

# Glaciers Melt as Mountains Warm: A Graphical Case Study

J. Hobbs · H. Wickham · H. Hofmann · D. Cook

Received: date / Revised version: date

**Abstract** For the 2006 ASA Data Exposition we created graphics that, in the legacy of John Tukey, tried to “force the unexpected upon us” (Tukey 1972). The data were geographic and meteorological measurements taken every month for six years on a coarse 24 by 24 grid covering Central America. Using conventional static graphics and some less conventional interactive graphics, we were able to find expected features in the data, such as seasonal patterns, spatial correlations, and El Niño events, as well as some more surprising results, several of which were corroborated by stories in the news.

**Keywords** 2006 ASA Data Exposition; interactive graphics; dynamic graphics; spatio-temporal data; temporal data; exploratory data analysis; climate change

## 1 Introduction

Data analysis is messy! Real data sets are rarely perfect and we may need multiple passes through the data to ensure that we have found all problems and that our findings are robust. At every stage, there are many possible next steps and our path was guided by intuition, domain knowledge and past experience. Many steps led to dead ends, and most plots ended up in the wastebasket. This may sound chaotic, but it had some structure—even highly improvisational data exploration has structure. Our general approach included the following components:

1. Refining questions and enumerating expectations
2. Investigating the data source
3. (Re-)Organizing the data
4. Plotting
5. Modeling
6. Incorporating other data
7. Presenting findings

The process was not linear and we returned repeatedly to earlier stages. Chatfield (1995) has more details on this general strategy.

The Expo data was daunting because it was complicated by multiple contexts: spatial, temporal, multivariate. We could have simplified the data by focusing the study to one variable, one time point or one location, but instead we tackled everything. Our analysis took many hours of labor, with a lot of data re-formatting, re-processing, many plots, much discussion, and extensive sharing of code and discoveries amongst the team. Finally, for presentation in this paper, all of the plots, including the interactive graphics, were reproduced to a publication quality for a uniform appearance.

This paper covers more than a simple presentation of our findings – it also describes the process that led to our discoveries in the Expo data. We think that the process is interesting reading, both as an example of exploratory data analysis and as a story of the chase for discovery.

## 2 Data Description

The Data Expo dataset consists of monthly observations of several atmospheric variables, which come from the International Satellite Cloud Climatology Project (ISCCP). The observations are derived from satellite-based measurements of the atmosphere’s radiative characteristics. The dataset includes observations over 72 months (1995-2000) on a 24 x 24 grid stretching from 113.75°W to 56.25°W longitude and 21.25°S to 36.25°N latitude. The following variables were recorded:

- Ozone in atmospheric column (**ozone**) in Dobsons.
- Surface pressure (**pressure**) in millibars.
- Near-surface temperature (**surftemp**) in degrees Kelvin.
- Surface temperature from clear sky composite (**temperature**) in degrees Kelvin.
- Cloud cover at low (**cloudlow**), medium (**cloudmed**) and high levels (**cloudhigh**), in fractions of 100.

## 3 Getting started

### 3.1 Refining questions and enumerating expectations

The initial announcement of the Data Expo provided limited general information about the data set and the objectives for analysis. Accompanying the description was a set of four general questions about the data set:

- What are the important relationships between variables?
- Are there important trends?
- Are there important groupings or clusters?
- Are there any unusual locations or time periods?

Armed with these questions, we wrote down some hypotheses about what we expected to see in the data, on the basis of what we knew or had heard about climate in the study region:

- The sampling area includes the equatorial Pacific Ocean, so we will see El Niño and La Niña effects on temperature, and perhaps other variables, for those events that occurred between 1995 and 2000.
- Seasonal trends will be different for sea versus land areas and the northern hemisphere versus the southern hemisphere.
- Data from the same month will show correlations between years (seasonal patterns).
- Data from neighboring spatial locations will be correlated.

To further refine these hypotheses, as well as to develop new ones, we investigated how the variables were defined in more detail.

### 3.2 Investigating the data source

From the variable descriptions, it appeared that the observations came from at least two satellite products from the International Satellite Cloud Climatology Project (ISCCP). Pressure, ozone and near-surface air temperature were derived from the TIROS Operational Vertical Sounder (TOVS), while cloud cover and clear-sky composite temperature were probably derived from the Gridded Cloud Product (Rossow and Schiffer 1991).

Unfortunately, many of the variables in the data can not be derived directly from the satellite measurements. For example, we can not observe surface temperature directly, but only brightness temperature,  $T_B$ , which is a function of the surface temperature,  $T_{sfc}$ , and the temperature profile,  $T(z')$ , through a vertical column of the atmosphere. The relationship is described by the “basic equation of radiometry” (e.g. Ulaby et al (1981)) For a specified temperature profile  $T(z')$ , the surface temperature,  $T_{sfc}$ , can be obtained from the brightness temperature,  $T_B$ .

Cloud-top temperatures are found in a similar fashion to give information about cloud height, with higher clouds being colder than lower clouds. Brightness temperature can be combined with the ideal gas law to provide information about pressure. In any case, values of coefficients and the vertical temperature profile vary with varying land surface characteristics, creating a possible source of error.

### 3.3 (Re-)Organizing the data

The data were provided as a set of individual files, one file for each of the variables for a single month—72 files for each variable—each formatted as follows:

```
VARIABLE : Mean low cloud amount (%)
FILENAME : ISCCPMonthly_avg.nc
FILEPATH : /usr/local/fer_data/data/
SUBSET   : 24 by 24 points (LONGITUDE-LATITUDE)
TIME     : 16-JAN-1995 00:00

      113.8W 111.2W 108.8W 106.2W 103.8W 101.2W 98.8W ...
           27   28   29   30   31   32   33 ...
36.2N / 51: 7.50  7.00  7.00  7.00 11.00 14.50 25.50 ...
33.8N / 50: 11.50 11.50  9.50  8.50 12.50 17.50 27.50 ...
...
```

It was interesting that the four of us arrived at different ways to organize the data, triggered by the different ways each of us was thinking about the data, and each emphasizing one of the spatial, temporal or multivariate aspects. The data organization heavily influenced the choice of methods to analyze the data.

The main approaches used data in a 2d form, where rows correspond to observations and columns to variables. While most statistical software permits the creation of 3d cubes, with two dimensions for space and one for time, most statistical analysis procedures do not expect data in this form. With a 2d form, there are many possible formats for this type of data. There are two that we used most often, which can be described as “short and wide” and “long and thin”.

In the “short and wide” format the focus is on the locations. It has 576 ( $24 \times 24$ ) rows, one for each location, and all the remaining data, such as elevation and all monthly measurements of ozone, temperatures, and cloud cover, are encoded in columns:

```
x y  lat    long temperature_1 temperature_2 ...
1 1 -21.2 -113.80      296.9      297.8 ...
2 1 -21.2 -111.29      296.5      296.9 ...
...
```

(If you are surprised by the high values of temperature, it is because they are measured in degrees Kelvin). It might also be thought of as the spatial context form—one observation for each spatial location. Each of the original files corresponds to one column of this data set. This version of the data allows the use of parallel coordinate plots to emulate time series for each location, as seen in Figures 2, 4, 6 and 5.

The “long and thin” format has 41472 ( $72 \times 576$ ) rows, i.e. each location is repeated for every time index (all 72 months). Elevation, which is constant over time, is repeated to match the other variables. We also created a new variable, `time`, to temporally index January 1995 to December 2000. The long and thin format looks like this:

```
x y lat    long    elev    time year month airtemp surftemp ...
1 1 36.25 -113.75 1526.25 1    1995 1    272.7  272.1 ...
1 2 33.75 -113.75 612.94  1    1995 1    279.5  282.2 ...
```

This form can be considered to be the spatiotemporal context—one observation for each spatial location and time. This version of the data allows us to study the multivariate aspects of the data and is the format underlying Figures 3, 8 and 9.

To keep the geographic context of the data in the forefront, we pieced together a map of the region from Google maps, printed it out, and pinned it to the wall. A small version of this map is shown in Figure 1.

## 4 Plotting the data

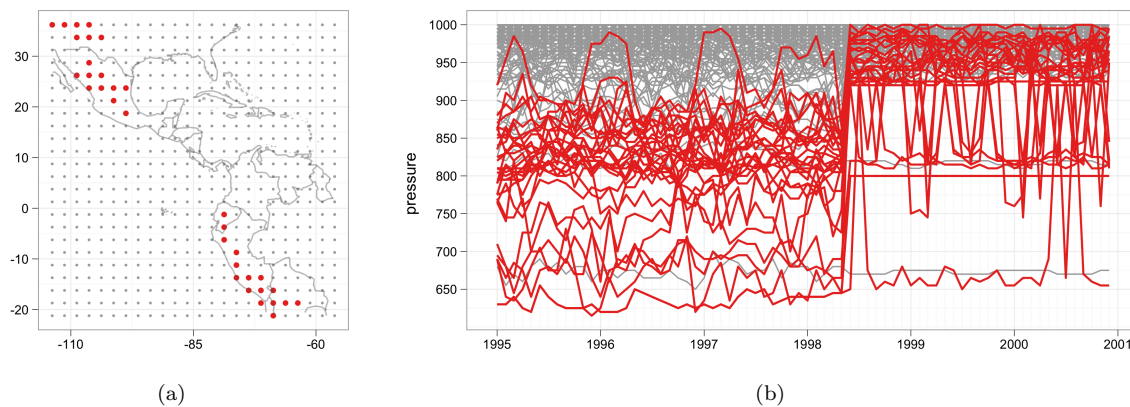
### 4.1 High altitude anomalies

Data exploration often uncovers problems with the data, and this data set is no exception. A look at a time series of pressure for each location reveals a problem: between May and June of 1998, a number of locations exhibit a



**Fig. 1** Map of the geographic area where the Expo data was collected, created from Google maps.

dramatic increase in pressure, with several showing an increase of over 100 millibars. Holding all other variables constant, this would be equivalent to a change in elevation of over 1 kilometer! Assuming that land doesn't rise by a kilometer in a month, puts the blame squarely on a disparity of the pressure variable. Using linked brushing (McDonald 1982; Becker et al 1988) between a plot of spatial location and a parallel coordinate plot of pressure measurements in time (Figure 2), it can be seen that all of the problematic locations are at high elevations, in the Rocky Mountains and Andes.

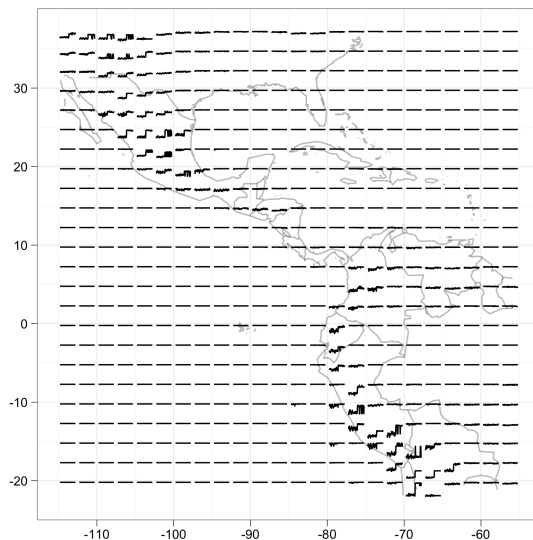


**Fig. 2** Spatial location plot (a) and time series of pressure (b). Locations with problematic pressure changes (a change of more than 80 millibars) are colored red and enlarged in both plots. Problematic locations are in high-altitude areas of the Rocky Mountains and Andes.

Another way to look at the pressure data is to draw small multiples of the pressure time series while maintaining the spatial structure, as in Figure 3. In the exploratory stages, this was done using manual controls and

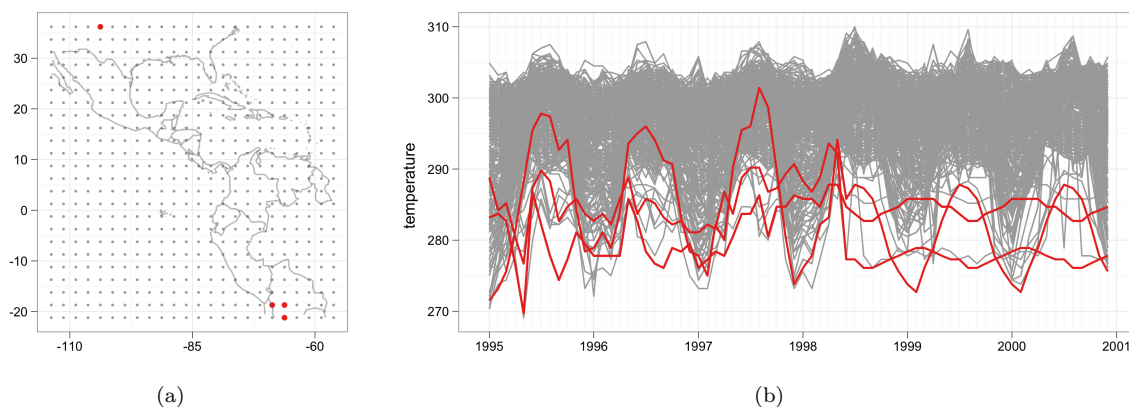


the correlation tour (Buja et al 1988; Cook and Buja 1997; Cook et al 2006). It works by making separate linear combinations of variables in the horizontal and vertical directions. Here, longitude is linearly combined with time, horizontally, and latitude is linearly combined with pressure, vertically, to produce small time series at each spatial location.



**Fig. 3** Small multiples of pressure time series with spatial structure. Pressure values are flat over much of the spatial domain, but in the high altitude areas there is a strange jump or split in values around the middle of the time series.

The manual tour allows rapid exploration of different variables. An examination of air temperature revealed unusual patterns at just a few locations. Figure 4 shows the time series of temperature with four locations highlighted. For these locations the temperature profiles are conspicuously regular after June of 1998. Closer inspection revealed that the monthly temperature values are repeated in the subsequent years, making us suspect that these air temperature values were imputed, rather than observed. We also learned that there is something odd about the relationship between the two temperature variables in the high altitudes, and that cloud patterns change in the Pacific during an El Niño event. The accompanying video, found at <http://had.co.nz/dataexpo>, illustrates these findings.

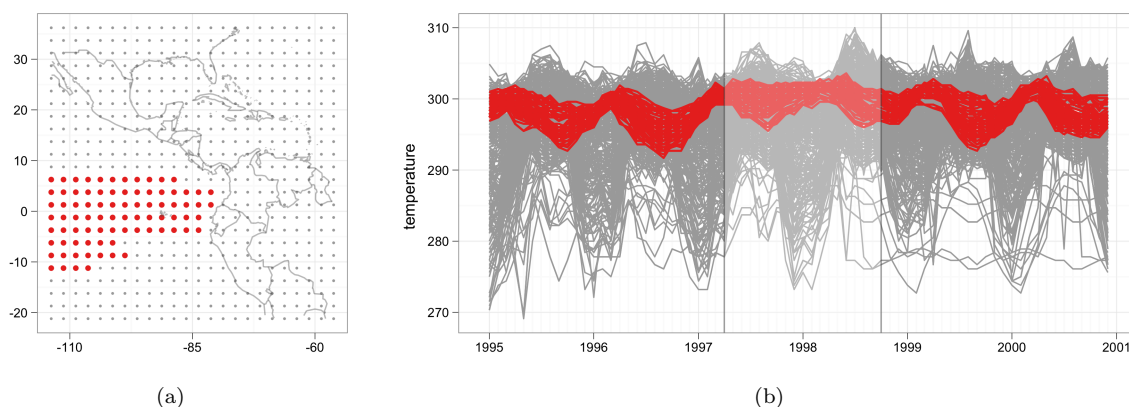


**Fig. 4** Spatial location plot (a) and temperature time series (b). The four suspicious locations are colored red and enlarged. There are only three lines visible in the time series because of over-plotting. The regular pattern of values after June 1998 are an indication that these values are imputed and not observed.

Because we did not have access to the underlying raw data, we could not investigate these data problems further, or fix them, and so for the rest of the analysis we ignored pressure, and treated temperature somewhat cautiously for some locations.

#### 4.2 The El Niño effect

In late 1997 and early 1998, sea surface temperatures in the equatorial Pacific remained unseasonably high. This phenomenon is known as El Niño, and is of interest because it signals profound changes in the weather patterns of the entire world. In Figure 5 these locations are highlighted, along with the corresponding time series of surface temperature. The El Niño event is visible in the time series as the flat section during 1997-8, where the temperatures do not drop to the usual seasonal lows.



**Fig. 5** Spatial location plot (a) and time series of surface temperature (b). The locations in the equatorial Pacific are colored red and enlarged. In these locations, temperatures stay unseasonably high during the winter of 1997 and spring of 1998 (shaded). This phenomenon is known as El Niño.

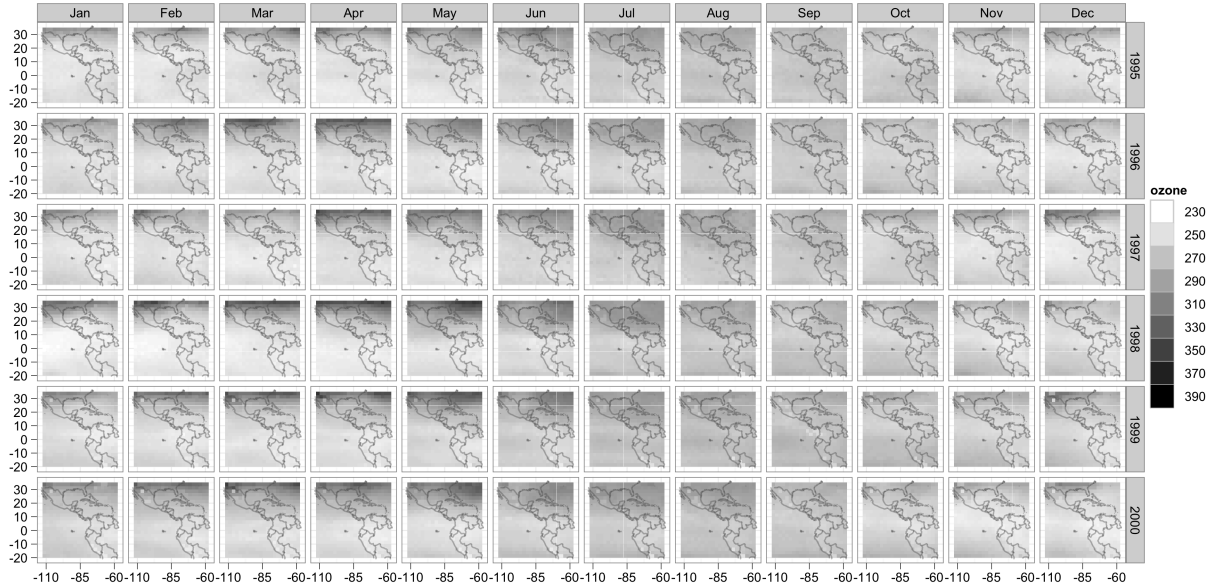
#### 4.3 Spatiotemporal trends

The classical approach to displaying spatial data is the image plot, where a color gradient is used to represent the value of a variable measured on a regular grid. Figure 6 shows image maps for ozone, by month and year. The gray values are scaled uniformly across all the maps, allowing us to both explore patterns within a plot, and to make comparisons across plots.

Looking at each plot we can see a spatial trend: the maps are darker at the top and bottom and lighter in the middle, which says that high values are found farther from the equator.

Across the plots, horizontally, monthly or seasonal trends can be compared, and vertically the yearly trends can be compared. Reading across a row, we can see a distinct seasonal trend: in the months of June-October, ozone levels increase from the north towards the equator, and to some extent in the south also. Reading down columns, we can see differences between years: comparing May in each year we see that ozone levels in the north were lower in the first few years and jumped dramatically in 1998 and stayed high through 2000. March 1997 is somewhat of an anomaly, with low ozone in the north compared to both February and April of the same year as well as the March values for the other years.

A different, yet similar, picture is attained by examining temporal patterns in ozone conditioned on location. Figure 7 shows a different type of graphic, a star glyph of ozone at each location. Each star glyph is a time series of ozone in polar coordinates. This graphic allows us to compare temporal patterns at spatial locations. Different sizes and shapes represent global patterns: large circular icons indicate high values at all time points; flower shapes indicate seasonal variability, with peaks and dips at the same time each year. Global patterns can be seen by looking at the glyphs en masse; try squinting your eyes or looking at the figure from a distance. Ozone levels are highest furthest from the equator and really small just north of the equator. They have more seasonal variability in the middle latitudes, both north and south.



**Fig. 6** A classical approach to plotting spatiotemporal data: ozone values shown using gray scale on the spatial coordinates with separate plots for each month and year. Ozone levels in the north are high January-June, but more evenly spread throughout the region July-November. In 1997, March saw lower ozone levels than in other years.

Figure 8 explores trends in two spatial groups: northern latitudes, (a) and (c), and the equatorial Pacific, (b) and (d). We select points in the spatial domain and examine the changes in the time series of air temperature, surface temperature, ozone and low cloud. Ozone, air and surface temperatures show more seasonal variation in the northern latitudes, while low cloud is more variable in near the equator. Interestingly, the seasonal trend in low cloud is noticeably absent during the 1997-8 El Niño event.

#### 4.4 Multivariate relationships

Figure 9 and Table 1 help to examine the pairwise relationships between variables for all locations and time points. As expected, the two temperature variables are strongly associated with a correlation equal to 0.81. There are a couple of points where the two temperature values differ substantially, which further investigation revealed to occur on the edges of the spatial region: these are probably errors created during the original data processing.

The relationships between other pairs of variables are more complex. Temperature has little association with ozone. The clouds have something of a constrained relationship, for example if there is a lot of high cloud cover there tends to be little low cloud cover. To learn more about these associations we need to include the spatiotemporal dependence, for example, by conditioning on time or location to investigate small neighborhoods.

	temperature	surftemp	ozone	cloudlow	cloudmed	cloudhigh
temperature	1.00	<b>0.81</b>	-0.35	-0.20	-0.09	0.22
surftemp	<b>0.81</b>	1.00	-0.21	-0.07	-0.31	0.04
ozone	-0.35	-0.21	1.00	-0.01	-0.03	-0.08
cloudlow	-0.20	-0.07	-0.01	1.00	-0.42	<b>-0.54</b>
cloudmed	-0.09	-0.31	-0.03	-0.42	1.00	<b>0.62</b>
cloudhigh	0.22	0.04	-0.08	<b>-0.54</b>	<b>0.62</b>	1.00

**Table 1** Pearson correlation matrix for the measured variables.

## 5 Modeling

The modeling has two components, identifying the strong patterns, and then examining the remains when these are removed.

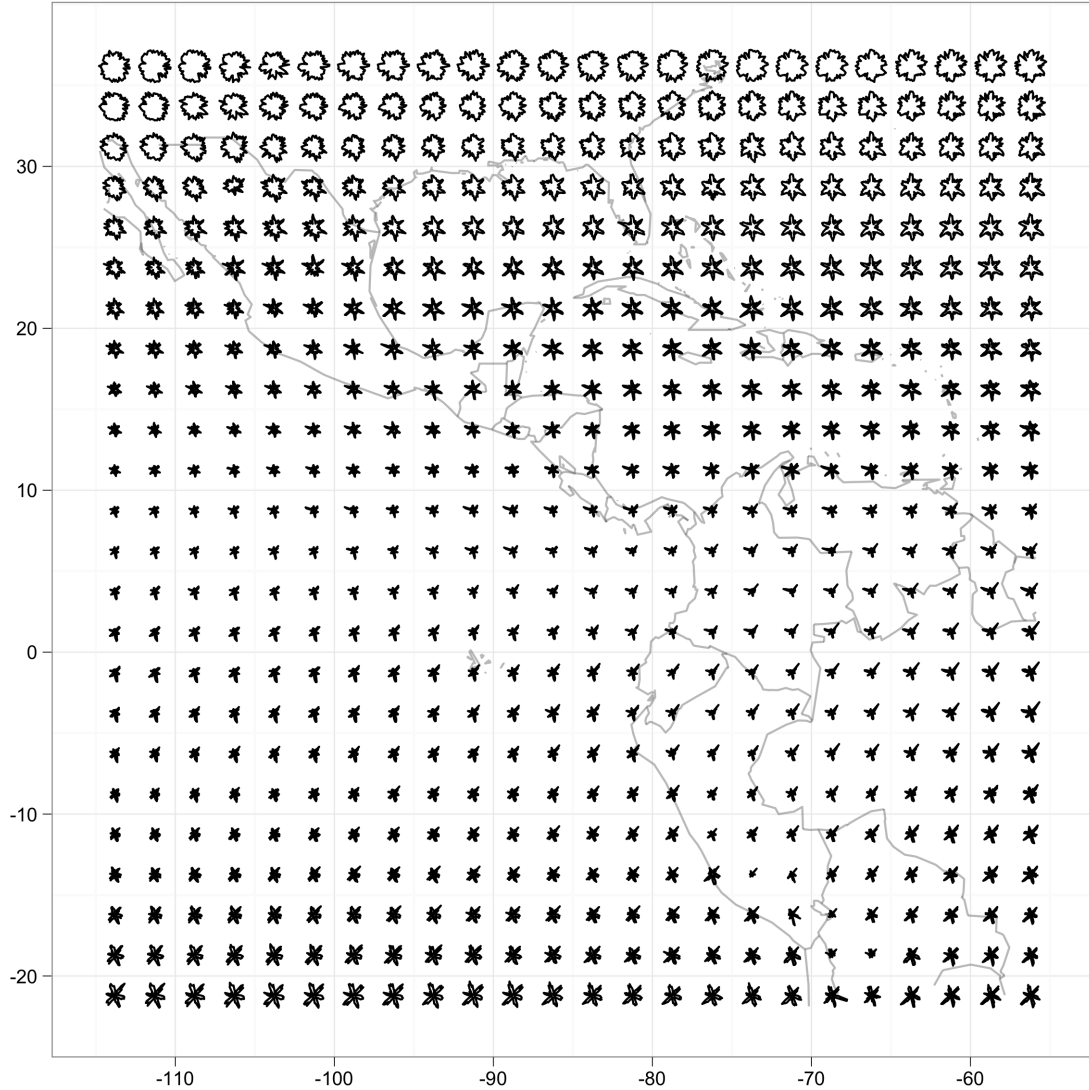


Fig. 7 Star glyphs of monthly ozone values between 1995 and 2000. Larger glyphs at higher latitudes suggest that ozone is generally higher in this region. Flower-like glyphs indicate strong seasonal patterns, which is strongest at middle latitudes.

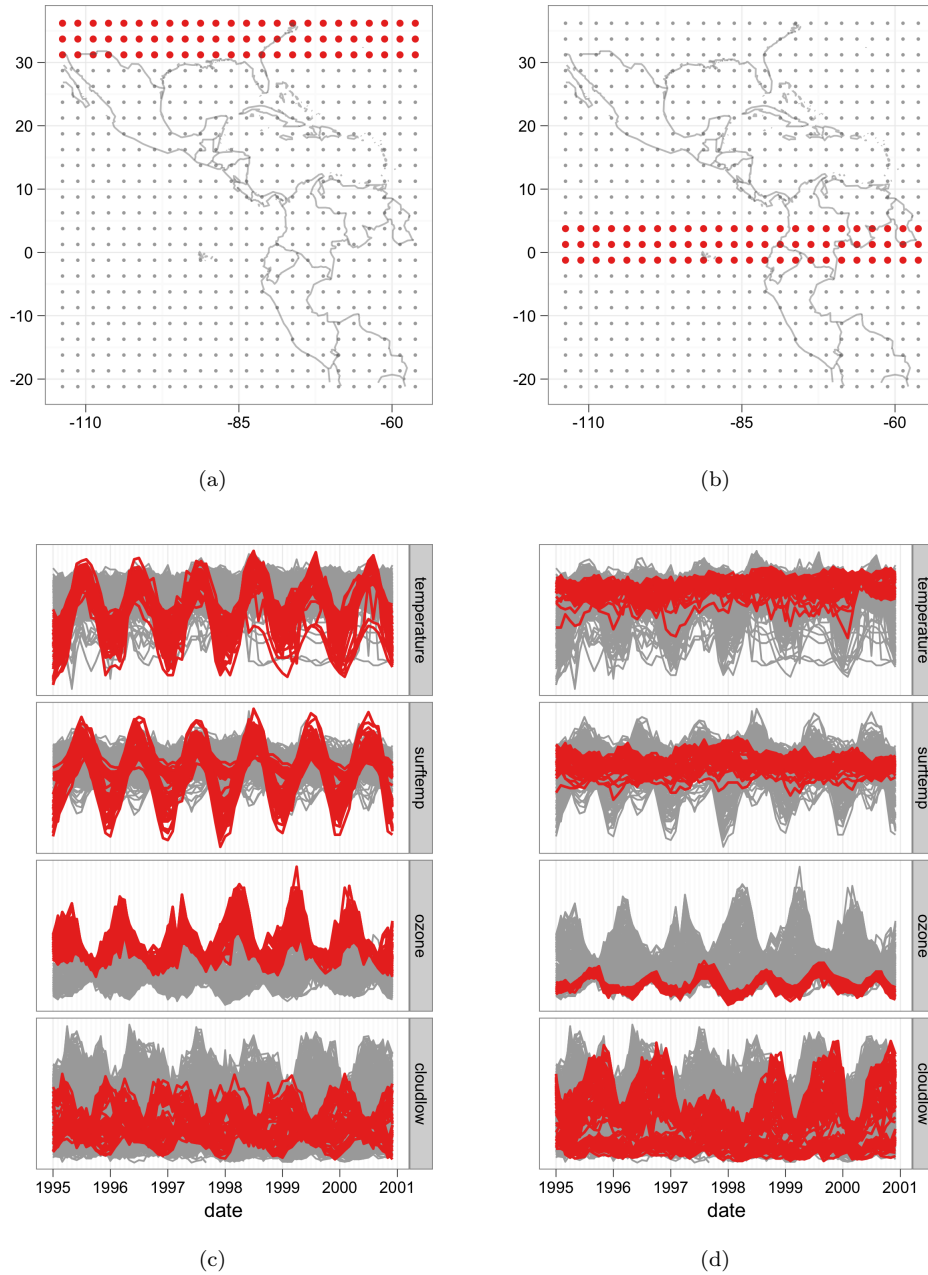
### 5.1 Removing seasonal temperature trends

Motivated by patterns apparent in the raw data, a model with an intercept, a linear temporal trend, and a seasonal component using a combination of sine and cosine was fitted for the data at each spatial location:

$$Y = \mu + \alpha t + \beta_1 \sin(m) + \beta_2 \cos(m) + \varepsilon, \quad \varepsilon \sim MVN(0, \sigma^2 I)$$

where  $Y$  surface temperature (in K)  
 $t$  time index,  $t = 1, \dots, 72$   
 $m$  month,  $m = 2i\pi/12$ ,  $i = 1, \dots, 12$   
 $\mu$  average temperature  
 $\alpha$  monthly increase in average temperature (in K)  
 $\beta_1, \beta_2$  amplitude of seasonal variations

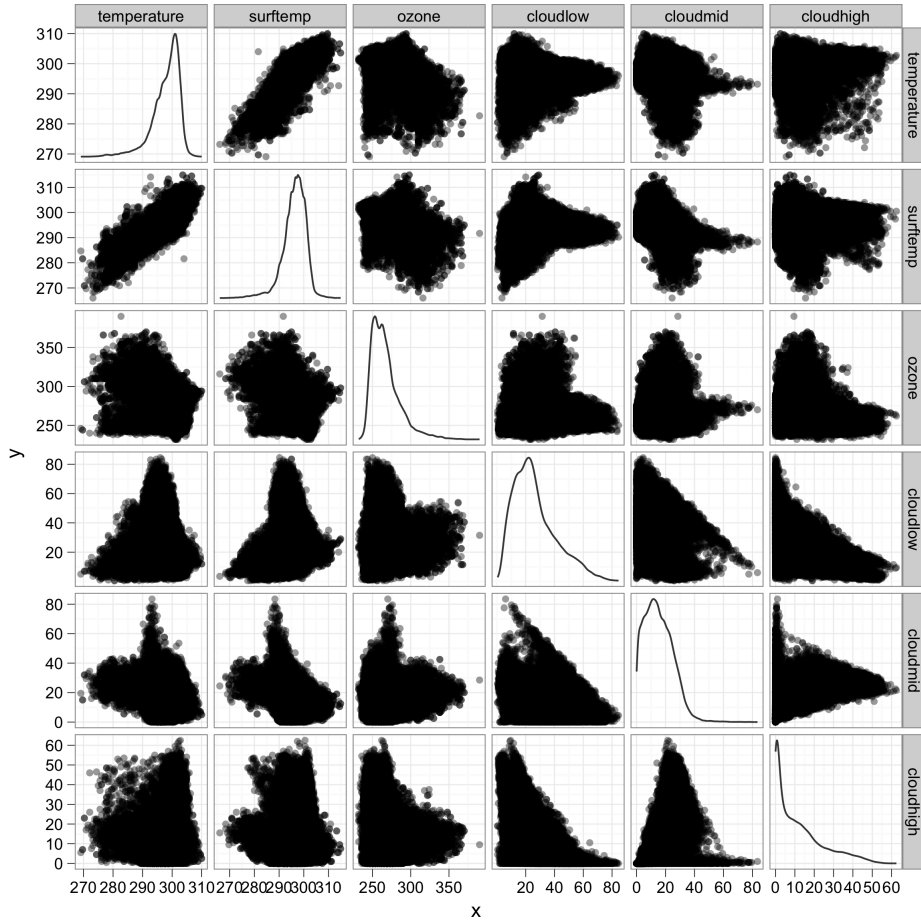
Figure 10 shows the (de-seasonalized) linear trend (a), and residuals (b), at each location. There are a number of interesting patterns. Many locations in South America, in the western parts along the Andes Mountains, and to some extent Mexico, exhibit an increasing trend over time, with some locations showing increases of several degrees



**Fig. 8** Spatial location plot, (a),(b), and time series of air temperature, surface temperature, ozone and low cloud cover, (c),(d). At more northern latitudes (a),(c) there is more seasonal trend in air and surface temperatures and ozone. Low cloud cover exhibits more seasonal trend near the equator (b),(d), and it is noticeably lower during the 1997-8 El Niño event.

per year! Large increases in temperature are associated with high altitude, and some of the highest increases are near glaciers (marked by the solid dark circles). Now, Section 4.1 noted that there are a few locations where we should be suspicious about the temperature values, and some, but not all, of these locations with the dramatic increase coincide with these. They do however occur at similar locations to the pressure anomalies observed earlier. So, to investigate further we collected temperature recordings from ground stations in several South American locations, and trends in these measurements are discussed in Section 6.2.

The El Niño event can also be seen. Many locations in the equatorial Pacific (middle to lower left of the spatial region) exhibit a “bump” in the residual pattern in the middle of the time series, which would correspond to late 1997 and early 1998.



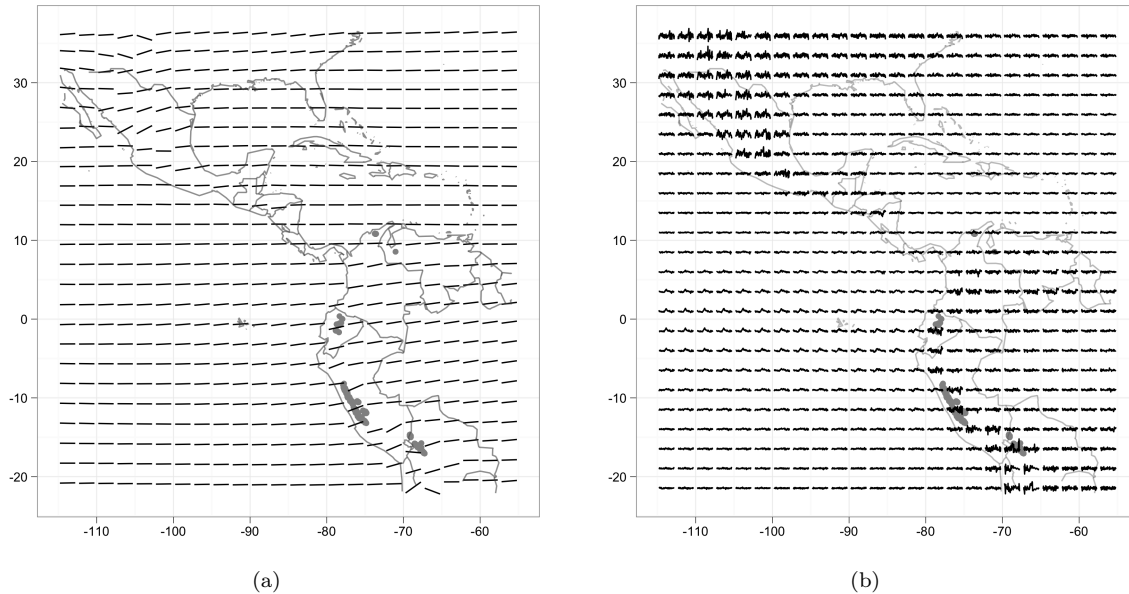
**Fig. 9** Examining all pairwise relationships using a scatterplot matrix for all locations and time points. Plots along the diagonal show density estimates for each variable. The two temperature variables are strongly associated, with the exception of a few points. The relationships between other variables are more complex.

## 5.2 Clustering locations into local climates

Cluster analysis can help us learn about local climates in the larger geographic region. A hierarchical cluster analysis was conducted using all meteorological variables, except pressure, in the “short and wide” form of the data. Variables were standardized to have zero mean and unit variance, and distance was defined with the Euclidean metric and Ward’s linkage (Johnson and Wichern 2002). Ten clusters were chosen, based on the dendrogram and from our own limited knowledge about the geographical regions. The cluster classification of each location is shown in Figure 11(a). The measurements in each cluster are shown as time series plots in Figure 11(b). Spatial location was not included, but the resulting ten clusters are geographically localized, which lends weight to the notion of different climate regions over the spatial domain.

Clusters 1 and 2 have highly varied levels of mid-level and high cloud cover and small seasonal trends in ozone and temperature. Cluster 1 has a big peak in high cloud cover during the middle of the time period, corresponding to the El Niño event. This peak is evident in cluster 5, and to some extent cluster 6, too—both of these are locations in the equatorial Pacific where the El Niño occurs. Interestingly, there is a small peak around this time in cluster 10, which is in the Caribbean. It has been suggested that El Niño has some effect on hurricanes in the Caribbean, so perhaps this provides a link between the two geographical regions. Section 6 explores this further. Clusters 9 and 10 in the north have the most seasonal variability in ozone. Cluster 3, which contains the coastal mountain range in South America, shows increasing temperatures. Cluster 4, covering most of South America and much of Mexico, is like the miscellaneous cluster, with large variation on most of the variables.

We could spend a lot of time digesting the intricacies of these clusters, but the main message is that the clusters corresponding to similar climates roughly match what we know about geography: Pacific regions, Caribbean,



**Fig. 10** Long-term trends (a) and residuals (b) from seasonal surface temperature model. Large increases in temperature are associated with high altitude, particularly near glaciers (marked by dark grey points). El Niño can be seen as a mid-period bump in the residuals in the Pacific region.

land/sea. However, we should be cautious about over-interpretation, as we may have actually partitioned a smooth, continuous, climate gradient into arbitrary chunks.

## 6 Incorporating other data

### 6.1 Where are the glaciers?

We researched the locations of glaciers in South America and overlaid these in black on the map view in Figure 10. The most dramatic increases in temperature occur at high elevations and near the locations of glaciers. Searching for more information on glaciers and warming on the internet also uncovered the photos of the Qori Kalis glacier in Peru taken in 1978 and 2000, which show that this glacier has retreated substantially over two decades.

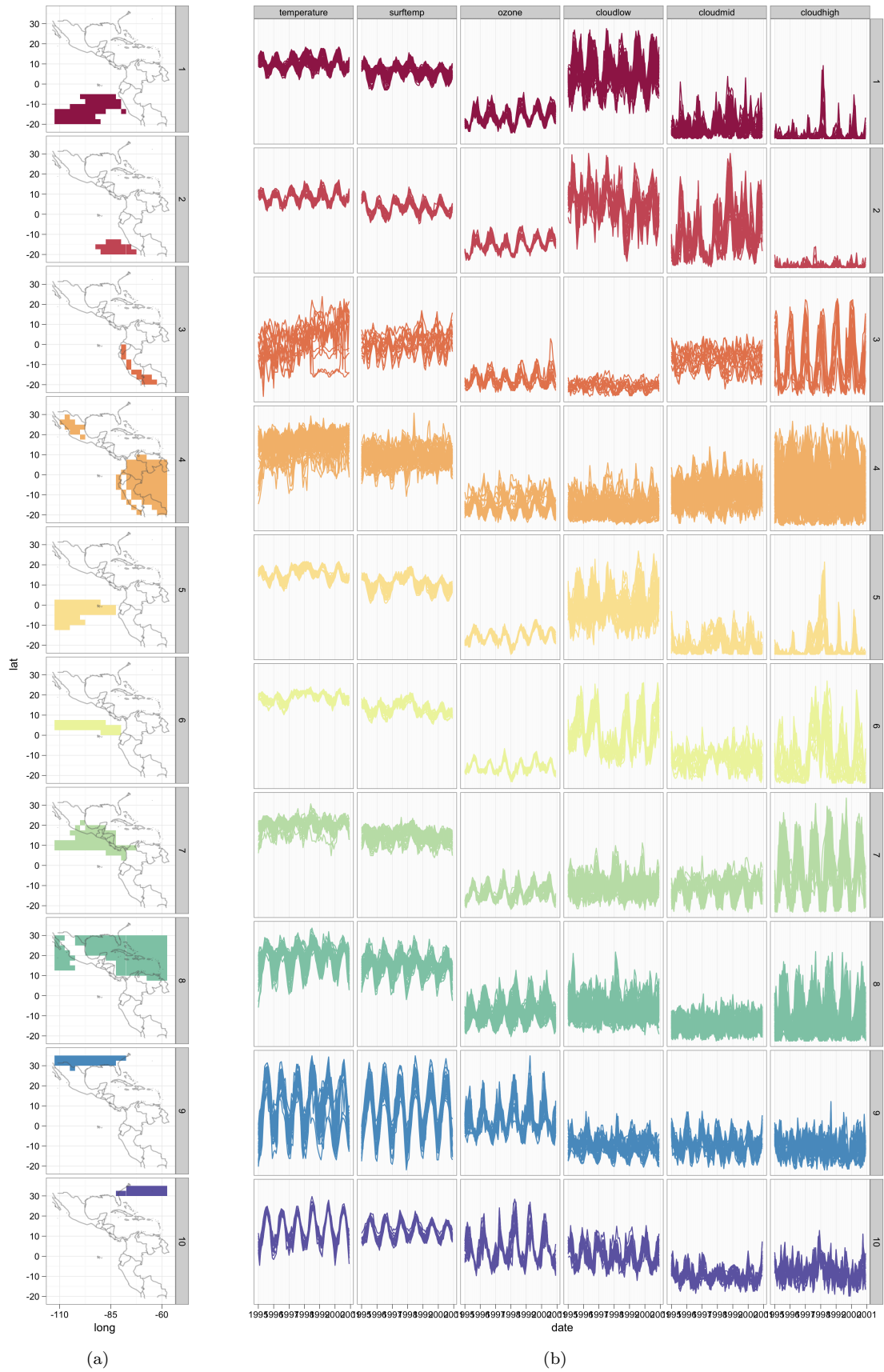
### 6.2 Is the satellite data verified by ground stations?

To investigate the apparent increase in temperatures at high altitudes we searched for information from ground-recording stations in the vicinity. The closest we could find are in Colombia, near Bogota, Cali, and Antonio. (There are glaciers in Colombia which are also reported to be receding.) Temperature data for these stations are shown in Figure 12. The measurements cover varied time frames, all of which start earlier than the Expo time period. In each case a loess curve is fit to the data. Measurements for Bogota (mean temperatures only) are erratic, but values for Antonio and Cali, where we have minimum, maximum and average values, do indicate some increasing trend over the longer time period, although not as strong as in the satellite data.

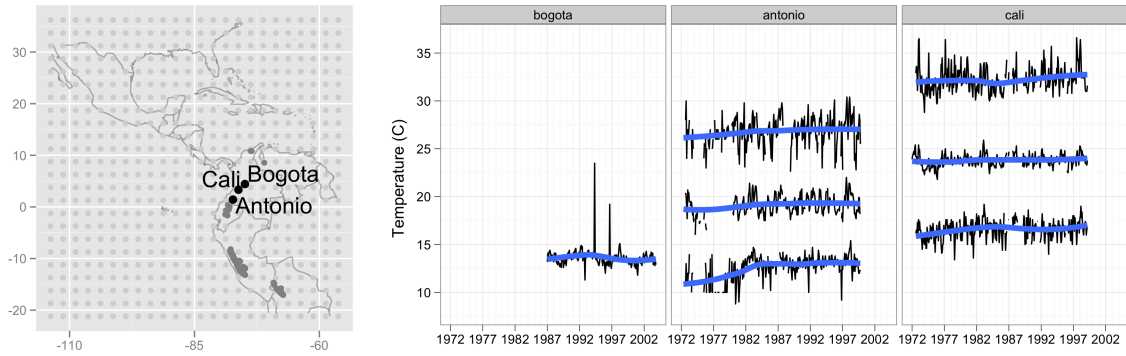
### 6.3 How are El Niño and tropical cyclones related?

As we have shown, the El Niño of 1997-98 was evident in some of the variables in the Expo data (temperature, cloud cover). In general, El Niño has been linked to climate anomalies all over the world. To investigate the relationship between El Niño and tropical hurricanes, in particular, we constructed tropical storm track data from the National Hurricane Center's (NHC) archive of tropical cyclone reports for the Atlantic, Caribbean and Gulf of Mexico. The best storm data records the latitude/longitude position of the storm center, the surface air pressure, maximum sustained wind speed and storm stage every six hours. Storm stage was classified as





**Fig. 11** Results of hierarchical clustering using Ward's method on temperature, ozone and cloud values: (a) geographic location of clusters and (b) time series of variables for each cluster, helps to understand the differences between clusters.



**Fig. 12** Monthly temperature data from recording stations in Colombia: Bogota (Elev. 2548m), Antonio (Elev. 1826m), Cali (Elev. 969m). Minimum, mean and maximum are shown for Antonio and Cali, mean only for Bogota. Loess curves are overlaid on the data. Small increasing trends can be seen for Antonio and Cali.

Hurricane, Tropical Storm, Tropical Depression or Extratropical, and an individual storm could and often did go through all four stages. Only named storms (those that had Tropical Storm or Hurricane status at some point) were included in our analysis.

Despite reports in the news that tropical storms increase in response to El Niño events, the storm tracks data suggests a marked decline of hurricane activity (Figure 13). This is validated in the findings of Bove et al (1998), a study of the historical frequency of tropical hurricanes in the Atlantic basin. They concluded that hurricane activity is reduced during El Niño events.

## 7 Summary of discoveries

Our initial expectations were mostly found in the data:

- The El Niño event was striking in time series plots of temperature.
- All variables show seasonal trend in some locations, mostly in the higher and lower latitudes: the same month is correlated between different years. The exception is the El Niño event, 1997-1998, in the equatorial Pacific, where temperatures and cloud patterns do not follow similar patterns as the previous year.
- There are clear differences between land and sea. Even though we did not focus on these differences in our analysis, the difference between land and sea jumps out in the cluster analysis and has about as much effect as the differences in latitude.
- All variables exhibit spatial dependence: generally, neighboring areas have similar values. There are some large jumps going from land to sea and also in the high altitude areas where a small spatial difference corresponds to a big difference in elevation.

What was more interesting, and what we spent more time on, were the surprising features of the data. We actually learned some things about climate change, things that were only rumors to us prior to the analysis. Each of us is now more attuned to the stories of glacial melt and fracturing of ice shelves. Here is a summary of our unexpected findings:

- There is obviously a problem with the pressure variable, so we left this variable out for much of our analysis.
- Temperatures at high altitudes have been increasing over the time period; in some locations as much as several degrees per year, instead of following the usual seasonal peak and decline. These locations match locations of glaciers, and ground truth data somewhat supports this pattern.
- In the equatorial Pacific, the high cloud peaks during El Niño, as does low cloud, and medium cloud is lower than normal. And, interestingly, there is some change in cloud patterns in the Caribbean during these events, also, which may provide a link to the El Niño events and severity of storms in the Caribbean.
- Storm track data in the Atlantic basin exhibit a decline in the number and severity of storms near the one El Niño event in the study period. This contradicts some media reports claiming El Niño increases the number and severity of storms in the Caribbean, but is consistent with the scientific literature.

Ideally we would get some feedback from the data experts at NASA about these findings.

## 8 Tools and additional data sources

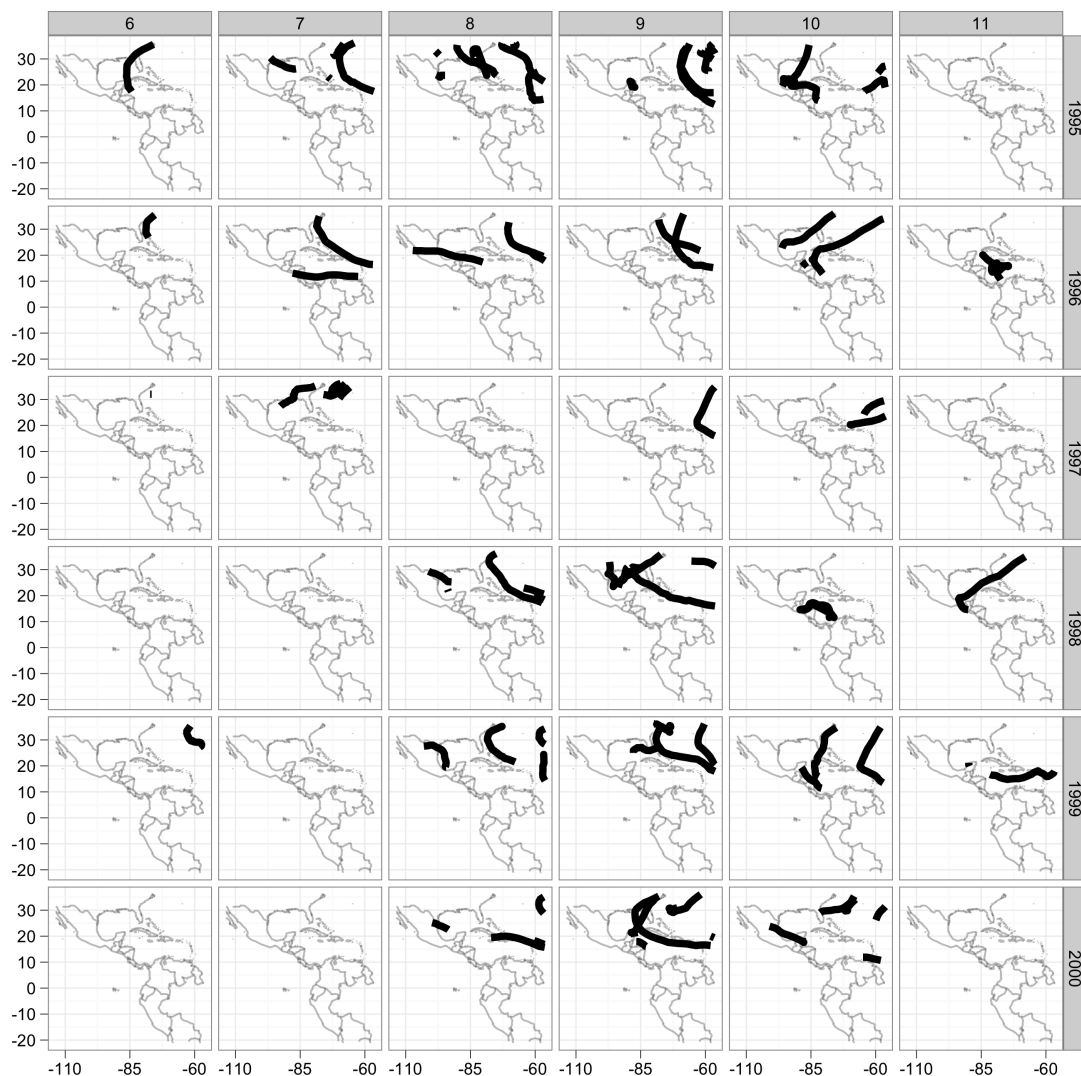
The full 4' x 8' poster, and associated movies are accessible on our data expo website: <http://had.co.nz/dataexpo>.

R (R Development Core Team 2006) and the packages `ggplot` (Wickham 2009) and `rggobi` (Lawrence and Wickham 2006), were important tools for our statistical analysis and for producing graphics. MANET (Unwin et al 1996), GGobi (Swayne et al 2003), and Gauguin (Gribov et al 2006) were used for dynamic and interactive graphics.

We maintained working documents describing our findings and re-structured data sets in a central storage location so that all of the team could readily access each other's work. Since each of us tended to use different software, we periodically used this exercise as a chance to educate others, giving software tutorials on the Expo data exploration for each other and to colleagues in the department.

We supplemented the data provided by the organizers with additional data available on the web:

- Positions of glaciers are provided in the form of the World Glacier Inventory by the National Snow and Ice Data Center at Boulder, CO. [http://nsidc.org/data/glacier\\_inventory/](http://nsidc.org/data/glacier_inventory/)
- Qori Kalis glacier photos, 1978 and 2000. <http://researchnews.osu.edu/archive/andespics.htm>



**Fig. 13** Hurricane and tropical storm tracks by month (June–November) for each year 1995–2000. Hurricane activity seems to slow down during the time of the El Niño in Fall 1997 to Summer 1998. There are early storms but few late storms in 1997 and mostly late storms in 1998.

- Temperature data from recording stations in Colombia are available from the National Climatic Data Center at the National Oceanic and Atmospheric Administration (NOAA). <http://gis.ncdc.noaa.gov/website/ims-cdo/gsom/viewer.htm>
- Storm track data were provided by the NOAA National Hurricane Center. <http://www.nhc.noaa.gov/pastall.shtml>
- ISCCP is the International Satellite Cloud Climatology Project, part of the World Climate Research Program. <http://isccp.giss.nasa.gov/>

**Acknowledgements** This work has been partly supported by the National Science Foundation Vertical Integration of Graduate Research and Education grant 0091953, and grant DMS0706949.

## References

- Becker RA, Cleveland WS, Wilks A (1988) Dynamic Graphics for Data Analysis. In: Cleveland and McGill (1988)
- Bove MC, O'Brien JJ, Eisner JB, Landsea CW, Niu X (1998) Effect of El Niño on U.S. Landfalling Hurricanes, Revisited. *Bulletin of the American Meteorological Society* 79(11):2477–2482
- Buja A, Asimov D, Hurley C, McDonald JA (1988) Elements of a Viewing Pipeline for Data Analysis. In: Cleveland and McGill (1988), pp 277–308
- Chatfield C (1995) *Problem Solving: A Statistician's Guide*. Chapman and Hall, <http://www.crc.com>
- Cleveland WS, McGill ME (eds) (1988) *Dynamic Graphics for Statistics*. Wadsworth, Monterey, CA
- Cook D, Buja A (1997) Manual Controls For High-Dimensional Data Projections. *Journal of Computational and Graphical Statistics* 6(4):464–480, also see [www.public.iastate.edu/~dicook/research/papers/manip.html](http://www.public.iastate.edu/~dicook/research/papers/manip.html).
- Cook D, Lee EK, Buja A, Wickham H (2006) Grand Tours, Projection Pursuit Guided Tours and Manual Controls. In: *Handbook of Data Visualization*, Springer, <http://www.springer.com>, p To appear
- Gribov A, Unwin A, Hofmann H (2006) About Glyphs and Small Multiples: Gauguin and the Expo. *Statistical Computing and Graphics Newsletter* 17(2):14–17, URL <http://www.amstat-online.org/sections/graphics/newsletter/newsletter.html>
- Johnson RA, Wichern DW (2002) *Applied Multivariate Statistical Analysis* (5th ed). Prentice-Hall, Englewood Cliffs, NJ
- Lawrence M, Wickham H (2006) rggobi: Linking R and GGobi. R package version 2.1.4
- McDonald JA (1982) *Interactive Graphics for Data Analysis*. Tech. Rep. Orion II, Statistics Department, Stanford University
- R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>, ISBN 3-900051-07-0
- Rosow WB, Schiffer RA (1991) ISCCP Cloud Data Products. *Bulletin of the American Meteorological Society* 72(1):2–20
- Swayne DF, Temple Lang D, Buja A, Cook D (2003) Ggobi: Evolving from XGobi into an Extensible Framework for Interactive Data Visualization. *Journal of Computational Statistics and Data Analysis* 43:423–444, <http://authors.elsevier.com/sd/article/S0167947302002864>
- Tukey JW (1972) Exploratory Data Analysis: As Part of a Larger Whole. In: *Proceedings of the 18th Conference on Design of Experiments in Army Research and Development I*, Washington, DC, p 1010
- Ulaby FT, Moore RK, Fung AK (1981) *Microwave Remote Sensing: Active and Passive*, vol 1. Addison-Wesley, Reading, MA
- Unwin A, Hawkins G, Hofmann H, Siegl B (1996) Interactive Graphics for Data Sets with Missing Values - MANET. *Journal of Computational and Graphical Statistics* 5(2):113–122
- Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer