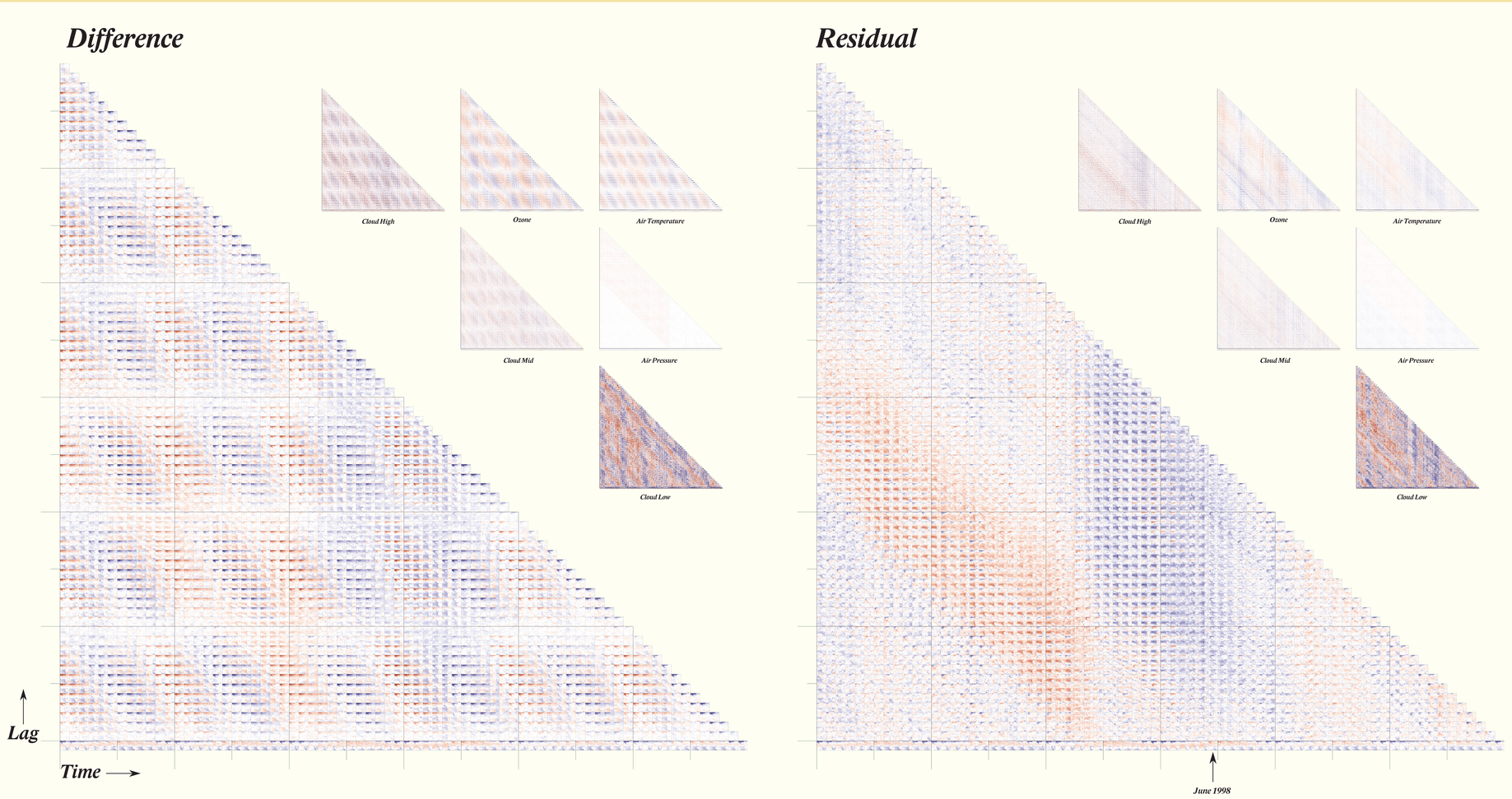
Multiple lagged differences in spatial time series

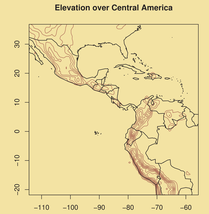
Jeffrey Horner, Systems Analyst, BS, Rafe Donahue, Research Associate Professor, PhD

Department of Biostatistics, Vanderbilt University Medical Center



Introduction

How can one graphically summarize 7 spatial time series data sets, each one spanning a 24 by 24 grid? That's 4,032 univariate time series graphics! How do you visually arrange the data in a single viewing area? This is exactly the problem we tackle with the exploratory graphics to the left. We have in our possession temperature (surface and air), ozone, air pressure, and cloud cover (low, mid, and high) measurements averaged monthly from January 1995 to December 2000 over the coarse 24 by 24 grid region covering Central America, and we're not afraid to plot them.



Method

Using the original spatial time series data, we employ a lagged difference operator [1] to construct our graphics. Our method reveals seasonality along with other interesting features. Each tiny box in the matrix graphics on the left is a small multiple [2] of the 24 by 24 grid map shown above, with map lines, axes, and labels removed. A divergent color scale of blue to white to red is used to highlight the data, where blue relates to negative changes or lower measurements, red relates to positive changes or higher measurements, and white relates to either no change at all or some notion of mean or median measurement, depending on the graphic.

Difference graphic

The large triangular graphic to the far left is our first attempt to get a handle on the raw surface temperature data. The 71 lag-differenced time series graphics are plotted one on on top of another. The horizontal axis measures increasing time, and the vertical axis measures increasing lag. The skinny graphic tucked just beneath the horizontal axis is a graphic of the original spatial time series. This serves as a visual label for each point in time. The lag-differenced color scale is constructed by using the larger of the maximum and minimum difference across all lags and times and region. For instance, if the maximum difference was 10 and minimum difference was -5, blue would relate to -10, white to 0, and red to 10. The original time series color scale is constructed similarly, yet with its own minimum, median (to serve as 0), and maximum.

This graphic serves as a model for all the triangular graphics at left. We consider it a first step in exploratory analysis when one wants to look at any spatial time series data. Tick marks are eliminated in this graphic since, in general, one may not know in advance where to place them in regards to seasonality; however, our surface temperature data reveals yearly seasonality, so we take that into account for the next graphic.

Residual graphic

The second step in our exploratory analysis is an ambitious attempt (at best!) to subtract the seasonality. We construct the differenced mean across time and lag by using all pairwise combinations of region measurements, but only looking forward in time. For example, the January to February mean contains 15 entries, since we are examining the forward change. They are Jan 95 to Feb 95, Jan 95 to Feb 96, ..., Jan 95 to Feb 00, Jan 96 to Feb 97, Jan 96 to Feb 98, ..., and Jan 00 to Feb 00. Note that we don't include January 96 to February 95 since this difference looks backwards and inclusion of such differences would result in the zero mean change overall. Taking the data from the first plot and subtracting the mean reveals an interesting graphic; it reveals a notable event.

Other measurements

We construct the graphics for the other measurements and arranged them on the left. Only the low cloud data contain some missing data, but missing data are ignored. While these graphics are an exploratory process, and we reserve their interpretation for the experts, June 1998 seems to be a very interesting month in the Pressure graphics. In fact, most all the graphics express the similar "V" shaped motif with the bottom point of the "V" indicating June 1998.

Conclusion

"For us the most useful plot will be one that might reveal the unexpected or the unobvious. Sometimes a plot 'in the large' will do this. Usually, however, it is the plot of the residuals that has the greatest use and the greatest impact."

---John W. Tukey [3]

We feel these statements made nearly 30 years ago still hold, and in the present analysis, both the plot "in the large" (our difference graphic) and the plot of the residuals (our residual graphic), seem to hold equal weight. For instance, the Pressure difference graphic conveys the "V" shaped motif stronger than the Pressure residual graphic. However, more of the residual graphics convey the "V" shaped motif than do the difference graphics.

Bibliography

[1] Brockwell, Peter J., Davis, Richard A. 1987. Times Series: Theory and Methods. Springer-Verlag, New York.

[2] Tufte, Edward R. 1990. Envisioning Information. Graphics Press, Cheshire, CT.

[3] Tukey, John W. 1977. Exploratory Data Analysis. Addison-Wesley, Reading, MA.