

# **Student Projects on Topics in Statistical Analysis and Interpretation of Geophysical Data sets**

Edited by Sergey Kravtsov\*

University of Wisconsin-Milwaukee  
Department of Mathematical Sciences  
Atmospheric Science Group

2008 Spring Semester

May 12, 2008

---

\* Department of Mathematical Sciences, Atmospheric Science Group, University of Wisconsin-Milwaukee, P. O. Box 413, Milwaukee, WI 53201. *E-mail:* [kravtsov@uwm.edu](mailto:kravtsov@uwm.edu); *URL:* [www.uwm.edu/~kravtsov](http://www.uwm.edu/~kravtsov); *Fax:* 414-229-4907.

## **Preface**

This volume documents a series of studies done by atmospheric science graduate students as term projects for the Spring 2008 ATM SCI 950 seminar. All of these studies are directly related to students' thesis projects and address a variety of meteorological/climatic topics. The first three papers are based on application of the discriminant analysis to three different problems: (1) identification of robust modes of low-frequency variability in global surface temperature instrumental records; (2) classification of Atlantic hurricane images and prediction of hurricane intensities; and (3) cluster analysis of air-parcel trajectories associated with extreme pollution episodes. The next set of three papers deals with climatic variability over the North Atlantic and continental US regions: (4) paper #4 studies decadal-to-interdecadal climate signals presumably originating as a result of oceanic intrinsic variability and/or air-sea interaction in the North Atlantic; (5) next, possible regional manifestations of these modes in the Great Lakes water levels are discussed; while (6) paper #6 concentrates on synoptic time scales and studies the relationship between weather conditions and air pollution. Finally, the last paper (7) discusses the issue of long-range weather predictability by looking at anomalously persistent weather patterns using a combination of statistical techniques. The result is a very interesting sample of projects, which at the same time give a general idea of widely adopted statistical methodologies currently used in atmospheric/climate research.

*Sergey Kravtsov*

# 1. Interdecadal Climate Variability In Global Sea-Surface Temperature Observations

Dawn Thielke

*Atmospheric Sciences Program, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin*

May 5, 2008

## ABSTRACT

More than a century of monthly sea-surface temperature observations on a global grid are massaged to diagnose the interdecadal variability. A discriminant analysis was used to separate the variability on interdecadal timescales from variations that occur over time periods of less than a decade. Canonical variates are computed by maximizing the ratio of interdecadal to intradecadal variance; the spatial patterns associated with the canonical variates are called discriminating patterns.

The leading canonical variate shows a significant warming trend during the 20th century and the associated spatial pattern illustrates a warming of the Southern Hemisphere relative to the Northern Hemisphere. Both the location of the observed warming and the temporal rise in temperature are consistent with previous studies that have identified the highest signal of warming in the tropical oceans. The second canonical variate has an oscillatory character and its associated spatial pattern is dominated by basin-wide uniform temperature anomalies in the North Atlantic, with teleconnections to the Pacific and southern oceans. This is suggestive of an internal mode of climate variability known as the Atlantic Multidecadal Oscillation (AMO).

## I. Introduction

Earth's climate is complex and involves interactions between the atmosphere, ocean and land surfaces, even animals and humans. One element of climate, temperature, has varied over the course of Earth's history due to internal variability within the climate system as well as natural, external forcings such as solar irradiance and variations in Earth's orbital characteristics. Alterations to the composition of the atmosphere due to natural causes or from influence by humans can result in changes within the climate system too (Cubasch et al. 2001). Since the late 18th century, the burning of fossil fuels and modifications in agriculture and manufacturing have significantly increased the levels of greenhouse gases and sulfate aerosols in our atmosphere, likely impacting global temperature patterns (Schneider & Held 2001).

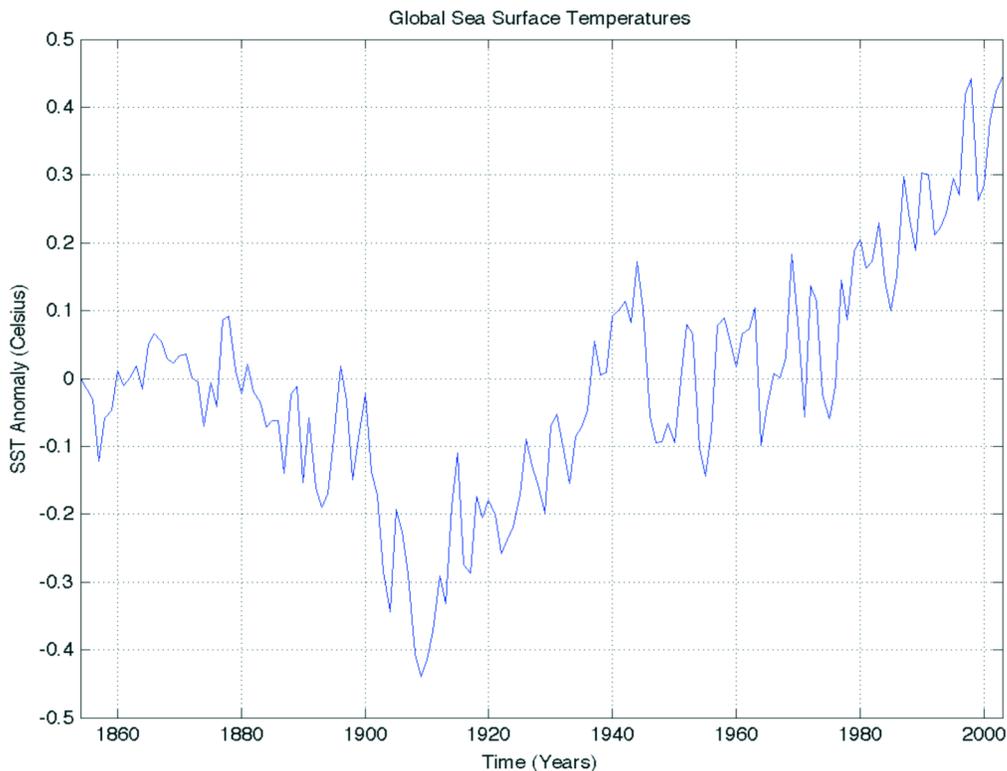


Figure 1. Global sea-surface temperature anomalies as a function of time for the years 1854 to 2003.

The variation in the global sea-surface temperatures from 1854 to 2006 is depicted in Figure 1, illustrating how temperatures have alternated between warm cool phases for the majority of the temperature record. However, in the last few decades a notable warming trend can be observed which is difficult to associate with natural climate variability alone. The ability

to separate changes in the global temperature pattern into natural and forced variations is imperative to understanding the effects of increases in greenhouse gas concentrations on our climate (Ghil & Vautard 1991).

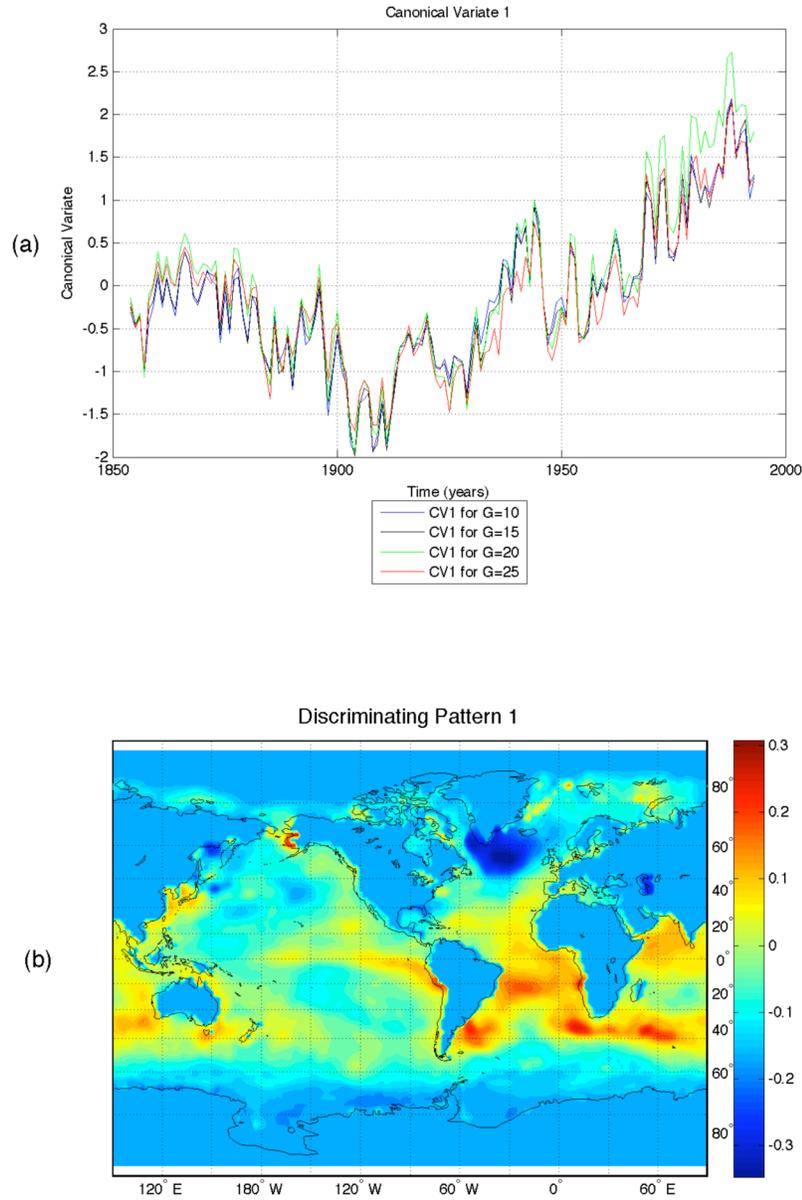


Figure 2. (a) The leading canonical variate for group sizes of 10, 15, 20 and 25 is shown and represents interdecadal temperature variations within the sea-surface temperature data. (b) Discriminating pattern as a spatial representation of the interdecadal sea-surface temperature variations for the years 1854 to 2003. The color shading represents the sea-surface temperature anomalies (in degrees Celsius) for each point on the global grid.

However, changes in temperature trends can occur over a few years (intradecadal) or on timescales of a few decades (interdecadal) and fluctuations in temperature patterns over a short time can obscure trends that are occurring over longer timescales. For that reason, it is useful to separate the low frequency variability (signal) within the temperature record from the high frequency variability (noise). In this study, we differentiate between the interdecadal and intradecadal variations in temperature by means of discriminant analysis.

## **II. Data and Methods**

Since we desired to examine the structure between two pre-defined data groups, a discriminant analysis was used as a temporal filter to diagnose the interdecadal variability within a temperature dataset. We used sea-surface temperature data from 1854-2006 and restricted the data to the first 150 years. Anomalies were formed, thus the seasonal cycle was removed and the temperatures were then area-averaged over the spatial grid defined by the region 60S to 60N. The temperature data was then divided into equally populated groups. Since the choice of group size is subjective, various group sizes of 10, 15, 20 and 25 were used and several discriminant analyses were performed.

Canonical variates, which are the linear combinations of the temperature variables that maximize the ratio of interdecadal to intradecadal variance between groups, were calculated through eigenvector decomposition. The discriminating patterns were found by projecting the original temperature data onto the canonical variates, giving us a spatial representation of the interdecadal temperature variations. High correlation was found among the different canonical variates, indicating that the results are robust and they are not dependent on group size (Figures 2a, 3a). For that reason, we chose to focus on the canonical variates and discriminating patterns found with a group size of 15, meaning that a group consists of one decade of temperature observations. The first and second canonical variates yielded variance ratios, on average, of about 6 and 2.5 respectively, with the remaining canonical variates producing variance ratios of less than one. As a result, analysis was limited to the first two discriminating spatial and temporal patterns as these represent the maximal variance between the previously determined data groups.

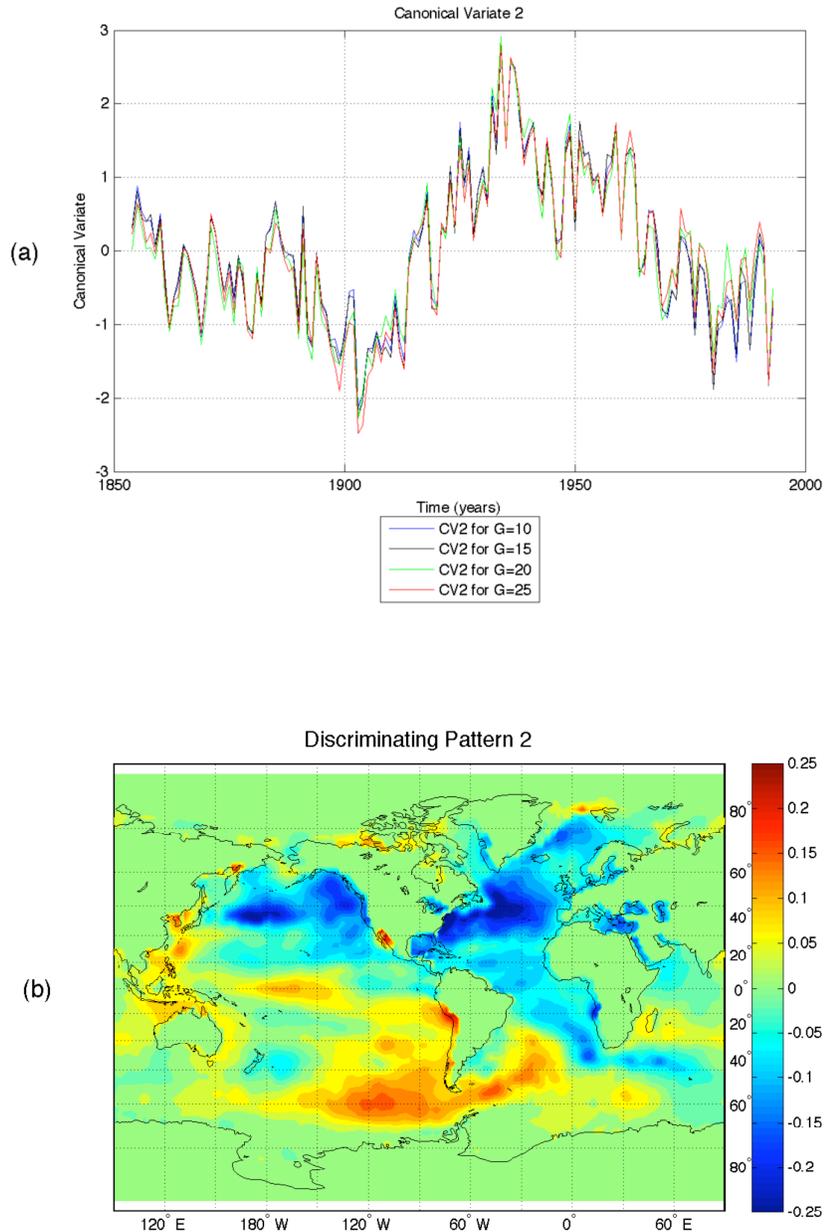


Figure 3. (a) and (b) As in Figure 2a,b but for the second canonical variate and discriminating pattern, respectively.

### III. Results

The first canonical variate (Figure 2a) shows considerable warming over the last several decades. Since we removed the variability on timescales of a decade or less, this variate shows a significant interdecadal variation. Similarly, the first discriminating pattern (Figure 2b) shows

warming in the Southern Hemisphere and localized cooling in the northern North Atlantic. The warming in the Southern Hemisphere relative to the cooling in the Northern Hemisphere agrees with previous climate studies which have shown that the strongest warming signal can be found predominantly in the tropical oceans (Trenberth et al. 2007). The localized cooling in the northern North Atlantic is consistent with what recent climate models have shown to be a deceleration of the thermohaline circulation as a result of greenhouse gases increasing in concentration, resulting in a reduced level of heat transport poleward. Although the slow down of the thermohaline circulation influences temperature patterns in both the Northern and Southern Hemispheres, the effects are most significant over the North Atlantic (Figure 2b) (Cubasch et al. 2001).

The second canonical variate and its associated discriminating pattern mimic internal modes of variability within the climate system. The second canonical variate has an oscillatory nature, with a timescale of several decades (Figure 3a). The discriminating pattern is dominated by basin-wide uniform anomalies in the North Atlantic, with teleconnections to the Pacific and southern oceans. Since the middle of the 19th century, North Atlantic sea-surface temperatures have fluctuated between warm and cool phases over multidecadal timescales (Trenberth, 2007). Therefore, the uniform temperature anomalies in the North Atlantic can be attributed to an internal mode of variability within the climate system known as the Atlantic Multidecadal Oscillation (AMO), which describes long term variations in the sea-surface temperatures over the North Atlantic (Enfield et al. 2000).

#### **IV. Conclusions**

The interdecadal variability within observational sea-surface temperature data over the past 150 years was revealed. The leading canonical variates and discriminating patterns provided observational evidence for both natural and anthropogenic influences on the climate over interdecadal timescales. However, a more thorough understanding of the natural variability that occurs on interdecadal timescales is crucial to determining the anthropogenic impacts on global climate patterns. Furthermore, because the leading discriminating pattern (Figure 2b) showed localized cooling, it should also be investigated as to whether interdecadal climate variability can be detected on smaller spatial scales. For instance, one could choose to investigate the temperature changes in a particular region and look at carbon dioxide and sulfate

aerosol concentrations independently to infer any relationship between temperature fluctuations and increases in greenhouse gas and sulfate aerosol concentrations.

Furthermore, a solid understanding of natural, multidecadal modes of variability within the climate system is essential to establishing anthropogenic effects on climate changes. The fact that this analysis was able to minimize the frequent, intrinsic variability and reproduce interdecadal variations in the sea-surface temperature pattern quite well, could be useful in identifying successful climate models based on whether or not a model is able to duplicate these results. If a model is able to mimic the natural, multidecadal variability within the climate system, then it could be used to identify forced trends in the climate.

## References

- Cubasch, U., G.A. Meehl, G.J. Boer, R.J. Stouffer, M. Dix, A. Noda, C.A. Senior, S. Raper and K.S. Yap, 2001: Thermohaline Circulation Changes. *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* [J.-W. Kim and J. Stone (eds.)].
- Enfield, David B., A.M. Mestas-Nunez, P. J. Trimble, "The Atlantic Multidecadal Oscillation and its Relation to Rainfall and River Flows in the Continental U.S." *Geophysical Research Letters* 28 (2001): 2077.
- Ghil, M. and R. Vautard. "Interdecadal Oscillations and the Warming Trend in Global Temperature Time Series." *Nature* 350 (1991): 324.
- Latif, M. and T. P. Barnett. "Decadal Climate Variability over the North Pacific and North America: Dynamics and Predictability." *Journal of Climate* 9 (1996): 2420.
- Schneider, Tapio, and Isaac M. Held. "Discriminants of Twentieth-Century Changes in Earth Surface Temperatures." *Journal of Climate* 14 (2001): 249-254.
- "Science-El Nino/La Nina & PDO". *NASA Jet Propulsion Laboratory*. 18 Sep. 2006. California Institute of Technology. 10 Apr. 2008 <topex- www.sealevel.jpl.nasa.gov/science/pdo.html>.
- Smith, T.M., and R.W. Reynolds, Improved reconstruction of SST (1854-1997). *J. Clim.* 17, 2466-2477 (2004). Trenberth, K. E., P. D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. K. Tank, D. Parker, F. Rahimzadeh, J. A. Renwick, M. Rusticucci, B. Soden and P. Zhai, 2007: Observations: Surface and Atmospheric Climate Change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)].

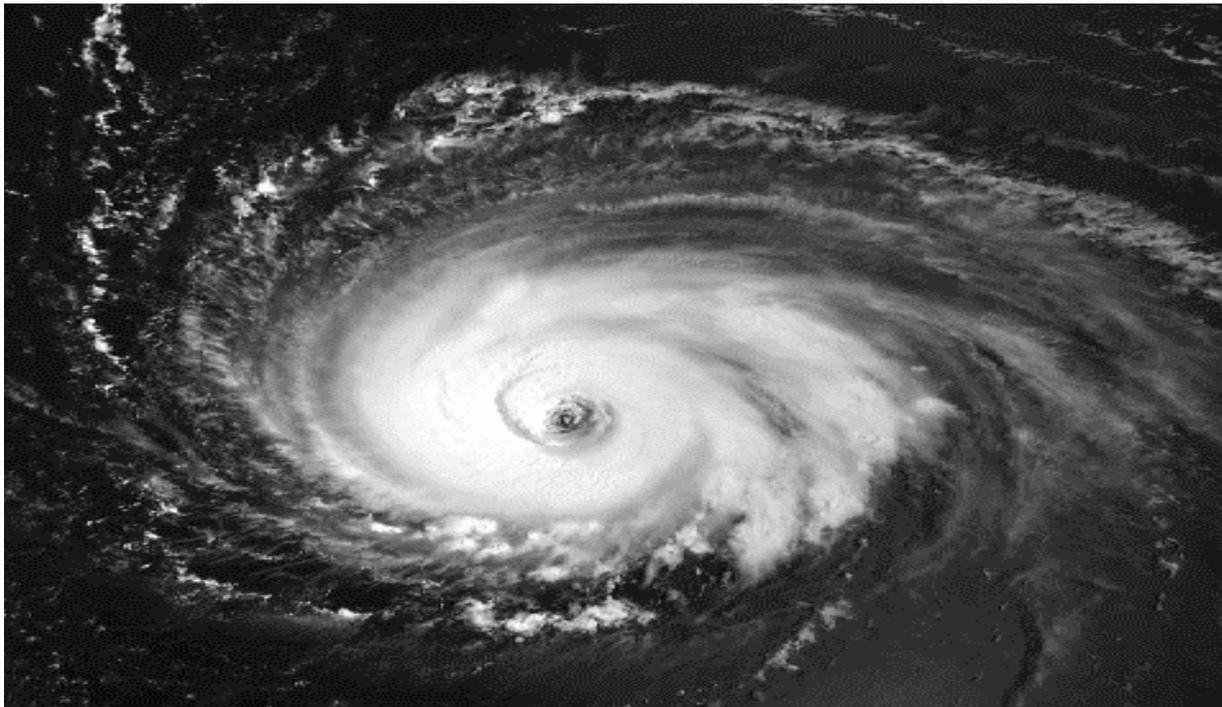
## 2. Linking Hurricane Patterns to Intensity Trends

**Abstract:** Infrared hurricane images are analyzed using matrix methods to associate their intensity tendencies with hurricanes' specific flow-pattern anomalies.

Jennifer Knapkiewicz

May 5, 2008

Hurricane Iris, 1995



**Introduction:** For thousands of years, hurricanes have tormented man with a significant loss of life and property. From hurricanes on Columbus' voyage, to the Great Galveston Hurricane of 1908, and to hurricane Katrina in 2005 people have attempted to forecast intensities and track hurricanes with modest success. Although forecasting technique has come a long way from watching sea swells and clouds to using hurricane satellite images, the forecast skill, for both hurricane intensity and path still remains quite low (Bottger et al., 1975). In the present study we will concentrate on now-casting hurricane intensities based solely on the storm's satellite image. Our method provides an appealing alternative to measuring the wind speed directly by flying reconnaissance planes into the hurricane, which is costly and time consuming.

Dvorak (1975) developed a method of analyzing visible satellite images to determine current and future hurricane intensity. His system was used extensively in the past and is still used today in a somewhat more advanced form. The method is a variation of a constructed analog (CA) prediction, in which a current satellite image of a hurricane is compared and matched to hurricane images, either observed in the past or simulated in the models, at various stages of their life cycles. The forecaster would then choose the image that matches best with the current hurricane and apply the best-match past flow chart to the current conditions to infer hurricane intensities and short-term (24-hr) trends. Although operationally useful, Dvorak's method lacks accuracy and objectiveness. In the present project, we analyzed a dataset which tracks the evolution of six different hurricanes, and applied matrix methods to construct an objective scheme for inferring hurricane intensity based solely on a hurricane's spatial pattern.

**Data:** We analyzed infrared images of six different hurricanes (Marilyn, Humberto, Iris, Karen, Felix, and Noel) archived in the MeteoSat data set (MeteoSat data archives). This data set also provides the measured hurricane intensities in terms of the maximum wind speed. All of the hurricanes analyzed came from the active 1995 Atlantic Basin season and were chosen based on track length, data accessibility, location, and life cycle. The original 1,210 images were then limited to those that stayed within the Atlantic basin south of 30°N and east of 90°W. Each image was cropped to a square 80 × 80-km region centered on the hurricane's eye. These remaining images were then limited once again to only include hurricanes with a maximum wind speed greater than 64 knots. This final step limited the original data set to M=141 individual hurricane snapshots on a 41 x 41 grid with a 2-km spatial resolution.

**Analysis Procedure:** The data was formed into a  $M \times N$  matrix whose  $M=141$  rows represented spatial maps (total of  $N=1681$  data points) associated with a single hurricane snapshot. The rows were then sorted based on maximum wind speeds so that the first row corresponded to the snapshot of the least intense hurricane, the second row – to that of the second least intense hurricane, and so on. The column-wise means were then removed; the resulting ensemble-averaged hurricane image is shown in Fig. 1.

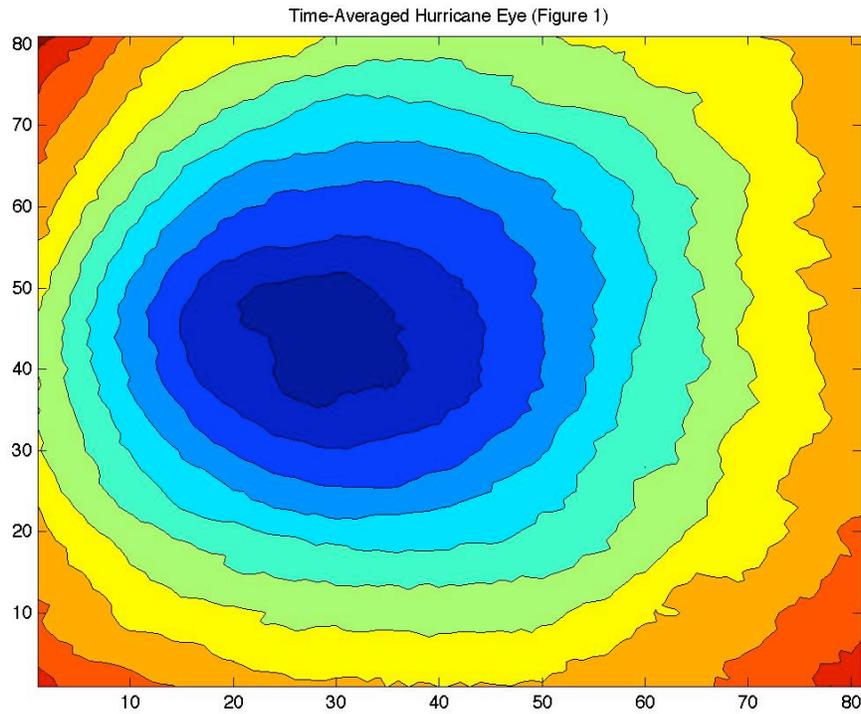
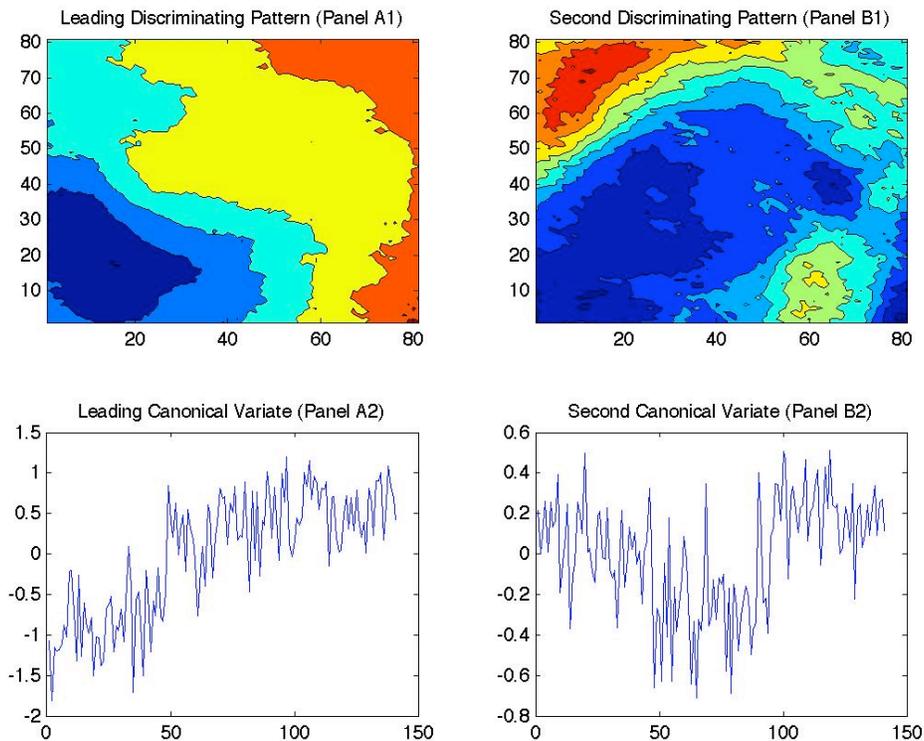


Figure 1: Ensemble-averaged hurricane image on the 80 x 80 km grid centered on the hurricane’s eye.

We then used discriminant analysis (Schneider and Held 2001) to identify spatial patterns in the hurricane-image anomalies (deviations from ensemble-mean hurricane image), which optimally characterize the differences in maps of hurricanes with different intensities. To do so, the data matrix of sorted hurricane image anomalies is first split into  $G$  equal-sized groups to form a  $G \times N$  matrix of hurricane-anomaly group means. The discriminant analysis finds a set of  $N$  weights, called the leading discriminating pattern, to form a series of length  $M$  of the leading canonical variate ( $CV-1$ ) by adding weighted columns of our original  $M \times N$  sorted hurricane-

anomaly matrix. The  $G$ -valued time-series of  $CVg-1$  of weighted group means can be obtained in the same way (and using the same weights) from the matrix of hurricane-anomaly group means. The weights are chosen so that the ratio of the variance of  $CVg-1$  to the variance of  $CV-1$  is maximized. The next discriminating pattern and canonical variate are found in the same way subject to the  $CV-2$  being orthogonal to  $CV-1$ , and the total of  $G-1$  discriminating patterns and canonical variates can be obtained analogously.

The discriminating patterns thus identify directions, in the  $N$ -dimensional space, in which the ratio  $R$  of intergroup-to-intragroup variability is maximized. The patterns characterized by  $R$ -values exceeding unity are optimal in describing differences between hurricanes of various intensities in the sense of being least affected by the sampling (intragroup) variability. We will use multiple linear regression of hurricane intensity  $I$  onto the canonical variates associated with leading discriminating patterns to construct an objective now-casting scheme of inferring hurricane intensity based on the satellite snapshot of the hurricane.



**Figure 2:** Leading discriminating patterns (panels A1 and B1) and associated canonical variates (panels A2 and B2).

**Results:** The method was applied using  $G=3, 5, 7,$  and  $10$  groups corresponding to group sizes of  $47, 28, 20, 14$  hurricane images. Figure 2 shows two leading discriminating patterns,  $PAT-1$  and  $PAT-2$ , as well as canonical variates  $CV-1$  and  $CV-2$  computed for  $G=3$  (the  $R$  ratios for the two patterns are  $3.66$  and  $1.43$ , respectively). Both discriminating patterns exhibit symmetry about southeast–northwest axis, with dipolar  $PAT-1$  and tri-polar  $PAT-2$ . Both  $CV-1$  and  $CV-2$  are dominated by fairly monotonic long-term trends, which make them potentially useful predictors of hurricane intensity (recall that the hurricane images were sorted according to intensities and therefore the intensity  $I$  monotonically increase over the set of  $141$  images).

In order to construct our predictive scheme to now-cast hurricane intensities, we use the canonical variates as predictors in a multiple linear regression model

$$I = b_1 CV_1 + b_2 CV_2 + r, \quad (1)$$

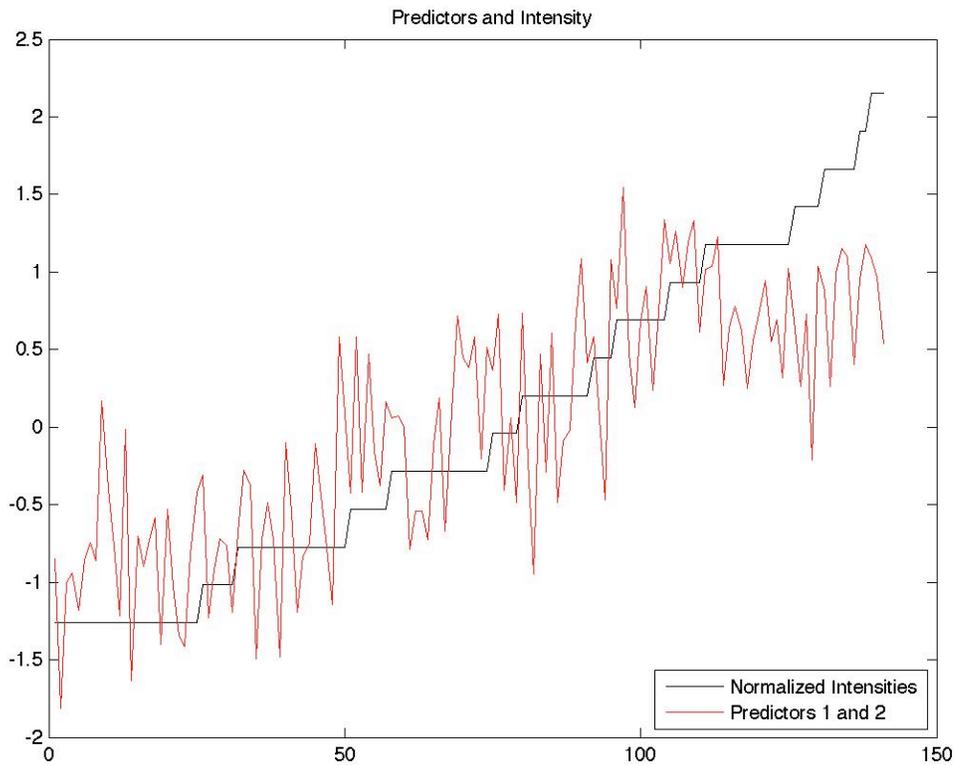
where  $r$  is the regression residual, and the coefficients  $b_1, b_2$  are found by the method of least squares. The fraction of variance in hurricane intensity explained by the fit (1) is found as

$$v^2 = (1 - \text{Var}(r)/\text{Var}(I)) \times 100\%. \quad (2)$$

We show in Fig. 3 the actual hurricane intensities along with the prediction of the regression model (1), that is, the quantity  $b_1 CV_1 + b_2 CV_2$ . For this particular fit, the fraction of explained variance  $v^2=62.3\%$ . Note that the intensity in the low-to-medium range are predicted better than the strongest intensities, which are consistently underestimated by the regression fit (1).

We have checked the robustness of the above results using various group sizes and various number of canonical-variate predictors used in the regression model. In each case, the leading two canonical variates were similar to those shown in Fig. 2, while the increase in the forecast skill due to using additional predictors was fairly minor, across the whole range of hurricane intensities.

The above results identify the  $G=3$  prediction scheme as optimal. The scheme can be applied as follows: A real-time hurricane image is projected onto our leading two discriminating patterns (see Fig. 2), thus determining the values of  $CV-1$  and  $CV-2$ . The regression model (1) is then used to provide an estimate of the current intensity as  $I = b_1 CV_1 + b_2 CV_2$ . Thus, we have constructed a method for objective estimation of hurricane intensities using satellite imagery.



**Figure 3:** Normalized intensities (solid black line) and our regressed predictors (solid red line).

**Summary and future work:** Current methods of determining hurricane intensities are often costly, subjective, and unreliable. In this project an efficient and objective discriminant-analysis method for determining hurricane intensity was developed and tested using a set of six hurricanes from the 1995 Atlantic hurricane season. The method now-casts hurricane intensity based solely on the hurricane's satellite image and is comparable with that of alternative techniques.

The major goal of this research is to move forward toward a more accurate twenty-four hour prediction for maximum wind speed using a current satellite image. This can be achieved by regressing the intensity tendencies rather than the current intensities.

An additional possibility is to apply our procedure to Pacific hurricanes. The dynamics that drive Pacific hurricanes differ, sometimes greatly, from those that drive Atlantic Basin storms and intensity estimates are often determined using the same methods globally. Our method can be redeveloped using a database of Pacific storms and tested for accuracy when determining the intensity of these storms.

Applying these methods to historic hurricane data sets may also yield other results. It has long been wondered if global-warming has affected the strength and frequency of hurricanes. We will be able to apply our technique to storms from differing decades to determine if there is an intensity increase. Inferences can then be made as to the cause of this increase; either it be actual, or an artifact of the technique used to determine the intensity.

## References

- Bottger, H., Eckardt, M., and Katergiannakis, U., 1975: Forecasting Extratropical Storms with Hurricane Intensity Using Satellite Information. *Mon. Wea. Rev.*, **103**, 1259.
- Dvorak, V., 1975: Tropical Cyclone Intensity Analysis and Forecasting from Satellite Imagery. *Mon. Wea. Rev.*, **103**, 420.
- MeteoSat Data Archives: <http://badc.nerc.ac.uk/data/meteosat/>
- Schneider, T., and Held, I., 2001: Discriminants of Twentieth-Century Changes in Earth Surface Temperatures. *Journal of Climate*, **14**, 249–254.

### **3. Backward trajectory analysis and its statistical significance using discriminant analysis**

Laura Carnahan

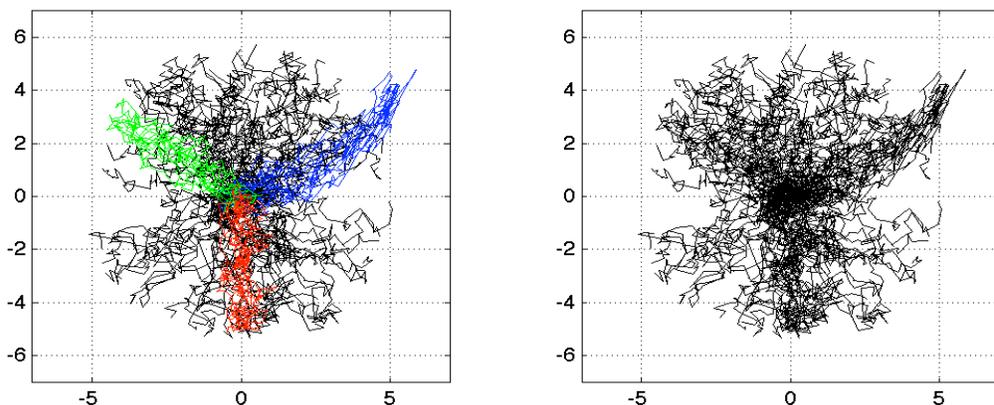
May 5, 2008

**Abstract:** The ability to attribute certain ending temperatures to specific air paths using backwards trajectories from the NOAA's HYSPLIT website is explored through the technique of discriminant analysis. Trajectories are grouped into ten groups, maximizing the variation between groups while minimizing the variation within each group. This method identifies a specific path warm parcels take to the receptor site during the summer, but shows that anomalously cold winter days are associated with a wide variety of synoptic conditions.

**Introduction:** It is often important to know from where specific pollutants have come to a given location. This information can be helpful in determining if companies meet Environmental Protection Agency standards, protecting wildlife and other historical relics, and in making sure the population of an area is safe. The idea of tracing pollution back to a source requires knowledge of the concentration and composition of pollution at the receptor site, as well as knowledge of the path the air took to get to the receptor. One of the widely used and important techniques to do so is cluster analysis. In cluster analysis, one groups trajectories of air parcels according to certain characteristics, such as the distance between the trajectories, the length of the trajectory, or the curvature of the trajectories (Moody 1991, Harris 1992, Sirois and Bottenheim 1995, and Kahl 1997). This study uses a novel clustering technique, which identifies recurrent parcel trajectories based on the values of air temperature at a particular receptor in southeastern Wisconsin.

**Data Sets:** In this study, we used temperature data sets from the National Climatic Data Center (NCDC)'s "Cooperative Summary of the Day TD3200" records. The location of the receptor was chosen to be located in Lake Geneva, Wisconsin. The associated air trajectories tracking the location of air parcels for the period of 72 hours prior to their arriving to the receptor site, with

one-hour resolution, were obtained from the HYSPLIT website, run by the National Oceanic and Atmospheric Administration (NOAA) group. We have analyzed one hundred independent trajectories for (i) the winter season, and (ii) the summer season. In each case, fifteen of the trajectories represented anomalously warm (in summer) and anomalously cold (in winter) cases, while the rest of the trajectories were chosen randomly. Our objective was to establish whether the anomalous receptor temperatures are associated with a preferred air path.

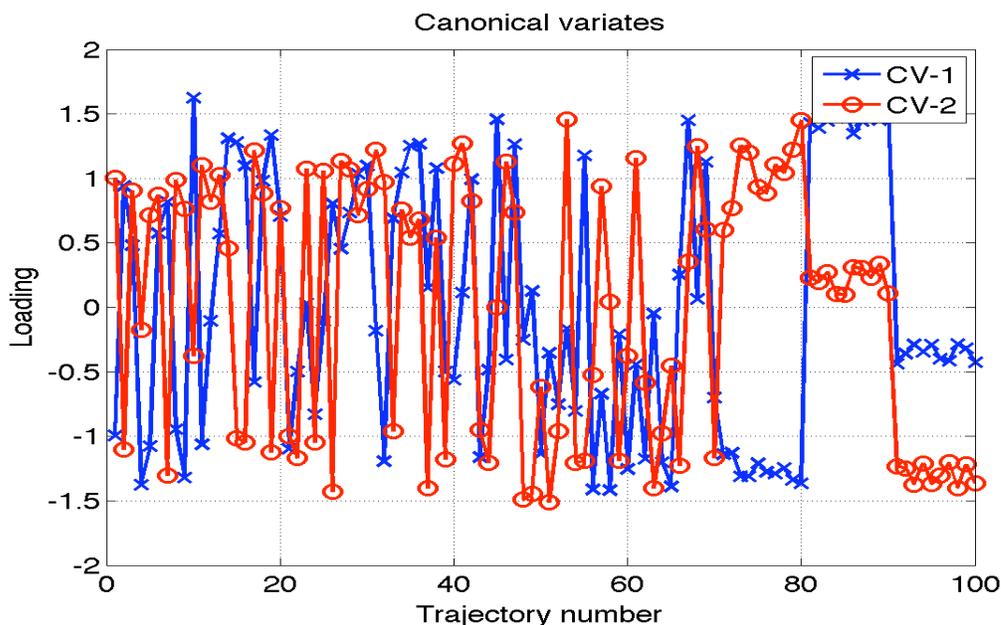


**Fig. 1.** Synthetic data set. Right panel: “clustered” trajectories 71-80 in red, 81-90 in blue, and 91-100 in green. Left panel: plot of all the synthetic trajectories in black.

**Method:** For a given season, the 2-D trajectories (latitude and longitude) were sorted according to the temperature of the trajectory at the receptor site to form a 100 x 146 data matrix whose first row contained the vector of parcel coordinates  $(x_1, x_2, \dots, x_{73}, y_1, y_2, \dots, y_{73})$  for the trajectory with the coldest (among 100 cases) receptor temperature. The second row contained coordinates of the "second coldest" trajectory, etc. Here,  $x_1$  and  $y_1$  are the longitude and latitude of the air parcel 72 hours prior to arrival at the receptor,  $x_2$  and  $y_2$  are the longitude and latitude of the parcel 71 hours before arrival, and so on, with  $x_{73}$  and  $y_{73}$  being the longitude and latitude of the receptor.

We then applied to the matrix of sorted trajectory cases, the technique of discriminant analysis (Schneider and Held 2001) by dividing the cases into ten groups (of ten cases each) and identifying the groups of trajectories with small within-group variability, which are, at the same time, significantly different from neighboring groups. The technique involves finding the linear

combination of 146 columns of trajectory-case matrix, which form a vector  $\mathbf{c}$  (of size 100) of leading canonical variates such that the ratio of its intergroup-to-intragroup variance is maximized, which is equivalent to maximizing the ratio of intergroup-to-total variance. The next canonical variate was found in the same way, subject to being orthogonal to the first one, etc. We then identified the groups which were characterized by minimum intragroup variability and stood out in the background of neighboring groups' intragroup spread.



**Fig. 2.** Two leading canonical variates one for the synthetic trajectories. The last three groups (trajectories 71-80, 81-90, and 91-100) are identified as trajectory clusters.

## Results:

**A. Synthetic Data Set.** In order to test our technique, we first created a set of 100 synthetic trajectories in which trajectories 1-70 were random, while trajectories 71-80, 81-90, and 91-100 had specific paths, slightly perturbed by red noise deviations (see Fig. 1). The results of discriminant analysis applied to this set of trajectories are shown in Fig. 2 and Fig. 3.

The two leading canonical variates are shown in Fig. 2. It is immediately obvious that the loadings of the last three groups are very similar within each group, and at the same time, significantly different from the neighboring groups (intragroup spreads between groups 8, 9, and 10 do not overlap). Another way to visualize the same result is used in Fig. 3, which shows the

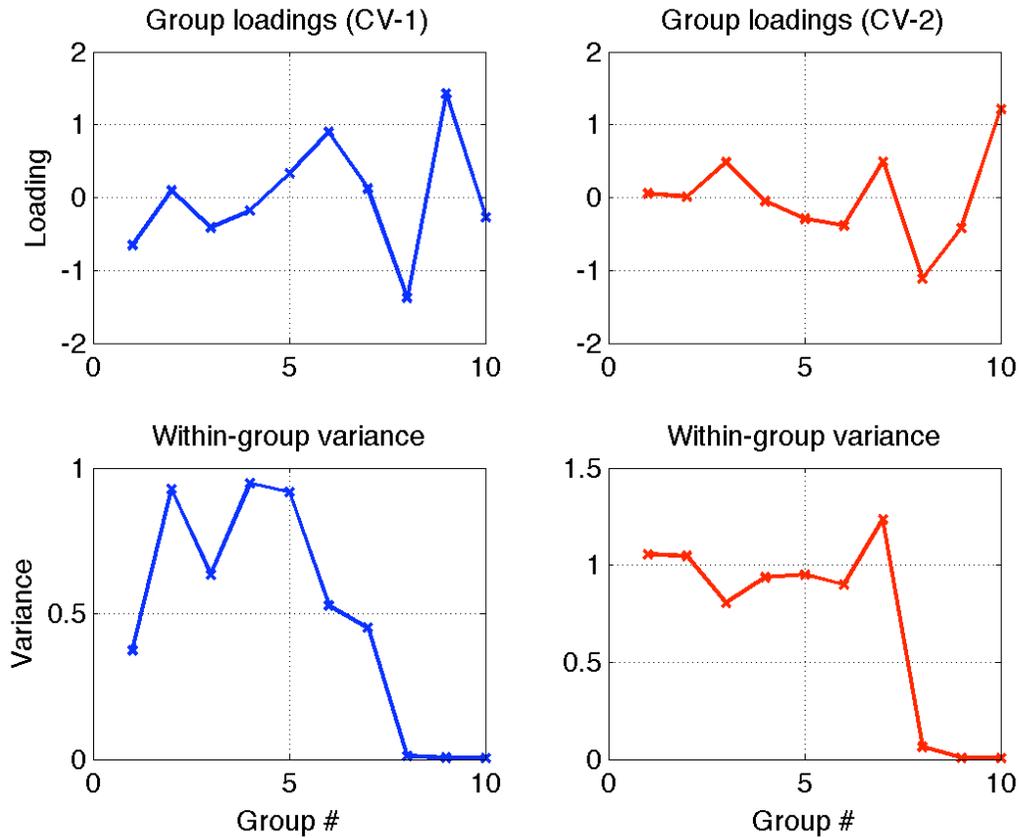


Fig. 3. Group loadings (upper row) and variances (lower row). Left column: CV-1; Right column: CV-2.

canonical variates averaged within each group (upper row) and each group's intragroup variances (lower row), for the two leading canonical variates. Once again, the last three groups are seen to have the loadings much different from those of the neighboring groups (upper row), whereas their intragroup variance is much smaller than that of other groups (lower row). The information contained in Fig. 2 and Fig. 3 is best visualized as an errorbar plot (Fig. 4), which shows group-averaged loadings with errorbars representing intragroup variances, for each of the leading canonical variates. This plot also shows that the last three groups represent distinct trajectory clusters (with small intragroup variances, not overlapping with the variable loadings of neighboring groups), thus indicating the method's success for our synthetic data set analysis.

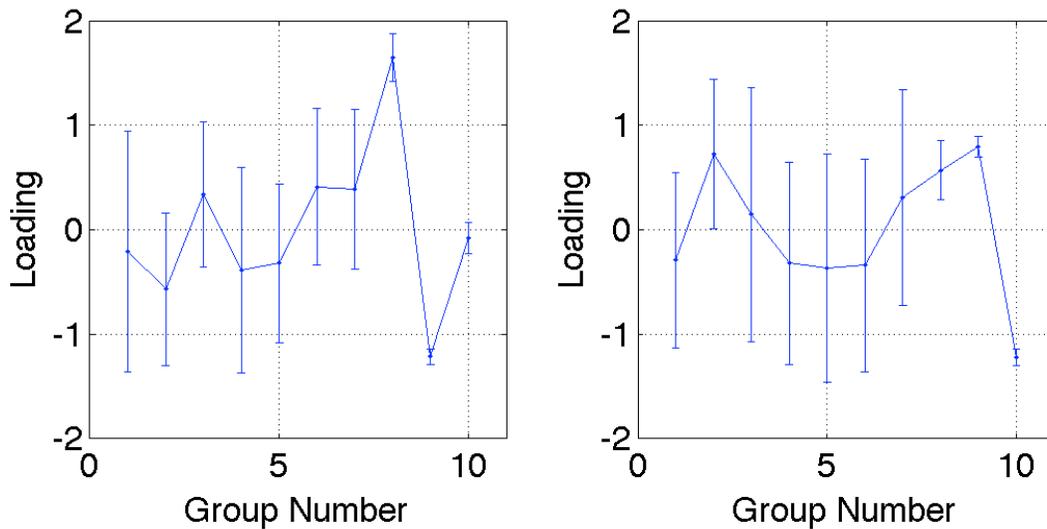


Fig. 4. Errorbar plot for the synthetic data set for canonical variate 1 (left) and canonical variate 2 (right).

**B. Observed data set of summer cases.** The same technique was then applied to summertime trajectories arriving at Lake Geneva, Wisconsin. In Fig. 5, the anomalously warm trajectories are shown in red, while the trajectories from randomly selected dates are plotted in black. It appears that the anomalously warm air typically originates in the Gulf of Mexico. Again, group-averaged loadings with errorbars are plotted in Fig. 6. The intragroup variability for groups 9 and 10 is smaller than that of other groups based on the size of the errorbars.



Fig. 5. Plot of anomalously warm trajectories (shown in red) and randomly selected summertime trajectories (shown in black).

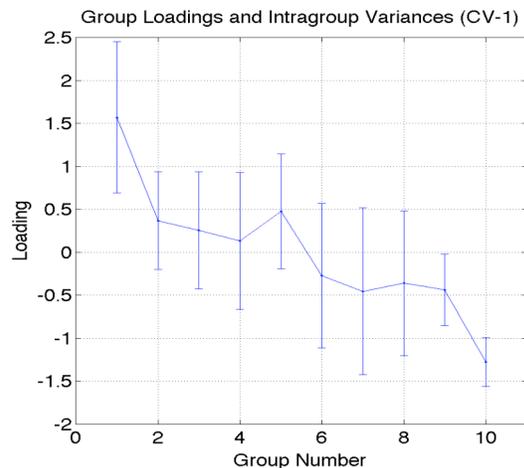


Fig. 6. Plot of group-averaged loadings and errorbars of CV-1 for summertime trajectories.

At the same time, the group loadings of groups 9 and 10 are significantly different (errorbars do not overlap). This identifies group ten as a cluster, while group nine is not, wince its loading is statistically indistinguishable from that of group eight. These results lead to the conclusion that, during the summer, anomalously warm air typically comes from the same area (Gulf of Mexico) to arrive in southeastern Wisconsin.

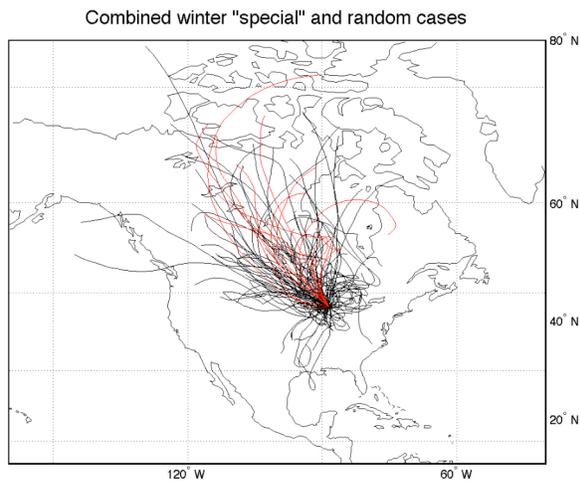


Fig. 7. Plot of anomalously cold trajectories (red) and randomly selected winter trajectories (black).

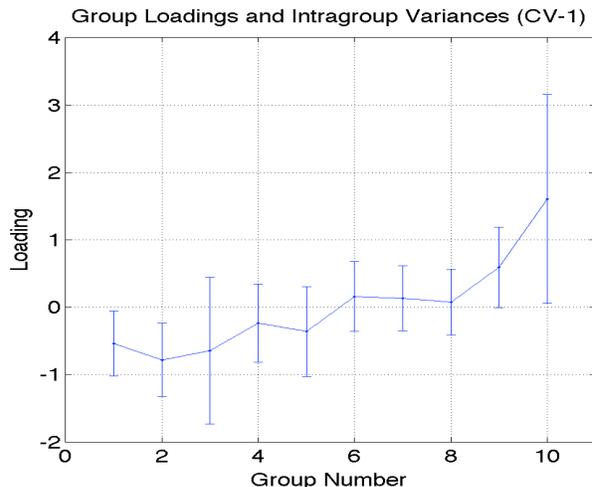


Fig. 8. Plot of group-averaged loading and error bars of CV-1 for the winter trajectories.

**C. Observed data set of winter cases.** Just as was done for fifteen anomalously warm summer trajectories, 14 anomalously cold winter trajectories were considered and discriminant analysis was applied. The “special” trajectories and randomly selected trajectories, when plotted on the same map, all seem to come from the north, but from a wide variety of directions to the receptor site (see Fig. 7). In comparing the first group of ten trajectories to the other groups, a significantly different group is unidentifiable. The plot of group-averaged loadings and errorbars (see Fig. 8) shows that the loadings for the first group were statistically indistinguishable from all other loadings (group loadings are all similar in magnitude). In addition, there is not a significantly low within-group variance for either canonical variate one or two, which is evident in the fact that the error bars overlap and are all of similar size. This leads to the conclusion that, during the winter, anomalous trajectories, as opposed to most trajectories, do not come from a single, unique location.

**Summary and discussion:** We used discriminant analysis to identify clusters in air-parcel trajectory data based on the value of the tracer (air temperature in this study) at the receptor point. Both warm summer cases, and cold winter cases were analyzed in this study. Based on our analysis, we concluded that the warmest days in Lake Geneva, Wisconsin during summer do correspond to distinctive air paths, originating over the Gulf of Mexico. On the other hand, the coldest days in Lake Geneva during the winter are not associated with a specific air route.

Our method should be contrasted to a traditional bootstrap method, in which the ensemble-mean air path associated with a certain range of tracer values at the receptor site is compared to the means of randomly chosen trajectories. These randomly chosen trajectories, however, are not associated with any specific values of tracer concentrations. In contrast, our method effectively chooses synthetic trajectories within a given range of tracer values to mimic the range associated with the original trajectories, thereby providing a more objective estimate of statistical significance.

**What Can Be Done?** We plan to use this technique to identify routes of air pollution in the United States and will develop further the statistical significance testing aspect of the method. This will be done through blending methods, such as bootstrap, with the current method, which will provide more robust estimates of robust clusters of trajectories.

## References

- Harris, J.M., 1992: Variations in atmospheric methane at Mauna-Loa Observatory related to long range transport, *J. Geophysical Research*, **97**, 6003 – 6010.
- Kahl, J.D.; et. Al., 1997: Air mass trajectories to Summit, Greenland: a 44-year climatology and some episodic events, *J. Geophysical Research*, **102**, 26,861 – 26,875.
- Moody, J.L.; et. Al., 1991: Precipitation composition and its variability in the southern Indian Ocean – Amsterdam Island, *J. Geophysical Research*, **90**, 20,769 – 20,786.
- Schneider, T.; Held, I.M., 2000: Discriminants of twentieth-century changes in earth surface temperatures, *J. Climate*, **14**, 249 – 254.
- Sirois, A.; Bottenheim, J.W., 1995: Use of backward trajectories to interpret the 5-year record of pan and O-3 ambient air concentrations at Kejimikujik National Park, Nova-Scotia, *J. Geophysical Research*, **100**, 2867 – 2881.

## 4. Decadal sea-surface temperature variations in the mid-latitude North Atlantic Ocean

*Nate Jamison*

### I. Introduction

Long-term records of sea-surface temperature (SST) provide a valuable source of information about global climate trends. Such records indicate a pronounced warming trend on a global scale in the last 150 years. This warming trend has not been consistently linear, however, as periods of cooling have been interspersed among the warming periods.

This irregular pattern of warming is somewhat difficult to accurately detect and predict, largely due to the presence of natural climate variability on a wide range of timescales. Traditionally, this natural variability has been interpreted as red noise, but recently much work has been done on establishing the presence of regular variability imbedded in the relatively “noisy” SST record. Separation and identification of these regular trends allows for better analysis of the effect of anthropogenic or other external factors in the warming trend experienced over the last several decades.

This project will use Multi-channel Singular Spectrum Analysis (M-SSA) to extract long-term trends, as well as quasi-periodic variation patterns, from the record of global sea surface temperatures. This analysis is an update of a similar study by Moron et al. (1998), with the focus on 8 and 13-year signals reported there as well as a 22-year signal.

### II. Data and Methodology

This project used an updated version of the Northern Hemisphere extended SST anomaly data. (Kaplan 1998). This consists of monthly data on a  $5^{\circ} \times 5^{\circ}$  grid, covering the period from December 1856 to November 2007. While data is available over the entire Northern Hemisphere, we have focused on the North Atlantic region ( $10^{\circ}$  to  $80^{\circ}$ N,  $120^{\circ}$  to  $0^{\circ}$ E).

The SST data was weighted by the square root of cosine of latitude to account for meridional convergence. The wintertime mean values for each year were computed by averaging only the monthly SST values for December through March. The purpose of taking only wintertime values was to capture the strongest signal, since the summertime anomalies are generally weaker than the wintertime ones.

We compressed the original SST data via Empirical Orthogonal Function (EOF) analysis and used the leading six Principal Components (PCs) as the “channels” of the vector time series to be analyzed by M-SSA. M-SSA diagonalizes the lag-covariance matrix of this multi-channel time series, with lags ranging from 0 to  $M-1$ , where  $M$  is the window width. The eigenvectors of the resulting matrix are called time-EOFs (T-EOFs), and their associated coefficients, which consist of scalar time series, are called time-PCs (T-PCs). M-SSA permits the identification of particularly prominent oscillations, which manifest themselves as a pair of T-PCs of similar (and large) variance in phase-quadrature and corresponding to the same frequency. Monte Carlo significance testing at a 95% confidence level is used to determine which oscillatory pairs are statistically significant relative to a red-noise null hypothesis.

By using a subset of the obtained T-EOFs and T-PCs, we created a reconstructed time series of the same length of the original time series to get a clearer picture of the variability associated with quasi-periodic signals identified by M-SSA. A reconstruction using all of the T-EOFs and T-PCs, as opposed to some subset of them, would be identical to the original time series. The quasi-periodic signal reconstructions can be used to identify the magnitudes and spatial relationships of the oscillatory patterns.

For each oscillation identified, we checked for robustness by varying the M-SSA window size,  $M$ , and ensuring that the analysis captures, for each  $M$ , the statistically significant pairs at the same frequencies, whose T-PCs are in quadrature. We then visualized typical evolution associated with each oscillatory signal using M-SSA reconstruction, and performed EOF analysis on the reconstructed spatial patterns to determine the leading modes of the oscillations.

### III. Results

Figure 1 shows the variance percentages accounted for by individual EOF modes, as well as cumulative percentages. The leading six modes account jointly for about 77% of the total SST variance.

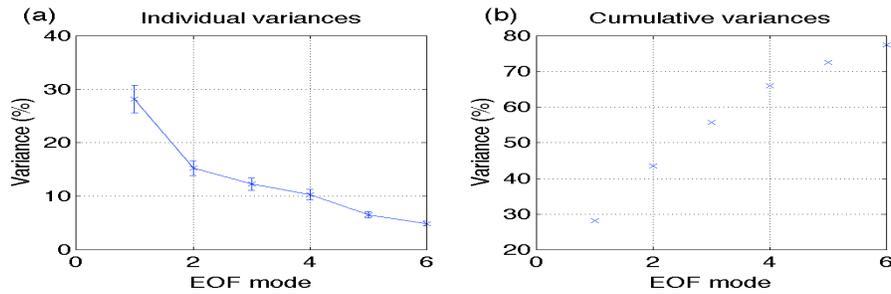


Fig. 1. EOF spectra of SST data: (a) Percentage of individual variance accounted for by individual modes; (b) cumulative percentages.

The three leading EOFs are shown in Fig. 2. The first EOF is dominated by a mono-polar pattern off the east coast of the United States, co-located with the time-mean Gulf Stream position, while EOFs two and three exhibit tri-polar and dipolar patterns in the North Atlantic.

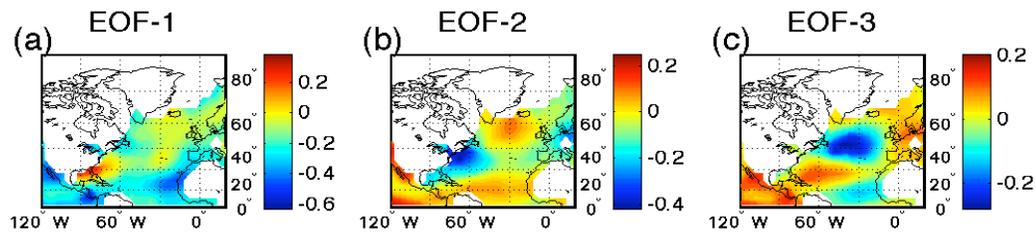


Fig. 2. Leading dimensional ( $^{\circ}\text{C}$ ) EOFs of the North Atlantic SST data.

The M-SSA spectrum of the leading six PCs of SST (Fig. 3) reveals three dominant, statistically significant pairs, corresponding to periods of 22, 13 and 7.5 years. The latter two oscillations were found to be robust with respect to the window size  $M$ , for  $20 \leq M \leq 60$ , while the identification of the former, lowest frequency oscillation required the window size exceeding  $M \geq 30$ .

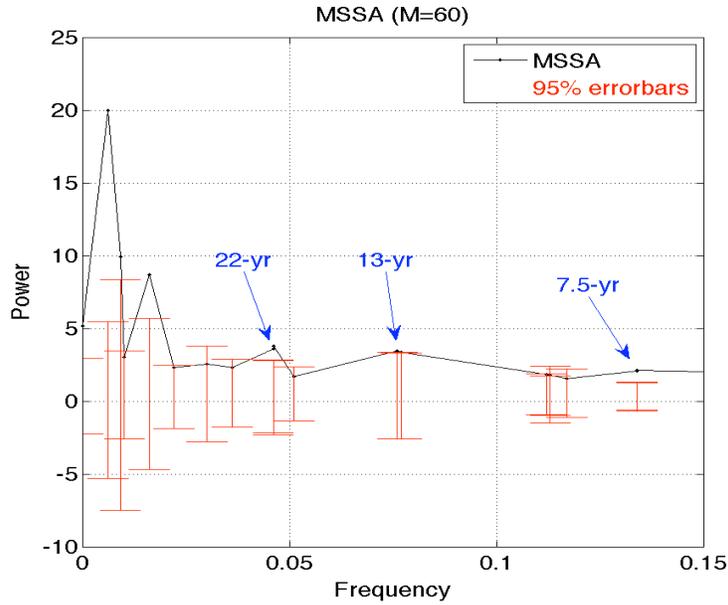


Fig. 3. M-SSA spectrum of the leading six PCs of wintertime North Atlantic SST data.

Figure 4 shows the reconstructed time series using RCs based on the 7.5-year oscillation from all 6 channels. It shows that channels 3 and 4 predominate, except for the tail end of the time period where channel 2 is more pronounced. The magnitude of the oscillations in channel 2 increase over time, while the magnitudes for channels 3 and 4 do not show a similar increase. Additionally, the oscillations of the channels 3 and 4 are in phase with each other, while channel 2 is out of phase with 3 and 4. Channels 1, 5 and 6 account for little of the variance.

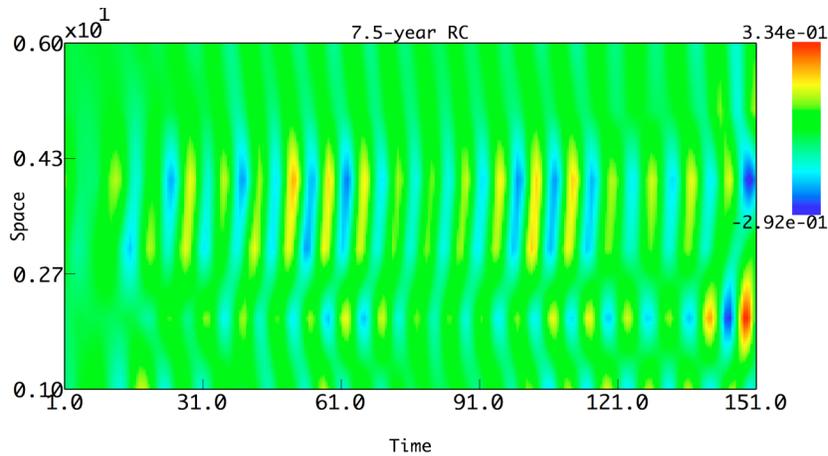


Fig.4. Reconstruction of 7.5-year oscillation for all six channels.

The corresponding reconstruction for the 13-year oscillation, shown in Fig. 5, indicates that the magnitude of the oscillatory signal increases over time, which appears as more intense blobs of

red and blue. As can be seen, the first three channels account for a majority of the variance as the intensity of the colors decreases towards the top of the plot where channels 4-6 are plotted. In addition, the oscillations of the first and second channels are out of phase with each other, as are the oscillations of channels 2 and 3.

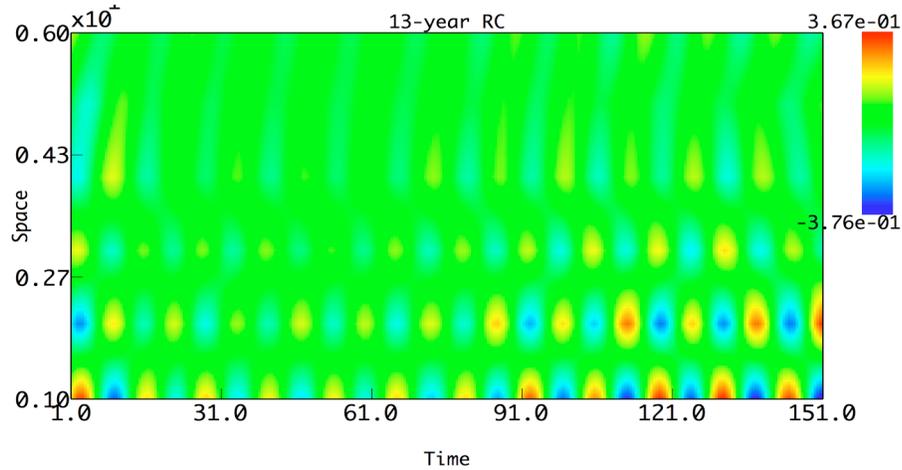


Fig. 5. Reconstruction of 13-year oscillation for all six channels.

Figure 6 shows the reconstruction for the 22-year oscillation. As can be seen, channels 2, 3 and 4 account for most of the variance, while channels 1, 5 and 6 account for little of the variance. The magnitude of the oscillations in all channels was strong at the beginning and end of the time period with a weakening of the oscillations in the middle of the time period.

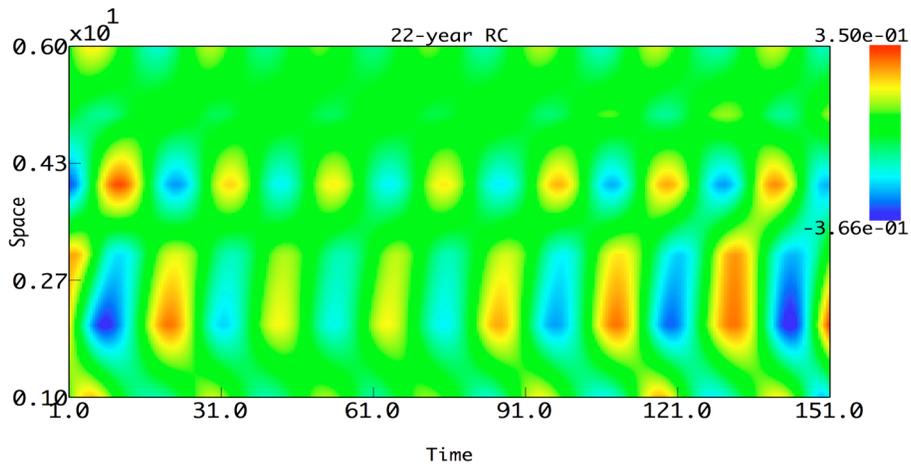


Fig. 6. Reconstruction of 22-year oscillation for all six channels.

After identifying the oscillations and creating reconstructed time series of the oscillatory signals, we plotted the spatial evolution of the oscillatory patterns by projecting our original spatial EOFs

on the reconstructed time series. The resulting reconstructed EOFs represent the spatial EOFs of the data filtered to isolate the features associated with the oscillatory pattern.

Figure 7 shows the spatial evolution of the 7.5-year oscillation over the period of 9 years. As can be seen, the primary feature of the oscillation pattern is an area of warming that begins off the southeast coast of the United States and propagates northeastward along the coast and towards Iceland. Also, during the portion of the period when the warming is in the central North Atlantic, there is a simultaneous warming along the west coast of North Africa.

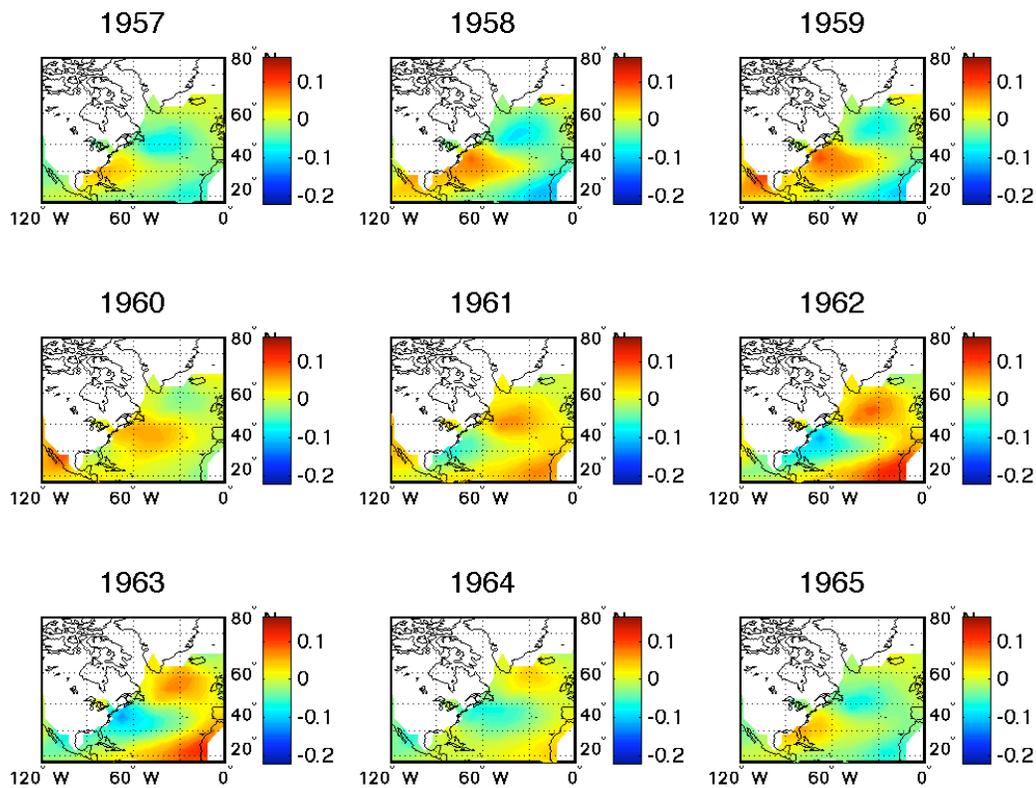


Fig. 7. Projection of 7.5-year reconstruction on spatial EOFs.

We performed EOF analysis on this reconstructed EOF to determine the predominant spatial pattern associated with this oscillation. Figure 8 shows the variance percentages accounted for by individual EOF modes, as well as cumulative percentages. The leading three modes account jointly for about 97% of the total SST variance.

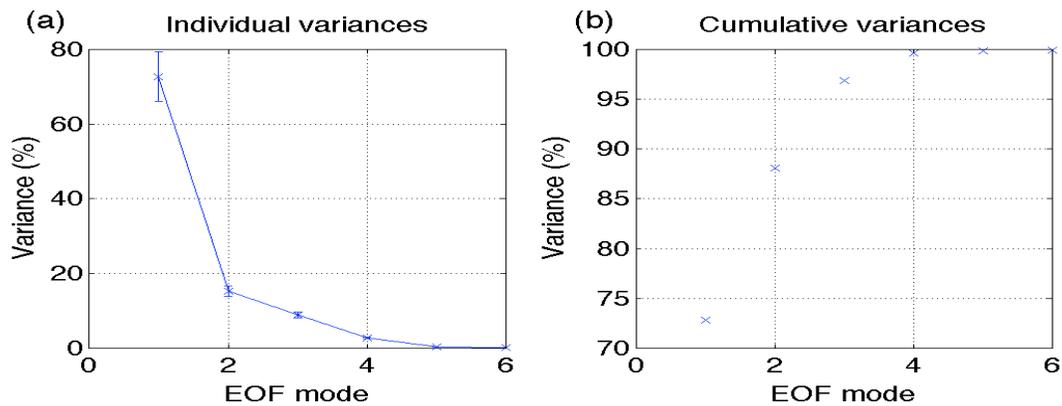


Fig. 8. EOF spectra of reconstructed 7.5-year data: (a) Percentage of individual variance accounted for by individual modes; (b) cumulative percentages.

The three leading EOFs of the reconstructed data are shown in Figure 9. The first EOF is dominated by a tri-polar pattern with positive values south of Greenland and Iceland and off the west coast of North Africa, and negative values off the east coast of the United States. The second EOF exhibits a similar tri-polar pattern, except the second area of positive values is located in the Caribbean rather than off of Africa. The third EOF exhibits a mono-polar pattern in the central North Atlantic.

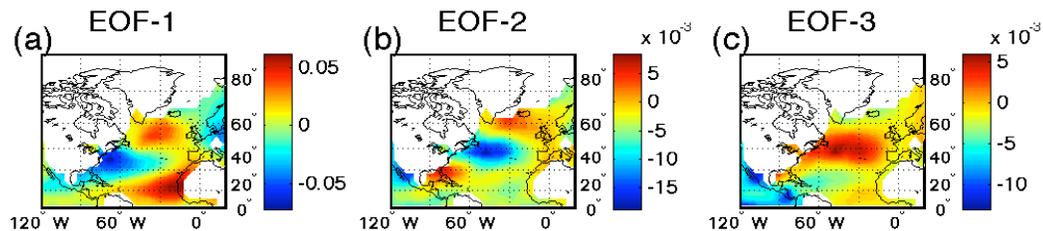


Fig. 9. Leading EOFs of the reconstructed 7.5-year data.

We created similar reconstructions for the 13-year and 22-year oscillations as well. Figure 10 shows the spatial evolution of the 13-year oscillation over a period of 15 years. As can be seen, this oscillation exhibits a stationary dipolar pattern with one pole off the coast of Newfoundland and another off the west coast of North Africa.

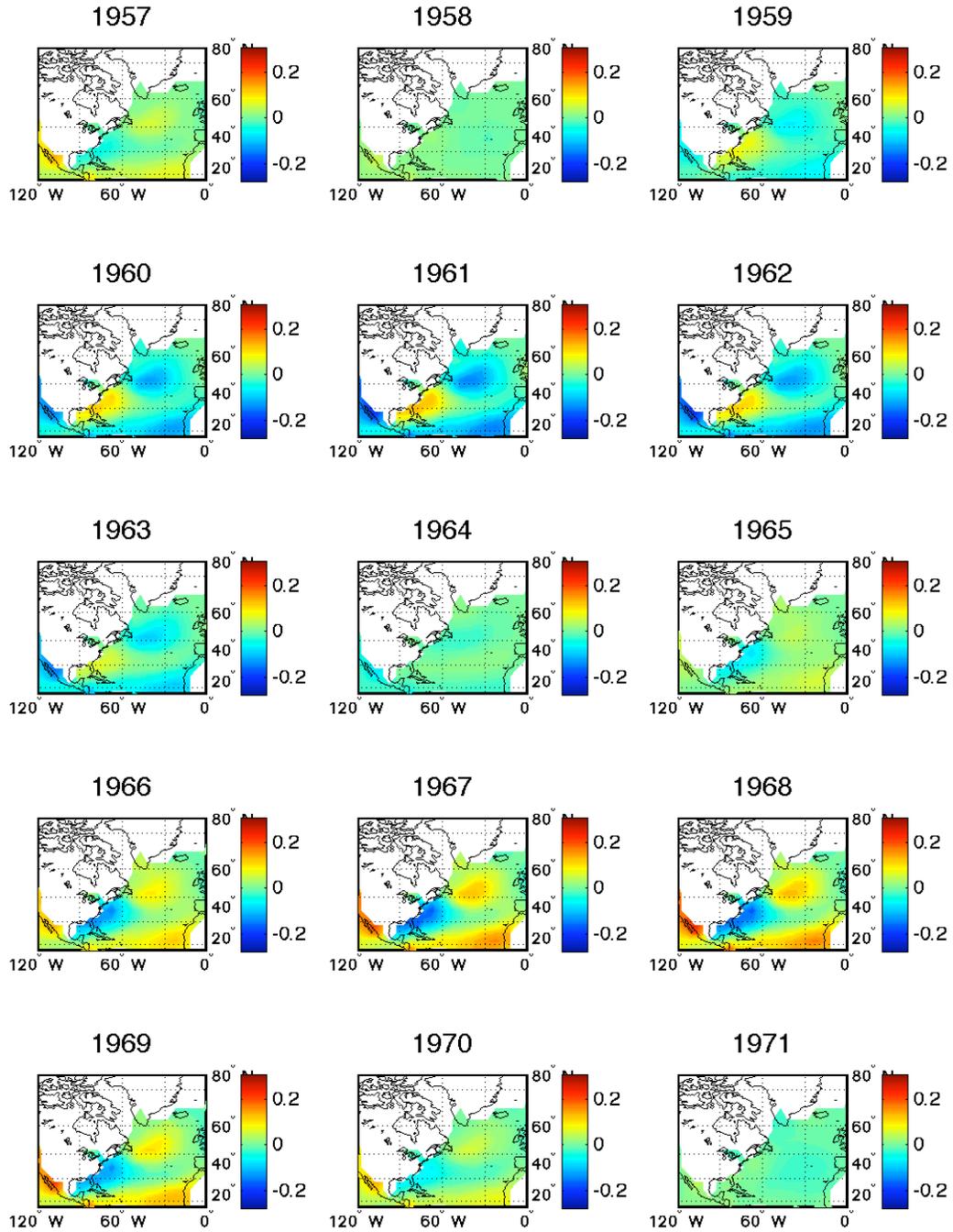


Fig. 10. Projection of 13-year oscillation on spatial EOFs.

We again performed EOF analysis on the reconstructed EOF to determine the predominant spatial pattern associated with this oscillation. Figure 11 shows the variance percentages accounted for by individual EOF modes, as well as cumulative percentages. The leading mode alone accounts for almost 97% of the total SST variance.

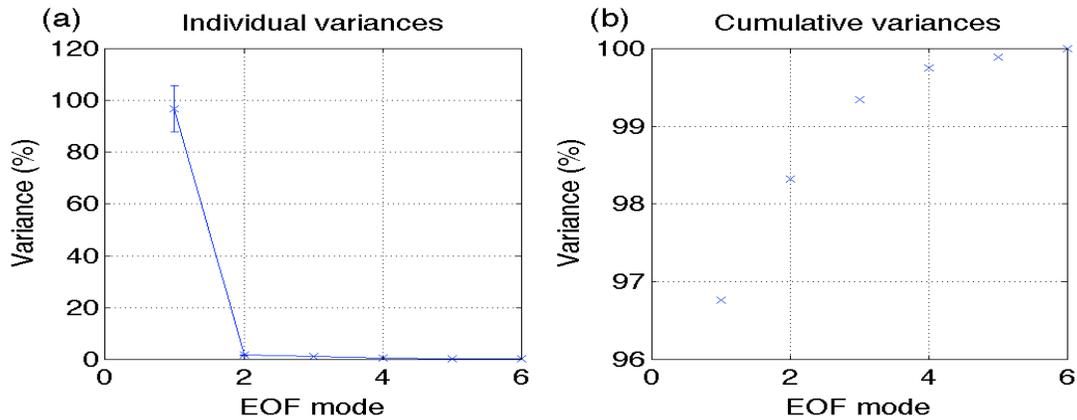


Fig. 11. EOF spectra of reconstructed 13-year data: (a) Percentage of individual variance accounted for by individual modes; (b) cumulative percentages.

The three leading EOFs of the reconstructed data for the 13-year oscillation are shown in Figure 12. This leading EOF exhibits a pair of regions with negative values off the west coast of North Africa and Newfoundland, and a pair of regions with positive values north of the European continent and off the southeast coast of the United States.

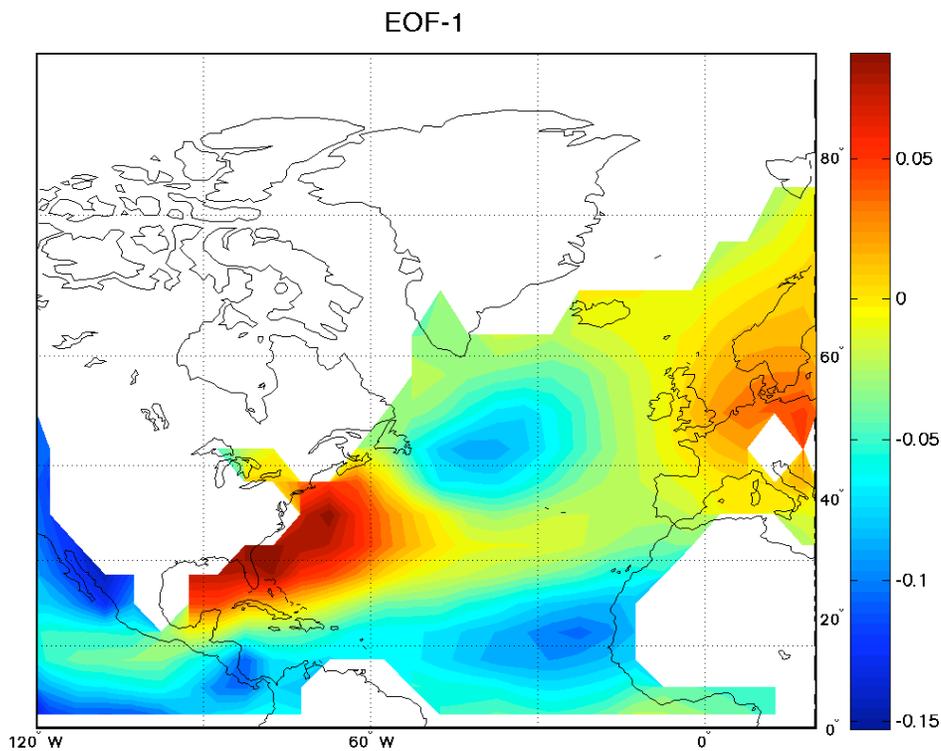


Fig. 12. Leading EOF of the reconstructed 13-year data.

Figure 13 shows the spatial evolution of the 22-year oscillation over a period of 15 years. The primary feature of this oscillation is a stationary region located off the coast of Newfoundland that alternates positive and negative values.

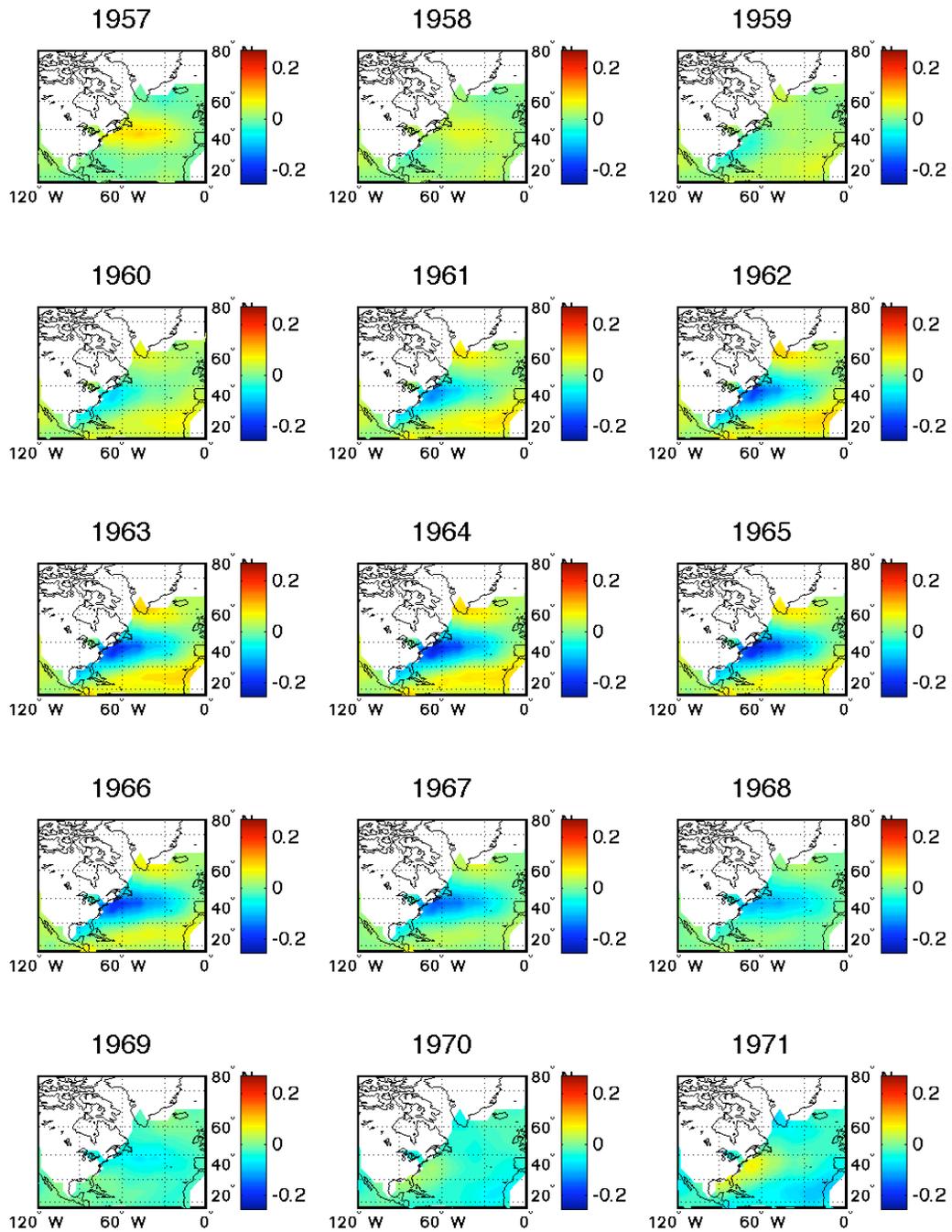


Fig. 13. Projection of 22-year oscillation on spatial EOFs.

The resulting EOF analysis on the reconstructed EOF shows that the leading two modes account jointly for about 96% of the total SST variance, as can be seen in Fig. 14.

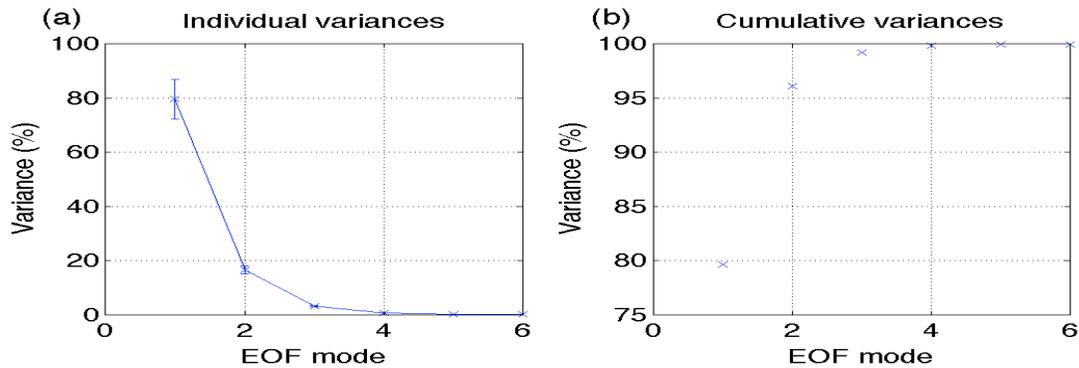


Fig. 14. EOF spectra of reconstructed 22-year data: (a) Percentage of individual variance accounted for by individual modes; (b) cumulative percentages.

Finally, Fig. 15 shows the leading two EOFs associated with the reconstruction for the 22-year oscillations. The first EOF exhibits an area of negative values off the New England coast with positive values over the rest of the North Atlantic, while the second EOF shows a similar area of negative values, this time farther out to sea, and an area of positive values off the southeast coast of the United States.

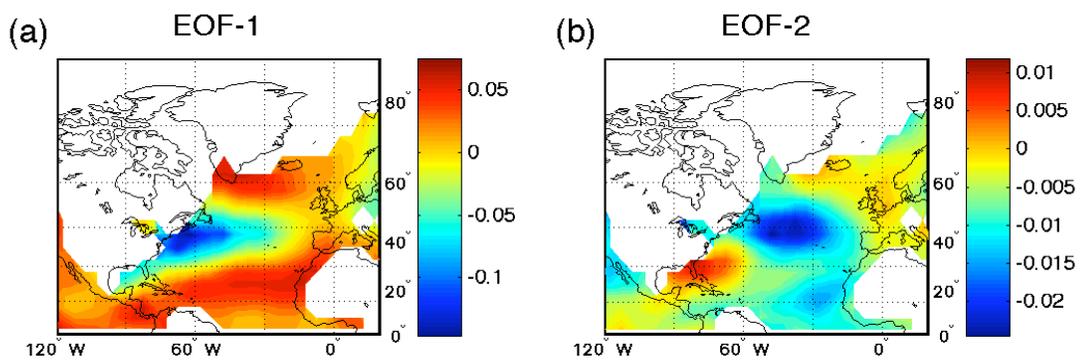


Fig. 15. Leading EOF of the reconstructed 22-year data.

## IV. Summary and future work.

We used M-SSA analysis on wintertime mean SST temperatures in the North Atlantic to identify regular oscillations. We discovered three oscillatory patterns with periods of 7.5, 13 and 22 years respectively. We then created reconstructed time series from these oscillatory patterns to identify the nature of the oscillations, both in terms of the amount of variance explained and the spatial structure of the oscillatory patterns.

The 7.5-year oscillation is characterized by an area of warming that propagates northeastward from the southeast coast of the United States to northern Europe, while the 13-year and 22-year oscillations are characterized by more stationary patterns. The 13-year oscillation exhibits a dipolar structure with poles off the Newfoundland coast and North Africa, while the 22-year oscillation exhibits a mono-polar structure off the Newfoundland coast and extending in the central North Atlantic.

The different amount of variance accounted for by the leading EOFs of each oscillation as well as the different spatial patterns of the oscillation indicates that these oscillations are distinct features. More analysis, possibly including comparison of the oscillations in historical data with idealized models of particular dynamic features to determine the physical processes associated with each oscillation.

This M-SSA technique can be used on higher resolution datasets to perhaps provide a more detailed picture of the spatial nature of these oscillations. And of course we do not necessarily need to limit analysis to the North Atlantic; other regional or global analyses could be performed. Another use of this technique would be in verifying global climate models to determine if they, in fact, reproduce the oscillations identified using M-SSA.

## References

- Kaplan A, Cane M, Kushnir Y, Clement A, Blumenthal M, and Rajagopalan B, (1998) Analyses of global sea surface temperature 1856-1991, *Journal of Geophysical Research*, **103**: 18,567-18,589.
- Moron V, Vautard R, and Ghil M, (1998) Trends, interdecadal and interannual oscillations in global sea-surface temperatures, *Climate Dynamics*, **14**: 545-569.

## 5. Causes of quasi-periodic cycles in levels of lakes Michigan and Huron

**Abstract:** *While Great Lakes' seasonal water-level variations have been researched and well documented previously, no studies thus far addressed longer-term, decadal cycles contained in the 150-yr lake-level instrumental record. Paleo-reconstructions based on Lake Michigan's coastal features, however, hinted to an approximate 30-yr quasi-periodic lake-level variability. In the present study, spectral analysis of 1865 – 1996 Lakes Michigan/Huron historic levels revealed 8 and 12-yr periodicities; these time scales match those of large-scale climatic signals previously found in the North Atlantic. It is suggested here that the 30-yr cycle evident in paleodata is due to intermodulation of the two near-decadal signals. Furthermore, water budget analysis argues that the North Atlantic decadal climate modes translate to the lake levels primarily through precipitation, runoff and outflow components.*

### Introduction

The Great Lakes The Great Lakes provide transportation for shipping, hydroelectric power, sustenance and recreation for the more than 30 million people that live in its basin. Understanding and predicting lake-level variations is therefore a problem of great societal importance due to their immediate and profound impact upon the economy and environment (Polderman and Pryor 2004). A lack of understanding of how or when the levels will change often results in expensive and drastic human-induced alterations that, more often than not, produce undesirable results (Egan 2006). In addition, it has recently been determined that attempts to regulate the levels have proven insignificant next to uncontrollable factors, such as climate changes (USACE, 2002). These findings underscore the notion that the best way to minimize the impact of lake-level fluctuations is to gain a better understanding, and then learn to better forecast, the inevitable level changes.

Lake-level gauges have been placed in many locations and have since provided very accurate records of the Great Lakes' water levels which, for some basins, extend back more than 150 years. Lakes Michigan and Huron, in particular, have been monitored since 1865. As seen in Fig. 2 (green line), the levels have continually fluctuated throughout this recorded history, and

many attempts have been made to develop a timeline over which the level changes occur. Unfortunately, the short-term availability of historic records has thus far disallowed conclusive evidence for any long-term definitive cycles (Thompson and Baedke 1997).

Ridge Data To complement instrumental records, Thompson and Baedke (1997, 2000a,b) used a technique that analyzes coastal features of Lake Michigan at various locations and combines it with radiocarbon dating to track the approximate lake levels for the past 4,700 years. The data show two distinct quasi-periodic fluctuations. The first averages around 33 yr and ranges from 0.5 to 0.6 m in amplitude, while the second averages about 160 yr and ranges from 0.5 to 1.5 m. Both cycles date back at least 3,400 years and the accuracy of the paleodata has been confirmed by comparison with the 150-yr historic data set.

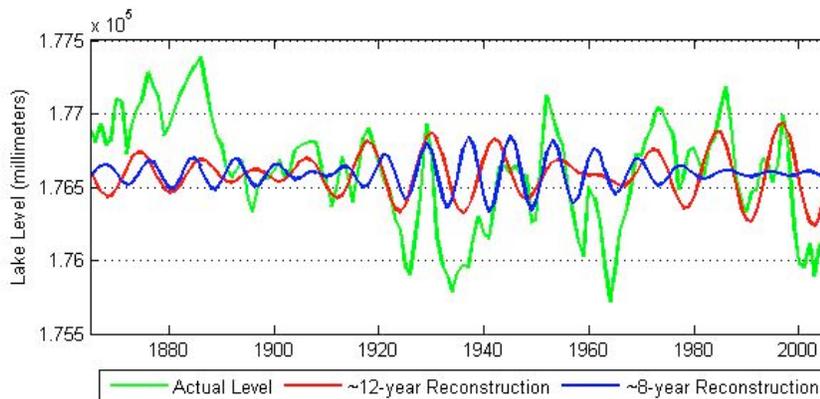


Fig. 1. Michigan-Huron lake-level variations. Green line: Annual averaged Michigan-Huron lake levels; red and blue lines: reconstructions of 12-yr and 8-yr oscillations found by the MTM spectral analysis (see text and Fig. 2).

The Michigan-Huron System Lakes Michigan and Huron behave hydraulically as one lake (Bishop 1990), and it can therefore be assumed that they experience synchronous long-term lake-level fluctuations. In addition, the Michigan-Huron system displays the greatest range of levels (Argyilan and Forman 2003) and is the least affected by human induced modifications (Changnon 1987). Lake-level changes for these lakes can therefore be safely primarily attributed

to climatological changes (Polderman and Pryor 2004). These considerations, as well as the availability of the geologic data, motivated the present study's focus on Lakes Michigan and Huron only, rather than on the whole Great Lakes system.

## **Data**

For our study, we used beginning-of-month Michigan–Huron water levels, made available online for the period between 1900 and 2006 by the Great Lakes Environmental Research Laboratory (GLERL). Additional levels were obtained from the U.S. Army Corps of Engineers (USACE) that date back to 1865, also with a monthly sampling. Lakes Michigan, Huron and Georgian Bay have been combined previously to produce a single hydrograph. We have also used GLERL monthly totals for precipitation, runoff, evaporation, St. Marys River inflow, Chicago Diversion and St. Clair River outflows, for the period between 1948 through 1996, for which continuous monthly record for all of the above components is available.

## **Identification and Interpretation of Lake-Level Periodicities**

We first annually averaged the beginning-of-month lake-levels to concentrate on interannual and longer time scales. We then used Multi-taper method (MTM: Thomson 1982, 1990; Mann and Lees 1996) and Monte Carlo Singular Spectrum analysis (SSA: Broomhead and King 1986; Fraedrich 1986; Vautard and Ghil 1989) to identify significant periodicities in the data relative to the red-noise null hypothesis. Both analyses revealed  $\sim 8$  and  $\sim 12$ -yr cycles, statistically significant at the 99% *a priori* confidence level; the MTM spectrum is shown in Fig. 2. The time scales of these cycles are of particular interest because of their similarity to the time scales of known climatic signals previously found in the North Atlantic (Deser and Blackmon 1993; Moron et al. 1998). We used MTM-based reconstructions to compute two narrowband filtered time series approximating the quasi-periodic evolution of the lake levels, for both 8-yr and 12-yr signals (see Fig. 1, blue and red lines, respectively).

In a nonlinear system, the interaction between two quasi-periodic signals centered at frequencies of  $f_1$  and  $f_2$  may result in the occurrence of secondary periodicities, with frequencies  $f_1 \pm f_2$ ; this

phenomenon is referred to as intermodulation. In our case, the intermodulation of the 8-yr and 12-yr periodicities will result in the oscillation with a period in the neighborhood of 30 years, thus providing an appealing interpretation of confirming the ridge-data periodicities found by Thompson and Baedke (1997, 2000a,b). Note that while long-period peaks do appear in the spectral analysis of historic lake levels (see Fig. 2), the historic data is too short to assure statistical significance of such signals. On the other hand, the ridge data set's resolution is too coarse to resolve the periods shorter than a couple of decades, which may explain why the near-decadal periods have not been found in the proxy lake-level data analysis.

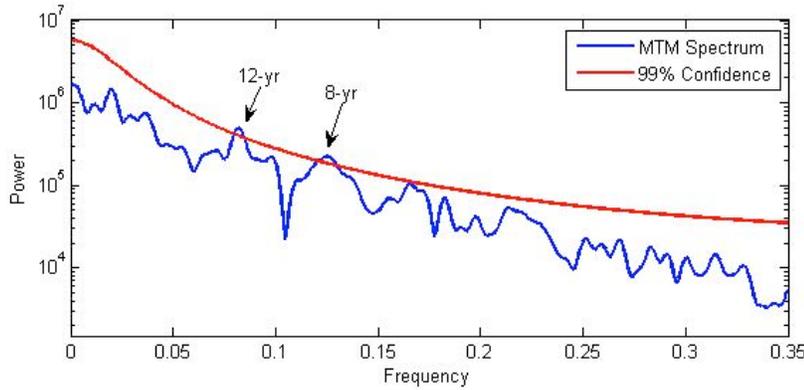


Fig. 2. MTM spectrum of the historic lake-level data (blue) and the 99% confidence level based on red-noise null hypothesis (red). The 12 and 8-yr spectral peaks are statistically significant.

### Attribution of Lake-Level Cycles to Individual Sources

In order to determine which components of lake-level water budget dominated to the decadal cycles identified above, we first evaluated lake-level anomalies caused by individual source terms in the water balance equation:

$$L_{m,y} = L_{m-1,y} + NBS_{m-1,y} - O_{m-1,y} + I_{m-1,y} + \varepsilon_{m-1,y} \quad (1)$$

Here  $NBS = P_{m-1,y} + R_{m-1,y} - E_{m-1,y}$  denotes the net basin supply,  $NBS = P + R - Em$  is the month,  $y$  is the year,  $L$  is the beginning-of-month lake level,  $O$  is the total monthly outflow,  $I$  is

the total monthly inflow,  $P$  is the total monthly precipitation,  $R$  is the total monthly land runoff,  $E$  is the total monthly evaporation, and  $\varepsilon$  is the residual error due to omitted lake-level sources and measurement uncertainties. To obtain the lake-level time series due to a single source in (1), we integrated this equation in time while artificially suppressing all other sources. For example, the lake-level anomalies resulting from precipitation alone can be determined by following equation:

$$L_p = L_{m-1,y} + P_{m-1,y}, \quad (2)$$

and the contributions to the lake-level time series due to other individual sources (that is, evaporation, runoff, outflow, inflow, and error) were obtained in the same way. The resulting six linearly detrended individual lake-level curves appear in Fig. 3 (see figure legend); their sum exactly reproduces the observed lake-level hydrograph (black).

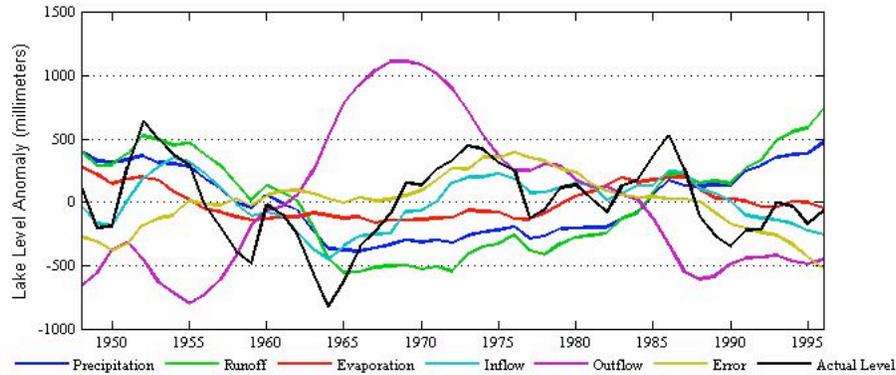


Fig. 3. Annually average levels and calculated levels resulting from individual lake-level sources (see text).

We next used the six individual-source lake-level time series as predictors in the multiple linear regression of the 8-yr and 12-yr signal reconstructions shown in Fig. 1 (blue and red curves, respectively), via partial least squares (PLS: Abdi 2003). PLS regression finds a set of new predictors — latent variables — as such linear combinations of original predictors that maximize the covariance between these new predictor variables and the response variable. Using all of the latent variables is equivalent to employing regular least squares, but using a subset of latent variables may produce more stable regression coefficients, with a much-reduced uncertainty.

Table 1. Results of PLS regression of 8-yr and 12-yr lake-level reconstructions (Fig. 1) onto the multiple lake-level predictors associated with individual lake-level sources (Fig. 3). Only significant contributors are shown. The uncertainties are estimated as half the range between the 10<sup>th</sup> and 90<sup>th</sup> percentiles based on 100 synthetic realizations of multi-variable red-noise model of the individual lake-level time series.

	8-yr reconstruction	12-yr reconstruction
Precipitation	0.5332 ± 0.314	0.532 ± 0.506
Runoff	0.3412 ± 0.241	2.172 ± 1.587
Outflow	1.0464 ± 0.688	1.930 ± 1.191

The regression coefficients significantly different from zero, at the 10% level, are shown in Table 1, which thus identifies the same triplet of main contributors to both of the near-decadal lake-level cycles: precipitation, runoff and outflow. Since the outflow data is strongly related to, and therefore is calculated as a function of lake level, it is reasonable to conjecture that any periodicities detected in the lake-level itself would also manifest in the outflow, which is consistent with the results of Table 1. The outflow cannot therefore be considered the driving source of the oscillation. We thus identify the large-scale precipitation as the primary driver of the lake-level periodicities related to the North Atlantic climatic signals.

Our findings that precipitation and runoff are the primary drivers of the fluctuations are also in agreement with previous studies. Our finding that precipitation and runoff are the primary contributors to the lake-level cycles are also consistent with earlier analyses of Brinkmann (2000), who found that monthly *NBS* anomalies are most sensitive to variability in precipitation in the upper lakes, and runoff in the lower lakes.

### Future Work

We plan to enhance the present analyses in three ways. First, we will employ the synthetic ridge data based on a red-noise null hypothesis to ascertain statistical significance of the 30-yr cycle identified in the lake-level paleodata. Second, we will study the ways in which the synoptic environment is altered to allow transmission of the North Atlantic climatic signals to the lake

levels through precipitation fluxes. Finally, we will study secular changes of North Atlantic decadal signals over the past century to diagnose possible effects of global warming on these oscillations, and therefore on past and future lake-level variations.

## References

- Abdi, H., 2003: Partial Least Squares (PLS) regression. In *Encyclopedia of Social Sciences Research Methods*, ed. Lewis-Beck, M., Bryman, A., Futing, T. Sage: Thousand Oaks, CA.
- Argyilan, E.A., and Forman, S.L., 2003: Lake level response to seasonal climatic variability in the Lake Michigan-Huron system from 1920 to 1995. *J. Great Lakes Res.*, **29**(3), 488–500.
- Bishop, C. T., 1990: Historical variation of water levels in Lakes Erie and Michigan-Huron. *J. Great Lakes Res.*, **16**(3), 406–425.
- Brinkman, W.A.R., 2000: Causes of variability in monthly Great Lakes water supplies and lake levels. *Climate Research.*, **15**, 151–160.
- Broomhead, D. S., and G. P. King 1986: Extracting qualitative dynamics from experimental data. *Physica D*, **20**, 217–236.
- Changnon, S.A. Jr., 1987: Climate fluctuations and record-high levels of Lake Michigan. *Bull. Amer. Meteor. Soc.*, **68**:1394–1402.
- Deser, C., and Blackmon, M. L., 1993: Surface climatic variations over the North Atlantic Ocean during winter: 1900–1989. *J. Climate*, **6**(9), 1743–1753.
- Egan, D., 2006, September 17: Is Lake Michigan draining? *Milwaukee Journal Sentinel*, pp. 1A, 23A.
- Fraedrich, K., 1986: Estimating the dimension of weather and climate attractors. *J. Atmos. Sci.*, **43**, 419–432.
- Mann, M. E., and J. M. Lees, 1996: Robust estimation of background noise and signal detection in climatic time series. *Clim. Change*, **33**, 409–445.
- Moron, V., Vautard, R., and Ghil, M., 1998: Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Climate Dynamics*, **14**, 545–569.
- Polderman, N.J., and Pryor, S. C., 2004: Linking synoptic-scale climate phenomena to lake-level variability in the Lake Michigan-Huron basin. *J. Great Lakes Res.*, **30**(3), 419–434.
- Thomson, D. J., 1982: Spectrum estimation and harmonic analysis. *IEEE Proc.*, **70**, 1055–1096.
- Thomson, D. J., 1990: Quadratic-inverse spectrum estimates: Application to paleoclimatology. *Philos. Trans. R. Soc. London*, **A332**, 539–597.
- Thompson, T. A., and Baedke, S. J., 1997. Strandplain evidence for late Holocene lake-level variations in Lake Michigan. *Geological Society of America Bulletin*, **109**(6), 666–682.
- Thompson, T. A., and Baedke, S. J., 2000a: A 4,700-year record of lake level and isostasy for Lake Michigan. *J. Great Lakes Res.*, **26**(4), 416–426.
- Thompson, T. A., and Baedke, S. J., 2000b: A geologic perspective on Lake Michigan water levels. *Great Lakes Update*. **140**, 1–6, U.S. Army Corps of Engineers.
- U.S. Army Corps of Engineers (USACE). 2002: Are Great Lakes water levels recovering? *Great Lakes Update*, **148**, 1–4.
- Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D*, **35**, 395–424.

## **6. Diagnosis of relation between synoptic disturbances and air pollution in the US**

**Abstract:** Using daily data of sea-level pressure on a regular grid and air-pollutant concentrations at irregularly spaced stations, establish weather conditions that favor extreme air-pollution episodes.

**K. G. Thishan Dharshana**

**Spring 2008**

## 1. Introduction

Over past decades, air pollution has become a greater threat all over the world. Among many pollutants, Particulate Matter (PM) is considered to be particularly dangerous for humans due to its identified affiliation to respiratory system related illnesses. For that reason, identification of PM sources has become one of the most important areas of air pollution meteorology. While a large number of studies have considered the contribution of stationary and mobile source emissions to PM levels at a particular location or a region, a substantially smaller number of studies have focused on meteorological factors which may affect local PM concentrations. In particular, higher air pollution is thought to be associated with anticyclonic conditions, due to stable air column and air descent, which decrease the dispersion of pollutant and increase air stagnation. In a similar way, cyclonic conditions may be associated with lower air pollution. An example supporting the above view is shown in Fig. 1, which plots a snapshot of biweekly anomalies of sea-level pressure (SLP) and PM<sub>2.5</sub> (Particulate Matter of less than 2.5  $\mu\text{m}$  in diameter): The two fields exhibit substantially positive spatial correlation.

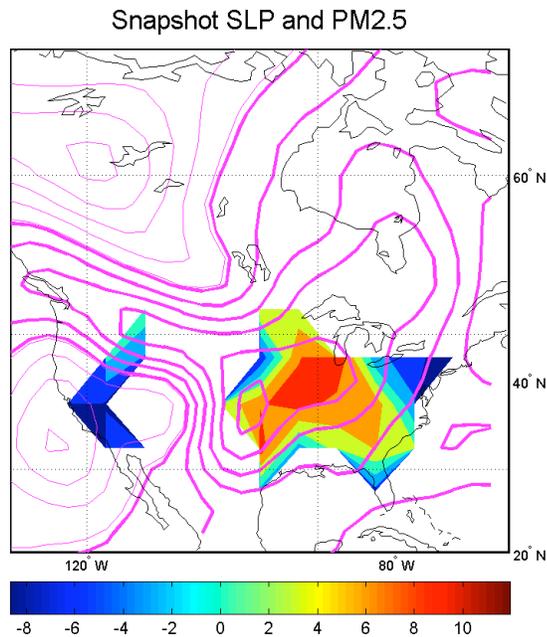


Fig. 1. Snapshot of biweekly anomalies of SLP (contours; heavy lines show positive and zero anomalies, light lines — negative anomalies) and PM<sub>2.5</sub> (color shading).

The present study aims at establishing the degree to which the above association between air pollution and weather conditions amounts to air pollution variability. In other words, we are trying to establish how typical the positive spatial correlations seen in Fig. 1 are. We do so by applying matrix methods to multivariate data sets of SLP and PM<sub>2.5</sub> concentrations.

## **2. Methodology**

We analyzed daily-mean SLP data on a 2.5° regular grid in the Northern Hemisphere (NH) taken from the NOAA NCEP-NCAR Reanalysis Project (Kalnay et al., 1996) and PM<sub>2.5</sub> data from the U. S. Environmental Protection Agency's (USEPA, 2008), over the period between January 1, 2000 and December 31, 2006. This period was selected based on PM<sub>2.5</sub> data availability considerations. The available PM<sub>2.5</sub> data was averaged spatially within 5°×5° bins and 3-day time averaged prior to analyses; the grid boxes in which data coverage was less than 50% were discarded. The remaining missing data points were filled using available covariance based imputation techniques. Continuous daily time series at every available location were then computed from 3-day data by cubic splines. The resulting SLP and PM<sub>2.5</sub> data sets were high-pass filtered by removing nonoverlapping 15-day boxcar averages to concentrate on synoptic time scales.

We next computed regional Empirical Orthogonal Functions (EOFs; Preisendorfer, 1988) of each data set in the region between 125°W to 70°W and 25°N to 50°N. The leading 11 EOFs of SLP and 22 EOFs of PM<sub>2.5</sub> account for about 95% of total variance of the respective fields. We then performed combined EOF analysis and Maximum Covariance Analysis (MCA; Bretherton et al., 1992; Wallace et al., 1992) using the individual EOF subspaces identified above to study co-variability of SLP and PM<sub>2.5</sub>. The individual-field principal components (PCs) used in our analysis were normalized by the standard deviation of their respective PC-1. We performed our analysis using the whole data set, as well as its winter and summer subsets. We will not show the seasonal results, since they were qualitatively analogous to those based on all-year data.

### 3. Results

The leading three combined EOFs of SLP/PM2.5 (Fig. 2) account for 13, 11 and 8% of total variance, respectively.

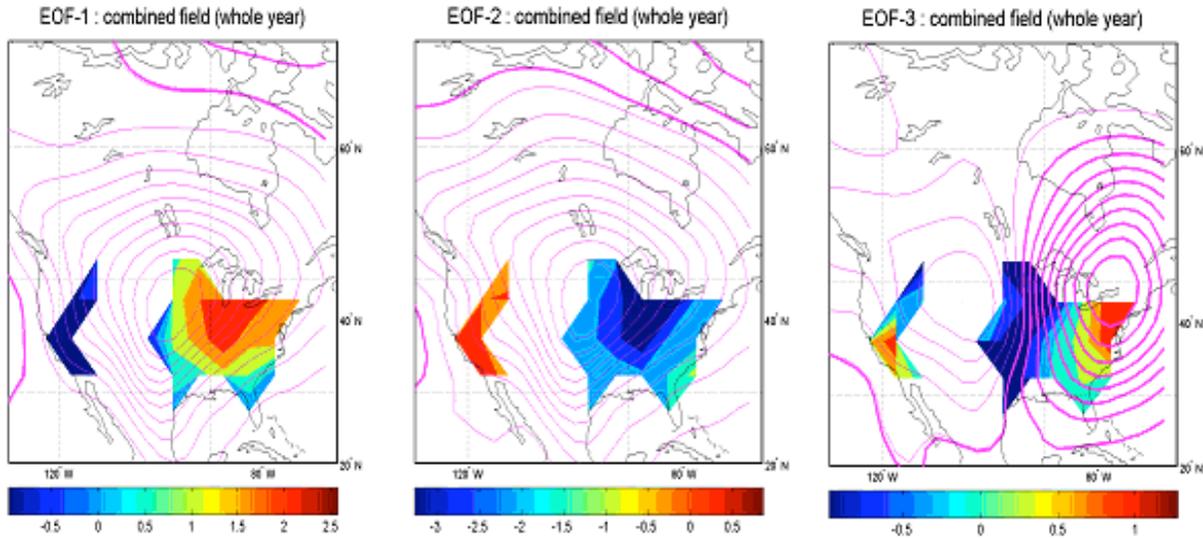


Fig. 2. Leading combined EOFs of SLP/PM2.5. Same plotting conventions as in Fig. 1.

The leading two EOFs display the same SLP pattern, but PM2.5 patterns are nearly perfectly anti-correlated. The latter patterns are similar to those of the leading two EOFs of individual PM2.5 field (not shown). We thus argue that from combined EOF analysis alone it is impossible to unambiguously conclude whether any of these two combined EOFs represents a genuine relation between SLP and PM2.5 fields, since (i) PM2.5 variability cannot physically influence SLP variability; (ii) the combined EOFs may be dominated by variability of PM2.5 component alone; and (iii) the same synoptic (SLP) pattern cannot result in opposite PM2.5 patterns.

Similar considerations apply to EOF-3, which shows the pattern partially consistent with our hypothesis about SLP/PM2.5 connection, as positive SLP anomalies over the east coast are superimposed on relatively high PM2.5 values there\*. Is this mode, or any of the leading two modes, indeed representative of true relation, or is it an artifact of the EOF analysis? In order to

---

\* The opposite relation is seen in the west coast, but it turns out that analogous summertime combined EOF shows spatial correlation over the whole region.

address the above question, we analyzed our pair of fields via MCA analysis to identify the SLP and PM2.5 patterns characterized by strongest co-variability. If some of the combined EOF patterns identify true relationship between SLP and PM2.5, we would expect to find similar patterns by the MCA analysis.

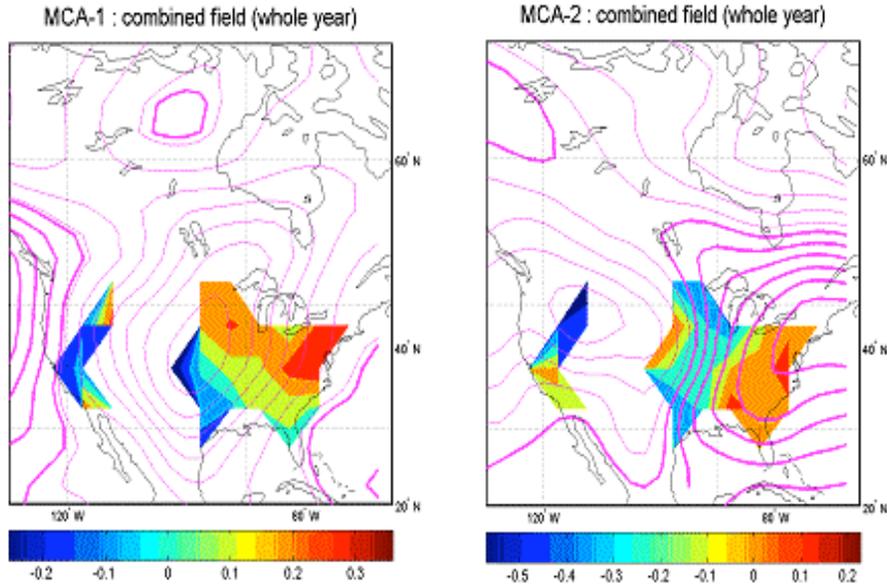


Fig. 3. Leading MCA patterns: Same plotting conventions as in Fig. 1.

The three leading MCA modes account for 53, 25 and 8% of total squared covariance, respectively; the two leading modes are shown in Fig. 3. The MCA-1 mode has similarities to combined EOF-1, with SLP anomaly over the central US, and opposite-sign PM2.5 anomaly to the east. In the same way, the MCA-2 is analogous to EOF-3: both are characterized by west-to-east oriented SLP dipole, positively spatially correlated with its companion PM2.5 pattern. Finally, combined EOF-2 has no MCA analog, and we therefore conclude it to be dominated by PM2.5 variability alone, since it does not account for any substantial fraction of the two fields' co-variability.

While the leading patterns of combined EOF and MCA analysis bear general similarities, they are not perfectly correlated. In discussing possible dynamical meaning of these patterns below, we concentrate on the MCA patterns, since they are not affected by the within-field covariance information (whereas combined EOF modes are) and look directly into the structure of the two

fields co-variability. We argue that the leading MCA modes do show a type of relationship between synoptic disturbances and air pollution we would expect from qualitative physical considerations. This relationship is most evident in MCA-2 pattern, which is dominated by positively correlated west-east oriented dipoles of SLP and PM<sub>2.5</sub> over US. However, this in-phase relationship is less obvious in MCA-1 pattern. We interpret this discrepancy by noting that synoptic disturbances captured in MCA analysis may not be stationary, but rather tend to propagate from west to east. The high/low air-pollution concentrations are likely to be seen in areas which have been experiencing high/low SLP conditions for a few days, so, in general, upstream (westward) of SLP maxima/minima. The MCA-1 pattern shows negative SLP anomaly over US and positive SLP values to the east of US. Accordingly, the PM<sub>2.5</sub> values are negative in the central/western part of the US and positive in the east, consistent with our interpretation above.

#### **4. Summary and future work**

This study focused on quantifying the relationship between synoptic conditions and air pollution over the continental US, using regional combined EOF analysis and MCA analysis of Sea-Level Pressure (SLP) and concentrations of Particulate Matter of less than 2.5  $\mu\text{m}$  in diameter (PM<sub>2.5</sub>). Both analyses produce consistent results, which support the existence of positive correlation between SLP and PM<sub>2.5</sub> concentrations, so that anticyclonic conditions are associated with high air pollution and vice versa. The MCA mode that emphasizes this relationship accounts for about 25% of the total squared covariance. However, the leading MCA mode, which accounts for nearly 50% of the total squared covariance, argues for a more general interpretation, which differentiates between stationary and propagating anomalies. Anomalous air-pollutant concentrations clearly occur in locations experiencing anomalous SLP conditions for a few days. Therefore, for synoptic patterns propagating from west to east, one would expect highest (lowest) PM<sub>2.5</sub> concentrations westward of high (low)-pressure centers. We thus interpret the leading two MCA modes as the mixture of the above “propagating” and “stationary” responses, with the MCA-1 dominated by the former, and MCA-2 by the latter.

Several possible extensions of the present work include: (i) explicitly accounting for time-lag information in analyzing the two fields co-variability [for example, by applying Multi-channel Singular Spectrum Analysis – M-SSA (Vautard, 1999)] — this would distinguish between stationary and propagating covariance; (ii) considering low-frequency SLP anomalies rather than high-pass filtered anomalies to emphasize stationary anomalies and address long-range air-pollution prediction issues; (iii) analyzing emission data sets to quantify fraction of PM<sub>2.5</sub> variability associated with variable emissions; (iv) employing health-related data bases and using the above analyses to provide foundation for establishing objectively motivated environmental standards.

## References

- Bretherton, C. S., Smith, C., and Wallace, J. M., 1992: An intercomparison of methods for finding coupled patterns in climate data series. *J. Climate*, **5**, 541–560.
- Kalnay, E., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Preisendorfer, R. W. 1988: *Principal component analysis in Meteorology and Oceanography*. Elsevier, Amsterdam, 425pp.
- United States Environmental Protection Agency (USEPA), 2008: Air-pollution data for this study is retrieved from the official website of USEPA  
<http://www.epa.gov/ttn/airs/airsaqs/detaildata>.
- Vautard, R., 1999: Patterns in time: SSA and MSSA in Analysis of climate variability-Applications of statistical techniques. 2-nd edition, H. Von Storch and A. Navarra, Eds., Springer-Verlag, Berlin, 265–286.
- Wallace, J. M., Smith, C., and Bretherton, C. S., 1992: Singular value decomposition of wintertime sea-surface-temperature and 500-mb height anomalies. *J. Climate*, **5**, 561–576.

## 7. Persistent Anomalies of Mid-latitude Atmospheric Flow

By *Nicholas Schwartz*

**Abstract:** Output of a realistic atmospheric model is analyzed to establish relationship between regionally intensified and predominantly zonally symmetric persistent wind patterns.

### 1. Introduction

Over the past few decades, there have been numerous studies regarding the low-frequency variability of mid-latitude atmospheric flows. One way to define low-frequency variability is in terms of persistent large-scale flow patterns that occur on time scales of at least 10 days. These structures are referred to as “regimes.” Identifying these regime patterns may have implications for improving medium to long-range weather and climate prediction (Marshall and Molteni 1993; Kondrashov et al. 2004; Kravtsov et al. 2008). The goal of this study is to examine the present relationship between regionally intensified and predominantly zonally symmetric regimes using a novel technique for identification of anomalously persistent wind patterns.

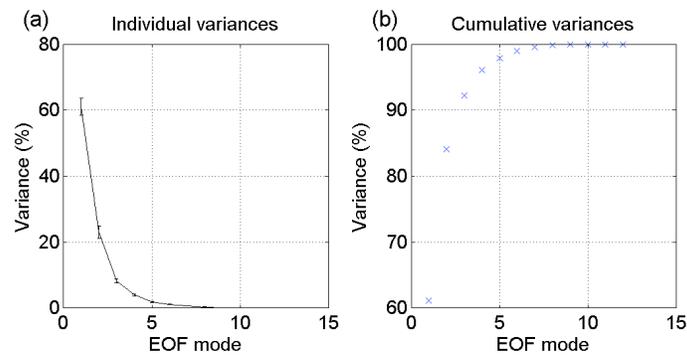


Fig. 1. Percentage of variance accounted for by EOF modes of the zonal-mean zonal wind. (a) Individual variances: Errorbars for each EOF mode represent one standard deviation of scaled variance computed over 100 synthetic realizations of a 10-dimensional red-noise model (see text). (b) Cumulative variances for each EOF mode.

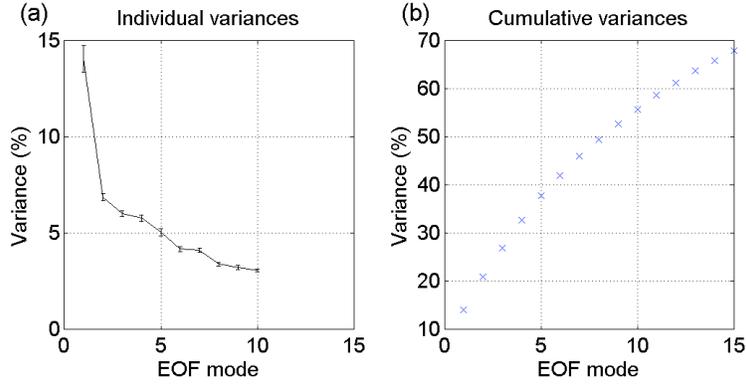


Fig. 2. As in Fig. 1, but for the 200-mb streamfunction EOFs.

## 2. Data set

We analyzed output from a perpetual-winter integration of a quasi-geostrophic 3-layer (QG3) model of Marshall and Molteni (1993). This model describes the time evolution of QG potential vorticity  $q$ :

$$\frac{\partial q}{\partial t} = -J(\psi, q) - D(\psi) + S,$$

where  $\psi$  is the streamfunction,  $D(\psi)$  a linear operator that represents dissipative terms,  $S$  is a source term, and  $J$  the Jacobian; the equations are discretized on a T21 triangular grid. In the following analysis, we considered a 10000-day sample of the 200-mb streamfunction anomalies, with respect to climatology, in the Northern Hemisphere. In addition to the streamfunction anomalies, we also computed and analyzed 200-mb zonal-mean zonal velocity anomalies.

## 3. Methodology

### *a. EOF Analysis*

We used Empirical Orthogonal Function (EOF) analysis (Preisendorfer 1988) primarily as a data compression tool to reduce the model-space dimension (of about 1500 variables),

while capturing the majority of the variance contained in the original data set. EOF analysis was performed on the anomalies for both the streamfunction and zonally averaged zonal wind data sets, weighted by the square root of the cosine of latitude to account for meridian convergence. The EOF spectra for the 200-mb zonal-mean zonal wind metric and 200-mb streamfunction metric are shown in Figs. 1 and 2, respectively. The eigenvalue uncertainties are estimated by the Monte-Carlo method using a 10-dimensional red-noise model, which is introduced and further explained in section 3b. For the zonal wind metric the first four EOF modes are statistically distinct, and account for 61%, 23%, 8%, and 4% of the total variance, respectively (96% of the total variance cumulatively). For the streamfunction metric, modes 2, 5, as well as the mode pairs 3–4 and 6–7 are statistically significant, while the first 15 modes account for about 68% of the total variance.

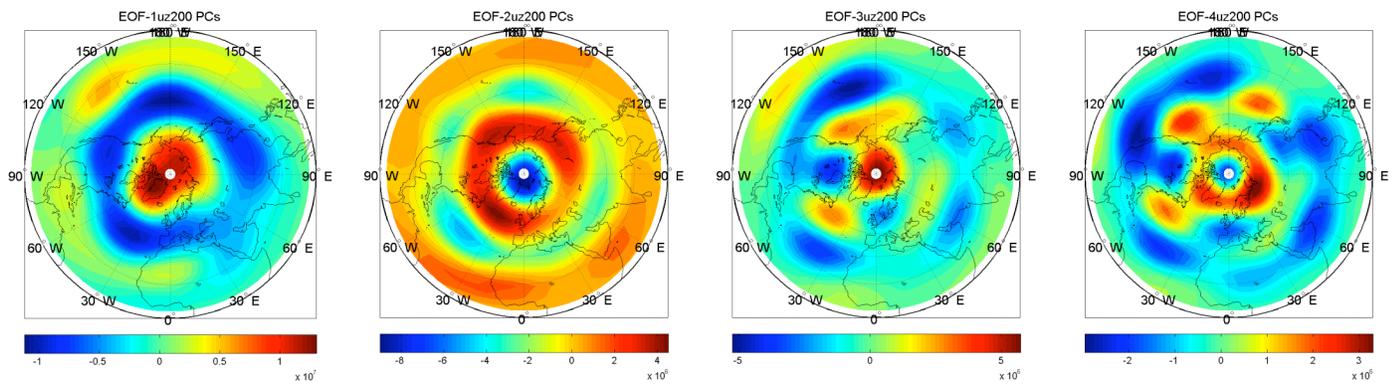


Fig. 3. The projections of normalized PCs of the zonally averaged zonal-wind onto the 200-mb streamfunction ( $\text{m}^2/\text{s}$ ).

The projection of the streamfunction anomalies onto normalized principle components (PCs) associated with the zonal-wind and streamfunction EOFs are shown in Figs. 3 and 4, respectively. The projections of the leading two PCs of the zonal wind have a pronounced zonally-symmetric component (Fig. 3, left two panels) and describe meridional displacements and intensification/weakening of the zonal-mean zonal jet. This variability is often referred to as being associated with the Arctic Oscillation (AO) teleconnection pattern (Wallace and Gutzler 1981). Note that projections of zonal-wind PCs three and four onto the streamfunction time series are much less zonally symmetric compared to the two leading modes.

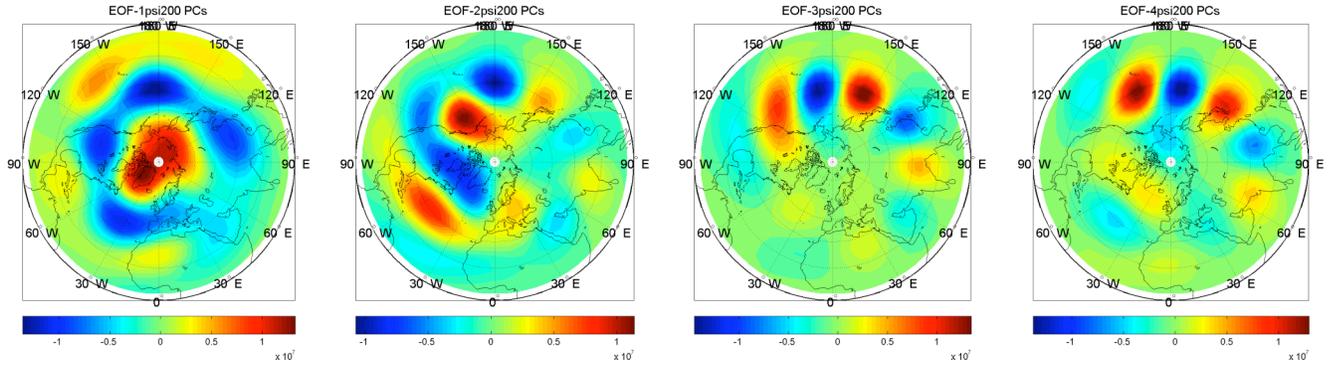


Fig. 4. EOFs of the 200-mb streamfunction ( $\text{m}^2/\text{s}$ ).

The leading EOF of the streamfunction (Fig. 4, left panel) is very similar to that of the leading EOF of the zonal wind (Fig. 3, left panel), but the remaining modes are different. The second mode is a combination of a dipolar pattern characteristic of the North Atlantic Oscillation (NAO) and the Pacific/North American Pattern (PNA), while EOFs 3 and 4 are dominated by a wave train in the Eurasian/Pacific sector. For both sets of EOFs, largest-scale lowest-frequency behavior is captured by the two leading EOFs.

#### *b. Linear stochastic model*

In assessing statistical robustness of EOF modes described in section 3a, we employed a linear stochastic model constructed using the methodology of Kravtsov et al. (2005, 2006). Two different 10-dimensional models were constructed in the phase subspaces of the ten leading zonal-wind modes and streamfunction EOFs, respectively. These statistical models are multi-level extensions of the linear inverse models (LIMs; Penland 1989); the multi-level structure allows the models to better capture the time dependence contained in the original multivariate time series [see Kravtsov et al. (2005, 2006, 2008) for further details]. Using these models, we produced 100 surrogate realizations of the leading ten PC time series of each data set. To illustrate how well our linear stochastic models capture the variability of the original data sets, we plot the one-dimensional probability density function (PDF) and autocorrelation functions of the leading PCs for both the zonal wind and streamfunction (Figs. 5 and 6), as well as the 3<sup>rd</sup> and 97<sup>th</sup> percentile of these quantities associated with the stochastic surrogates (dashed lines).

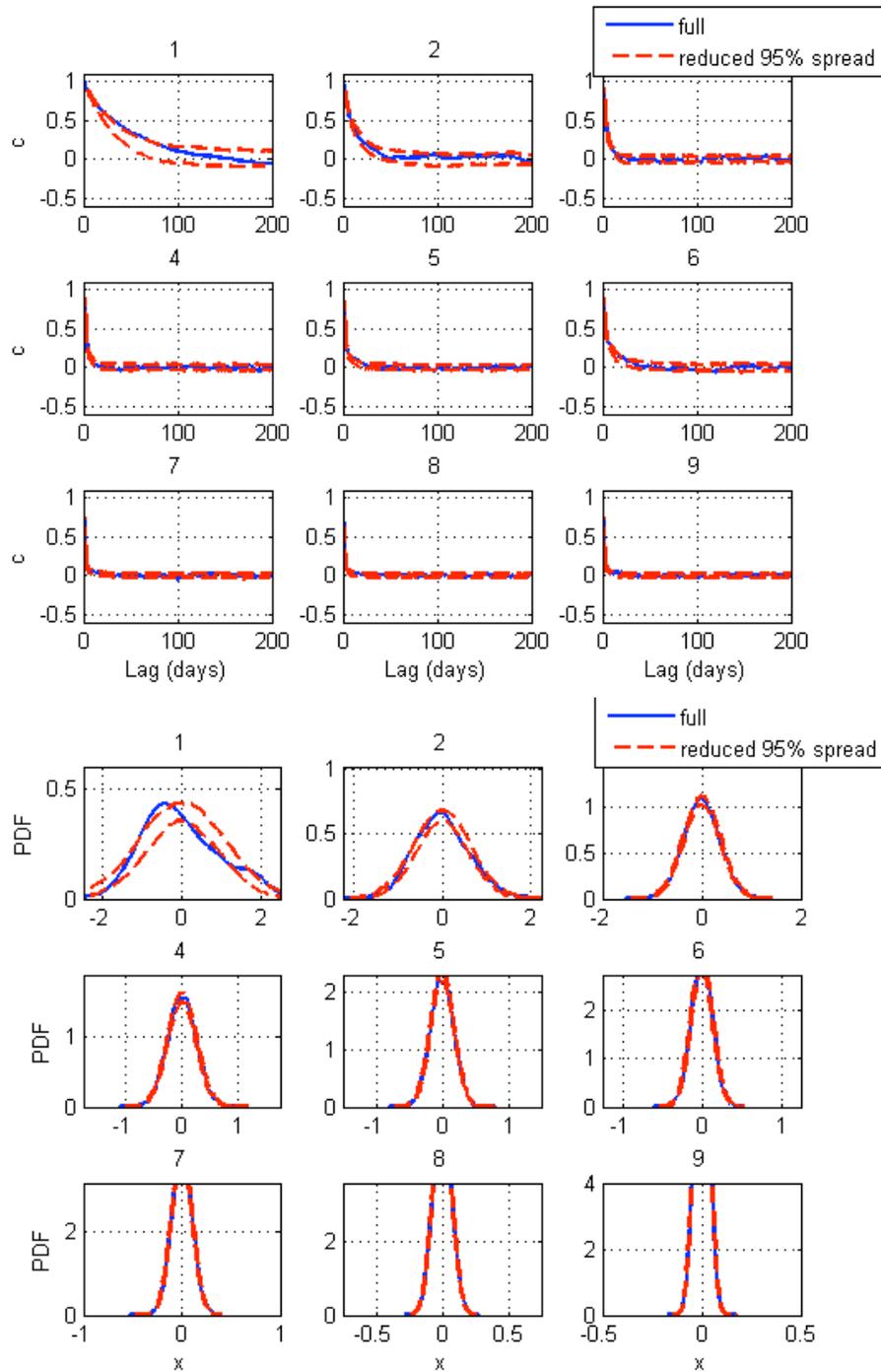


Fig. 5. Autocorrelation functions (top) and one-dimensional probability density functions (bottom) for the first 9 zonally-averaged zonal wind PCs. The PC indices are shown in panel captions. The blue solid lines represent the PDF and autocorrelation for the actual data, and the red dashed lines represent the 95% spread from the linear stochastic model. Principle components have been normalized by the standard deviation of the first PC prior to analysis.

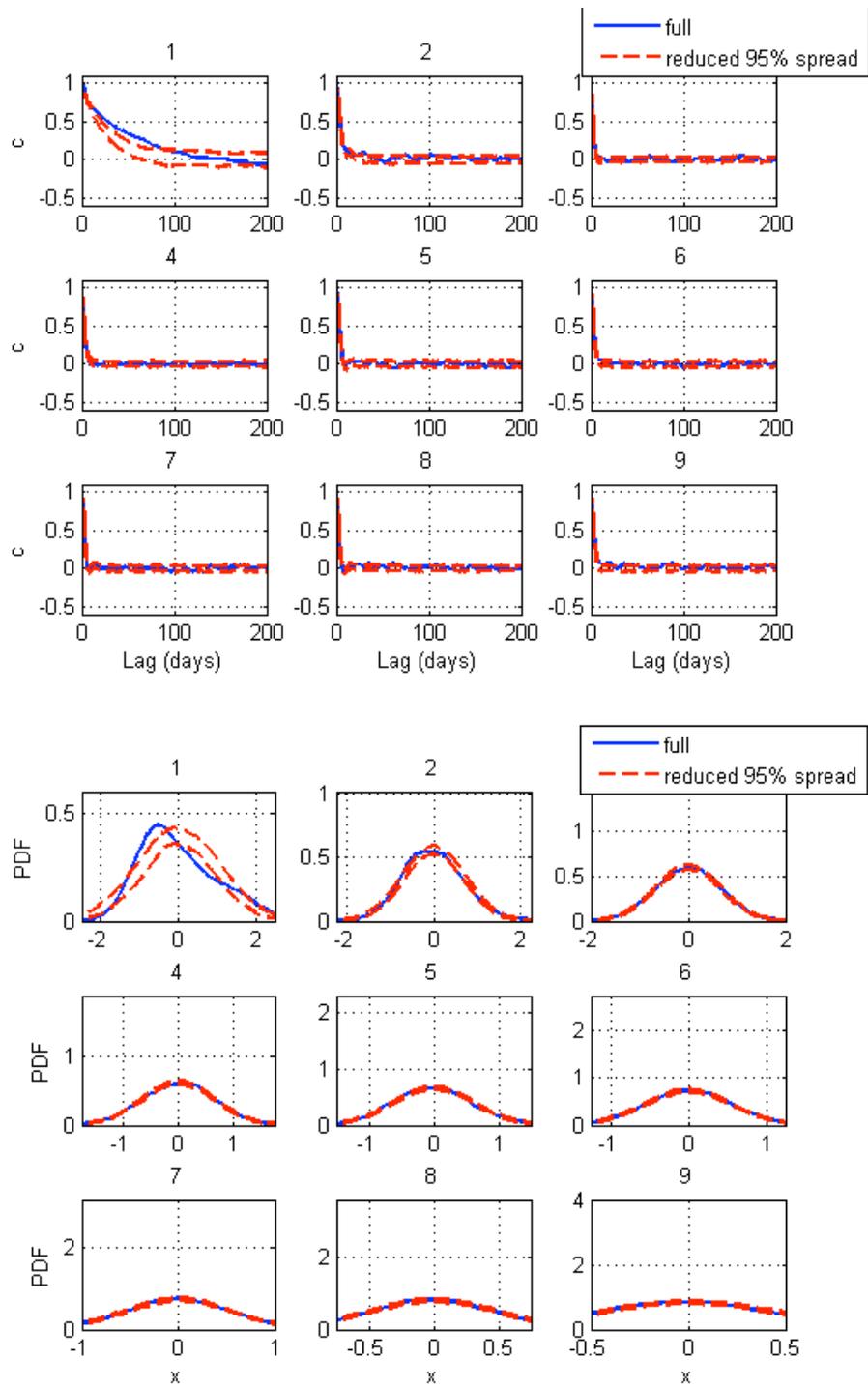


Fig. 6. Same as in figure 3, but for the steamfunction-based stochastic model.

For the majority of the trailing PCs, the actual PDF and autocorrelation functions are within the bounds predicted by the linear stochastic model. However, the leading PC of both the zonally averaged wind and streamfunction data sets exhibit somewhat more substantial deviations from the linear null hypothesis. Most notably, these modes are characterized by non-gaussian PDFs, which indicates that nonlinear processes may be instrumental in causing and maintaining these modes. We next look more closely at the structure of the model trajectories in the phase space of the leading two EOFs for both the zonally averaged wind and streamfunction metrics. Our objective is to identify anomalously persistent flow patterns (regimes), which are possibly associated with workings of nonlinear processes, as well as to compare the regimes defined using zonal-wind and streamfunction metrics.

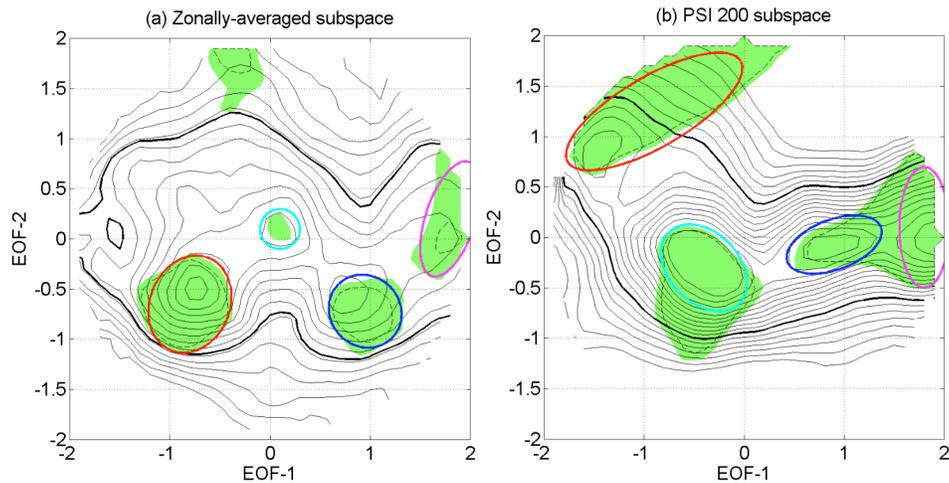


Fig. 7. Regime regions in the (a) zonal wind phase space; and (b) the streamfunction phase space. Contours ( $[0.7, 1.5]$ , contour interval 0.05) show the ratio of low-pass filtered PCs' PDF to raw PCs' PDF; the PDFs were computed in the subspace of the two leading EOFs. The shaded region represents the areas, in which the PDF ratio exceeds the 95% confidence level of the corresponding linear stochastic model. Potential regimes are marked by ellipses (Regime-1 – magenta, Regime-2 – blue, Regime-3 – red, Regime-4 – cyan). These ellipses are chosen to capture statistically significant PDF-ratio maxima (see text).

### *c. Identification of Regimes*

We adopted here the regime-identification procedure of Kravtsov et al. (2008). Their method involves computing 2-dimensional PDFs using original and low-pass filtered PCs of the fields considered (we used 10-day running means to compute the low-pass filtered PCs). We first computed, for every location in the EOF-1–EOF-2 phase space, the ratios of low-pass filtered data PDF to the unfiltered data PDF, for both zonally averaged wind and

streamfunction-based PC data sets, as well as for each of their corresponding 100 surrogate PCs from the stochastic model simulations. We then defined regime regions as areas of local PDF-ratio maxima where this ratio exceeds the 95% confidence level based on the linear stochastic model. These areas are dominated by anomalies of abnormally (nonlinearly) high persistence (relative to the linear-model ratios). The PDF ratios for the zonal-wind and streamfunction data sets are shown in Figs. 7a and b, respectively. Four regimes are identified in each case, and the regime regions are approximated by elliptical areas subjectively chosen to capture the local maxima and surrounding area of a statistically large PDF ratio (shaded regions).

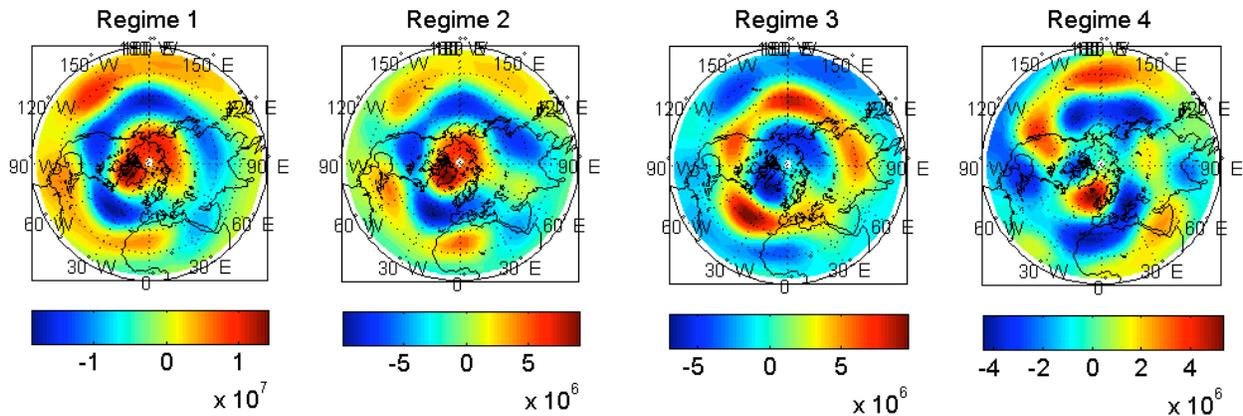


Fig. 8. Composite streamfunction anomalies associated with the 4 regimes (see Fig. 7a) based on the zonally averaged zonal wind analysis.

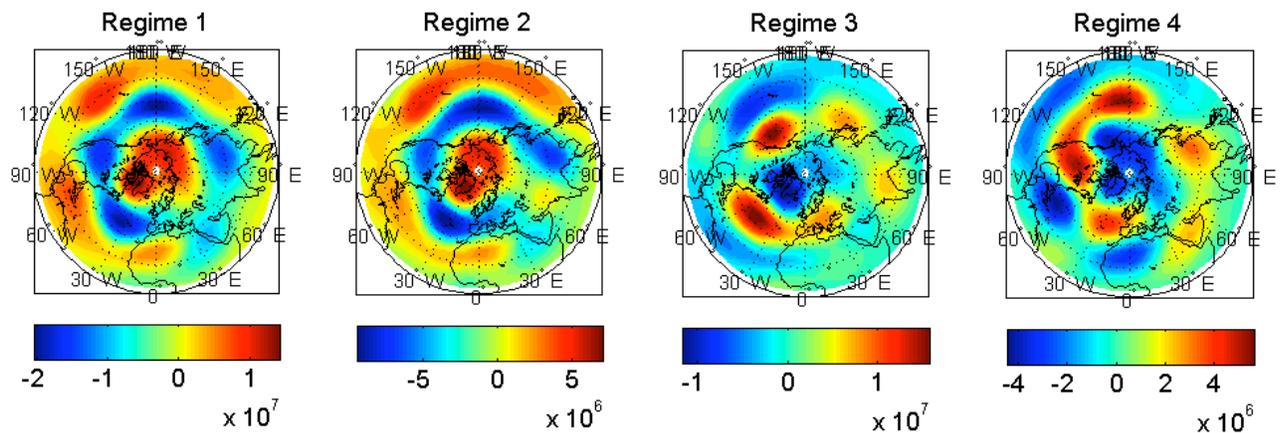


Fig. 9. Same as in Fig. 8, but for the streamfunction-norm regimes (Fig. 7b).

Finally, we computed the regime-day indices by marking the days during which the two-dimensional low-pass filtered PC trajectory was passing through a given regime area. Out of 10000 days, the regime populations were as follows: for the zonally averaged zonal wind metric, Regime-1 population was 333 days, Regime-2 – 400 days, Regime-3 – 941 days, and Regime-4 – 225 days. For the streamfunction metric, Regimes 1–4 contained 403, 452, 608, and 1341 days respectively. We computed the composite streamfunction anomalies (regime patterns) by averaging the anomalies corresponding to regime-day indices, for each of the four regimes, and for both the zonally averaged wind and streamfunction-based analyses.

#### 4. Findings and future work

The composite patterns for the zonal-wind-norm and streamfunction-norm regimes are shown in Figs. 8 and 9, respectively. While the regime populations computed using the two types of analysis exhibit substantial differences\* (see above), these analyses identify essentially the same regime patterns (compare the corresponding panels of Figs. 8 and 9). Regimes 1 and 2 in both cases are characterized by a pronounced zonally symmetric component, and we refer to them as  $AO^-$  and  $AO^-$  regimes, since they share the same zonally symmetric pattern, but Regime-1 patterns possesses larger-magnitude anomalies compared to that of Regime-2. Regime-3 patterns for zonal-wind-based and streamfunction-based analyses are well correlated, but the latter has a much less pronounced zonal symmetry than the former. We conclude, therefore, that the zonal symmetry of this regime in the zonal-wind-based analysis is an artifact of zonal averaging, and call this regime the  $NAO^+$ , since its pattern is dominated by a dipole over the North Atlantic Ocean. The fourth regime is not zonally symmetric for both types of analysis.

The above regimes are more or less identical to those found by Kondrashov et al. (2004), who used the output from the same QG3 model, but very different methodology of regime identification. Furthermore, these regimes have patterns which bear many similarities with the observed ones (Wallace and Gutzler 1981; Smyth et al. 1999). One of the emerging conclusions from the present work concerns dynamical distinctness of AO and NAO teleconnection patterns. In particular, we couldn't identify  $AO^+$  or  $NAO^-$  regimes, which

---

\* These differences may be due to a relatively short data record analyzed here.

points to nonlinear differences between these two entities; this contrasts an alternative view that NAO is merely a regional manifestation of AO.

In terms of future work, we will further quantify the similarities between the zonal-wind and streamfunction regimes, using such quantities as, for example, the joint probability of regime occurrence. We will then apply the method described here to a larger data set to ensure statistical significance. Finally, we will apply our methodology to the analysis of the observed atmospheric flow in the Northern Hemisphere.

## References

- Kondrashov, D., K. Ide, and M. Ghil, 2004: Weather regimes and preferred transition paths in a three-level quasigeostrophic model. *J. Atmos. Sci.*, **61**, 568–587.
- Kravtsov, S., D. Kondrashov, and M. Ghil, 2005: Multi-level regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J. Climate*, **18**, 4404–4424.
- Kravtsov, S., A. W. Robertson, and M. Ghil, 2006: Multiple regimes and low-frequency oscillations in the Northern Hemisphere's zonal-mean flow. *J. Atmos. Sci.*, **63**, 840–860.
- Kravtsov, S., J. Ten Hoeve, S. B. Feldstein, S. Lee, and S. -W. Son, 2008: The relationship between statistically linear and nonlinear feedbacks and zonal-mean flow variability in an idealized climate model. *J. Atmos. Sci.*, submitted.
- Marshall, J. and F. Molteni, 1993: Toward a dynamical understanding of atmospheric weather regimes. *J. Atmos. Sci.*, **50**, 1792–1818.
- Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.*, **117**, 2165–2185.
- Preisendorfer, R. W., 1988: *Principal Component Analysis in Meteorology and Oceanography*. Elsevier: New York, 425 pp.
- Smyth, P., K. Ide, and M. Ghil, 1999: Multiple regimes in Northern Hemisphere height fields via mixture model clustering. *J. Atmos. Sci.*, **56**, 3704–3723.
- Wallace, M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812.