Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 9 (2012) 877 – 886

Procedia Computer Science

International Conference on Computational Science, ICCS 2012

Practical Application of Parallel Coordinates for Climate Model Analysis

Chad A. Steed[a,*], Galen Shipman[a], Peter Thornton[a], Daniel Ricciuto[a], David Erickson[a], Marcia Branstetter[a]

[a]*Oak Ridge National Laboratory, One Bethel Valley Rd., Oak Ridge, TN 37830*

Abstract

The determination of relationships between climate variables and the identification of the most significant associations between them in various geographic regions is an important aspect of climate model evaluation. The EDEN visual analytics toolkit has been developed to aid such analysis by facilitating the assessment of multiple variables with respect to the amount of variability that can be attributed to specific other variables. EDEN harnesses the parallel coordinates visualization technique and is augmented with graphical indicators of key descriptive statistics. A case study is presented in which the focus is on the Harvard Forest site (42.5378N Lat, 72.1715W Lon) and the Community Land Model Version 4 (CLM4) is evaluated. It is shown that model variables such as land water runoff are more sensitive to a particular set of environmental variables than a suite of other inputs in the 88 variable analysis conducted. The approach presented here allows climate scientists to focus on the most important variables in the model evaluations.

Keywords: information visualization, climate model, sensitivity analysis, visual analytics
2000 MSC: 62P12, 68U05

1. Introduction

One of the most challenging tasks in exploring multivariate data is the identification and quantification of associations between a set of interrelated variables. For climate model data, this task is even more daunting due to the complexity and number of variables that are considered. Often times, the scientist has some understanding of expected relationships but unexpected discoveries are nearly impossible with conventional toolkits. In spite of the difficulty, these tasks are paramount to uncertainty quantification and more general exploratory analysis of climate model data. In general, the goal of these activities is to improve the accuracy and understanding of climate models which translate into better comprehension of climate change and prediction.

Although the volume and sophistication of climate modeling and related high performance computing have progressed rapidly in recent years [1, 2, 3], the methods to display and explore the information in useful and meaningful ways have not kept pace. For example, researchers typically rely on scatter plots and histograms to analyze multiple

*Corresponding author
Email addresses: csteed@acm.org (Chad A. Steed), gshipman@ornl.gov (Galen Shipman), thorntonpe@ornl.gov (Peter Thornton), ricciutodm@ornl.gov (Daniel Ricciuto), ericksondj@ornl.gov (David Erickson), branstetterm@ornl.gov (Marcia Branstetter)
URL: http://cda.ornl.gov/steed/ (Chad A. Steed)

doi:10.1016/j.procs.2012.04.094

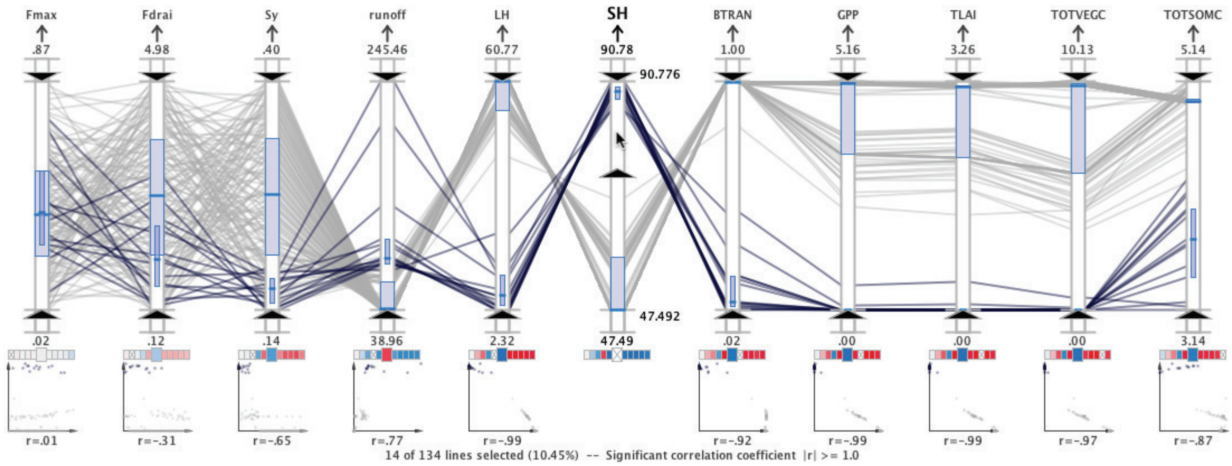


Figure 1: The parallel coordinates panel in EDEN provides an interactive canvas that guides the scientist to significant associations in multivariate data. In this case, the Harvard Forest ensemble data set of CLM4 parameters (3 leftmost axes) and outputs (8 rightmost axes) is shown after loading the data and using the query sliders on the *SH* axis to highlight 14 outliers. Each polyline in the plot represents an individual CLM4 model run. The plot shows that outliers are restricted to the upper range of the *runoff* variable, the lower range of vegetation productivity variables (*BTRAN*, *GPP*, *TLAI*, and *TOTVEGC*), and the lower range of the *Sy* parameter.

variable relationships. However, these graphics require the use of multiple separate plots since scatter plots are limited to at most three variables. This approach is also problematic due to human perceptual issues, especially when searching for combinations of conditions. At the same time, a number of effective multivariate visual analysis techniques have been introduced in the information visualization literature. The fact that relatively few of these approaches have been brought to bear on scientific problems where they are needed reflects a growing gap between viable information visualization techniques and real-world scientific analysis. To bridge this gap, experts from both fields must work closely together to create practical tools for today's most pressing problems.

It is in this spirit that a team consisting of climate model scientists and an information visualization researcher at Oak Ridge National Laboratory (ORNL) have collaborated to improve the visual evaluation of the Community Land Model version 4 (CLM4) [2]. A new system, called the Ensemble Data analysis ENvironment (EDEN), has been developed through this effort to provide an efficient framework to discover knowledge in complex CLM4 data sets. As shown in Fig. 1, EDEN is designed to harness the high bandwidth human visual channel with interactive parallel coordinates that guide the scientist to potentially significant associations in the data. Through a case study (see Section 5) with climate model scientists who are also co-authors of this work, the notion that interactive parallel coordinates can provide a more effective environment for climate model analysis is corroborated. Furthermore, the research reported here addresses an important point brought out in the NIH/NSF Visualization Challenges Report [4], which encourages visualization researchers to "collaborate closely with domain experts who have driving tasks in data-rich fields to produce tools and techniques that solve clear real-world needs." The tool in the current work is EDEN, the techniques are interactive parallel coordinates and statistical analytics, and the real-world need is the understanding of climate model parameters and outputs.

2. Related Work

Although there have been many approaches for the visual analysis of multivariate data shared in the literature [5], the techniques utilized for climate data analysis are generally still constrained to non-interactive, basic graphics using methods developed over a decade ago; and it is questionable whether these methods can cope with the complexity of today's data. For example, scientists often rely on simple scatter plots and histograms which require several separate plots or layered plots to study multiple attributes in a data set. However, the use of separate plots is not an ideal approach in this type of analysis due to perceptual issues such as the extremely limited memory for information that can be gained from one glance to the next [6]. These issues are illustrated through the perceptual phenomenon of *change blindness* [7] and they are exacerbated when searching for combinations of conditions.

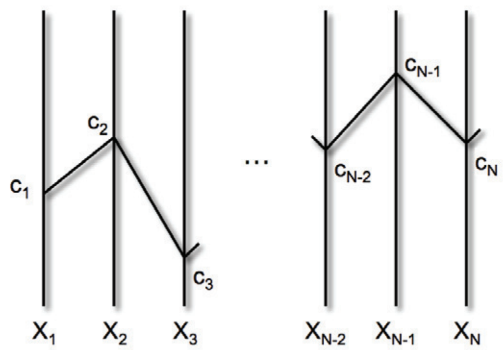


Figure 2: The polyline in parallel coordinates maps the N-dimensional data tuple C with coordinates $(c_1, c_2, \dots, c_N)$ with points on N parallel axes which are joined with a polyline whose N vertices are on the X_i-axis for $i = 1, \dots, N$.

Scientists often rely on the scatter plot matrix (SPLOM), which presents multiple adjacent scatter plots for all variable comparisons in a single display [8]; but the SPLOM requires a large amount of screen space and forming multivariate associations is still challenging. Wilkinson et al. [9] used statistical measures for organizing SPLOMs to guide the viewer through an exploratory analysis of high-dimensional data sets. Although the organization methods improve the analysis, the previously mentioned perceptual issues with SPLOMs remain to some degree. Another approach is to use layered plots, which condense the information into a single display; but there are significant issues with these techniques due to layer occlusion and interference as demonstrated by Healey et al. [6].

Parallel coordinates is a popular multivariate visualization technique that can alleviate said issues and it is the basis for the interactive display panel in the EDEN framework. The parallel coordinates technique was initially popularized by Inselberg [10] as a novel approach for representing hyper-dimensional geometries, and later demonstrated in the direct analysis of multivariate relationships in data by Wegman [11]. In general, the technique yields a compact two-dimensional representation of even large multidimensional data sets by representing the N-dimensional data tuple C with coordinates $(c_1, c_2, \dots, c_N)$ by points on N parallel axes which are joined with a polyline (see Fig. 2) [12]. In theory, the number of attributes that can be represented in parallel coordinates is only limited by the horizontal resolution of the display device. But in a practical sense, the axes that are immediately adjacent to one another yield the most obvious information about relationships between attributes. In order to analyze attributes that are separated by one or more axes, interactions and graphical indicators are required. To this end, several innovative parallel coordinates extensions that seek to improve interaction and cognition have been described in the visualization research literature. For example, Hauser et al. [13] described a histogram display, dynamic axis re-ordering, axis inversion, and some details-on-demand capabilities for parallel coordinates. In addition, Siirtola [14] presented a rich set of dynamic interaction techniques and Johansson et al. [15] described informative line shading schemes for parallel coordinates. Furthermore, several focus+context implementations for parallel coordinates have been introduced by Fua et al. [16], Artero et al. [17], and Novotńy and Hauser [18]. More recently, Qu et al. [19] described a method for integrating correlation computations into a parallel coordinates display. The EDEN visualization approach described in the remainder of the current work uses variants of these extensions to the classical parallel coordinates plot.

3. Data

The data analyzed in the current work has been generated using the CLM4 which represents the land component of the Community Climate System Model version 4 (CCSM4) [1]. Both global model output data from a single simulation and single point data in an ensemble of simulations have been generated for analysis of model output sensitivity to input parameter variation. In the current work, the focus is on analysis of the ensemble data sets. The monthly CLM4 outputs were post-processed into comma separated value (CSV) files with a selected number of model parameters (inputs variables that are used for initialization) and a number of outputs. Each column in the file represents a parameter or model output and each row represents a particular model run. Therefore, a model run is a

Table 1: Variable descriptions for the Harvard Forest site ensemble of CLM4 simulations averaged for the month of May over 10 years (1995–2004). An analysis of this data set using EDEN is presented in Section 5.

CLM4 Parameter Variables

Parameter	Units	Description
Fmax	*none*	maximum fractional saturated area
Fdrai	$m\ s^{-1}$	decay factor for subsurface runoff with depth
Sy	*none*	average specific yield

CLM4 Output Variables

Output	Units	Description
runoff	$mm\ s^{-1}$	total runoff
LH	$W\ m^{-2}$	latent heat flux
SH	$W\ m^{-2}$	sensible heat flux
BTRAN	*none*	transpiration scaling factor
GPP	$gC\ m^{-2}\ s^{-1}$	gross primary productivity
TLAI	$m^2\ m^{-2}$	total leaf area index
TOTVEGC	$kg\ C\ m^{-2}$	total vegetation carbon
TOTSOMC	$kg\ C\ m^{-2}$	total soil organic matter carbon

data tuple that contains both the parameters and outputs which can be used to evaluate the sensitivity of the model outputs to changes in parameters when simultaneously analyzed with other model runs. In the parallel coordinates plot, the data tuple is graphically displayed as a polyline. In Section 5, a case study is presented that uses a particular ensemble of CLM4 simulations for a single site (42.5378N Lat, 72.1715W Lon) averaged for the month of May over 10 years (1995–2004). This data set contains 134 model runs performed at the model location of the Harvard Forest eddy covariance flux tower, a temperate deciduous forest in Massachusetts, USA. In each model simulation, 3 model parameters were varied: *Fmax*, *Fdrai*, and *Sy*. In addition to the input parameters, 8 output variables are analyzed: *LH*, *SH*, *BTRAN*, *GPP*, *TLAI*, *TOTVEGC*, and *TOTSOMC*. A total of 11 variables are visualized in the case study (see Table 1).

4. Interactive Parallel Coordinates

The primary display in EDEN provides an interactive parallel coordinates visualization. Parallel coordinates is an effective method for the CLM4 data because it facilitates the display of large multivariate data sets. In fact, the number of variables that can be displayed is only limited by the horizontal resolution of the display device. A common set of parallel coordinates capabilities such as movable axes, details on demand, and axis inversion are available in EDEN. The parallel coordinates plot has been extended with a number of capabilities that facilitate exploratory data analysis and guide the scientist to the most significant relationships in the data. In the following sections, these features are summarized to provide context for the case study in section 5, but the reader is encouraged to explore the author's prior publications for a more detailed explanation of these extensions [20, 21, 22].

4.1. Descriptive Statistical Indicators

In EDEN, the parallel coordinates axes are enhanced with visual cues that guide the scientist's exploration of the information space. This approach is akin to the concept of the *scented widget* described by Willett et al. [23]. Scented widgets are graphical user interface components that are augmented with an embedded visualization to enable efficient navigation in the information space of the data items. In EDEN, each axis in the parallel coordinate display is scented with graphical encodings of several key descriptive statistics to summarize the variable distributions (see Fig. 3). The median, interquantile range (IQR), and frequency information are calculated for the data within the focus area of each axis. The user can switch between mean and standard deviation range as an alternative to the median and IQR display.

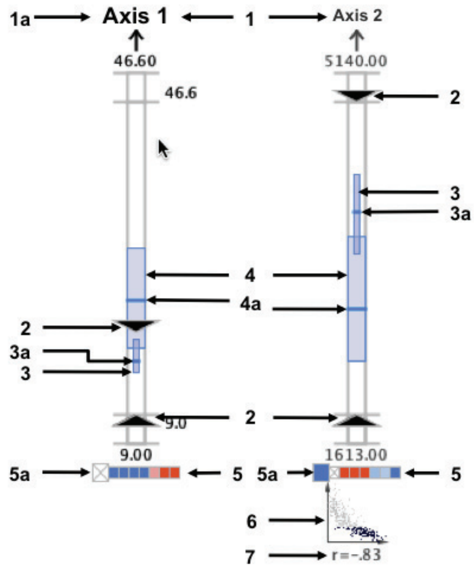


Figure 3: The axes in the parallel coordinates display are augmented with graphical indicators of key descriptive statistical quantities, correlation measures, and query range sliders. In this annotated figure, the numbered call-outs highlight specific features of the axis widgets which are described in detail in the remainder of this paper.

This variability information is graphically encoded in the boxes on each axis interior. The wide boxes (see 4 in Fig. 3) represent the statistics for all the axis samples, while the more narrow boxes (see 3 in Fig. 3) capture the samples that are currently queried with the query sliders (see 2 in Fig. 3). The thick horizontal lines (see 3a and 4a in Fig. 3) that divide the variability boxes vertically represent either the median or the mean value.

4.2. Dynamic Visual Queries

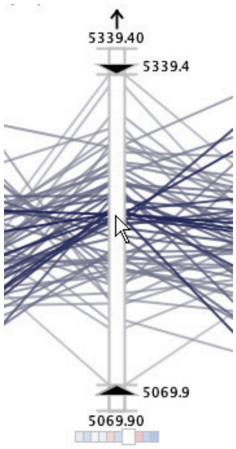
The concept of dynamic visual queries focuses on interactive displays that facilitate rapid, incremental, and reversible actions, selection by mouse gestures (rather than typing), and immediate and continuous display of results [24]. In EDEN, dynamic visual queries are realized through interactive axis scaling and query sliders

4.2.1. Dynamic Axis Scaling

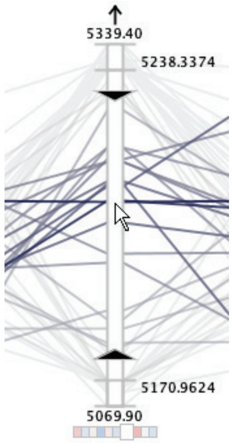
Each parallel coordinate axis can be interactively scaled to facilitate zooming into specific regions of interest. As shown in Fig. 4, the axis is partitioned into three sections delineated by horizontal tick marks: the central focus area and the top and bottom context areas. When the mouse hovers over the focus area, an upward scroll gesture expands the display of the focus area outward and pushes lines outside the new upper and lower limits into the context areas. A downward scroll gesture causes the opposite effect: focus area compression. Alternatively, the user may elect to use the scrolling gesture over either of the two context areas independently to alter either the maximum or minimum limits. This technique is helpful to the scientists when the parallel coordinate plot contains a densely packed group of lines and to remove outliers from the display.

4.2.2. Query Sliders

On each parallel coordinate axis, a pair of sliders (see 2 in Fig. 3) are displayed that can be dragged with the mouse to rapidly filter the data. By dragging either the top or bottom slider, the scientist can adjust the maximum and minimum limits of the current query, respectfully. For example, in Fig. 1, the query sliders are set to highlight the lines crossing the upper range of the *SH* axis. When sliders on multiple axes are set, the scientist is given the ability to construct Boolean AND queries. That is, lines that lie between the sliders of each axis are shaded more prominently. Lines that are not queried are still displayed in the plot for context; but these lines are shaded with less contrast to the background color of the panel. Through these sliders, mouse gestures are the primary means to select lines of interest as opposed to exiting to external packages or command line scripts.



(a) Before scaling



(b) After scaling

Figure 4: Image sequence illustrating dynamic axis scaling capability before (a) and after (b) axis scaling has been performed. In this example, scaling is performed by an upward mouse scroll gesture in the focus area of the axis which moves the values of the upper and lower limits closer in value, zooming into the central axis region.

4.3. Correlation Indicators

EDEN also enables a visual form of correlation mining to judge the strength of relationships between pairs of variables. As the user interacts with the system, the correlation information is updated in real-time and used to augment the parallel coordinates display. For each possible pairing of axes, the system automatically calculates the Pearson product-moment correlation coefficient, r. This computation yields a correlation matrix where each i, j element is equal to the r value between the i and j variables. The rows from this matrix are displayed graphically beneath each axis as a series of color-coded blocks (see 5 in Fig. 3,). In each block, the color encodes r between the axis directly above it and the axis that corresponds to its position in the set of blocks. When the mouse hovers over an axis, the axis is highlighted (see 1a in Fig. 3) and the correlation coefficient blocks corresponding to it below the other axes are enlarged (see 5a in Fig. 3). The correlation indicator blocks are shaded blue for negative correlations and red for positive. The stronger the correlation, the more saturated the color so that the stronger correlations are more prominent. The r value of an axis with itself is always equal to one and the corresponding indicator block is shaded white.

In addition to the correlation indicator blocks, EDEN displays small scatter plots below the correlation indicators for each axis when an axis is highlighted (see 6 in Fig. 3). The scatter plots are created by plotting the points with the highlighted variable as the y axis and the variable directly above the scatter plot as the x axis. Each scatter plot also shows the numerical r value associated with the pair of axes below the scatter plot (see 7 in Fig. 3). The scatter plots provide a visual means to confirm the type of correlation (positive or negative) and the strength. The type of correlation is also visually detectable in the line configuration of the parallel coordinates plot. Polylines that cross in an 'X' pattern (see the line crossings between *BTRAN* and *SH* in Fig. 5) are characteristic of a negative correlation while lines that appear to be more parallel indicate a positive correlation (see the line crossings between *GPP* and *TLAI* in Fig. 5). Unlike the other correlation indicators, the scatter plot is useful for discovering nonlinear relationships between variables. For example, a nonlinear relationship can be observed in a scatter plot even if r is zero.

5. Case Study: CLM4 Sensitivity Analysis

To illustrate the effectiveness of EDEN, the analysis of an ensemble of CLM4 simulations with 11 variables at the location of the Harvard Forest eddy covariance flux tower (see Section 3) is described in this section. In Fig. 5, the parallel coordinates panel is shown after loading the data set. From this initial plot, several interesting features in the data are revealed. First of all, the box plots on the first 3 axes (*Fmax*, *Fdrai*, and *Sy*) reveal nearly normal distributions, reflecting the sampling strategy used to generate the model input parameters. This sampling strategy

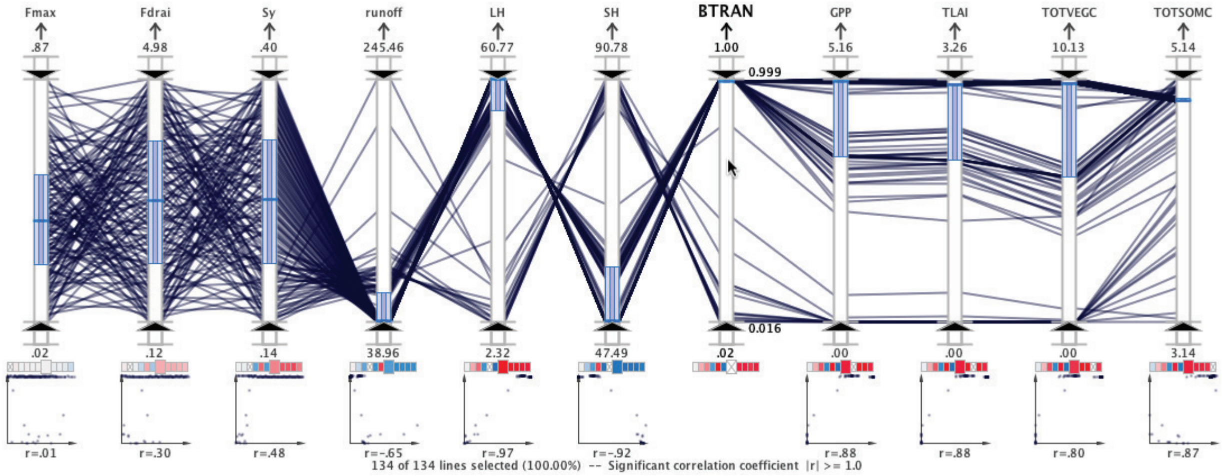
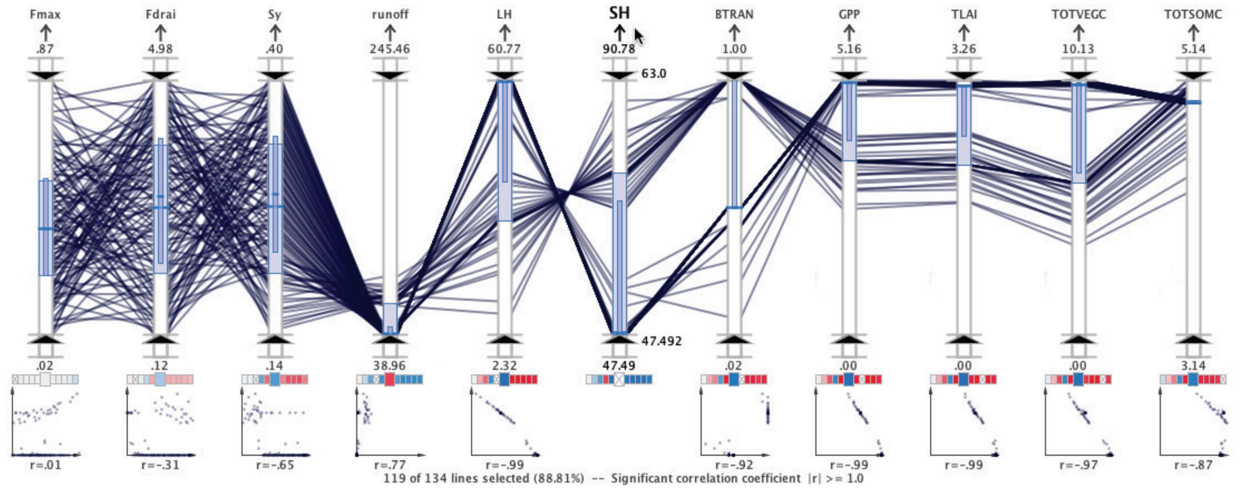


Figure 5: The initial view of the May Harvard CLM4 ensemble data set is shown after its loading into EDEN. With the mouse over the *BTRAN* axis, the view reveals very strong correlations with the other 6 output variables and a moderately strong correlation with the *runoff* output variable.

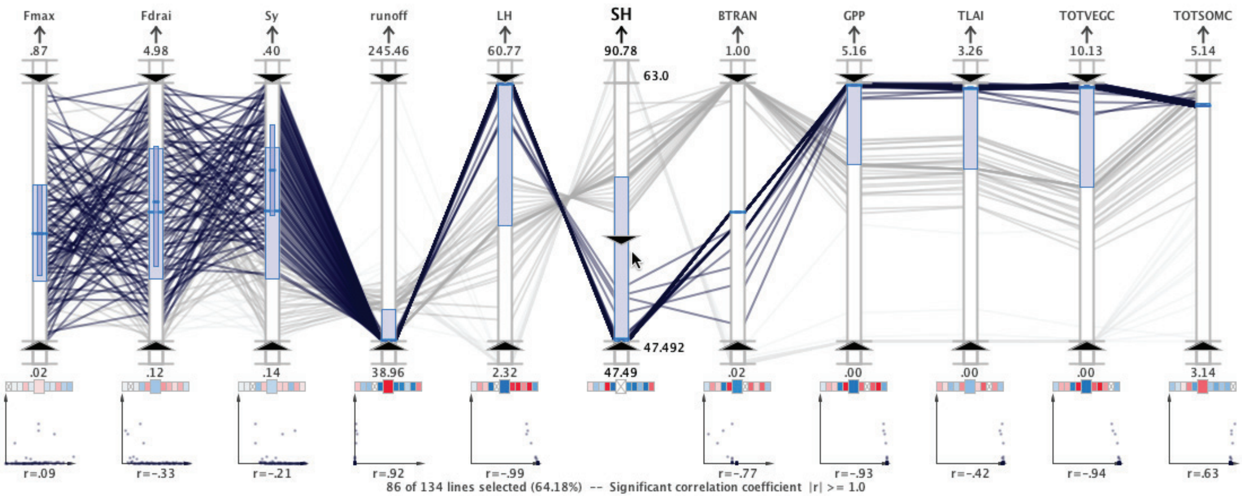
has been selected to represent the existing knowledge about values of these parameters and to maximize coverage of the samples over the parameter space. The box plots on the other axes suggest skewed distributions toward either the minimum (e.g. *runoff* and *SH*) or maximum values. In Fig. 5, the mouse is over the *BTRAN* axis revealing strong correlations (notice the highly saturated correlation indicator blocks enlarged beneath each axis) with the other 6 output variables (*LH*, *SH*, *GPP*, *TLAI*, *TOTVEGC*, and *TOTSOMC*). These strong correlations reflect the interrelated nature of the model outputs. For example, low values of *runoff* and *SH* as described above occur in simulations in which there are low amounts of foliage (low *TLAI* and *TOTVEGC*). Low foliage means low transpiration (a component of *LH*), leaving more soil water available for runoff and more net energy available for sensible heat (*SH*). These strong feedbacks between vegetation and soil hydrology are one example of many in the CLM4 where potentially unexpected relationships exist between indirectly connected model parameters and outputs.

Although Figure 5 is useful for representing broad relationships among the variables, outliers exercise too much influence on the display making it difficult to form more detail judgments on the correlation and distribution patterns. That is, the outliers expand the range of values for several axes which results in a mass of lines at the bottom or top of the parallel coordinate axes (see *runoff* and *BTRAN* in Fig. 5). In Fig. 1, the query sliders on the *SH* axis have been adjusted to highlight a group of 14 outliers. This plot also shows that these lines are outliers for the other 7 output variables. The outliers appear to be related to the upper range of values in the *runoff* axis. Furthermore, the outliers are contained in the lower range of the *Sy* axis, but unlike the *runoff* axis, the lower range of *Sy* also contains a number of polylines that are not outliers (notice the presence of a number of light gray non-query lines behind the queried lines on the *Sy* axis). The outlier lines are rather widely dispersed on the *Fmax* and *Fdrai* axes suggesting weak correlations. This analysis strongly suggests that *Sy* is controlling most of the variability in the output variables and that the *Fmax* and *Fdrai* variables are only minor contributions.

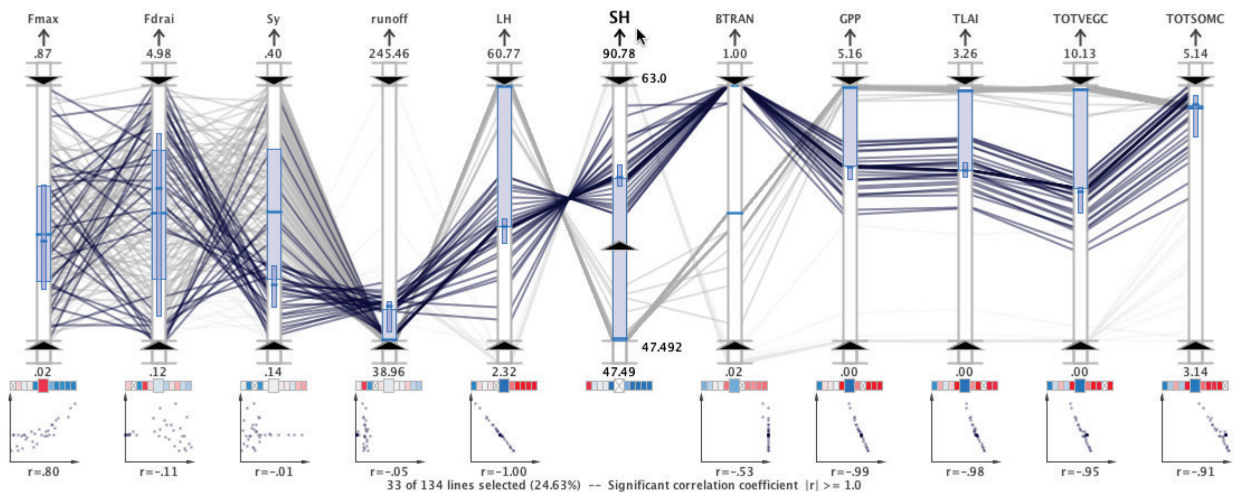
To gain more insight into the data set, the dynamic axis scaling capability (see Sec. 4.2.1) is used to push these outliers into the context area of the display. In Fig. 6(a), the focus area limits have been adjusted for the *LH*, *SH*, and *BTRAN* axes. The influence of the outliers on the display is reduced so that more structure can be observed in both the parallel coordinates plot as well as the scatterplots. It is important to note that the relationships described above are preserved when said outliers are removed. The *SH* axis is highlighted in this plot, showing that the correlation with *LH* is even stronger ($r = -0.99$) now that the outliers have been removed. In addition, the configurations of the *SH* and *LH* axes in Fig 6(a) suggest bimodalities. In Fig. 6(b) the group of lines from the lower range of the *SH* axis are selected using the query sliders. From this plot, it is clear that simulations with low *SH* correspond to high values for *LH* and low values for *BTRAN*. Conversely, the lines crossing the upper range of *SH* have a negative correlation pattern with *LH* (notice the 'X' shaped crossing between the axes). This pattern is shown in Fig. 6(c), and the correlation between *LH* and *SH* for the queried lines is a perfect negative correlation. This bimodality appears to



(a) After axis scaling on the *LH*, *SH*, and *BTRAN* axes



(b) Highlighting the cluster of lines in the lower range of *SH*



(c) Highlighting the cluster of lines in the upper range of *SH*

Figure 6: Dynamic axis scaling is used on the *LH*, *SH*, and *BTRAN* axes to remove the outliers and reveal more structure in the display (a). Two clusters of lines are observed for the lower range (b) and the upper range (c) of *SH*.

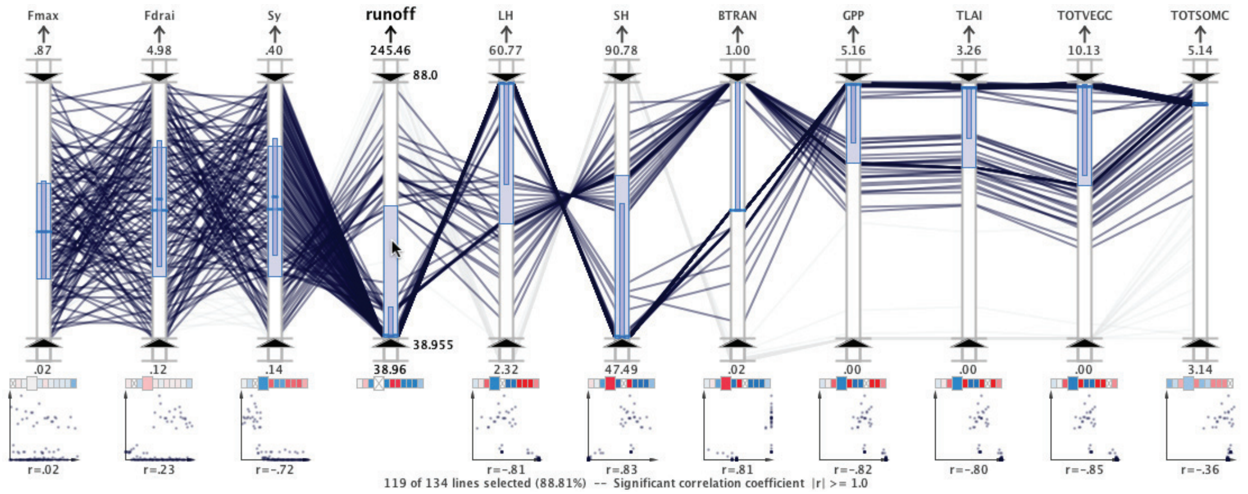


Figure 7: The two clusters of lines are shown for the *runoff* axis after scaling its focus area to remove additional outliers at the upper portion of the display.

be driven by a threshold value of *Sy* near 0.2, above which *runoff* is near the minimum value and below when there is increased *runoff* and also increased variability of this and other output variables.

In Fig. 7, the mouse is used to highlight the *runoff* output variable. The correlation indicators below the axes reveal strong correlations with all but the first two parameters (*Fmax* and *Fdrai*) and the last output variable (*TOTSOMC*). For this plot, the focus range has been dynamically scaled for the *runoff* axis to push a number of upper outliers into the context area. In this plot, two distinct groups of lines, one for the upper range of and one for the middle to lower range, are apparent. These two groups correspond to the groups highlighted in Figs. 6(c) and 6(b). The lines in the upper range of *runoff*, correspond to the lower range of *LH*, the upper range of *SH* and *BTRAN* and the lower ranges of *GPP*, *TLAI*, and *TOTVEGC*.

6. Conclusions

CLM4, the land-surface model used in this study, contains over 100 input parameters and over 300 output variables that vary spatially and temporally. Fully coupled climate models contain even more inputs and outputs. Discovering new relationships among these variables using traditional scatterplots or other graphical methods is not practical. In this work, the EDEN visual analytics toolkit is described for determining the significant relationships among many climate model variables in a practical climate model evaluation. During the course of this evaluation, it was noted that the best use by climate scientists occurs when they are involved in the initial selection of the climate variables to be examined. The parallel coordinates approach in EDEN facilitates the assessment of multiple variables simultaneously in a single display to ascertain the amount of variability that can be attributed to specific other variables. One limitation noted by the climate scientists is that domain expertise is necessary during the early stages of the analysis. When this input is not available, large amounts of model/analysis output was either not useful or difficult to interpret. This finding suggests a close collaboration of visualization designers and climate domain scientists throughout the analysis process. This limitation could also be resolved by combining EDEN with other data-mining approaches such that only potentially meaningful relationships are highlighted visually.

In the case study presented here, it is observed that model variables such as land model water runoff and other hydrologic variables are quite sensitive to a particular set of other environmental drivers and model parameters. In an earlier analysis of a 1,000 member ensemble covering 88 variables the relationships were less clear. This issue was likely a result of the insufficient sampling over the high dimensional space (81 parameters) and the difficulty in visualizing such a large number of variables. Focused analysis covering subsets of model parameters and outputs are necessary. Of particular interest is the variability in the surface latent and sensible energy heat fluxes that in turn influence atmospheric thermodynamics and climate. Initial inclusion in the climate model evaluation was required when breaking down critical variable combinations in practical applications to state-of-the-art climate model simulations.

Acknowledgments

This research is sponsored by the Climate Science for a Sustainable Energy Future (CSSEF) project that is funded by the U.S. Department of Energy, Office of Science. The authors wish to thank Zhangshuan Hou (Pacific Northwest National Laboratory) for the generation of the Harvard Forest site CLM4 ensemble data set. This paper was prepared by the Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, Tennessee 37831-6285, managed by UT–Battelle, LLC, for the U.S. Department of Energy, under contract DE-AC05-00OR22725. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Bibliography

[1] P. R. Gent, G. Danabasoglu, L. J. Donner, M. M. Holland, E. C. Hunke, S. R. Jayne, D. M. Lawrence, R. B. Neale, P. J. Rasch, M. Vertenstein, P. H. Worley, Zong-Liang Yang, M. Zhang, The community climate system model version 4, Journal of Climate 24 (19) (2011) 4973–4991.

[2] D. M. Lawrence, K. W. Oleson, M. G. Flanner, P. E. Thornton, S. C. Swenson, P. J. Lawrence, Zong-Liang Yang, S. Levis, K. Sakaguchi, G. B. Bonan, A. G. Slater, Parameterization improvements and functional and structural advances in version 4 of the community land model, Journal of Advances in Modeling Earth Systems 3 (M03001) (2011) 27 pp.

[3] J. T. Overpeck, G. A. Meehl, S. Bony, D. R. Easterling, Climate data challenges in the 21st century, Science 331 (6018) (2011) 700–702.

[4] C. Johnson, R. Moorhead, T. Munzner, H. Pfister, P. Rheingans, T. S. Yoo (Eds.), NIH/NSF Visualization Reserach Challenges, IEEE Press, 2006, http://tab.computer.org/vgtc/vrc/index.html (current 1 Jul. 2009).

[5] S. K. Card, J. D. Mackinlay, B. Shneiderman, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann, 1999.

[6] C. G. Healey, L. Tateosian, J. T. Enns, M. Remple, Perceptually-based brush strokes for nonphotorealistic visualization, ACM Transactions on Graphics 23 (1) (2004) 64–96.

[7] R. A. Rensink, Change detection, Annual Review of Psychology 53 (2002) 245–577.

[8] P. C. Wong, R. D. Bergeron, 30 years of multidimensional multivariate visualization, in: Scientific Visualization - Overviews, Methodologies, and Techniques, IEEE Computer Society Press, 1997, pp. 3–33.

[9] L. Wilkinson, A. Anand, R. Grossman, High-dimensional visual analytics: Interactive exploration guided by pairwise views of point distributions, IEEE Transactions on Visualization and Computer Graphics 12 (6) (2006) 1366–1372.

[10] A. Inselberg, The plane with parallel coordinates, The Visual Computer 1 (4) (1985) 69–91.

[11] E. J. Wegman, Hyperdimensional data analysis using parallel coordinates, Journal of the American Statistical Association 85 (411) (1990) 664–675.

[12] A. Inselberg, Parallel coordinates: Interactive visualization for high dimensions, in: E. Zudilova-Seinstra, T. Adriaansen, R. Liere (Eds.), Trends in Interactive Visualization, Springer-Verlag, London, UK, 2009, pp. 49–78.

[13] H. Hauser, F. Ledermann, H. Doleisch, Angular brushing of extended parallel coordinates, in: Proceedings of IEEE Symposium on Information Visualization, 2002, pp. 127–130.

[14] H. Siirtola, Direct manipulation of parallel coordinates, in: Proceedings of the International Conference on Information Visualisation, IEEE Computer Society, London, England, 2000, pp. 373–378.

[15] J. Johansson, P. Ljung, M. Jern, M. Cooper, Revealing structure within clustered parallel coordinates displays, in: IEEE Symposium on Information Visualization, 2005, pp. 125–132.

[16] Y.-H. Fua, M. O. Ward, E. A. Rundensteiner, Hierarchical parallel coordinates for exploration of large datasets, in: Proceedings of IEEE Visualization, 1999, pp. 43–50.

[17] A. O. Artero, M. C. F. de Oliveira, H. Levkowitz, Uncovering clusters in crowded parallel coordinates visualization, in: IEEE Symposium on Information Visualization, 2004, pp. 81–88.

[18] M. Novotńy, H. Hauser, Outlier-preserving focus+context visualization in parallel coordinates, IEEE Transactions on Visualization and Computer Graphics 12 (5) (2006) 893–900.

[19] H. Qu, W. Chan, A. Xu, K. Chung, K. Lau, P. Guo, Visual analysis of the air pollution problem in Hong Kong, IEEE Transactions on Visualization and Computer Graphics 13 (6) (2007) 1408–1415.

[20] C. A. Steed, P. J. Fitzpatrick, T.J. Jankun-Kelly, A. N. Yancey, J. Edward Swan II, An interactive parallel coordinates technique applied to a tropical cyclone climate analysis, Computers & Geosciences 35 (7) (2009) 1529–1539.

[21] C. A. Steed, P. J. Fitzpatrick, J. Edward Swan II, T.J. Jankun-Kelly, Tropical cyclone trend analysis using enhanced parallel coordinates and statistical analytics, Cartography and Geographic Information Science 36 (3) (2009) 251–265.

[22] C. A. Steed, T.J. Jankun-Kelly, P. J. Fitzpatrick, Guided analysis of hurricane trends using statistical processes integrated with interactive parallel coordinates, in: IEEE Symposium on Visual Analytics Science and Technology, IEEE Computer Society, Atlantic City, NJ, 2009, pp. 19–26.

[23] W. Willett, J. Heer, M. Agrawala, Scented widgets: Improving navigation cues with embedded visualizations, IEEE Transactions on Visualization and Computer Graphics 13 (6) (2007) 1129–1136.

[24] C. Ahlberg, B. Shneiderman, Visual information seeking: Tight coupling of dynamic query filters with starfield displays, in: Proceedings of Human Factors in Computing Systems, Association for Computing Machinery, Boston, MA, 1994, pp. 313–317, 479–480.