DEPARTMENT OF STATISTICS
University of Wisconsin
1300 University Ave.
Madison, WI 53706
Phone: (608) 262-2598
Fax: (608) 262-0032

TECHNICAL REPORT NO. 1128

August 22, 2006

Visualizing Abnormal Climate Changes in Central America from 1995 to 2000 - Data Expo 2006

Sang-Hoon Cho [1]
Department of Statistics, University of Wisconsin, Madison, WI

Hyonho Chun [2]
Department of Statistics and Department of Biostatistics and Medical Informatics,
University of Wisconsin, Madison, WI

This Poster Won First Prize in Data Expo 2006

The figures are best seen by enlarging them on the monitor.

Key Words and Phrases: Data Imputation, Smoothing Spline ANOVA, El Nino and La Nina, Ozone Depletion Areas, Cloud Effect on Temperature

[1]Supported in part by NIH Grant NCI CA075142.
[2]Supported in part by NSF Grant DMS0505636 and DMS0604572 and ONR Grant N00014-06-1-0095.

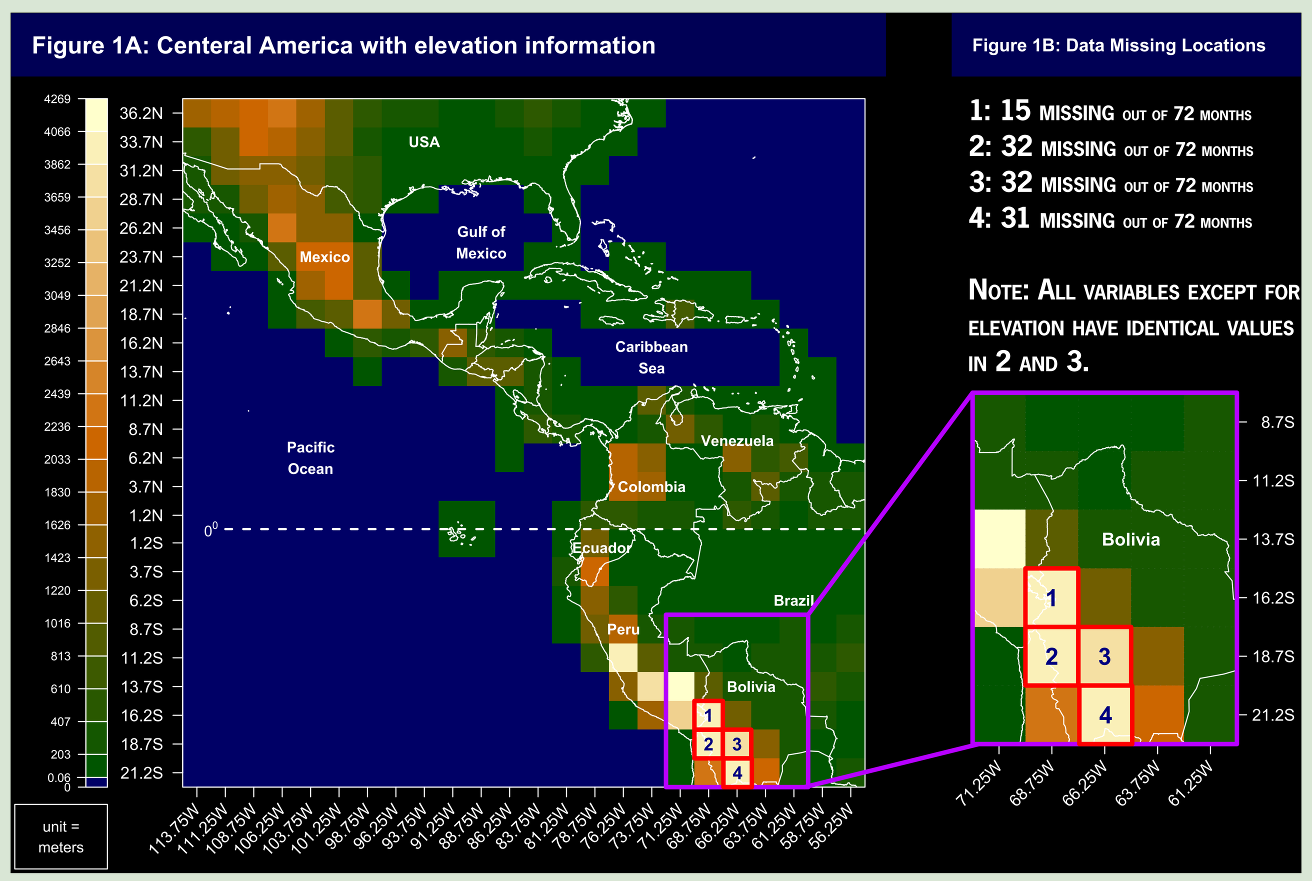
Visualizing abnormal climate changes in Central America from 1995 to 2000

Sang-Hoon Cho and Hyonho Chun

Statistics Department, University of Wisonconsin - Madison

2006 Data Expo

1. Introduction



Object : We aim to uncover and visualize abnormal climate changes.

Area : Central America within 115W ~ 55W and 37.45N ~ 22.45S on 24 by 24 grid [Figure 1A].

Variables and units : Variables are monthly averages from Jan 1995 to Dec 2000 except for elevation.

1. Elevation - Meter
2. Surface & Air Temperature - Kelvin
3. Ozone - Dobson
4. Cloud Low & Mid & High - %

Data Missing : Cloud low has missing observations [Figure 1B].

2. Data Imputation

We considered a nonparametric spatial model to impute the missing data in cloud low, using location and elevation information.

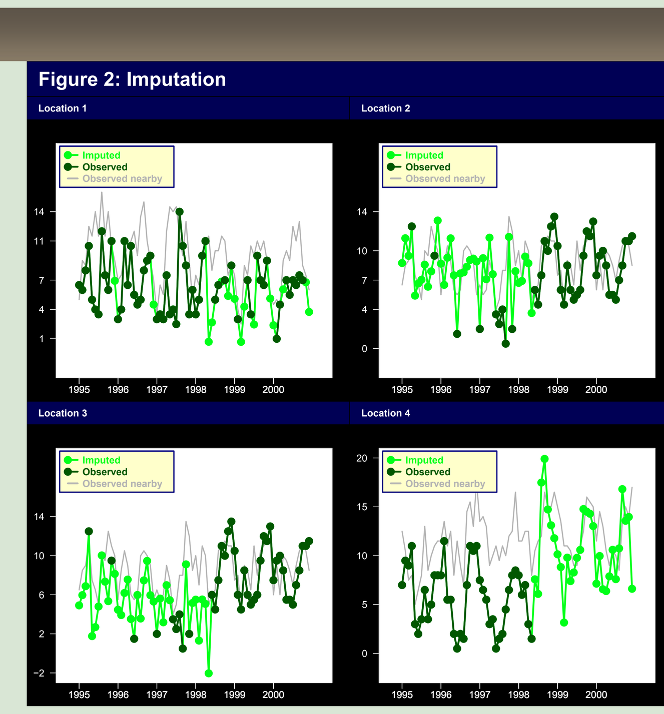
Model : Smoothing Spline ANOVA

$$\text{Cloud Low} = f_1(\text{lati}) + f_2(\text{long}) + f_{1,2}(\text{lati}, \text{long}) + f_3(\text{elevation}) + \text{Error}$$

where f_1, f_2, f_3 are nonparametric functions for main effects and $f_{1,2}$ for an interaction, respectively.

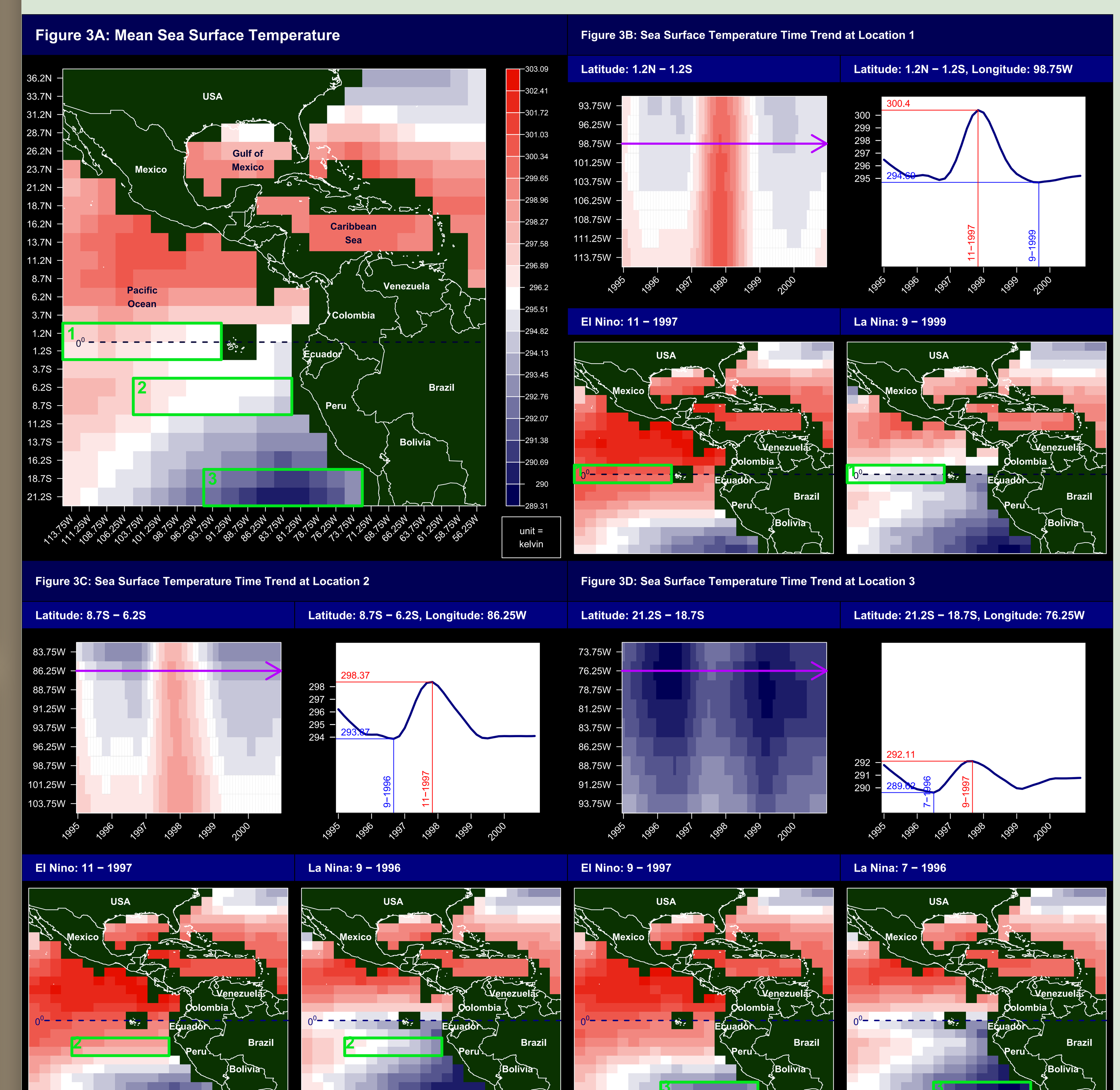
Procedure :

1. Fit a Smoothing Spline ANOVA for each month, excluding missing values.
2. Predict the missing values, using the fitted model.
3. Fit a Smoothing Spline ANOVA for each month again, including the imputed values.
4. Update the imputed values.
5. Repeat 3 and 4 step until imputed values converge.



Result : We selected nearby locations showing temporal trends similar to the trends at the missing locations during the observed time period.
The imputed values are close to the values at the selected nearby locations during the missing time period [Figure 2].

3. El Nino and La Nina



We utilized nonparametric time series and spatial models to obtain general trends of sea surface temperature (SST).

Models :

1. Seasonal Decomposition of Time Series by Loess

$$\text{SST}(n) = \text{Trend}(n) + \text{Seasonal}(n) + \text{Error}(n)$$

where $n = 1, 2, \cdots, 72$, each month from Jan 1995 to Dec 2000.

2. Smoothing Spline Anova

$$\text{Mean SST} = f_1(\text{lati}) + f_2(\text{long}) + f_{1,2}(\text{lati}, \text{long}) + \text{Error}$$

where f_1, f_2 are nonparametric functions for main effects and and $f_{1,2}$ for an interaction, respectively.

Procedure :

1. SST values were decomposed into three parts - trend, seasonal effect and error.
2. SST values adjusted for seasonal effect were averaged over time on each grid location and then were smoothed by Smoothing Spline ANOVA.

Results :

1. The time periods at the highest and lowest surface temperature levels on location 1, 2 and 3 correspond with El Nino (1997 - 98) and La Nina (1995 - 96, 1998 - 99) periods, respectively [Figure 3B, C, D].
2. The location 3 showed the lowest mean surface temperature and the constant trend over time [Figure 3A, D]. The location 3 corresponds with so-called cold- water upwelling areas along the Peru and Chile coasts.

4. Ozone Depletion Areas

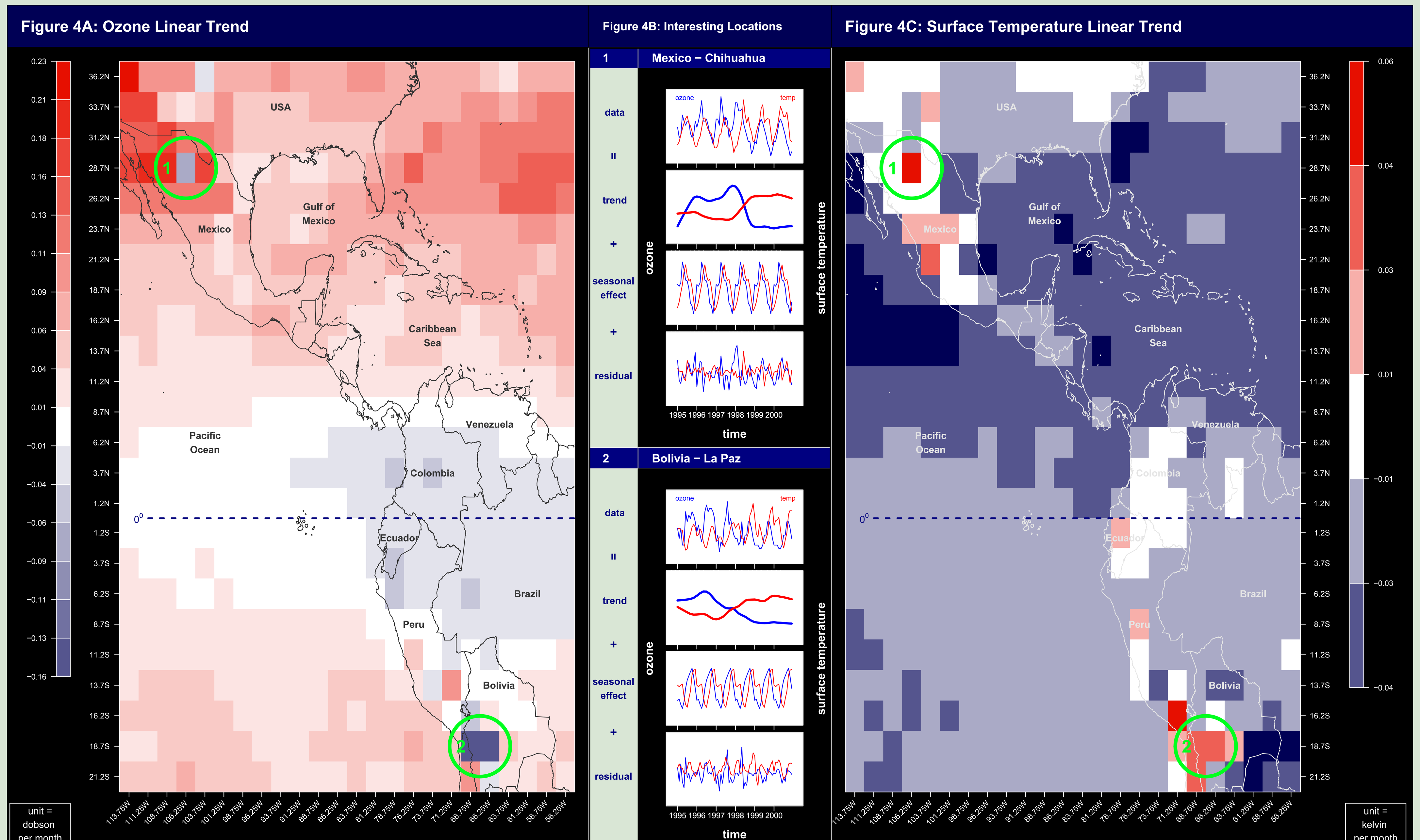


Figure 4A: Ozone Linear Trend | Figure 4B: Interesting Locations | Figure 4C: Surface Temperature Linear Trend

We considered a linear model and a nonlinear time-series model to find abnormal ozone trends.

Models :

1. Linear Model with AR(1) Error

$$\text{Adjusted Ozone} = \beta_0 + \beta_1 \text{Time} + \text{Error}$$

where $\text{Error} \sim \text{AR}(1)$.

2. Seasonal Decomposition of Time Series by Loess

$$\text{Ozone}(n) = \text{Trend}(n) + \text{Seasonal}(n) + \text{Error}(n)$$

where $n = 1, 2, \cdots, 72$, each month from Jan 1995 to Dec 2000.

3. K-means Clustering Algorithm

Procedure :

1. Linear models with AR(1) error were fitted to ozone and surface temperature adjusted for their seasonal effects on each grid location over time, and both linear trends were compared on location 1 and 2 [Figure 4A, C].
2. Nonlinear ozone trends were obtained by Seasonal Decomposition of Time Series by Loess, and K-means clustering algorithm was utilized to cluster these nonlinear ozone trends, minimizing within-cluster dissimilarity [Figure 4D].

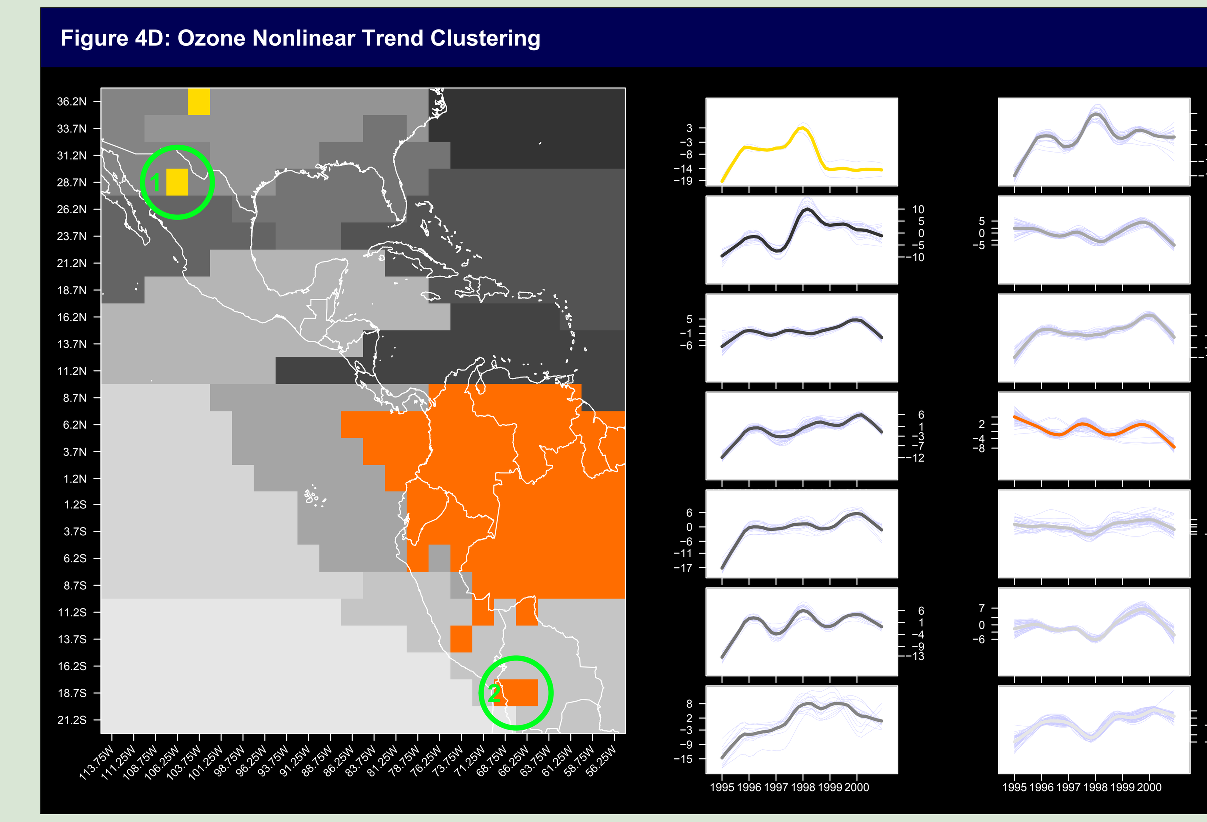


Figure 4D: Ozone Nonlinear Trend Clustering

Results :

1. Two interesting locations were found from Figure 4A and Figure 4C. Location 1, surrounded by locations in linear increasing trends, showed a linear decreasing trend. Location 2 showed the fastest linear decreasing trend.
Inside location 1, there is a city, Chihuahua, where air and water pollution have been severe, and inside location 2, there is a city, La Paz where an ozone hole has been reported.
In these two locations, surface temperature showed opposite linear trends [Figure 4C].
Figure 4B shows more details on trends.
2. Location 1 and 2 were also identified as distinct groups with nonlinear trends by using K-means clustering algorithm [Figure 4D].

5. Cloud Effect on Temperature

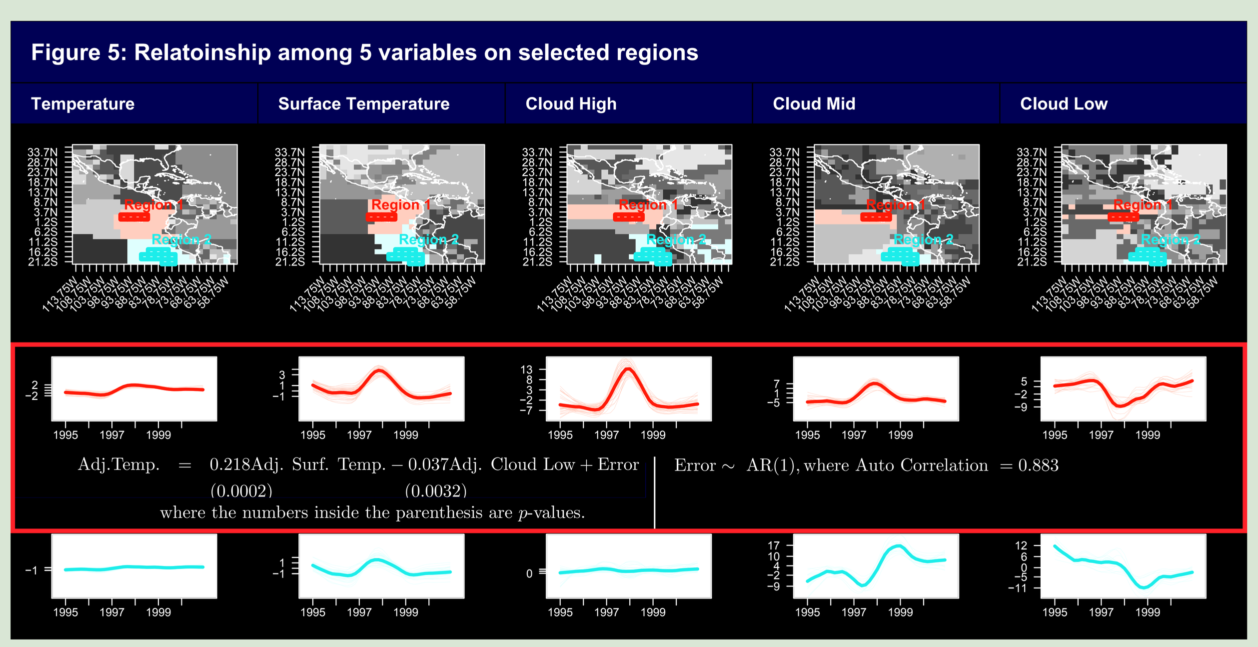


Figure 5: Relatoinship among 5 variables on selected regions

We tried to find cloud effect on temperature by considering the linear relationship among temperature, surface temperature and clouds.

Models :

1. Seasonal Decomposition of Time Series by Loess
2. K-means Clustering Algorithm
3. Linear model with AR(1) Error

Procedure :

1. For each variable, nonlinear trends were obtained by using Seasonal Decomposition of Time Series by Loess and were classified by using K-means clustering algorithm.
2. Two overlapped regions were selected where region 1 was most and region 2 least influenced by El Nino.
3. Fit a linear model with AR(1) error to values adjusted for seasonal effects to find a relationship between temperature and other variables on region 1 and 2 where values within each region were averaged.

Results :

1. Cloud mid and Cloud low showed the negative correlation on both regions.
2. On region 1, the positive linear relationship between temperature and surface temperature and the negative linear relationship between temperature and cloud low were found.
3. On region 2, no statistically significant model was found. Note that the variability of temperature is very small. Confounding may exist among covariates.

6. Summary and Conclusion

1. Missing values in cloud low were imputed by using spatial smoothness.
2. El Nino and La Nina were detected by using nonparametric trend estimation.
3. In ozone depletion areas, ozone and surface temperature showed opposite trends.
4. Statistically significant linear relationships among temperature, surface temperaure and cloud low were found.

7. Reference

1. Luo, Z, Backfitting in Smoothing Spline ANOVA, With Application to Historical Global Temperature Data, University of Wisconsin - Madison TR 964, July 1996. PhD. Thesis.
2. Paul Murrell, R Graphics, 2006, Chapman & Hall/CRC.
3. Trevor Hastie et al, The Elements of Statistical Learning, 2001, Springer.