USE OF HURST AND RENYI ANALYSIS TO DETECT AND CHARACTERIZE PACIFIC DECADAL OSCILLATION IMPACTS ON CLIMATE VARIABILITY IN ALASKA

By

Jean K. Talbot

RECOMMENDED:

_____

_____

_____

_____
Advisory Committee Chair

_____
Chair, Department of Atmospheric Sciences

APPROVED:

_____
Dean, College of Natural Sciences and Mathematics

_____
Dean of the Graduate School

_____
Date

USE OF HURST AND RENYI ANALYSIS TO DETECT AND CHARACTERIZE PACIFIC DECADAL OSCILLATION IMPACTS ON CLIMATE VARIABILITY IN ALASKA

A

THESIS

Presented to the faculty of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

By
Jean K. Talbot, B.A.
Fairbanks, Alaska
December 2011

ABSTRACT:

While climate systems are known be nonlinear, most statistical tools used to study climate are linear. Two nonlinear analyses are introduced for indicating predictability in climate studies: Hurst analysis and Renyi analysis, the advantages of which are illustrated by applying both to characterize Alaska climate time series 'dynamics' or temporal evolution. These methods are also applied to reanalysis and model data to compare with the observational analysis. Hurst analysis is used to calculate long term predictability in data on a scale of five to 15 years; Renyi analysis is used to quantify the degree of order on a time scale of two to 15 days.

The analyses revealed that temperature may be more statistically predictable in certain areas of Alaska during the positive phase of the Pacific Decadal Oscillation (PDO). Circulation effects associated with the PDO shift are found to plausibly cause the change in randomness of the SAT data.

Table of Contents Page

List of Figures Page

List of Tables Page

List of Appendices Page

Acknowledgements

Thanks to my advisor, my committee, and the rest of my coauthors for making this project come together. Many thanks to Barbara Day for helping me keep track of paperwork, deadlines, and my sense of humor. Thanks to the faculty of the Atmospheric Sciences department for their helpful discussions and enthusiasm. Thanks to D. Dammon, J. Mayfield, S. Stegall for helpful discussions on this project. Thanks to R. Bekryaev for useful discussions on the different methods of finding the Hurst exponent. Thanks to Flora Grabowska for help finding rare reference material. Adam Phillips and David Bailey are thanked for helping us obtain the appropriate CCSM4 data. This work is supported by funding from the Department of Energy through grant number DE-SC0001898. Support for the Twentieth Century Reanalysis Project data set is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office.

Use of Hurst Analysis and Renyi Information to Detect and Characterize Pacific Decadal Oscillation Impacts on Climate Variability in Alaska

1. Introduction

When nonlinear methods are applied to climate time series the results are often difficult to relate to weather and climate prediction, which is the main goal of climate studies. This study examines two nonlinear analysis techniques for their application to predictability in climate studies: Renyi analysis, a nonlinear measure of randomness on the short time scale of up to a few weeks, and Hurst analysis, a measure of memory on a time scale of five to 15 years.

Renyi analysis quantifies the randomness and order by finding the frequency of patterns in the data. Order in terms of Renyi analysis is directly related to predictability because a high frequency of certain patterns means those events are more statistically likely. Alekseev & Yakobson (1981) used symbolic dynamics, including Renyi analysis, to mathematically describe generated time series. Renyi analysis was employed by Voss et al. (1996) in an algorithm to classify high risk cardiac patients, and it was used by Krutzman et al. (2008) to show the mixing effect of the Antarctic polar vortex on southern hemisphere stratospheric mixing of methane. We use it to find patterns in consecutive days of near median data compared to extreme data.

Hurst analysis is a measure of persistence, or the probability of variability patterns. It was developed by Hurst (1951) to analyze the long term dynamics of the flow of the Nile river. Since then it has been used to analyze climate data, but often the interpretation of the results and connection to linear climate studies is left unclear. Mandelbrot and Wallis (1968) utilized Hurst analysis to quantify the “noah and joseph effects”, referring to the amplitude of climate extremes and the length of time that abnormal conditions can persist in hydrological data. Tsonis and Roebber (1999) found that the geopotential height has less persistence at higher latitudes than in the tropics on a

scale of a week to several years. Govindan et al. (2002), using Hurst on observational temperature data and climate model outputs with greenhouse and aerosol forcing, found that all of the climate models they tested underestimated the Hurst exponent. Furthermore, Zhu et al. (2010) argued that climate models may be improved for long term predictions in locations that exhibit long term persistence, relating the statistical probability of Hurst analysis to a predictability in climate models.

Global temperatures have spacial patterns of persistent and random long term behavior. Using the Detrended Fluctuation Analysis method for determining Hurst exponents, Fraedrich and Blender (2003) found that, on the time scale of 1-15 years, coastal stations exhibited long term persistence, whereas continental stations exhibited random behavior. They attributed the persistence at the coastal stations to the relatively long term consistency of maritime influence. In contrast, Király and Jánosi (2005) found that the pattern of random behavior far from coasts did not hold for Australia, and in their global study Király et al. (2006) found no connection between long term memory and distance from coast. However, the Király studies focused on a time scale of only a few weeks to 5 years. It is possible that their study focused on too short a time scale to be comparable with the Fraedrich and Blender study, and also possible that the interior of Australia has inherently different climate dynamics then other continental regions.

The main goal of this study is to demonstrate how Renyi analysis and Hurst analysis of observational climate data can be applied to climate data in terms of predictability and to provide hints on how to further explore the data and eventually lead to an understanding of the climate mechanisms that operate within the system.

The novel aspects of this study are the application of Hurst and Renyi analysis to climate data and the use of the results to outline plausible climate mechanisms. The key questions that this study explores are the following:

- Are Hurst and Renyi analyses applicable for investigating climate mechanisms?

- How does Hurst and Renyi analyses of Alaska station SAT and SLP compare with NCEP/NCAR Reanalysis? Does coarse graining of the station data have an impact on the results?
- Is persistence or order of the climate data consistent with time? If not, do the methods identify a climate shift?
- How do the Hurst and Renyi analyses compare before and after the climate shift? Are they consistent with each other, and with all data sources?
- Is the observed effect of the recent period of the PDO common to all periods of the PDO?
- Can this analysis be used for predictability for Alaska?

It is prudent to apply new methods in regions where some prior knowledge of the climate variability is known. To this end, data from observations, reanalysis and model output were analyzed and compared for Alaska. First Hurst and Renyi analyses were applied to the observational data and the results were used to identify climate variability mechanisms. The analyses were then applied to reanalysis data to evaluate whether similar variability operates in these data sets as found in the observational data. Finally, Twentieth Century Reanalysis and global climate model simulations were investigated to test whether the observed climate mechanisms are can be generalized in a longer climate data set.

2. Data and Methods

2.1 Data

Surface air temperature (SAT) data at 26 Alaska stations (Figure 1) was acquired from the National Climate Data Center (NCDC) Global Summary of the Day data set, in which each daily temperature is constructed from the daily mean of the hourly data. The stations considered had at least 50 years of consecutive data with no more than 10% total missing data, and no more than 30 consecutive days of missing data. Missing data were

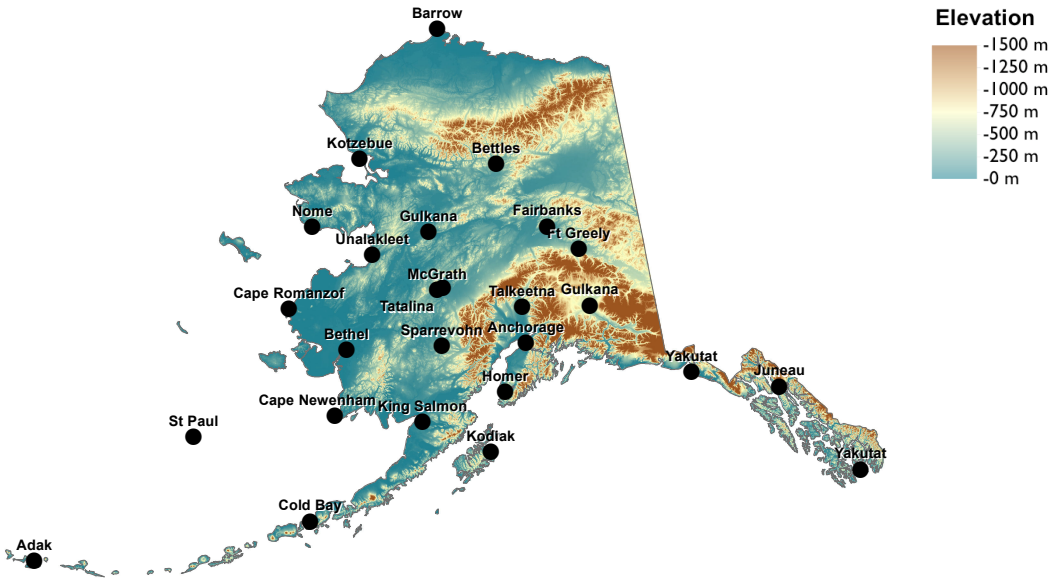


Figure 1: Locations of Alaska Stations

linearly interpolated. In addition, monthly SAT data from North Asian stations were used from Bekryaev et al. (2010) and are available online at http://climate.iarc.uaf.edu/geonetwork/srv/en/main.home under the name 'Data Sets of Monthly Mean Surface Air Temperature'. Gridded daily SAT and sea level pressure (SLP) fields from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) Reanalysis surface data set were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at http://www.esrl.noaa.gov/psd/

To test our methods on time series longer than our station data, SAT and SLP data from 1924 to 2007 were analyzed from the Twentieth Century Reanalysis project provided by the National Oceanic and Atmospheric Administration Office of Atmospheric Research Earth Systems Research Laboratory Physical Sciences Division (NOAA/OAR/ESRL PSD).

For comparison between model output and observed data, daily temperature and pressure were acquired and analyzed from the Community Atmosphere Model (CAM), from version 4 of the Community Climate System Model (CCSM4) project http://www.ccsm.ucar.edu(), supported by the Directorate for Geosciences of the National

Science Foundation and the Office of Biological and Environmental Research of the U.S. Department of Energy. The CCSM4 data were obtained from the Earth System Grid.

The seasonal cycle was removed from the SAT and SLP time series by subtracting the the long-term daily mean and normalizing by the daily standard deviation. Each data set was first analyzed as a whole, and then separately from 1948-1975 and 1977-2007 after it became clear that the time series dynamics (temporal evolution) was not stationary.

The analysis employs SLP-based storm track information that are constructed using the algorithm of Zhang et al. (2004) on NCEP/NCAR Reanalysis. Storm count, duration and central pressure were examined in various storm locations, and in particular the Bering Sea. The number and duration of storms per year and per season were calculated for the Bering Sea, defined as the area between 55^{o} and 71^{o} N, and 180^{o} and 197^{o} E.

Monthly climate index values for the Pacific Decadal Oscillation (PDO) were acquired from University of Washington's Joint Institute for the Study of the Atmosphere and Ocean and are available at http://jisao.washington.edu. The PDO is a climate index derived from the first empirical orthogonal function (EOF) of sea surface temperature anomaly patterns in the North Pacific, known for its low frequency cycles. The North Pacific (NP) index is related to the PDO; the NP index is derived from the sea level pressure in the North Pacific, which has a strong correlation with the PDO. Monthly climate index values for the NP index were acquired from the University Corporation for Atmospheric Research, Climate and Global Dynamics, Climate Analysis Section and are available at http://www.cgd.ucar.edu/cas/.

2.2 Hurst Analysis

Standard statistics cannot do not reveal much information about long term memory in a time series; although autocorrelation can give some idea of memory, it usually falls below significant correlations after several days. Hurst analysis, on the other

hand, can show useful nonlinear correlations as far out as the length of the data set allows. These correlations classify time series behavior as persistent, anti-persistent, or random: when the time series is persistent, the next period of time (the next decade, for example) is likely to exhibit similar characteristics to the previous period; when the time series is anti-persistent, the next period is likely to anti-correlated with the previous period, and when the time series is random, then there is no relation between one period of time to the next. These are correlations that exist over many time scales.

The degree of persistence is measured by the Hurst exponent, H. In this study, H was calculated using the rescaled range (R/S) analysis (Mandelbrot and Wallis 1968, 1969), which they used to quantify the Noah and Joseph effects, referring to extremes and persistence in climate. While many methods exist to calculate the Hurst exponent, the rescaled range method is used to calculate the H in this study and it yields results comparable to other methods such as detrended fluctuation analysis (DFA) (Woodard, 2004).

For certain time lags (T) characteristic within a time series, the data usually conforms to the power law

$$R/S = kT^H \tag{1}$$

where R/S is the rescaled range (defined below) of the data at a certain time lag T, and k is a scaling constant. When 0<H<0.5, the time series is anti-persistent, when 0.5<H<1.0, the series is persistent, and when H~0.5, the series is random.

The rescaled range (R/S) for a time lag T is calculated by considering ever larger segments of the data in the series and calculating the range, normalized by the standard deviation of the segment. In each segment, a running sum of the anomalies is calculated at each data point from the first data point in the segment up to that point. These running sum values are calculated

$$Z_i = \sum_{j=1}^{i} x_j \tag{2}$$

where Z_i is the value of the running sum at point i, x_j is the anomaly value of the segment at point j, and the sum is from the first point in the segment to the i^{th}. The lowest value the running sum reaches is subtracted from the highest value, which constitutes the range for that segment. This procedure is performed for each non-overlapping segment of length T, and the mean of the resulting range values for each of these segments is found. The mean of the standard deviation from each non-overlapping segment of length T is also found. Finally the mean range is divided by the mean standard deviation, which is the R/S value for time lag T.

The size of data sequences considered are increased, for this study by a factor of 1.2, from two days to the time lag that reaches half the length of the data set, here usually about 15 years. The resulting rescaled ranges (R/S) versus the time lag are plotted on a double logarithmic scale; when several points fall in a straight line, the range of corresponding time lags can be interpreted as periods of time for which the time series conforms to a power law; the slope of that line is the Hurst exponent. It is advisable to manually examine the R/S plot when finding this slope, because if the slope is not constant, then the data for that time lag does not conform to a power law and the Hurst exponent will have no physical interpretation.

Most data sets do not maintain the same Hurst exponent at all time lags; often there is a change of exponent at certain time lags; the Hurst exponent H is only valid for sequential time lags for which the power law is constant. If the data does not conform to a power law, or if $H>1$, then the long term dynamics on that time scale are changing within the time series and the Hurst exponent cannot be calculated. We found that the Alaska temperature and pressure data usually exhibited a characteristic Hurst exponent change between one to five years, so for standardization of our long term analysis we explore the Hurst exponent from five to 15 years.

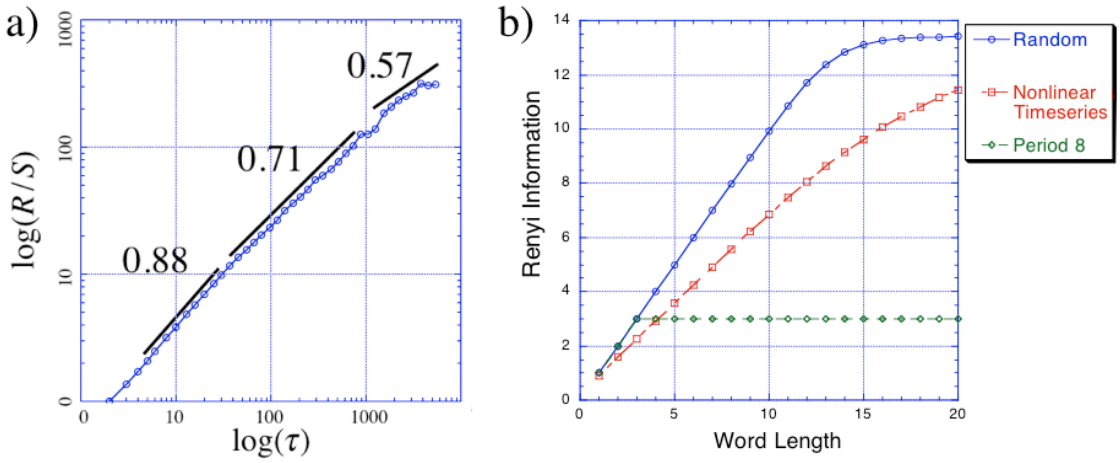


Figure 2: (a) Example plot of R/S versus time lag, Hurst exponents at different time lags are labeled. (b) Example plot of Renyi information versus word length for a random series (blue), periodic series (green), and nonlinear series (red).

Hurst Analysis can also reveal for what time lags the time series has a changing value of H, which was one of the main results of this study. For more information on the detection of non-stationarity in a time series, see Appendix B.

The method is exemplified in Figure 2a, which shows the R/S plots for the station SAT data at Nome, Alaska. The left frame includes all R/S points, and the right frame shows the points of interest from five to 15 years. It is clear that, on this time scale, the two periods exhibit very different long term memory: the temperature is random (H=0.52) from 1948-1975, and persistent (H=0.77) from 1977-2007. More persistence from 1977-2007 is typical behavior for the Hurst exponent of SAT data for most stations in interior and northwestern Alaska.

2.3 Renyi Analysis

To quantify the amount of order in a time series on a short time scale (up to 15 days in this case) Renyi information, a particular measure of Renyi analysis, is calculated as described by Wackerbauer, et al. (1994). First a partition must be chosen. For example, a common partition is the median value, leading to the binary set of symbols used {0,1};

every data point below the median is transposed to a '0' and every data point above the median is transposed to a '1'. The question being asked with this partition is whether the data changes randomly between below and above the median value, or whether it does so with some more frequent patterns. The set of symbols is called an 'alphabet', and consecutive patterns of these symbols are called 'words'.

Once a partition has been chosen, the frequencies of all the words of different length are counted. For words of length n made from an alphabet of k letters, there are k^n possible words. Overlapping words are counted. The Renyi information for a specific word length is calculated by

$$I(q) = \frac{1}{1-q} \log \sum_{i=1}^{N} p_i^q \tag{3}$$

where I is the Renyi information, N is the total number of words of the specific word length, p_i is the frequency of each word (if one of the frequencies $p_i=0$, it is not considered into the equation), and q is an integer. The integer q allows us to focus the information on more or less frequent words. For $q>>1$, only the most probable events contribute to the information; for $q<<0$, only the least probable words contribute to the information. For the limit $q \to 1$, the information becomes the Shannon information:

$$I(1) = -\sum_{i=1}^{N} p_i \log p_i \tag{4}$$

Shannon information (q=1) is often used because each word contributes to the information exactly as much as the frequency of the word.

Renyi information increases with word length; the amount it increases with word length can determine how much order is in a data set. On a plot of Renyi information versus word length, a random data set will have a slope of 1, and a repetitive data set will have a slope of 0. An example of these slopes is shown in Figure 2b. Most Alaska time series have a slope between 0 and 1; in contrast to Hurst analysis, the lower the slope, the more order there is in the time series.

There is a limit to the maximum calculable word length for the Renyi information which depends on alphabet size and data length. In a random data set, each of the words occur with similar frequency. Care must be taken with the interpretation of longer words, however: with an alphabet size k and a word length n, the number of possible words is k^n, and if the total length of the data is less the number of possible words, then it is impossible for every word to occur. After this characteristic word length, the slope of the information will flatten out. Ideally, the data length m, alphabet size k, and maximum considered word length n will be related by the equation

$$m = 5k^n \tag{4}$$

so that each possible word has a possibility of occurring 5 times.

3. Results

3.1 Station Results

The data were analyzed first from 1956-2007 and then from 1946-2007, when additional data to lengthen the time series were found. The Hurst exponents at various Alaska stations over the two time periods were notably different. This difference suggested a change in the long-term dynamics and lead to testing in order to determine dynamically similar periods. The difference in Hurst exponents was maximized when the time series was divided at 1976, coinciding with the phase shift of the Pacific Decadal Oscillation (PDO). The data were then analyzed separately for 1946-1975 and for 1977-2007 to ensure stationary Hurst exponents over the analysis period.

The PDO (Mantua and Hare 2002) is a climate index derived from sea surface temperature anomalies in the North Pacific Ocean. The PDO is known for its low frequency oscillation from 50 to 70 years (MacDonald and Case 2005). In 1976 it shifted from a negative phase that had lasted since 1946 to a positive phase which then lasted until 2007, as shown in Figure 3.

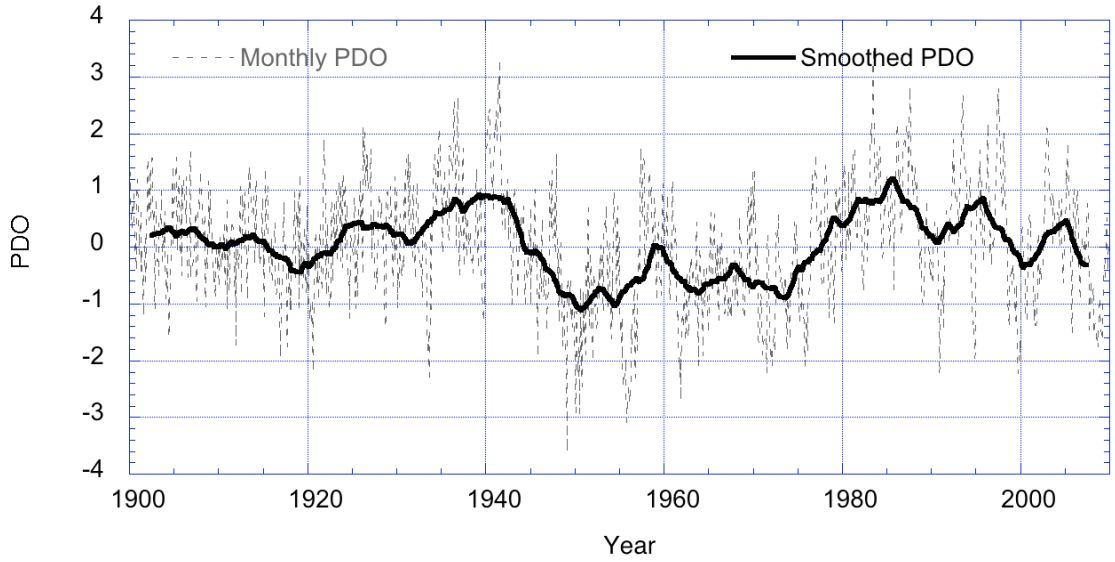


Figure 3: Monthly value of the Pacific Decadal Oscillation (grey) and the five year running average PDO (black) from 1900 to 2010.

Climate in Alaska is known to be greatly affected by the PDO. Niebauer (1998) showed that after 1976 the position of the Aleutian low changed, the frequency of El Niño events compared to La Niña events increased, and there was 5% reduction of winter sea ice in the Bering Sea. Papineau (2001) showed that the PDO affects Alaskan temperatures directly, as well as indirectly via the frequency of El Niño events. Neal et al. (2002) showed that during the positive PDO phase the majority of annual stream discharge in Southeast Alaska shifted to earlier in the year. For a complete assessment of the climatic effect of the PDO shift on each region of Alaska, see Hartmann and Wendler (2005). Because the 1976 PDO shift had such a profound effect on Alaska climate, it is not surprising that the event could also have triggered a change in the dynamic variability of Alaska temperature and pressure data.

To measure the difference in the five to 15 year Hurst exponent of SATs, the data from the 26 Alaska stations was calculated independently from 1946-1975, and from 1977-2007. Both daily and monthly mean station temperatures were analyzed. Because

the results were nearly identical, it was concluded that monthly time averaging has a minimal effect on the five to 15 year Hurst exponent.

From 1946-1975 the Hurst exponent of the temperature was generally random or weakly persistent in the interior and northwestern regions of Alaska; in Barrow it was weakly anti-persistent; in Southern Alaska the Hurst exponent was persistent or weakly persistent (Figure 4a). From 1977-2007 the Hurst exponent was considered to have changed if the difference in Hurst exponent was greater than 0.1. Out of 26 stations, 12 exhibited an increase of persistence in 1977-2007 compared to 1948-1975, all but one was located in interior and northwestern Alaska. Additionally, three stations exhibited a decrease of persistence, two in Southeast Alaska and one on the southern Bering Sea. However, all three of these decreases are slight enough for the station result to remain persistent during the positive PDO. It should also be noted that the southern coastal stations which exhibit no change remain persistent during both phases of the PDO. The change from random to persistent behavior in interior and northwestern Alaska implies that during the negative PDO phase, the behavior of the temperature is statistically unpredictable on the five to 15 year time scale, whereas during the positive PDO phase, the behavior of the temperature is statistically predictable on the five to 15 year time scale.

Next we used Hurst analysis to analyze the sea level pressure (SLP) time series at Alaska stations. We found that while SLP's undergo a notable change of Hurst exponent with the change of the PDO in some of the same locations as the change in SAT's, it is a qualitatively different change from that seen in temperature. Out of 19 stations, 12 in interior, western and southeastern Alaska showed an increase in Hurst exponent (Figure 5). Most of these changed from weakly anti-persistent during the negative PDO phase to weakly persistent during the positive PDO phase. Galena is the main exception, which changed from weakly persistent to persistent. Only the King Salmon station became less persistent. The change from weakly anti-persistent to weakly persistent SLP's in interior, western and southeastern Alaska implies that the SLP at most of these stations is no more

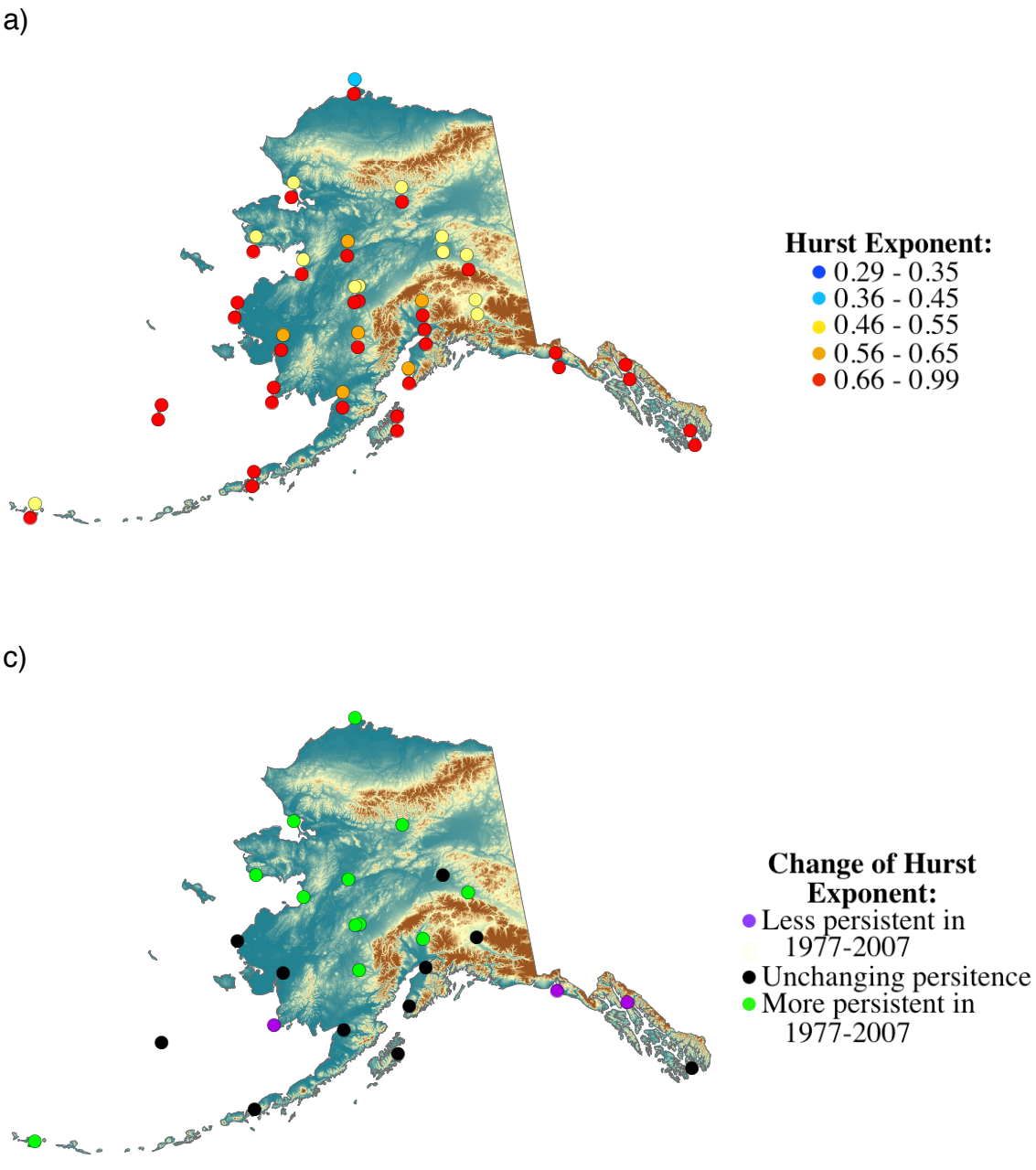


Figure 4: (a) Persistence during 1948-1975 (upper dot) and during 1977-2007 (lower dot) for station SAT. (b) Change of persistence during 1977-2007 compared to 1948-1975 for station SAT.

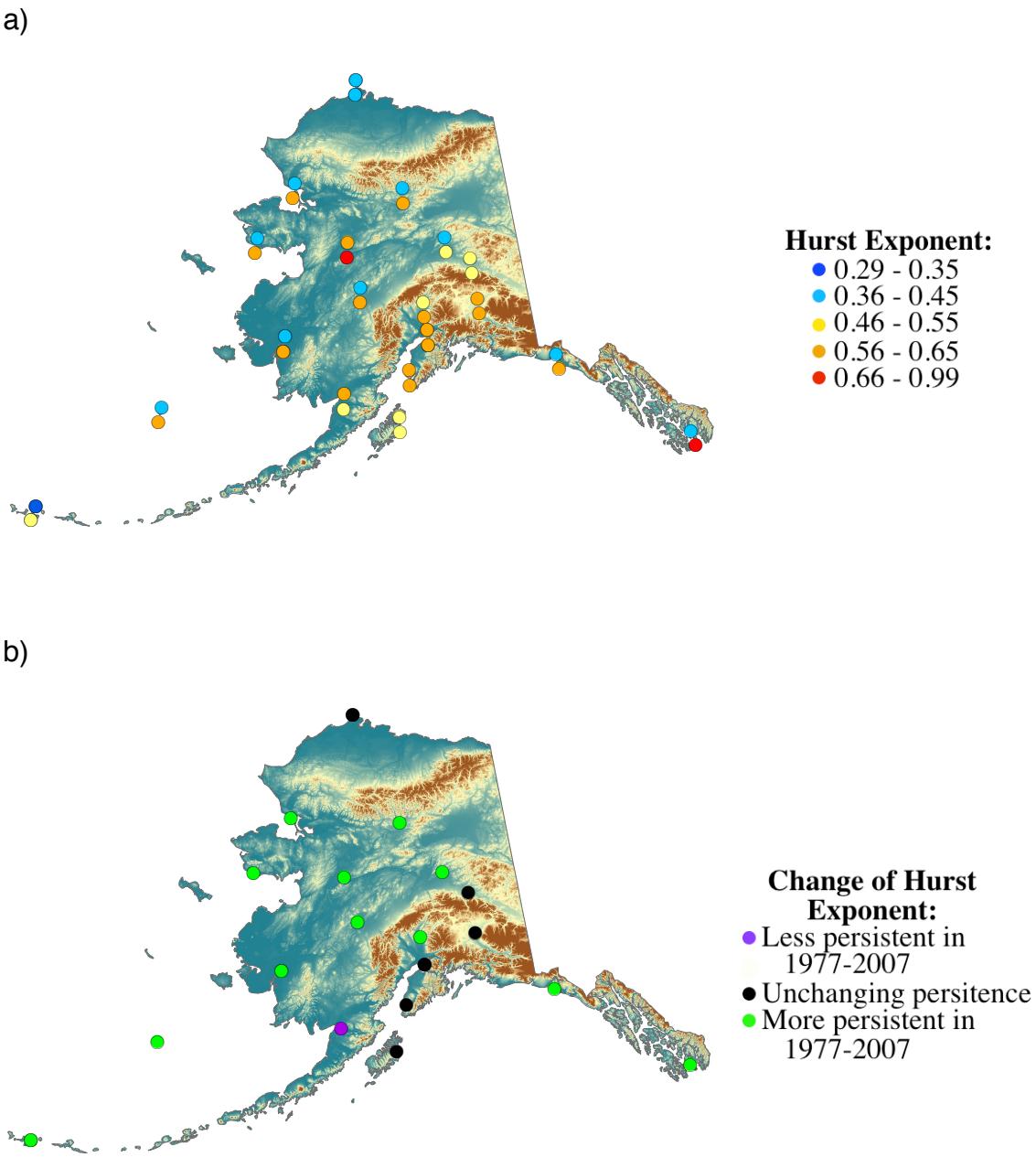


Figure 5: (a) Persistence during 1948-1975 (upper dot) and during 1977-2007 (lower dot) for station SLP. (b) Change of persistence during 1977-2007 compared to 1948-1975 for station SLP.

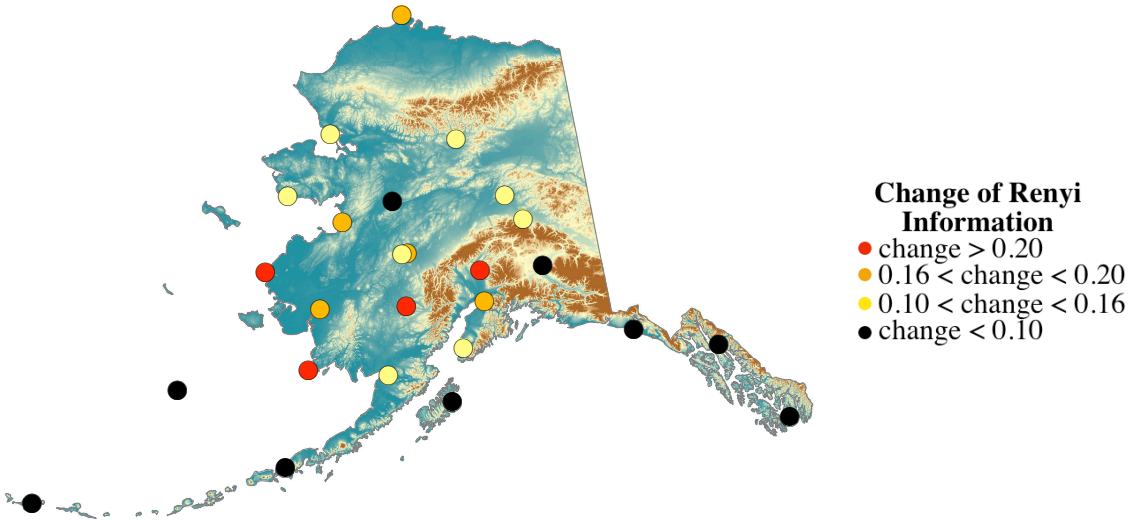


Figure 6: Change of autocorrelation lag 10 on temperature symbol set using above/ below average partition, from 1977-2007 compared to 1948-1975

predictable during either of the PDO phases, but that during the negative PDO phase the pressure is more likely to have alternating variability patterns, whereas during the positive PDO phase the pressure in the next five to 15 year time period is more likely to behave like the previous time period.

The changes found in the long term behavior of Alaska temperature and pressure leads to the question, is there any change in the short term behavior of these time series? We applied Renyi analysis to monthly and daily temperature and pressure series.

We found that daily temperatures at some stations become more ordered during the positive PDO. Daily pressure, on the other hand, exhibited no change with the PDO phase change. When Renyi analysis was applied to monthly averages, both the temperature andpressure time series were close to the random line for both phases of the PDO. The following is a discussion about the daily temperature result.

Because the Renyi analysis operates on time scales similar to the autocorrelation function, further comparison is necessary to determine whether the Renyi analysis is characterizing something more than the autocorrelation in the time series. Autocorrelation

is a measure of the linear short term memory in a data set, whereas Renyi information is a nonlinear measure. For data with a strong autocorrelation, care must be taken with the chosen partition so that the Renyi signal is not overwhelmed by the autocorrelation function. For the Alaskan stations, the binary alphabet based on a division between above and below median temperatures leads to Renyi information dominated by the autocorrelation signal. When the time series are replaced by their respective binary symbol sets, the autocorrelation at several stations at time lag 5 increases in 1977-2007 compared to 1948-1975, as shown in Figure 6. When the change of Renyi information is plotted, the corresponding map looks almost identical to Figure 6, with the exceptions of Ft Greely, Galena and Bettles, all of which showed slightly larger changes of Renyi information.

The reason for this similarity can be explained by an examination of the frequency of words; the frequency of each possible word of length 5 for the above/below mean partition is plotted for Fairbanks in Figure 7a. The most common words by far are the first and last, corresponding to temperatures that are all below the median or all above the median. Since the temperature anomaly time series mostly consist of long periods of above median temperatures followed by long periods of below median temperatures and vice versa, the autocorrelation on the symbol set is significant to a time lag of more than 10 days for most stations, and this strong autocorrelation drowns out other signals that might be seen from the Renyi information. If we want to see past the autocorrelation and find something new, we need to choose a different partition.

Because we wanted to maximize the possible word length of the information, the partition needed to be a binary alphabet. The chosen partition, which we call the quartile partition, was such that the 50% of the data closest to mean was symbolized as a 0 and the lowest 25% and highest 25% of the data was symbolized as a 1, as is shown in Figure 7d. This partition can be thought of as patterns of near mean versus extreme data.

The autocorrelation of the symbol sets for all stations falls off considerably faster using the quartile partition as compared to the above/below mean partition, dropping

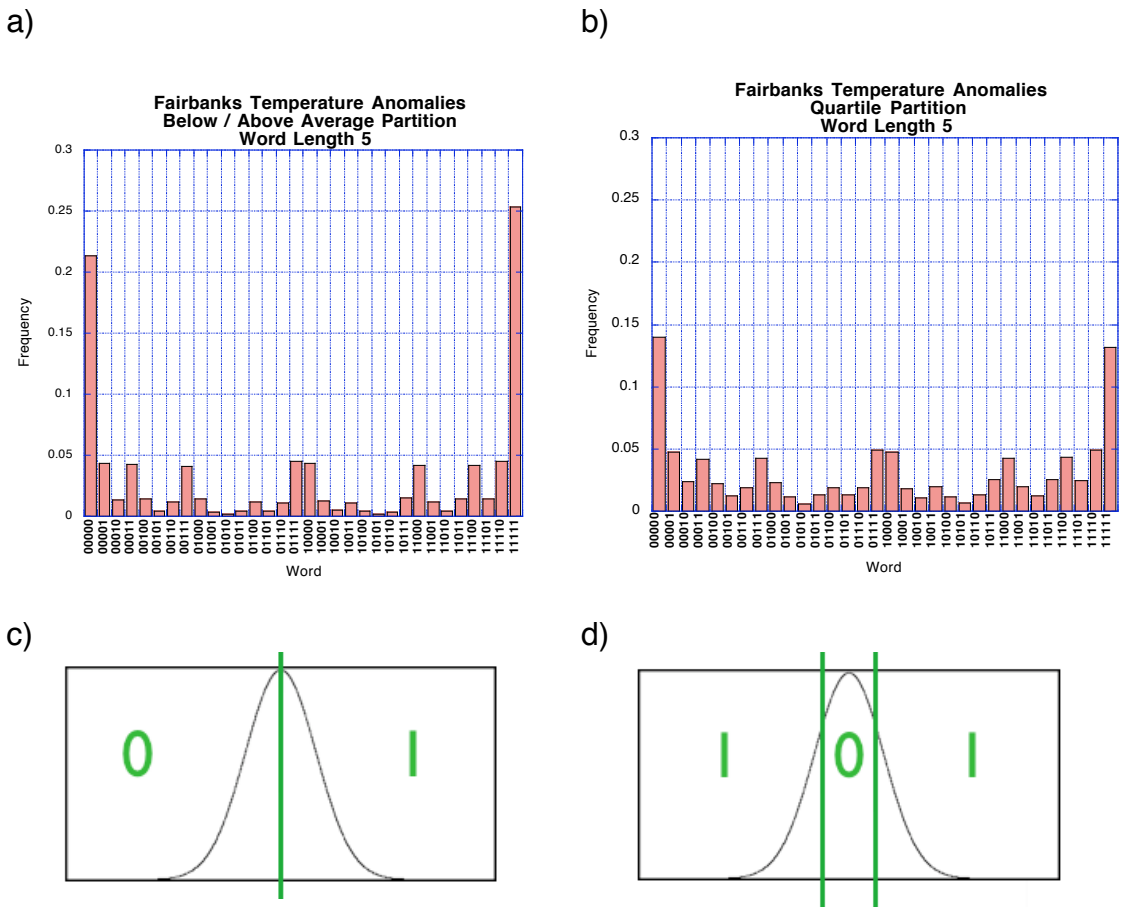


Figure 7: (a) Frequency of each possible word of length 5 in Fairbanks temperature symbol set based on the above/below average partition. (b) Frequency of each possible word of length 5 in Fairbanks temperature symbol set based on the quartile partition. (c) Definition of the above/below median partition based on a Gaussian data distribution. (d) Definition of the quartile partition based on a Gaussian data distribution.

below significance after a time lag of 4 days on average. The frequencies of the words of length 5 are shown in Figure 7b: the first and last words still dominate, but are only about half as frequent as with the median partition. At most stations, no change in autocorrelation at lag 5 was found after 1976, so we can be confident that the change seen in the Renyi information is a nonlinear signal that is different from the autocorrelation.

The Renyi information for Fairbanks SAT for the periods 1948-1975 and 1977-2007 is shown in Figure 8, with q=-1, q=1, and q=3 in order to examine the effect

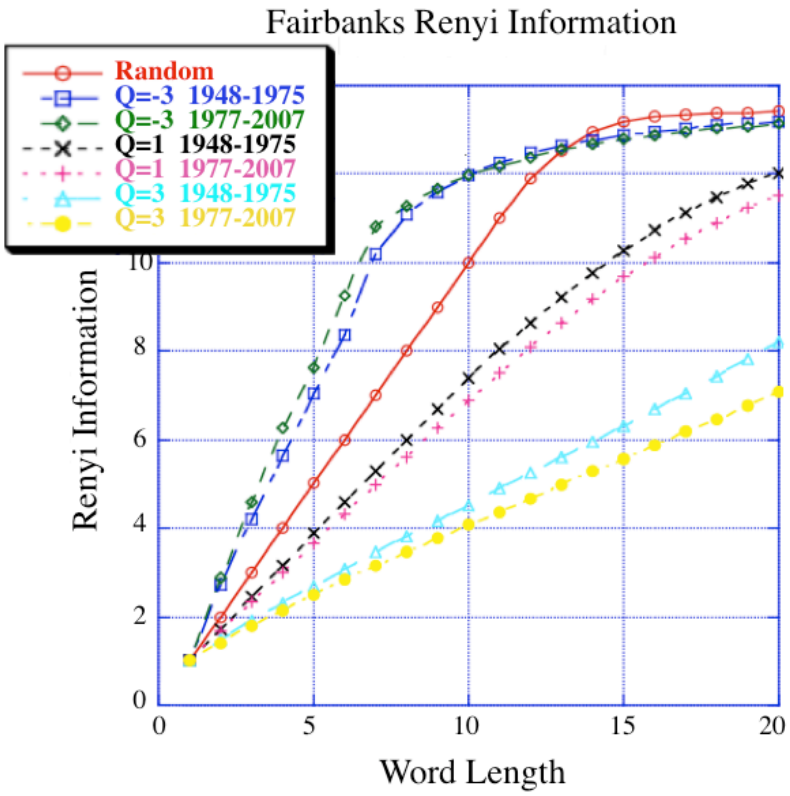


Figure 8: Renyi information on Fairbanks temperature data using the quartile partition for 1948-1975 and 1977-2007, q=-1, q=1, q=3. For each q, the information is less random from 1977-2007 than from 1948-1975.

of more frequent words (high q) versus less frequent words (negative q). For each of the q values, the Fairbanks Renyi information shows more order during 1977-2007 than during 1948-1975. For Fairbanks at least, this implies that the decrease in randomness was robust, and occurred for a high q value, which corresponds to the most common events, as well as for the negative q value, which corresponds to the less common events.

The change in Renyi information for SAT at all Alaskan stations is plotted in Figure 9 for q=1, q=3 and q=-1. The location of the large changes are the Northwest Coast of Alaska, as well as southwestern Alaska. It is interesting that, for many Alaskan locations, the stations that exhibit a change in Renyi information also exhibit a change in Hurst exponent; the observed increase of order or persistence after 1976 occurred over multiple time scales.

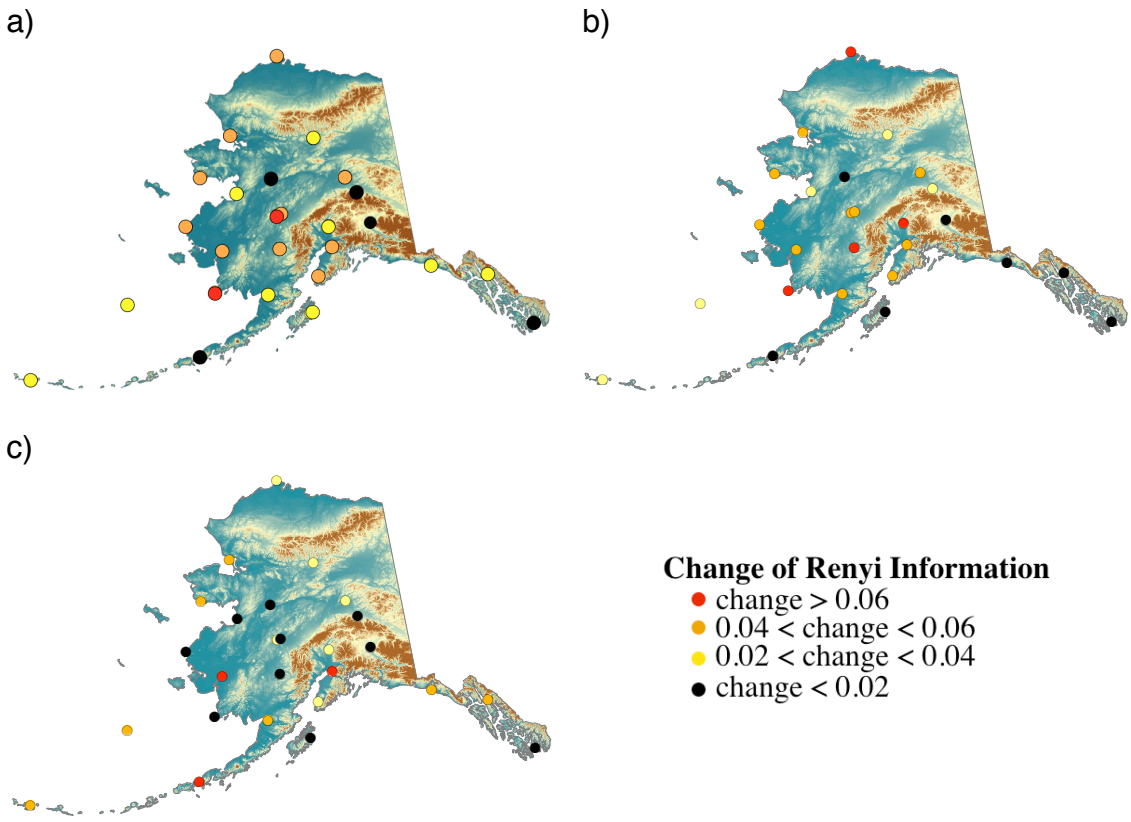


Figure 9: Renyi information change on Alaska stations temperature for 1977-2007 compared to 1948-1975, q=1 (a), q=3 (b), and q=-3 (c).

3.2 Synoptic Link Related to Persistence

The positive phase of the PDO creates less random or more persistent behavior of climate series in Alaska on both short and long time scales is most likely linked to the large-scale climate variability in the North Pacific.

The impact of the 1976 shift of the PDO on pressure is most pronounced in the winter, when after 1976 the Aleutian low became much stronger during winter months (Trenberth and Hurrell, 1994). In addition to the position and strength of the Aleutian low, Rodionov et al. (2005) showed that whether monthly mean SLP displays or a single or split Aleutian low is of primary importance in determining the impact of the Aleutian low on Alaska's climate. A strong, consolidated Aleutian low is associated with warmer

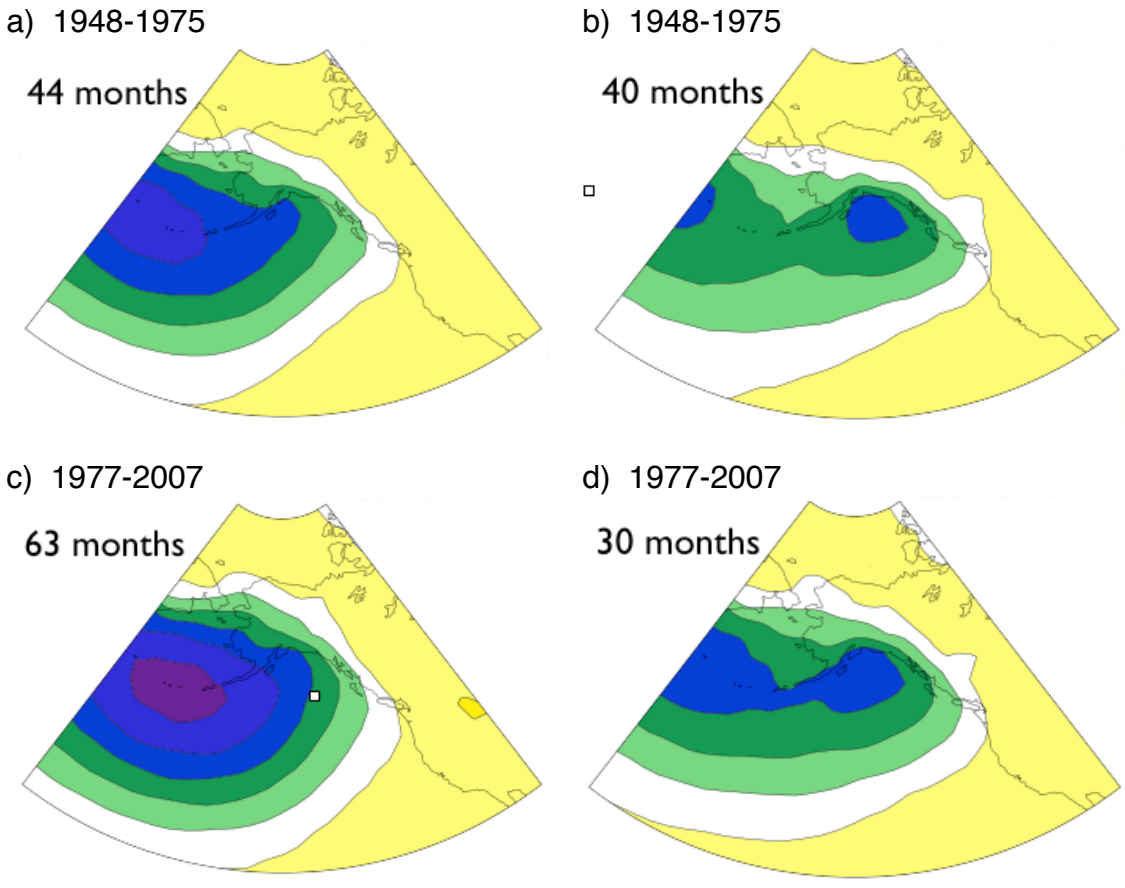


Figure 10: Average SLP during winter (DJF) months during the negative PDO phase with (a) a single Aleutian low and (b) more than one low pressure area. Average SLP during winter months during the positive PDO phase with (c) a single Aleutian low and (d) more than one low pressure area.

temperatures in Alaska and the Bering Sea (Niebauer, 1998, Stabeno et al. 2001), whereas a split Aleutian low has no discernible correlation with Alaskan temperatures (Rodionov et al. 2005).

The positive PDO phase is characterized by preferential behavior of both the El Niño southern oscillation (ENSO) and the Aleutian low, whereas the negative PDO phase had no preferential behavior of either phenomenon. Niebauer (1998) showed that, while during the negative PDO phase El Niño events occurred at the same frequency as La Niña events, during the positive PDO phase El Niño events occurred with 3 times the frequency of La Niña events.

While it El Niño/La Niña events are not correlated with split versus consolidated Aleutian lows, the occurrence of consolidated Aleutian lows increased significantly during the positive phase of the PDO. Figure 10 shows composite SLP patterns for a single versus split Aleutian low for winter months in both the negative and positive PDO phases. While during the negative PDO phase there are nearly an equal number of months with split and consolidated Aleutian lows, in the positive PDO phase the single Aleutian low occurs more than twice as often as the split Aleutian low. These synoptic preferences inherently make the atmosphere less random and more predictable on the short time scale of Renyi analysis.

For an explanation of the long term Hurst result, we turn to Fraedrich and Blender (2003), who proposed that persistent long term behavior of SAT's could be attributed to marine influence; if this is the case, then perhaps the affected areas of Alaska received more marine influence at the expense of continental climate patterns during the positive PDO.

In order to test the Fraedrich and Blender theory, Hurst Analysis was performed on monthly station SAT in Asia from 1948 to 2007. With the exception of the area around Novosibirsk, most stations located far from the coast have either random or weakly persistent behavior on a 3 to 15 year time scale, whereas stations located on the coast have persistent behavior. As Asia is the largest continent in the world, this served to confirm the Fraedrich and Blender theory for the purposes of this study, and began the search for ways in which interior and northwestern Alaska received more maritime influence during the positive PDO. Once again we turned to the effects of the stronger, more consolidated Aleutian low.

Stabeno et al. 2001 suggest that the anomalously high winter temperatures associated with the strong, consolidated Aleutian Low are due to an increase in storms passing into the Bering which originate far to the south of the Bering. During the winter, there is maximal temperature gradient northward, so when a southern storm is brought

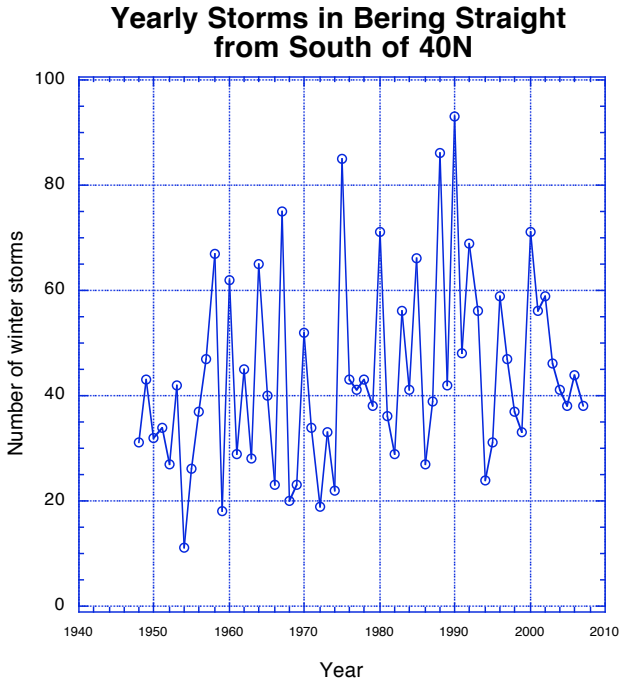


Figure 11: Number of Bering storms per year which began below 40 degrees North.

northward around the Aleutian Low and into the Bering, it can dramatically change the climate of Alaska stations which exhibit dependence on the Bering Sea.

The winter storms in the Bering Sea were examined for a possible connection with the observed change in long and short term randomness in Alaska stations. Figure 11 shows the yearly total of winter storms in the Bering Sea which originated south of 40^{o} north. From 1977-2007 there is an increase of, on average, 35 per cent of storms per year which meet this condition, compared to 1948-1975. Note that there are the same number of storms per year in the Bering Sea but more of the storms originate from south of 40^{o} north during the positive PDO phase than during the negative phase. This leads to enhanced advection of warm maritime air into the Bering Sea during the winter. These warm storms (i.e. pineapple express) occur more

often during the winter, potentially crossing a threshold to affect the multi-year long term memory of temperature in regions affected by the Bering Sea.

In order to test the impact of the Aleutian low on Alaskan stations, the correlation between the North Pacific (NP) index and the temperature data at each of the considered Alaskan stations was calculated (Table 1). The NP index is a measure of the strength of the Aleutian low: it is a area weighted average of the SLP from 30^{o}-60^{o}N, and from 160^{o}-140^{o}W (Deser et al. 2004). With the exception of Barrow, each of the stations which show an increase in long term persistence with the 1976 climate shift also show an increase in the negative correlation with the winter NP index. Furthermore, with the exceptions of Fairbanks and Cold Bay, each of the stations that do not show a change in persistence do not have a notable change in correlation with the winter NP index.

Most of the stations which exhibit an increase of long term temperature persistence also become more negatively correlated with the strength of the winter Aleutian low; the stronger winter Aleutian low forces more southern storms into the Bering Sea, warming the stations proximal to coasts to be impacted and enhancing the maritime influence on these dry winter climates. It is possible that this increase of marine influence originating from the strong southerly flow from the Pacific Ocean disrupts inner continental patterns, is responsible for creating more five to 15 year persistence for the stations in Alaska which undergo this change.

Table 1: Correlation coefficients for Alaska station SAT data with the NP index for winter months (December, January, February and March) during the negative PDO phase (1946-1975) and the positive PDO phase (1977-2007), and the difference between the two. Stations which show an increase in SAT Hurst exponent during the positive PDO phase are highlighted in green.

NP/SAT correlations DJFM	1946-1975	1977-2007	Difference
St Paul	0.1566	-0.1483	-0.3049
Galena	-0.2055	-0.4695	-0.2640
Tatalina	-0.3362	-0.5468	-0.2106
Nome	-0.2563	-0.4654	-0.2090
Unalakleet	-0.2433	-0.4422	-0.1989
Cape Romanzof	-0.1658	-0.3378	-0.1720
Sparrevohn	-0.2068	-0.3587	-0.1519
McGrath	-0.2368	-0.3883	-0.1514
Fairbanks	-0.3317	-0.4704	-0.1387
Kotzebue	-0.2797	-0.4041	-0.1244
Cape Newenham	-0.2444	-0.3622	-0.1178
Bethel	-0.2357	-0.3513	-0.1156
Ft Greely	-0.3188	-0.4282	-0.1094
Bettles	-0.3096	-0.4039	-0.0943
Talkeetna	-0.4280	-0.5036	-0.0756
Adak	0.2484	0.1830	-0.0654
Gulkana	-0.4343	-0.4735	-0.0392
Kodiak	-0.4807	-0.5173	-0.0366
Cold Bay	0.0618	0.0313	-0.0306
Anchorage	-0.5524	-0.5783	-0.0259
King Salmon	-0.4427	-0.4615	-0.0189
Yakutat	-0.0347	-0.0396	-0.0049
Homer	-0.5497	-0.5521	-0.0025
Barrow	-0.0940	-0.0915	0.0025
Juneau	-0.5230	-0.5156	0.0073
Metlakatla	-0.0642	0.2267	0.2909

3.3 Reanalysis & Model Comparison

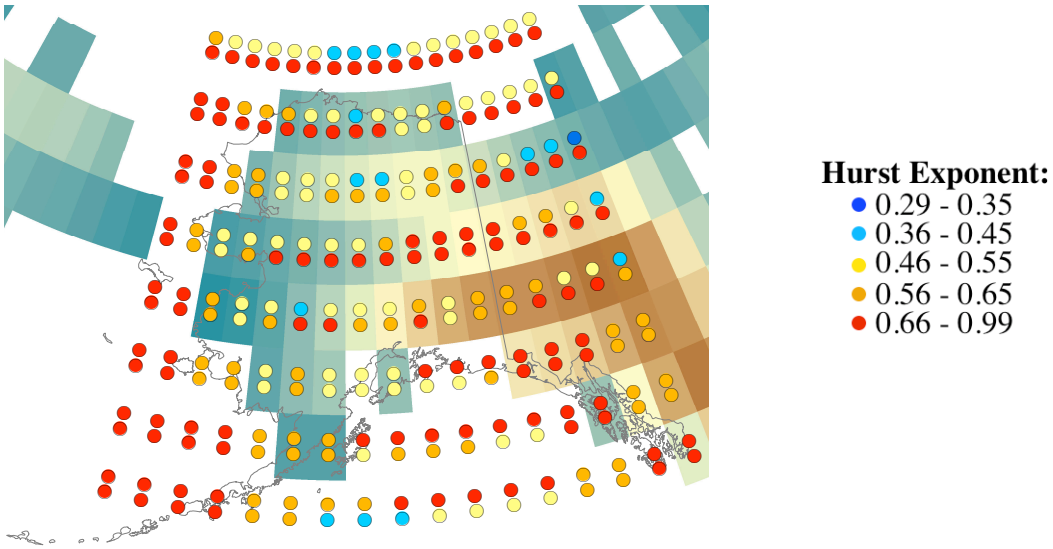
Since reanalysis data is the dynamically consistent coarse grained counterpart of station data, Hurst and Renyi analysis were applied to the NCEP/NCAR Reanalysis to compare the results with station data.

The Hurst exponents of the gridded reanalysis SAT data (Figure 12) tended to agree with those of the station data in most locations over Alaska with the exception of the area along the northwestern coast. The reanalysis showed more persistence during 1977-2007 than 1948-1975 for interior and northern Alaska. There is also an area in Southeast Alaska that became less persistent, which is near the two stations that showed a decrease in persistence. The major difference is that stations on Alaska's northwestern coast show increased persistence, whereas the reanalysis shows no change over that area. Despite this difference, the NCEP/NCAR Reanalysis compared favorably with the observations and displayed increased (decreased) persistence in the northern and interior (southern and eastern) Alaska of the Hurst exponent after 1976 for most locations in Alaska.

The NCEP/NCAR SLP Hurst exponents agree very well with the station data (Figure 13), which not unexpected since SLP is assimilated into the NCEP/NCAR Reanalysis.

The SLP data had similar Renyi information to that of the station SAT data, however no change was found between the two phases of the PDO, suggesting that the short term nonlinear randomness of the daily SLP in the reanalysis is not linked to the phase of the PDO.

a)



b)

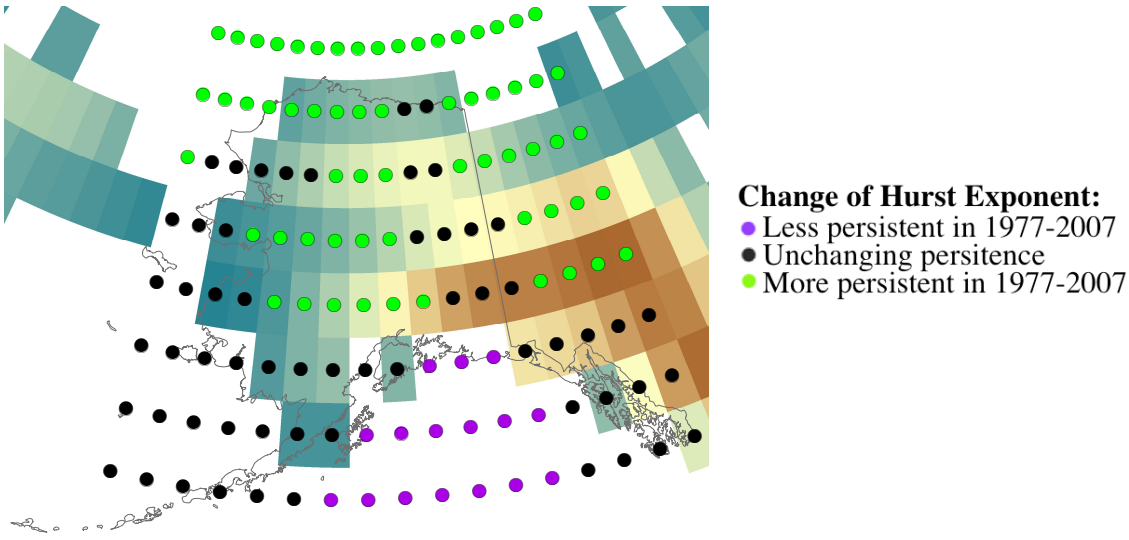


Figure 12: (a) Persistence during 1948-1975 (upper dot) and during 1977-2007 (lower dot) for NCEP/NCAR Reanalysis SAT. (b) Change of persistence during 1977-2007 compared to 1948-1975 for NCEP/NCAR Reanalysis SAT.

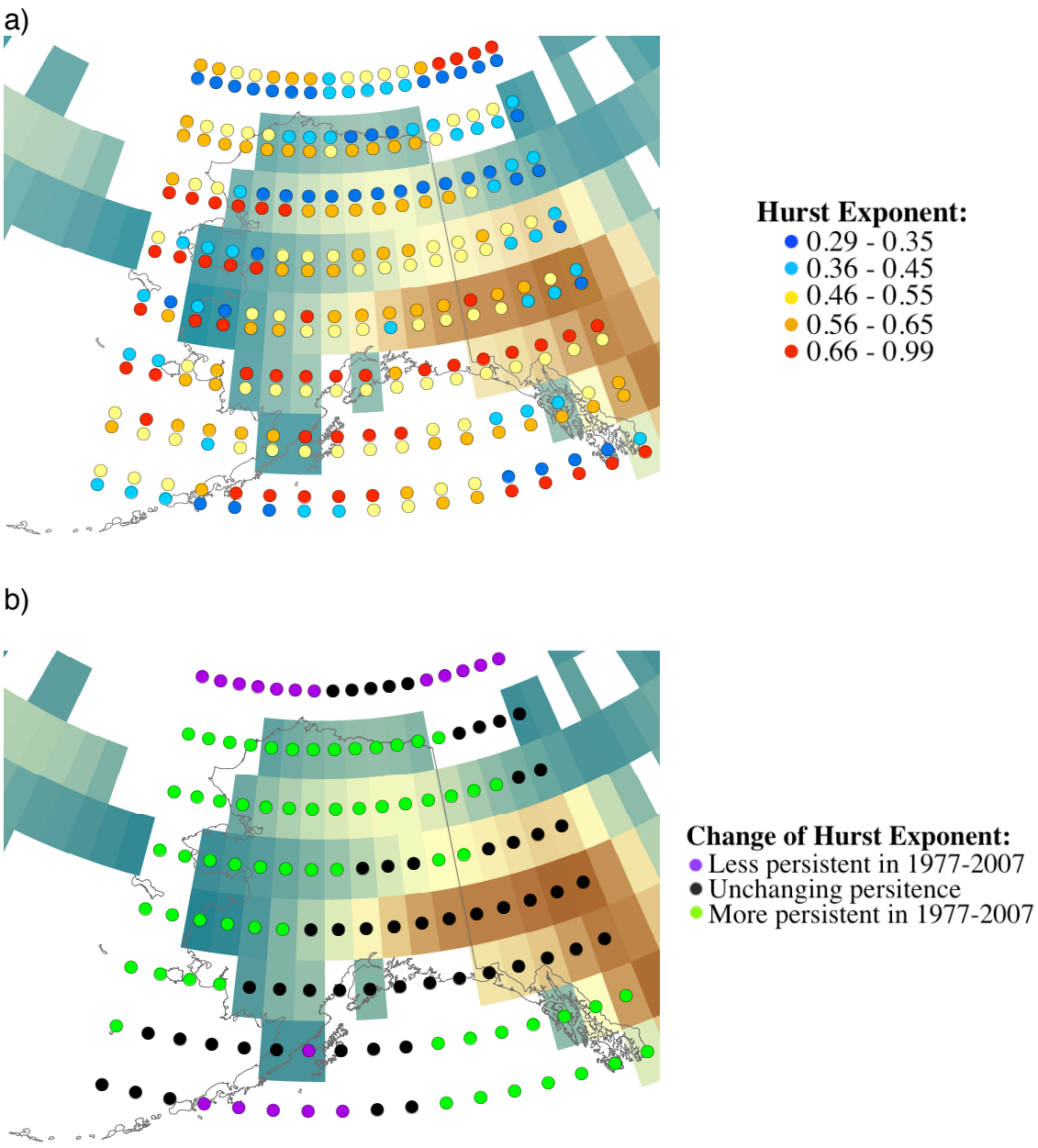


Figure 13: (a) Persistence during 1948-1975 (upper dot) and during 1977-2007 (lower dot) for NCEP/NCAR Reanalysis SLP. (b) Change of persistence during 1977-2007 compared to 1948-1975 for NCEP/NCAR Reanalysis SLP.

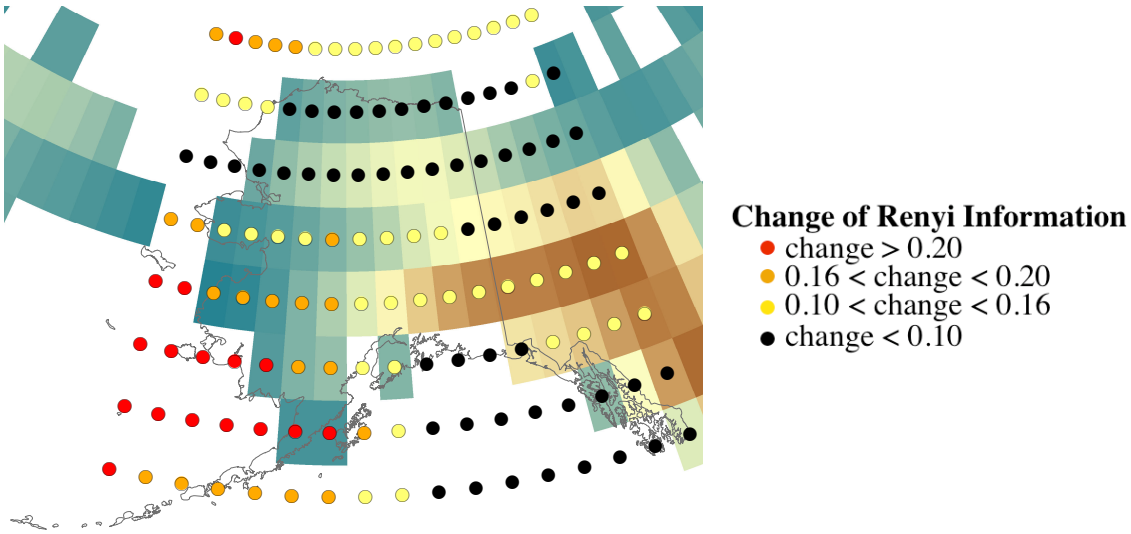


Figure 14: Shannon information (q=1) change on Alaska NCEP SATs for 1977-2007 compared to 1948-1975.

The reanalysis SAT, however, showed a much larger change in the Renyi information than the station SAT between the positive and negative PDO phases. Every grid point tested had a change of slope greater than 0.08. In order to find the spatial distribution of these changes, they are plotted in Figure 14, with a larger scale than that used in Figure 9. While every reanalysis grid point shows a much larger change compared to the station data, the largest change in the reanalysis is found in southwestern Alaska, which is consistent with the station result. The larger change over the northwestern coast found in the station data is not evident in the reanalysis data.

The increase of order in the short term and persistence in the long term Alaskan temperature data during the recent positive PDO has interesting implications for statistical predictability during positive PDO's. It would have stronger implications if we could show that this effect was common to all positive PDO phases; however, the station data does not extend back far enough to analyze the 1925-1945 positive PDO. Thus we turn to the 20th Century Reanalysis, a gridded data set that has the advantage of extending

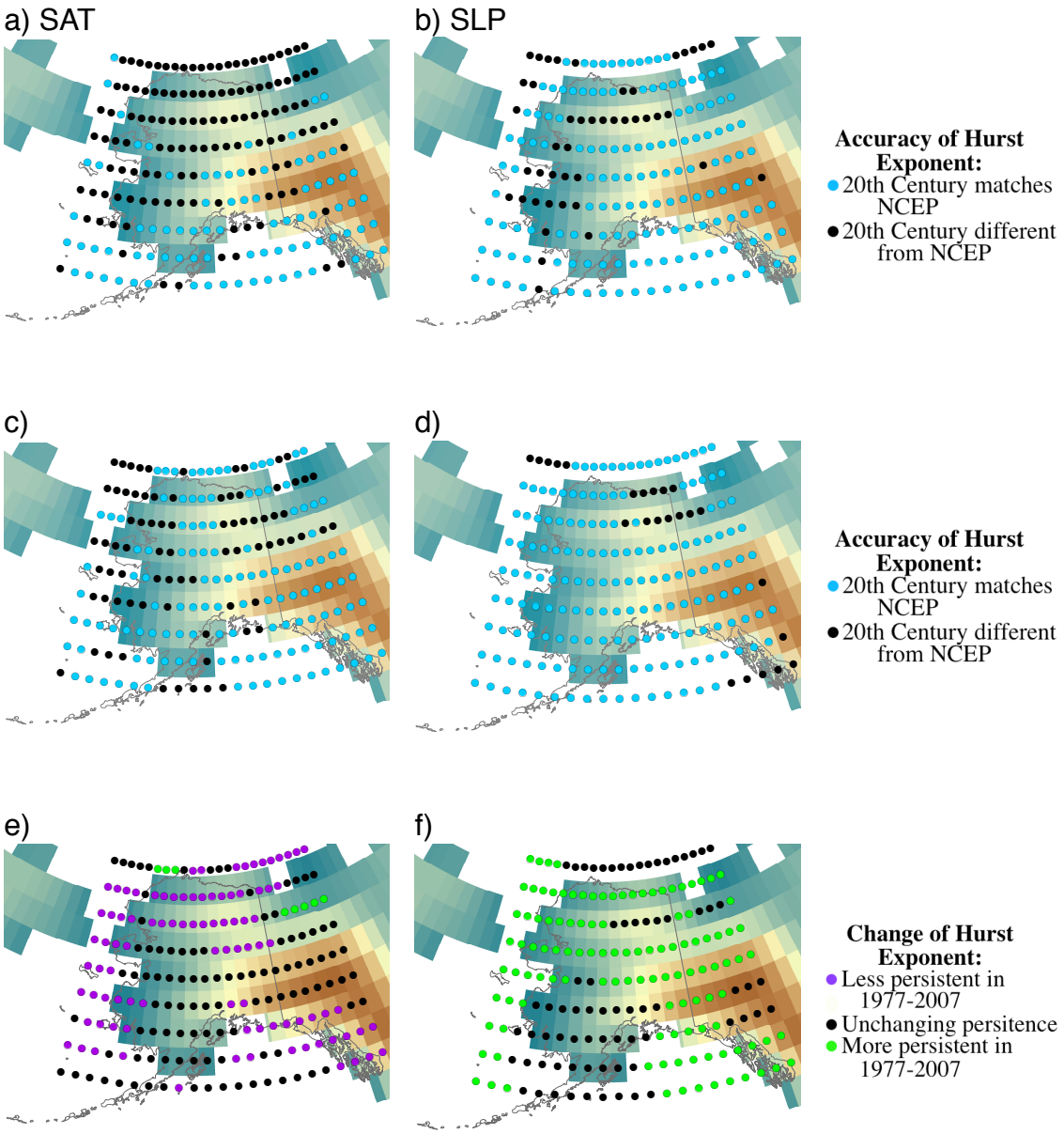


Figure 15: (a) Accuracy of 20th Century SAT data from 1948-1975 compared to NCEP/NCAR Reanalysis. (b) Accuracy of 20th Century SLP data from 1948-1975 compared to NCEP/NCAR Reanalysis. (c) Accuracy of 20th Century SAT data from 1977-2007 compared to NCEP/NCAR Reanalysis. (d) Accuracy of 20th Century SLP data from 1977-2007 compared to NCEP/NCAR Reanalysis. (e) Comparison of Hurst exponent of 20th Century SAT data from 1925-1945 compared to 1948-1975. (f) Comparison of Hurst exponent of 20th Century SLP data from 1925-1945 compared to 1948-1975.

back to 1871, though it has the disadvantage of only assimilating pressure and ice observations.

The 20th Century Reanalysis was analyzed during three PDO phases: 1924-1945, 1948-1975, and 1977-2007. First, the Hurst exponents from 1948-1975 and 1977-2007 were compared to those of the NCEP/NCAR Reanalysis data to ascertain the accuracy of the 20th Century Reanalysis (Figure 15). The SAT result was not promising; over the two periods where a comparison could be made, the Hurst exponents matched in less than half the grid points, and almost none of these matching locations were in Interior or Northwestern Alaska. In particular, the temperature persistence during the negative PDO phase was much higher than that obtained by station and reanalysis data. The SLP result was more successful, matching the NCEP/NCAR grid points over most of the land in the state; again this is likely because SLP is assimilated data in the 20th Century Reanalysis. The change from weakly anti-persistent to persistent behavior is a characteristic of the two most recent positive PDO's and may be common to all positive PDO's.

The Renyi analysis of the 20th Century SLP, like the Renyi analysis of station and NCEP/NCAR Reanalysis, showed no change with the PDO phases. The temperature did show an increase in order, although it was weaker than the effect found with the station data. The Renyi SAT result is shown in Figure 16. While Renyi analysis of the 20th Century Reanalysis SAT's is more accurate than the Hurst analysis, the result is still not similar enough to the station result to say whether less order is a characteristic of all positive PDO's.

In hopes of finding a general pattern of SAT behavior, we turn to a global climate model, the CCSM4. We applied the Hurst analysis as a diagnostic tool to analyze a twentieth century simulation of the Community Coupled Systems Model (CCSM4). Because the PDO in the model may not be synchronized with the observed PDO, we calculated the model NP index, which is highly anti-correlated with the PDO, using the method described by Deser et al. (2004) (Figure 17). The period 1979-2005 was chosen as the

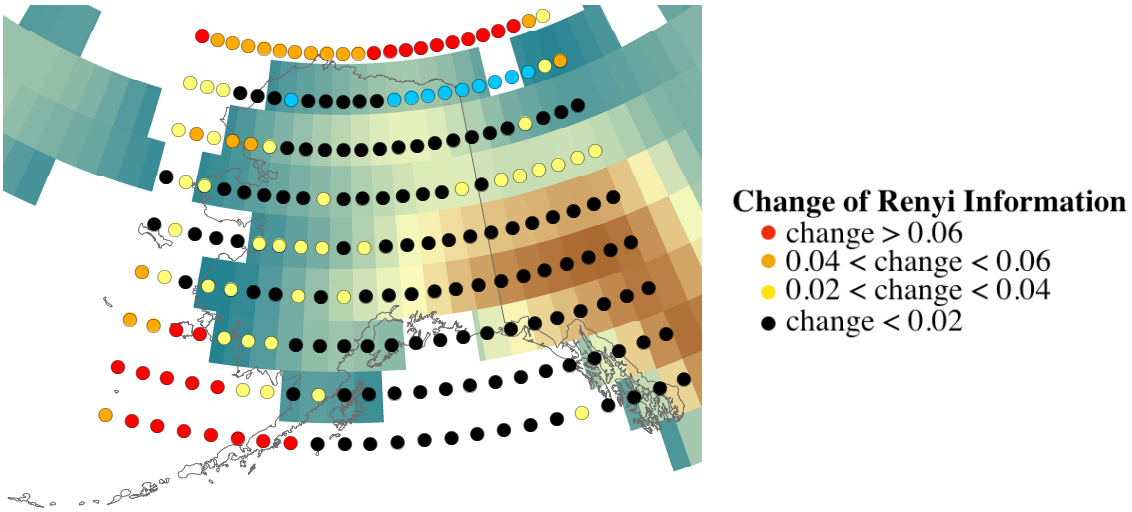


Figure 16: Change of Renyi information (q=1) of 20th Century Reanalysis SAT data from negative PDO phase to positive PDO phase.

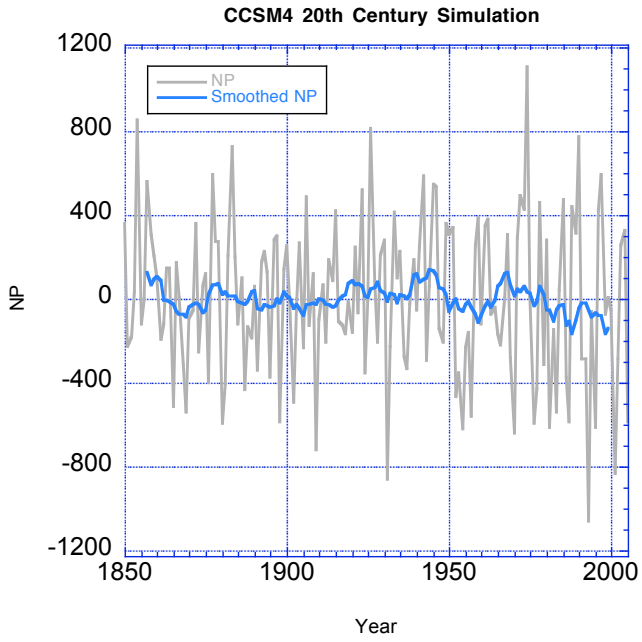


Figure 17: NP index derrived from CCSM 20th Century SLP data.

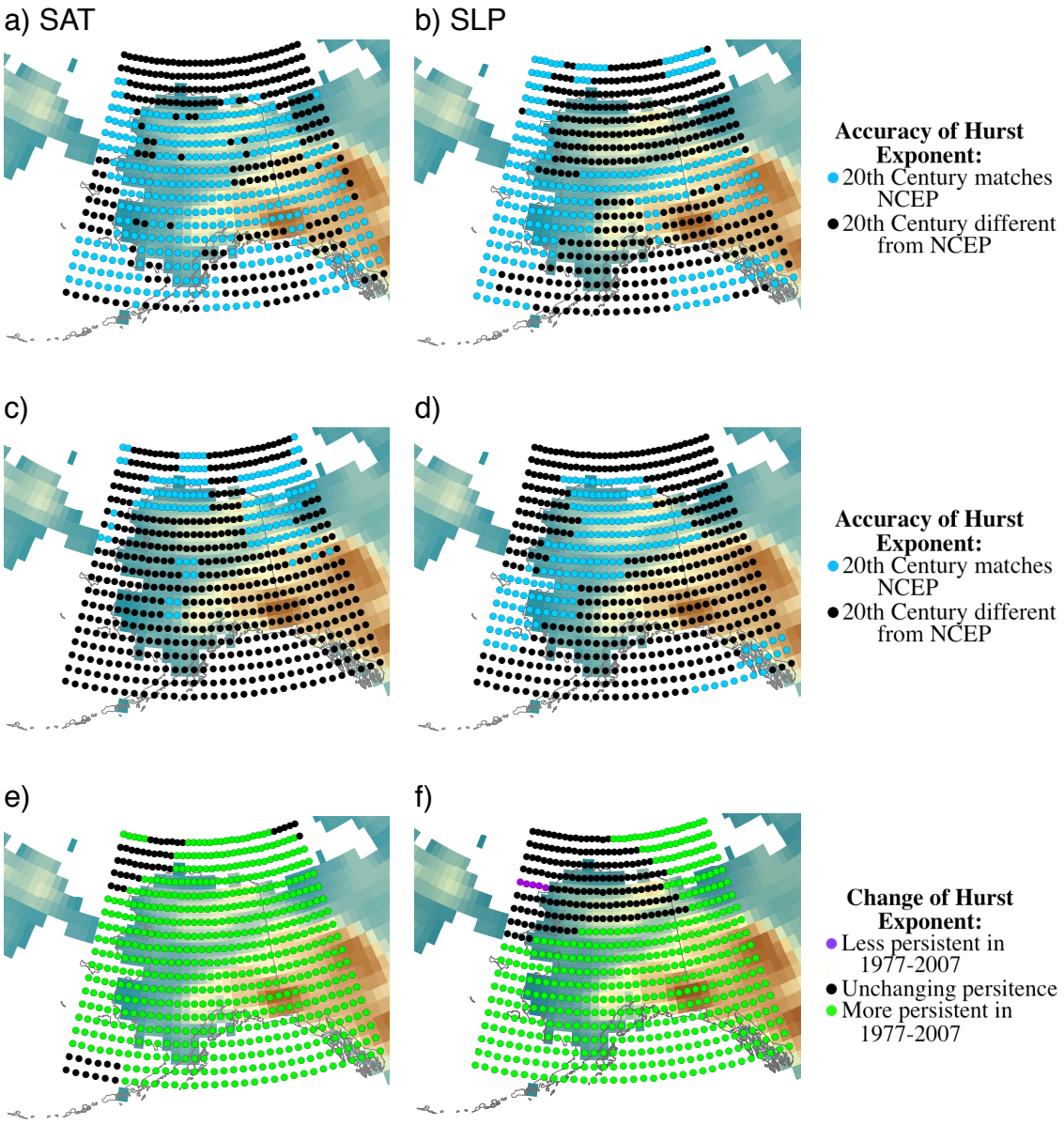


Figure 18: (a) Accuracy of CCSM 20th Century SAT data from 1948-1975 compared to NCEP/NCAR Reanalysis. (b) Accuracy of CCSM 20th Century SLP data from 1948-1975 compared to NCEP/NCAR Reanalysis. (c) Accuracy of CCSM 20th Century SAT data from 1977-2007 compared to NCEP/NCAR Reanalysis. (d) Accuracy of CCSM 20th Century SLP data from 1977-2007 compared to NCEP/NCAR Reanalysis. (e) Comparison of Hurst exponent of CCSM 20th Century SAT data from 1925-1945 compared to 1948-1975. (f) Comparison of Hurst exponent of CCSM 20th Century SLP data from 1925-1945 compared to 1948-1975.

positive PDO, and 1917-1949 as the negative PDO. Figures 18a and c show the accuracy of the CCSM SAT data as compared to NCEP/NCAR Reanalysis data during the negative and positive phases of the PDO respectively.

The Hurst exponents for the CCSM SAT data compare favorably during the negative PDO phase, but the increase of persistence is exaggerated in most of the state during the positive PDO phase. The exaggerated are of increased persistence is evident in Figure 8e, where all of Alaska except the Northwest corner is shown to have higher SAT persistence during the positive PDO phase.

The Hurst exponents for the CCSM SLP data are quite different from station SLP results (Figures 18b, d); no anti-persistence is found in the negative PDO phase, and both phases showed higher persistence than was calculated for the NCEP/NCAR Reanalysis data. Though Figure 18f shows an increase in SLP persistence over the entire state during the positive PDO phase, the CCSM Hurst exponents changed from persistent to extremely persistent.

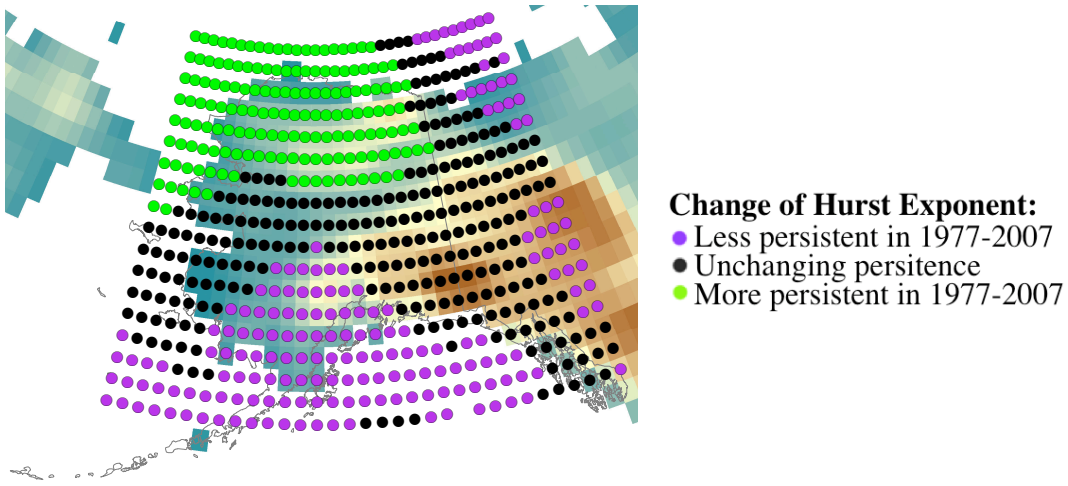


Figure 19: Comparison of Hurst exponent during the positive PDO compared to the negative PDO during the CCSM4 preindustrial control run.

The monthly CCSM data from the preindustrial control run from was then analyzed; There were no daily data available for the control run, and because there is so

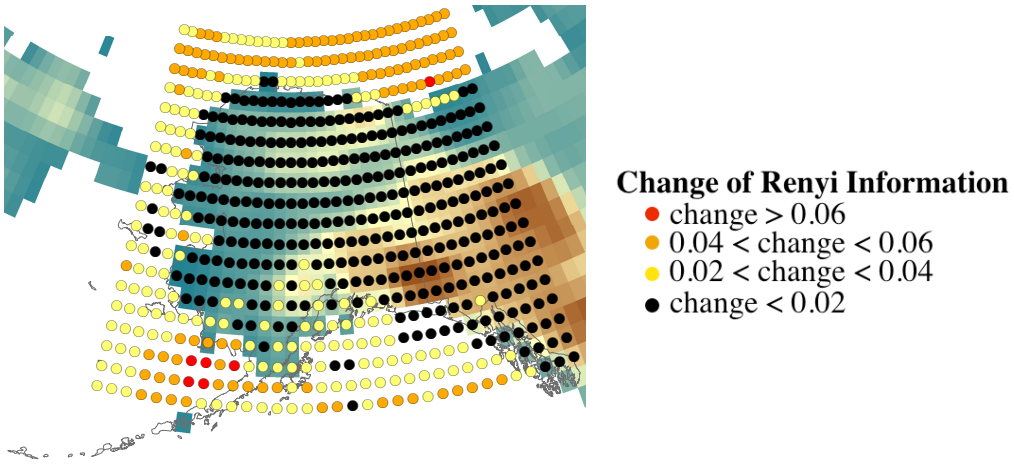


Figure 20: Change of Renyi information (q=1) of CCSM twentieth century SAT data from negative PDO phase to positive PDO phase.

little difference between the Hurst exponents from daily and monthly data, the results should be comparable. The predominant accuracy of the control run SAT data compared to NCEP/NCAR Reanalysis in both phases of the PDO leads to a very accurate calculation of the change of SAT persistence (Figure 19). The CCSM control simulation SLP data, on the other hand, did not compare well to the NCEP/NCAR Reanalysis, with exaggerated persistence at nearly all of the data points in both phases of the PDO, especially the positive phase. In both the 20th Century and control runs, CCSM data seems have more accurate 5 to 15 year persistence in its SAT data than its SLP data.

The Renyi analysis of the CCSM twentieth century SLP data, like station and NCEP/NCAR Reanalysis SLP, showed no change in conjunction with the PDO. The CCSM SAT data, however, showed more order at some points during the positive PDO, but only over the oceans (Figure 19). This pattern does not resemble the station or reanalysis results.

4. Summary

Differences of predictability of temperature data during opposite phases of the PDO were found in many Alaska stations, both on long (five to 15 years) and short (two

to 15 days) time scales. Hurst analysis was used to find differences in persistence on long time scales of five to 15 years, and Renyi analysis was used to find changes in order on short time scales of two to 15 days. These two time scales are unconnected, and represent different processes in the climate.

It was found that on long time scales, surface air temperature in interior and northwestern Alaska is random during the negative PDO, and persistent during the positive PDO. This implies that the long term variability of temperature for those regions of Alaska are statistically predictable during the positive PDO. Sea level pressure in interior, western and southeastern Alaska are for the most part weakly anti-persistent during the negative PDO, and weakly persistent during the positive PDO.

On short time scales it was found that while Renyi analysis of SLP did not change with the PDO, SAT in southwestern Alaska and along the northwestern coast became more ordered during the positive PDO compared to the negative PDO. The increase of order implies that the most frequent events happen even more frequently during the positive PDO, and statistical weather forecasts in those areas could be made more accurate during the positive PDO.

The more predictable behavior of the Alaskan temperature data on long and short time scales is possibly related to the behavior of the Aleutian low and storms. During the negative PDO, the Aleutian low was split just as often as it was consolidated, whereas during the positive PDO, it was twice as likely to be consolidated as split, and in addition the low pressure was stronger. This preferential synoptic pattern inherently creates a more ordered atmosphere, and explains the Renyi result.

The stronger, more consolidated Aleutian low associated with the positive PDO also drew more warm southern storms from south of 40˚N into the Bering Sea, though the total number of storms in the Bering did not significantly increase. This increase of warm, moist air brings increases the maritime nature of interior and northwestern Alaska. This maritime influence is consistent with increased long term persistence, and consistent with the results of Fraedrich and Blender (2003). Maritime influence can also be

measured linearly: during the positive PDO, most station temperature in interior and northwestern Alaska showed stronger correlation with the strength of the Aleutian low.

Overall the application of Hurst and Renyi analysis to the NCEP/NCAR Reanalysis compared well with station data. The Hurst analysis of the SLP was very similar to the station data results, and the SAT result was very similar in all locations except the northwestern coast of Alaska. The Renyi analysis of the SAT data showed much more of an increase in order than the station data all over Alaska. However, the largest change was seen in southwestern Alaska, which is one of the two places that showed the largest change in the station data. The spacial agreement of these analysis results between NCEP/NCAR Reanalysis and observations implies that, in terms of the quantities measured by Hurst and Renyi analysis, the short and long term dynamics (temporal evolution) of grid point time series of the NCEP/NCAR Reanalysis are similar to that of station data expect perhaps in northwestern Alaska.

Because of our analysis of the 20th Century Reanalysis data set, we are able to conclude that the weakly persistent Hurst exponent of SLP data is common to at least the last two positive PDO phases in interior, western and southeast Alaska. However, due to the poor match of Hurst and Renyi analysis of the SAT data, we were unable to conclude anything useful about SAT data in previous PDO phases from the 20th Century Reanalysis.

The time series dynamics of CCSM4 control and 20th century simulations did not compare favorably with the station or reanalysis results. The Hurst analysis showed much higher persistence in both temperature and pressure data in most Alaskan locations, particularly during the positive PDO. The widespread area where increased persistence is calculated suggests the PDO has an exaggerated area of influence on Alaska climate in CCSM4. The Renyi analysis showed an increase in order, but only over the oceans with no change over land.

This study has shown that the results from Renyi and Hurst analyses can be a useful part of climate studies in three distinct ways. First, these methods can show where

and under what circumstances the statistical predictability of weather and climate can be improved. Improved predictability was found in the station results of temperature time series: interior and northwestern Alaska showed persistence during the positive PDO, whereas in southwestern Alaska and along the northwest coast SATs showed more order during the positive PDO from two to 15 days. These results imply that during the positive PDO, the five to 15 year variability in interior and northwestern Alaska are statistically likely to behave similarly to the previous five to 15 years, and that in southwestern Alaska and on the northwest coast the most frequent patterns on the two to 15 day timescale are even more likely to occur and persist. Both on the short time scales and long time scales, the predictability of the temperature series is improved during the positive PDO.

These methods have also proven useful at finding relationships between climate time series and synoptic mechanisms. The improved predictability of the short term temperature variability in southwestern Alaska and along the northwest coast was plausibly linked to the more preferential behavior of the Aleutian low during the positive PDO. Also the improved predictability of the long term temperature variability in interior and northwestern Alaska was plausibly linked to the increase of warm storms from south of 40˚N in the Bering Sea during the positive PDO. Both the Renyi and Hurst analysis main results could be linked to circulation changes in the synoptic system.

A third use of these methods is the comparison of the behavior of different time series. When the results of the NCEP/NCAR Reanalysis data were compared to the station data results, it was found that while SLP data compared similarly in all areas of Alaska, and SAT data compared similarly in most areas areas of Alaska, both the Renyi and Hurst analyses found discrepancies in the behavior of NCEP/NCAR Reanalysis SAT data on the northwest coast of Alaska. Knowledge of these locational discrepancies are useful to help make informed decisions when considering the use of reanalysis data in place of observations, as well as when trying to find areas to improve reanalysis data.

5. Statement of Work:

The body of work above is being submitted to Journal of Climate by the following co-authors: Jean K. Talbot, Uma S. Bhatt, David Newman, Renate Wackerbauer, Igor V. Polyakov, Raul Sanchez, Heather Angeloff, Richard Thoman and Peter A. Bieniek. While I did all the calculations for this study and wrote the paper, this work is really the brain child of Drs. Bhatt, Newman, and Wackerbauer. Dr. Newman provided expertise with the rescaled range method of finding the Hurst exponent, Dr. Wackerbauer provided expertise with Renyi information. Drs. Bhatt, Polyakov, and Thoman helped with the large-scale climate and physical synoptic interpretation of the results. Dr. Sanchez provided insights in the interpretation of the Hurst exponent. Heather Angeloff acquired most of the data, and Peter Bieniek provided computing and editing help.

5.1 References:

Alekseev, V. M. and M. V. Yakobson, 1981: Symbolic dynamics and hyperbolic dynamic systems. *Physics Reports*, **75**, 290-325.

Bekryaev, R. V., I. V. Polyakov and V. A. Alexeev, 2010: Role of Polar Amplification in Long-Term Surface Air Temperature Variations and Modern Arctic Warming. *Journal of Climate*, **23**, 3888–3906.

Blender, R. and K. Fraedrich, 2003: Long time memory in global warming simulations. *Geophysical Research Letters*, **30**, 1769, doi:10.1029/2003GL017666.

Carvalho, L. M. V., A. A. Tsonis, C. Jones, H. R. Rocha, and P. S. Polito, 2007: Anti-persistence in the global temperature anomaly field. *Nonlinear Processes in Geophysics*, **14**, 723-733.

Cassano, E. N., J. J. Cassano and M. Nolan, 2011: Synoptic weather pattern controls on temperature in Alaska. Journal of Geophysical Reseach, **116**, D11108, doi: 10.1029/2010JD015341.

Deser, C., A. S. Phillips, and J. W. Hurrell, 2004: Pacific interdecadal climate variability: linkages between the tropics and the North Pacific during boreal winter since 1900. *Journal of Climate*, **17**, 3109-3124.

Fraedrich, K. and R. Blender, 2003: Scaling of atmosphere and ocean temperature correlations in observations and climate models. Physical Review Letters, **90**, 18501, doi: 10.1103/ PhysRevLet t.90.108501

Govindan, R. B., D. Vyushin, A. Bunde, S. Brenner, S. Havlin, H. J. Schellnhuber, 2002: Global climate models violate scaling of the observed atmospheric variability. *Physical Review Letters*, **89**, DOI: 10.1103 / PhysRevLett.89.028501.

Hartmann, B. and G. Wendler, 2005: Significance of the 1976 Pacific climate shift in the climatology of Alaska. *Journal of Climate*, **18**, 4824-4839.

Hurst, H. E., 1951: Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **112**, 780-808.

Király, A. and I. M. Jánosi, 2005: Detrended fluctuation analysis of daily temperature records: Geographic dependence over Australia. Meteorology and Atmospheric Physics, **88**, 119-128.

Király, A., I. Bartos, and I. Jánosi, 2006: Correlation properties of daily temperature anomalies over land. *Tellus*, **58A**, 593-600.

Krutzmann, N. C. and A. J. McDonald, S. E. George, 2008: Identification of mixing barriers in chemistry-climate model simulations using Renyi entropy. *Geophysical Research Letters*, **35**, L06806.

Lee, S., T. Gong, N. Johnson, S. B. Feldstein and D. Pollard, 2011: On the possible link between tropical convection and the northern hemisphere arctic surface air temperature change between 1958 and 2001. Journal of Climate, **24**, 4350-4367.

López-de-Lacalle, J., 2011: Trends in Alaska temperature data. Towards a more realistic approach. *Climate Dynamics*, online first, DOI 10.1007/s00382-011-1198-7

MacDonald, G. M. and R. A. Case, 2005: Variations in the Pacific Decadal Oscillation over the past millennium. *Geophysical Research Letters*, **32**, L08703

Mandelbrot, B. and J. R. Wallis, 1968: Noah, Joseph and operational hydrology. *Water Resources Research*, **4**, 909-918.

Mandelbrot, B. and J. R. Wallis, 1969: Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, **5**, 967-988.

Mantua, N. J. and S. R. Hare, 2002: The Pacific Decadal Oscillation. *Journal of Oceanography*, **58**, 38-44.

Neal, E. G., M. T. Walter, and C. Coffeen, 2002: Linking the pacific decadal oscillation to seasonal stream discharge patterns in Southeast Alaska. Journal of Hydrology, **263**, 188-197.

Niebauer, H. J., 1998: Variability in Bering Sea ice cover as affected by a regime shift in the North Pacific in the period 1946-1996. *Journal of Geophysical Research*, **103**, 27,717-27,737.

Papineau, J. M., 2001: Wintertime temperature anomalies in Alaska correlated with ENSO and PDO. *International Journal of Climate*, **21**, 1577-1592.

Rodionov, S. N., J. E. Overland, and N. A. Bond, 2005: The Aleutian low and winter climatic conditions in the Bering Sea. Part 1: Classification. *Journal of Climate*, **18**, 160-177.

Stabeno, P. J., N. A. Bond, N. B. Kachel, S. A. Salo and J. D. Schumacher, 2001: On the temporal variability of the physical environment over the south-eastern Bering Sea. *Fisheries Oceanography*, **10**, 81-98.

Talkner, P. and R. Weber, 2000: Power spectrum and detrended fluctuation analysis: Application to daily temperatures. *Physical Review E*, **62**, 150-160.

Trenberth, K. E. and J. W. Hurrell, 1994: Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics*, **9**, 303-319.

Tsonis, A. A. and P. J. Roebber, 1999: Long-range correlations in the extratropical atmospheric circulation: origins and implications. Journal of Climate, **12**, 1534-1541.

Voss, A., J. Kurths, H. J. Kleiner, A. Witt, N. Wessel, P. Saparin, K. J. Osterziel, R. Schurath, R. Dietz, 1996: The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death. *Cardiovascular Research*, **31**, 419-433.

Wackerbauer, R., A. Witt, H. Atmanspacher, J. Kurths, H. Scheingraber, 1994: A comparative classification of complexity measures. *Chaos, Solitons & Fractals*, **4**, 133-173.

Woodard, R., 2004: Building blocks of self-organized criticality. ProQuest Dissertations and Thesis, AAT 3150133.

Yeh, S. W., Y. J. Kang, Y. Noh, A. J. Miller, 2011: The North Pacific climate transitions of the winters of 1976/77 and 1988/89. J*ournal of Climate*, **24**, 1170-1183.

Zhu, X., K. Fraedrich, Z. Liu and R. Blender, 2010: A demonstration of long-term memory and climate predictability. Journal of Climate, **23**, 5021-5029.

Zhang, X., J. E. Walsh, J. Zhang, U. S. Bhatt and M. Ikeda, 2004: Climatology and interannual variability of arctic cyclone activity: 1948-2002. Journal of Climate, **17**, 2300-2317.

Appendix A: Detailed Renyi Analysis Method

Renyi analysis is a method which counts patterns in a time series by compressing the data series onto a symbol series based on the quantity of each data point. The set of symbols is called an 'alphabet', and the patterns of consecutive symbols are called 'words'. Using the methods outlined in Wackerbauer et al. (1994), first we will show how the symbol sequences are determined, and then we will define the quantities Renyi information and Renyi entropy.

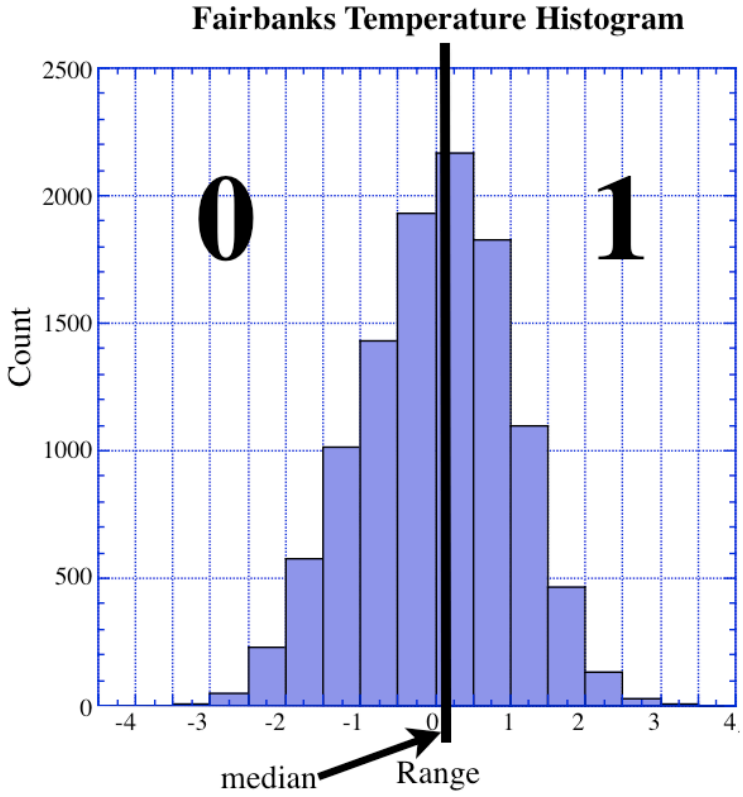


Figure A1: Histogram of Fairbanks temperature anomalies, with median partition.

To determine a symbol sequence, the time series must be divided into segments; this division is called a partition. While some theoretical data series have an inherent partition which captures all the dynamics of the data which is called a generating partition, data series from measurements have no such generating partition. This means

the chosen partition, or division of the data, has an important impact on the outcome of the analysis.

To illustrate the partitioning of a data series, we use a common example: division at the median. As shown in Figure A1, the symbol sequence generated by this partition would result in a 0 for every data point below the median, and a 1 for every data point above or equal to the median. So if the first few values of the data series were {0.23, 0.89, 0.65, 0.92, 0.13, 0.56, 0.47, 0.06, ...}, and the median were 0.5, the resulting symbol series values would be {0,1, 1, 1, 0, 1, 0, 0, ...}. The alphabet in this case is the binary set {0,1}. The chosen partition is not limited to the median but can be any set of divisions.

Once the partition has been chosen and the symbol series determined, the analysis can be performed. Renyi analysis includes two measures of order in a data set: Renyi information and Renyi entropy. Renyi information is defined by the equation

$$I(q,k) = \frac{1}{1-q} \log \sum_{i=1}^{N} p_i^q \qquad (a_1)$$

for p_i=0, where I is the Renyi information given words of length k, N is the total number of possible words of length k, p_i is the frequency of the i^{th} word, and q is an integer which allows manual focus of Renyi information on frequent or infrequent events. If q>>1 then only the most frequent words (events) will affect the value of the Renyi information, whereas if q<0 then only the most infrequent words (events) will affect the value of the Renyi information. Thus q is a tool by which the focus of inquiry can be tuned to rare or frequent events. If q=1, the limit of equation a_1 as $q \to 1$ is known as Shannon information, defined as

$$I(1,k) = -\sum_{i=1}^{N} p_i \log p_i \qquad (a_2)$$

Renyi information can be thought of as the percent of information needed to determine the next data point.

Renyi information increases with word length; the amount it increases with word length can determine how much order is in a data set. On a plot of Renyi information

versus word length, a random data set will have a slope of one, a repetitive data set will have a slope of zero, and a slope between these two values will have more order the closer to zero it lies.

Renyi entropy is derived from Renyi information; it is technically defined as the limit of Renyi information as word length $k \to \infty$:

$$E(q) = \lim_{k\to\infty} \frac{I(q,k)}{k} \qquad (a_3)$$

where E is the Renyi entropy and I is the information of word length k from equation a_1. However, since the infinite is beyond our grasp while using these tools for practical applications, the entropy can be calculated without the limit, as simply the Renyi information divided by the word length. This calculation of the Renyi entropy always shows a curved line, asymptotically approaching some value between zero and one, even if the Renyi information versus word length is linear. The asymptotic curve is caused by the discrepancy between the slope from the origin to the information at word length one, and the slope of the information from word length one to higher word lengths.

Perhaps a better calculation of a Renyi entropy dependent on word length could be calculated by first moving the axis of the Renyi information versus word length plot such that the point at word length one became the origin, and then divide subsequent information values by the adjusted word length. The equation for this transformation is

$$E(q,k) = \frac{I(q,k) - I(q,1)}{k-1} \qquad (a_4)$$

where E(q,1)=1. The adjusted Renyi entropy would get rid of the discrepancy of slopes for a resulting Renyi entropy that better represents the slope of the Renyi information and could detect at which word length the information ceases to increase linearly.

There is a limit to the maximum calculable word length for the Renyi information and entropy which depends on alphabet size and data length. In a random data set, each of the words occur with similar frequency. Care must be taken with the interpretation of longer words, however: with an alphabet size k and a word length n, the number of

possible words is k^n, and if the total length of the data is less the number of possible words, then it is impossible for every word to occur. After this characteristic word length, the slope of the Renyi information flattens out, and the Renyi entropy falls off to 0. Ideally, the data length m, alphabet size k, and maximum considered word length n will be related by the equation

$$m = 5k^n \tag{a5}$$

so that each possible word has a possibility of occurring five times.

It is also important to mention two caveats when choosing a partition. First, while any alphabet size is possible, due to the relation described by a4, the higher the alphabet, the lower the maximum reliable word length. Second, if a partition is chosen such that each letter represents an unequal portion of the data, the maximum slope of the Renyi information and the maximum value of the Renyi entropy can become higher than one. For this reason it is advisable to produce the Renyi information or entropy of a random series of length equal to the data series being analyzed.

Appendix B: Detailed Hurst Analysis Method

Hurst analysis measures how similar the variability of one period of time behaves compared to previous periods of time of similar length. While there are many ways to calculate the Hurst exponent, we use the rescaled range method described by Mandelbrot and Wallis (1969). This method is described in detail here.

The rescaled range method is easily influenced by trends, so care must be taken that all annual and linear trends be removed from the data series.

First the data is split into two segments. The length of each of these segments, also called the time lag, is τ. The mean of each segment is calculated, and then a running sum of the data points subtracted by the segment mean. The running sum values are

$$Z_i = \sum_{j=1}^{i} x_j - \bar{x} \tag{b1}$$

where Z_i is the value of the running sum at point i, i is the ith value of the segment, j runs from the first data point of the segment to the i^{th}, x_j is the data value at point j, and $\bar{x}$ is the segment mean. The rescaled range, or R/S is calculated by

$$(Z_{i_{\max}} - Z_{i_{\min}})/s \tag{b2}$$

where s is the standard deviation of the segment. The R/S value is calculated for all segments and the mean of these is the final R/S value associated with the time lag τ.

Next, the length of the segment or time lag is decreased; we decreased by a factor of 1.2. Each time τ decreases, a new R/S value is found, and the process continues until τ=2. Then we are left with a set of ordered pairs.

The ordered paris of R/S and τ conform to the power law

$$R/S = k \cdot \tau^H \tag{b3}$$

for certain consecutive values of τ, where k is a constant and H is the Hurst exponent. This is most easily seen on a double logarithmic plot of R/S versus τ, where power laws are visible as straight lines with a slope of H. It is important to note that R/S does not conform to the same power law for all values of τ. For example, in Figure B1 the Hurst

exponent from four days to three weeks is 0.88; after three weeks the Hurst exponent decreases to 0.71 until about 2.5 years, and from three to 15 years it becomes 0.57. It is important to visually examine each individual plot of R/S versus τ because the values of τ at which the Hurst exponent changes may be different for each data series.

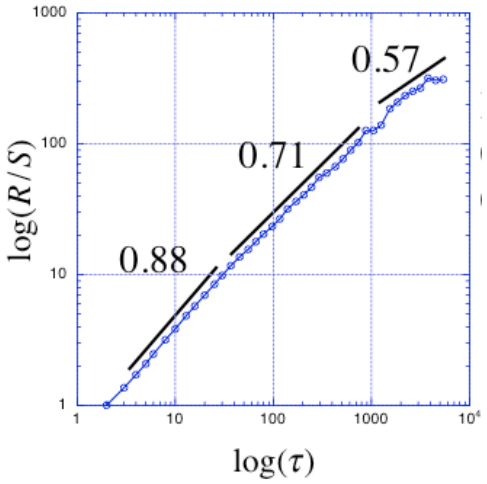


Figure B1: Double logarithmic plot of R/S versus τ for McGrath temperature data. Hurst exponents for different values of τ are labeled.

The value of the Hurst exponent shows how persistent the data series is for the characteristic values of τ. If H=0.5, then the data is random for the characteristic values of τ. If H>0.5, then the data is persistent, which is to say a period of time within the characteristic values of τ is statistically more likely to have similar variability patterns the previous periods of time. If H<0.5, then the data is anti-persistent, which indicates that a period of time within the characteristic values of τ is statistically more likely to have opposite variability patterns as the previous period of time.

Sometimes there is non-stationarity in the Hurst exponent, or that within the data series the Hurst exponent is changing for the characteristic values of τ; there are several different ways to look for non-stationarity. H should theoretically lie between zero and one, but sometimes the rescaled range algorithm provides Hurst exponents greater than one. This is an indication that there is non-stationarity in the data series. Another indication is a lack of power law, or a double logarithmic R/S versus τ plot that shows no

straight line for certain values of τ. On the other hand, sometimes non-stationarity exists but does not display either of the above features. If this is the case, the only way to detect the change of Hurst exponent is to test the R/S method on different parts of the data series and determine whether the outcome is the same for each part.

Appendix C: Lower 48 SAT Hurst Analysis

This research included the calculation of the Hurst exponents of GSOD observations and NCEP/NCAR reanalysis of SAT data from the lower forty-eight. Additional work beyond the scope of this project is required to provide a physical interpretation for the results and could be a project for future work.

From Figure C1 and C2, it can be seen that NCEP/NCAR reanalysis SAT's have similar Hurst exponents to those of the station observations.

The data was divided at 1976 because that is the year at which Alaska stations showed greatest difference in Hurst exponent. However, it is possible that this is nothing more than an arbitrary date for the lower forty-eight because there are significantly different climate processes operating in this domain. A key feature present in the Hurst analysis for SAT and SLP is decreased persistence in the later period compared to the earlier period in south central US (Figs. C.1 and C.2).

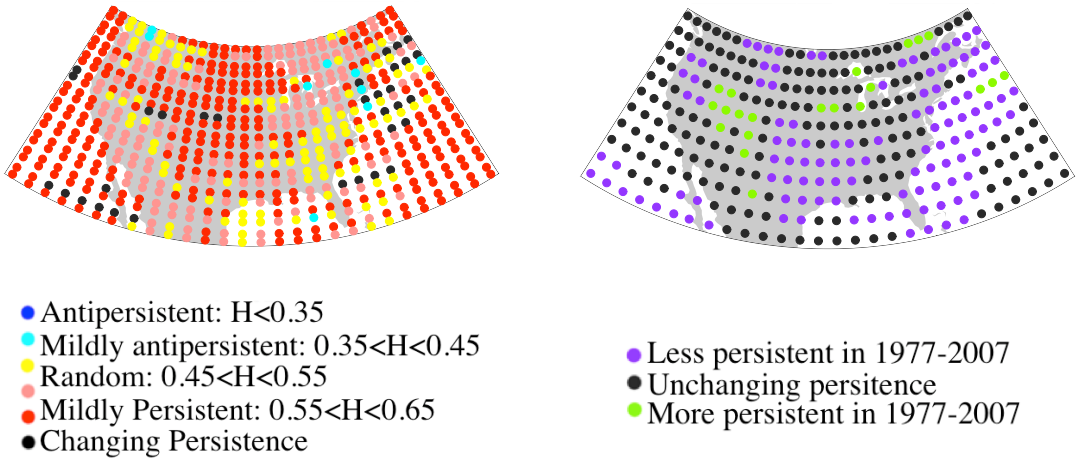


Figure C.1: Hurst Exponent of US lower forty-eight SAT reanalysis data. (a) Hurst exponent for 1946-1975 (top dot) and for 1977-2007 (bottom dot), and (b) the difference between hurst Exponents in 1977-2007 compared to 1946-1975.

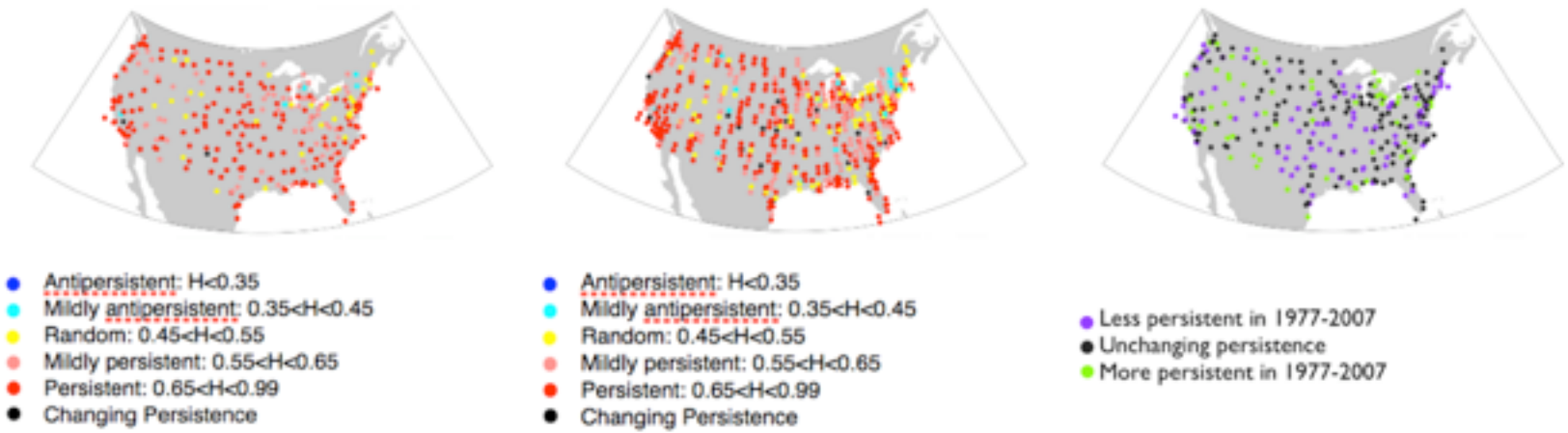


Figure C2: Hurst Exponent of the U.S. lower forty-eight SAT observational data. (a) Hurst exponent for the full period of record, (b) Hurst exponent for 1946-1975 (top dot) and for 1977-2007 (bottom dot), and (c) the difference between Hurst exponents in 1977-2007 compared to 1946-1975.