision study. Figure 2 showed results from one high-energy collision data set. Figure 4 displays data for the same timestep, but two different initial conditions of the simulation: Glauber b0 and Pcasc b0 (Figure 4(a) and (b), respectively). Both simulations are based on the same theoretical model, and therefore, should display the same distributions and correlations. Using our technique to

analyze and compare each data set, it became clear to our colleagues that the energy-density distributions and several energy-density correlations differed between the two simulations. Most notably, the correlation between energy-density and pressure (p) differs between each simulation. The energy-vs.-pressure scatter plot in figure 4(b) consists of two mostly-linear portions with a curve near the middle of the range, but the plot in figure 4(a) has two kinks, and they occur at a much lower value. When comparing the scatter-plot densities for e vs. v_x , v_y , and y_L between the two simulations, it is also apparent that pcasc b0 has a much larger distribution of low energy-density points as the points are almost all concentrated in the lowest 5% of the energy-density spectrum.

6.2 Weather Observation Data and its Principal Components

Our collaborators in weather simulation tried our techniques on a weather observation dataset consisting of 22 variables with values derived from 1-min averages of 10 HZ samples collected from the University of Houston's flux tower as part of the T-REX field campaign. The collection extended from March 4 to April 29 of 2006 at Owens Valley Radio Observatory. For the purposes of this paper, one day of the 57 has undergone analysis. Our statistics collaborators derived 22 principal-component axes from the original variables, for a total of 44 variables. Both the correlations between the original variables and the principal components are of interest. Because 99.9% of the variability in the data was explained with the first three principal components, they were the only candidates for one-vs.-all comparison. OVAPC enabled rapid exploration of the data and several strongly-correlated groups emerged within only a few OVA variable selections.

Data driven analysis—Of the strongly-correlated groups, the majority consisted of positively correlated variables. The first pair, soil heat flux 1 and 2 ($soil_1$ and $soil_2$) in Figure 5(a), while positively correlated, display two parallel, positively-sloped diagonal dot patterns. This reveals periodicity in the data that is otherwise difficult to notice with PC alone. Clear lack of periodicity is also a salient feature provided by SPO with LMPC. In this atmospheric case study, lack of periodicity indicates weather differences between the preceding and subsequent days. Most notably, in Figure 5(a) soil water content ($soil\ h_{20}$) corresponds to an actual precipitation event which did occur in the region and on the day studied.

The next positively-correlated group consists of temperature readings (specifically, sonic temperature, data logger reference temperature, and air temperature (T_{sonic} , T_{ref} , and T_{air})). This is an intuitive correspondence as the temperature of one region or entity affects the temperature of surrounding areas and entities. Another positively-correlated group was related to radiation, consisting of longwave incoming, longwave outgoing, shortwave incoming, shortwave outgoing, and net radiation (I_{in} , I_{out} , s_{in} , s_{out} , and net_{rad}). This correlation is also displayed in Figure 5(a), where each variable plotted in a row creates a succession of positive correlations. It is interesting to note that I_{in} displays a “u-shaped” SP, meaning there are two distinct, positive correlations between I_{in} and the other variables. This divergence could relate to night and day, times in which radiation patterns vary from each other. In theory, net radiation is mathematically related to longwave outgoing, longwave incoming, shortwave incoming, and shortwave outgoing by the equation: $net_{rad} = I_{in} - I_{out} + s_{in} - s_{out}$. Because I_{in} , I_{out} , s_{in} , and s_{out} are all positively correlated, it is likely that the magnitude of $I_{in} + s_{in}$ is consistently larger than $I_{out} + s_{out}$ since net_{rad} keeps in a positive relation with each variable. This is to say, there is probably more radiation coming into the system than out. If $I_{out} + s_{out}$ dominated, a negative correlation between each variable and net_{rad} would be expected. The variables relating to water comprise another positively-correlated pair: water content period and soil water content (h_{20_cont} and $soilh_{20}$).

The positive correlations within each aforementioned group can clearly be seen in Figure 5(a) as each group is a series of subplots of positive correlations. Several other strong correlations were identified. Battery voltage corresponds negatively to the pair of soil heat flux variables, the group of temperature variables, and wind speed (spd_{wind}). Perhaps high temperatures and wind currents drain the battery of the device. Relative humidity (rel_{hum}) exhibits an unexpected correlation to several variables. For mid to low values of relative humidity, other variables seem to increase drastically. This can be seen by SP patterns in which a dramatic vertical spike exist for certain variables *vs.* relative humidity, and also in luminance modulated lines in which a smoothly shaded “arm” juts from the low to mid region of relative humidity to the top of the other variable axis. Such variables include the soil heat flux variables, the temperature variables, and the radiation variables. In Figure 5(a), one such spike was seen in the plot of rel_{hum} *vs.* $soil_2$, and an inverted version was recognized in $batt_{volt}$ *vs.* rel_{hum} . Such spikes are difficult to see in traditional parallel coordinates when overplotting occludes the pattern, but with both SPO and LMPC they stood out immediately. Outliers were seen in the plot of spd_{wind} *vs.* $albedo$ in Figure 5(a), where several lines are drastically separated from the others. This was reaffirmed by the stranded points in the SP. The outliers present in this subplot were traced to a phenomenon that occurred in five minute periods of time in the middle of the night in which the value for $albedo$ would jump from 0% to 5 – 18%, perhaps due to intermittent low-angle artificial light coming from automobiles or workers in the area.

Principal component analysis—Each principal component was ordered by the percentage of variability it explains. Thus, determining which components are strongly correlated with each principal component helped determine which variables have the largest impact on the system. Figure 5(b) displays the correlation between principal component one (pca_1) and representatives from the above correlated groups that were seen in Figure 5(a). Note that a strong correspondence with a group member implies a strong correspondence with the group. From the plot, we saw a distinct positive correlation between pca_1 and each of the radiation variables (net_{rad} , I_{in} , I_{out} , s_{in} , and s_{out}). This indicated that radiation is responsible for the majority of variability in the data. Principal component two was revealed to be positively correlated with I_{in} , rel_{hum} , and $batt_{volt}$. However, pca_2 is negatively correlated with soil temperature, sonic temperature, data logger reference temperature, air temperature, soil heat flux 1, soil heat flux 2, water content period, and soil water content (T_{soil} , T_{sonic} , T_{ref} , T_{air} , $soil_1$, $soil_2$, h_2O_{cont} , and $soil_{cont}$ respectively). In fact, it seemed that almost all of the variables with a particularly low correlation to both pca_1 and pca_2 relate to wind: wind speed, wind direction, u -wind, v -wind, and w -wind. Interestingly, pca_3 displayed a positive correlation to u -wind and a negative correlation to wind direction. Thus, wind does have some impact on the variability of the system, but by an inappreciable margin as pca_3 explains only a fraction of variability.

6.3 Car Data

To compare against standard data used in the literature, we also applied our technique to the “cars” dataset,²⁰ plotting 392 points spanning 8 dimensions. Figure 6 displays two one-*vs.*-all parallel coordinate plots, with OVA variables horsepower and weight for Figure 6(a) and Figure 6(b), respectively. Each plot shows luminance modulated by its respective OVA parameter, with a corresponding OVASP overlaid. Five of the eight variables are displayed: mpg, weight, acceleration, displacement, and horsepower. These two plots were selected due to the clear correlations displayed.

Figure 6(a) reveals that horsepower is positively correlated with both weight and displacement. This can be seen from the distinct positively-sloped diagonal SP clusters, as well as by the smooth luminance transitions from dark to light in subplots *weight vs. HP* and

disp vs. HP. Both conclusions fit the data, because a heavier car probably requires more horsepower to maneuver as well as a larger engine (displacement). It is to be expected that a larger engine will add to the weight of the car. From this notion and the fact that weight and displacement both correspond to horsepower, the OVA variable, in similar ways, it is a reasonable assumption that weight and displacement are also positively correlated. The subplot of *disp vs. weight* in Figure 6(b) confirms this.

Negatively-sloped diagonal SP clusters coupled with a dramatic luminance cross-over in the center of subplots *mpg vs. HP* and *accel vs. HP* in Figure 6(a) imply horsepower is negatively correlated with miles-per-gallon and acceleration. Intuitively, an automobile with high horsepower, that is, high power output, uses more fuel. This implies a negative correlation between horsepower and gas mileage. Acceleration varying inversely with horsepower is less intuitive, but still valid. As mentioned, weight is positively correlated with horsepower and therefore, is negatively correlated with acceleration to some extent (confirmed by *accel vs. weight* subplot in Figure 6(b)). Thus, the weight of an automobile probably contributes strongly to the negative correlation between acceleration and horsepower.

The above correlations are only a subset of correlations found within the dataset. When expanding Figure 6 to 8 parameters, other positive correlations include weight and displacement; displacement and cylinders; year, mpg, and acceleration; and cylinders, weight, and displacement. Other negative correlations include mpg and displacement; mpg and cylinders; horsepower and year; horsepower and origin; weight and origin; acceleration and displacement; displacement and mpg; displacement and year; and displacement and origin.

7. DISCUSSION

We describe three extensions to parallel coordinates that increase the preattentive visual salience of relationships between variables in multivariate data sets. Each of these were applied to the case where one or a small set of variables are more likely to have relationships than the others; two of them generalize beyond this case to apply to PC displays in general. Each of these extensions combines well with the others, and with existing extensions. These techniques were evaluated singly and in combination using three data sets (two of which were large-scale real-world data sets from our collaborators) and in each case they increased the preattentive salience of relationships between variables.

7.1 Advantages

Luminance-Modulated PC—LMPC has been shown to extend the useful number of displayable lines using PC by creating distinctions between lines, particularly in overplotted space. This enables more densely covered plots while retaining effectiveness. When coupled to one of the variables, smooth and rough luminance transitions highlight strong and weak variable correlations, respectively. When modulating by a specific variable, correlations between that variable and all others are preattentively obvious since the luminance at any axis crossing maps that axis to the luminance variable. This is true even when using a traditional PC ordering as opposed to OVA. Luminance modulation does not change hue, so remains compatible with coloring extensions to PC such as line cluster brushing.

One vs All Plots—OVAPC and OVASP have been shown to reveal direct comparisons between one variable and all others at the same time, enabling rapid analysis of any one variable. For n parameters, all direct comparisons between all variables can be viewed with $O(n)$ ova variable selections, whereas $O(n^2)$ axis re-orderings are required for traditional PC. An OVA plot requires at most one more subplot than a traditional PC plot using the same

data, and can be reordered to be the same size as a traditional PC plot by moving the ova variable vs. ova variable subplot to the outside of the plot and clipping. Furthermore, OVA subplots are disjoint from each other, so axis reordering does not affect the structure of the graph. Since variables with similar correlations to the ova variable are likely to be strongly correlated with each other, inter-subplot variable correlations are apparent even though line tracing is not possible from subplot to subplot. Comparing subplots in this fashion can reveal groups of correlated variables.

Scatter Plots Overlay—SPO has been shown to reveal correlation patterns that may otherwise be occluded by dense and/or overlapping PC lines. It displays 2D projections of spatial data and lets the viewer use point clustering to determine correspondences. As points can be preattentively distinguished from lines, each SP subplot does not obstruct the corresponding PC subplot, but rather, enhances correlations by providing a supplemental variable comparison. At best, SPO requires no more space than an existing PC plot when overlaid and at worst, the same amount of space as a PC plot over the same data (with the addition of one subplot for ova). For the same number of points and lines, point density remain lower than line density. This ensures that each SP subplot will have a lower visual density than each PC subplot when using the same number of points and lines.

7.2 Limitations

Luminance-Modulated PC—LMPC cannot be directly applied in combination with density-based extensions to PC because its modulation of individual lines is incompatible with the need to accurately sum the total contribution. However, it might be used after the density has been computed to add back in detail within the high-density regions.

One vs All Plots—OVAPC requires twice the display space compared to traditional PC, so cannot be used when the number of variables is so large that its display takes up the entire screen space. Interactive control over display mode can mitigate this by enabling switching between traditional and OVA display modes.

Scatter Plots Overlay—For data sets with very large numbers of points (far beyond the number at which overplotting typically occurs in PC), SP can itself overplot the parallel coordinates, hiding them from view. Reducing the number of points by sampling can introduce small sampling artifacts. These artifacts could be mitigated by cycling the sample over time in an animated display.

8. FUTURE WORK

We plan to implement these techniques within the open-source VTK visualization framework and in the ParaView open-source visualization tool to provide wide accessibility to the scientists. We are currently working to apply these techniques to data sets from our biomedical and microscopy collaborators.

Acknowledgments

The authors wish to thank the collaborators on the MADAI project who have contributed time, support, and feedback on the visualizations. This work was supported by NIH 5-P41-EB002025, NSF's Cyber-Enabled Discovery and Innovation Program through grant no. 0941373 and Sandia National Laboratories contract number 979573.

References

1. Inselberg A. The plane with parallel coordinates. *The Visual Computer*. 1985; 1(4):69–91.

2. Heinrich J, Weiskopf D. Continuous parallel coordinates. *IEEE Transactions of Visualization and Computer Graphics*. 2009; 15(6):1531–1538.
3. Miller, JJ.; Wegman, EJ. Construction of line densities for parallel coordinate plots. Springer-Verlag New York, Inc; New York, NY, USA: 1991. p. 107-123.
4. Fua YH, Ward MO, Rundensteiner ED. Hierarchical parallel coordinates for exploration of large datasets. *Proceedings of Visualization*. 1999; 1999:43–50.
5. Novotný M, Hauser H. Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*. 2006; 12(5):893–900. [PubMed: 17080814]
6. Zhou, H.; Cui, W.; Qu, H.; Wu, Y.; Yuan, X.; Zhou, W. Splatting the lines in parallel coordinates. *Eurographics IEEE-VGTC Symposium on Visualization*; 2009. 8 pages
7. van Wijk, J. ACM SIGGRAPH '91. Las Vegas, Nevada: ACM SIGGRAPH; 1991. Spot noise, texture synthesis for data visualization; p. 309-318.
8. Cabral, B.; Leedom, L. Imaging vector fields using line integral convolution. *ACM SIGGRAPH '93*; Anaheim, California. 1993. p. 263-270.
9. Urness, T., et al. Effectively visualizing multi-valued flow data using color and texture. *Proceedings of IEEE Visualization*; 2003; Seattle. 2003. p. 115-122.
10. Johansson, J.; Ljung, P.; Jern, M.; Cooper, M. Revealing structure within clustered parallel coordinates displays. *Information Visualization*, 2005. INFOVIS 2005. IEEE Symposium on; oct. 2005; p. 125-132.
11. Steed CA, Fitzpatrick PJ, Swan JE, Jankun-Kelly TJ. Tropical cyclone trend analysis using enhanced parallel coordinates and statistical analytics. *Cartography and Geographic Information Science*. 2009; 36(3):251–265.
12. Ankerst, M.; Berchtold, S.; Keim, DA. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. *Proceedings of the IEEE InfoVis98*; 1998. pages 5260
13. Viau, C.; McGuffin, M.; Chiricota, Y.; Jurisica, I. The flowvizmenu and parallel scatterplot matrix: Hybrid multidimensional visualizations for network exploration. *Visualization and Computer Graphics*, *IEEE Transactions on*; nov.–dec. 2010; p. 1100-1108.
14. Cleveland, WC.; McGill, ME. *Dynamic graphics for statistics*. CRC Press, Inc; Boca Raton, FL, USA: 1988.
15. Li J, Martens JB, van Wijk JJ. Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization*. 2010; 9(1):13–30.
16. Yuan X, Guo P, Zhou H, Qu H. Scattering points in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*. 2009; 15(6):1001–1008. [PubMed: 19834165]
17. Wong PC, Bergeron RD. Multivariate visualization using metric scaling. *Proceedings of the IEEE Visualization*. 1997; 97:111–118.
18. Feng D, Lee Y, Kwock L, Taylor R. Matching visual saliency to confidence in plots of uncertain data. *TVCG*. 2010; 16(6):980–989. [PubMed: 20975135]
19. Petersen H, Steinheimer J, Bureau G, Bleicher M, Stöcker H. Fully integrated transport approach to heavy ion reactions with an intermediate hydrodynamic stage. *Phys Rev C*. Oct.2008 78:044901.
20. Vlachos, P.; StaLib, MM. data, software and news from the statistics community. [http://lib.stat.cmu.edu/\(2002\)](http://lib.stat.cmu.edu/(2002))

Figure 1. Comparison of Traditional Parallel Coordinates and Luminance Modulated Parallel Coordinates
 using T-REX (weather observations) dataset. (a) Traditional parallel coordinates with curved lines. (b) Luminance modulated parallel coordinates with curved lines. The luminance variable in (b) is from the h_2O_vap column. Overplotting occurs in the majority of subplots in (a), but once luminance modulation is applied in (b), local line details are restored.

Figure 2. Comparison of Traditional Parallel Coordinates to One-vs.-all Parallel Coordinates
 (a) Traditional parallel coordinates with curved lines. (b) One-vs.-all parallel coordinates with $\mu(1)$ selected as the one vs. all variable. A strong positive correlation between $n(1)$ and $\mu(1)$ is immediately apparent in (b), a correlation that is difficult to see in (a) even though $n(1)$ and $\mu(1)$ are separated by only two plots. A negative correspondence between T and $\mu(1)$ is also displayed in (b), but not obvious in (a). y_L vs. $\mu(1)$ and η vs. $\mu(1)$ in (b) are nearly identical plots, implying y_L and η are positively correlated. The plot y_L vs. η in (a) confirms this.

Figure 3. Comparison of One-vs.-all Parallel Coordinates and One-vs.-all Scatter Plots Overlay
 (a) One-vs.-all parallel coordinates. (b) One-vs.-all scatter plots. (c) Both overlaid. When correlating v_y and y using only OVAPC, the frequency of line crossovers makes it difficult to be sure there is a strong positive correlation, but when using OVASP it is immediately clear that such a correlation exists by the presence of a diagonal line with positive slope. Furthermore, OVASP reveals an elliptical correlation for x vs. y , one that is not obvious from OVAPC. However, when comparing $dsig_y$ and y , OVAPC shows two distinct line clusters, with the mid-range values of $dsig_y$ distributed evenly across y and the high and low values corresponding directly to y . These clusters are less apparent with OVASP.

Figure 4. Comparison of Two Heavy-Ion Collision Simulations

(a) LMOVAPC with SPO showing timestep 16 of the glauher b0 dataset with energy-density (e) selected as the OVA and LM variable. (b) LMOVAPC with SPO showing timestep 16 of the pcasc b0 dataset with e selected as the OVA and LM variable. Glauber b0 and pcasc b0 are the result of two difference simulations for a gold-gold collision at $E_{cm} = 200$ AGeV. Although based on the same theoretical model, the simulations display differences, primarily in their energy-density-*vs.*-pressure distributions.

Figure 5. Visualization for T-REX (weather observations) dataset

(a) SPO on top of LMPC. (b) SPO on top of luminance modulated OVAPC. The T-REX dataset contains twenty-two variables with values derived from 1-min averages of 10 HZ samples over a 57 day period. The data in (a) shows 17 variables, most chosen for their strong correlations to other variables. The ordering above has been chosen to display the majority of strong variable correlations, with clustered groups of correlated variables as can be seen by consecutive positively correlated plots. In (b), an luminance modulated OVAPC plot with SPO displays the first principal component derived for this dataset (pca_1) vs. 8 other variables. Many of the variables in (b) correspond to groups in (a), revealing correspondences between pca_1 and all other variables in each group.

Figure 6. OVASP on top of luminance modulated OVAPC
 using the 392 point "cars" dataset. (a) Horsepower selected as OVA variable. (b) Weight
 selected as OVA variable. Five of eight variables displayed.