Implications of global climate change on California's regional PM pollution during 1951-2008

Angadh Singh and **Ahmet Palazoglu***
University of California
Davis, California 95616

1. Introduction

Air pollution in California's Central Valley is largely attributed to local weather conditions, which are, in turn, driven by synoptic weather systems prevailing over the southwestern United States (Beaver and Palazoglu 2006). These synoptic systems are a part of a larger circulation pattern with complex interactions amongst its various components. An air pollution episode in Central California could thus be a result of complex interactions between the various components of a large scale circulation.

Accelerated changes in global circulation patterns in the recent past have altered the local climate in the Central Valley region. The implications of such changes could be to alter the frequency of meteorological conditions favorable to the build-up of pollutants and increased air pollution episodes in the valley. There has been a good deal of interest in understanding the effects of climate change on regional air quality and such efforts have involved conducting climate change simulations using general circulation models (GCM's) and the results obtained are currently subject to significant uncertainties (Vautard and Hauglustaine 2007) and (Jacob and Winner 2009).

The scope of data analysis methods to study the response of air quality to these climatological changes is limited due to the lack of good air quality and meteorological datasets. Though multivariate data modeling techniques are not prognostic in nature, the analysis of historical databases of climatological data can provide useful insight into such changes in the recent past and help understand future trends. Analysis of voluminous weather datasets can reveal low frequency trends in air quality which would be otherwise impossible to determine due to lack of data.

Particulate matter (PM) data for various sites located in the Central Valley is limited to the period between 1999–2008. The limited dataset is used to identify a unique set of days during the PM season each year (November-February) as severe and widespread exceedances using objective selection criteria. The daily averaged NCEP/NCAR reanalysis 500-hPa geopotential height weather maps for the target exceedance set are then used to identify areas on the contour maps which significantly differ from the overall seasonal mean by performing a statistical significance test at each data point for an expanded spatial domain (20° S to 75° N latitude and 100°E to 350°E longitude). Data points with significant differences were found to exist and correspond to key weather events in the northern hemisphere.

The reanalysis weather maps are analyzed using 2-D wavelet transforms to characterize variable gradients and the spatial extent of each of the interacting components (Strang and Nguyen 1996). This transformation enables the segregation of significant events on the contour maps and identifies relevant scales which contain meaningful information. The wavelet details are then analyzed at chosen scales for a comprehensive multiscale analysis of the reanalysis weather maps. The wavelet coefficients at each scale are subject to a significance test to identify a reduced set of statistically significant coefficients. Such dimension reduction is essential owing to a large number of variables and availability of limited target samples.

The principal step of the proposed methodology is to then scan historical database of the daily averaged NCEP/NCAR reanalysis weather maps (1951–2008) for similar events conducive to poor air quality in the past. Each day from a PM season in the past is compared to the identified exceedance set by first performing wavelet decomposition and then testing for similarity of chosen coefficients at the relevant scales. The matches returned by the algorithm are days having large scale weather conditions in the expanded domain which are conducive to widespread and in certain cases severe exceedances in Central Valley. The variation in the frequency of such events during the study period (1951-2008) is then examined to see the effects of the changing climate on California's Central Valley PM pollution potential for the past six decades.

2. Methodology

California's Central Valley presents high particulate matter PM concentrations during winter months. The current National Ambient Air Quality Standards (NAAQS) impose a threshold on the 24-hr (time-averaged) mass concentrations for airborne particulates in two important size ranges: $35\mu g/m^3$ for $PM_{2.5}$ levels and $150\mu g/m^3$ for PM_{10} levels. The NAAQS also imposes a threshold of $15\mu g/m^3$ for the annual average $PM_{2.5}$ concentration; no annual threshold exists for PM_{10}. Almost all the Central California Air Basins attain the current NAAQS for annual $PM_{2.5}$ and 24-hr PM_{10}, but do not attain the NAAQS for 24-hr $PM_{2.5}$. The analysis

*anpalazoglu@ucdavis.edu

is thus focused on days when $PM_{2.5}$ levels at sampling sites in the study domain exceed the $PM_{2.5}$ NAAQS threshold and result in an exceedance day. Most PM exceedances occur during the months November through February each year and this period is designated as the extended PM season. Exceedances occurring outside this period are very few and atypical to capture relevant meteorological patterns contributing to PM dynamics and transformation. It is important to eliminate effects of measurement uncertainties, local emissions and microscale meteorological effects not representative of regional (synoptic) conditions prevailing over the study domain. Exclusive exceedances at one or more locations may not be associated with large scale weather signatures. Current analysis emphasizes the influence of large scale weather associated with Central Valley's PM levels to draw out relationships between changing global climate patterns and wintertime air quality.

$PM_{2.5}$ levels used in the analysis were obtained from California Air Resources Board (2009 Air Quality DVD). We focus our attention to the a study domain encompassing Bay Area, southern Sacramento Valley region and parts from northern and central San Joaquin Valley region. 18 sites were chosen to represent regional PM concentrations in the study domain (Figure 1). The target sample chosen had at least one site with a $PM_{2.5}$ level of 60 μg/m^3 and more than 75% of sampling sites with available measurements recording exceedances. Data is available on 1-day, 3-day and 6-day irregular sampling schedules and results in lack of comprehensive spatial and temporal coverage. Days in the target samples are chosen so that $PM_{2.5}$ levels are available for 4 sites or more on the particular day. This measure serves to segregate days with severe and widespread PM exceedances throughout the study domain representing a regional response as opposed to local effects. A target sample of 58 days fulfilling the aforementioned conditions is thus identified and used in subsequent data analysis.

Global gridded weather data obtained from NOAA's NCEP/NCAR reanalysis project characterizes the varying 3-dimensional state of the atmosphere at 17 pressure levels (mb). We chose geopotential heights at the 500-hPa pressure level as a good indicator of large scale weather systems affecting regional weather patterns in Central California. A 500-hPa geopotential height weather map may be viewed as an imaginary surface in the troposphere with the value at any grid point corresponding to its height above the mean sea level. Events on this constant pressure surface (or topography) are summarized in a 2-D matrix with its elements recorded at fixed latitude/longitude coordinates. Any event (weather pattern) on a geopotential height map is uniquely characterized by varying characteristics such as spatial location, size and its associated pressure gradients. In addition, gridded weather data present copious spatial information contained in a large set of reconstructed variables. For small sample sizes, the problem of characterizing individual features and complex interactions between these features using statistical data analysis methods becomes intractable. Analysis of small target samples for a large spatial domain requires advanced dimension reduction techniques to summarize relevant information in less number of variables; preserving the multi-scale and multivariate nature of these atmospheric datasets at the same time.

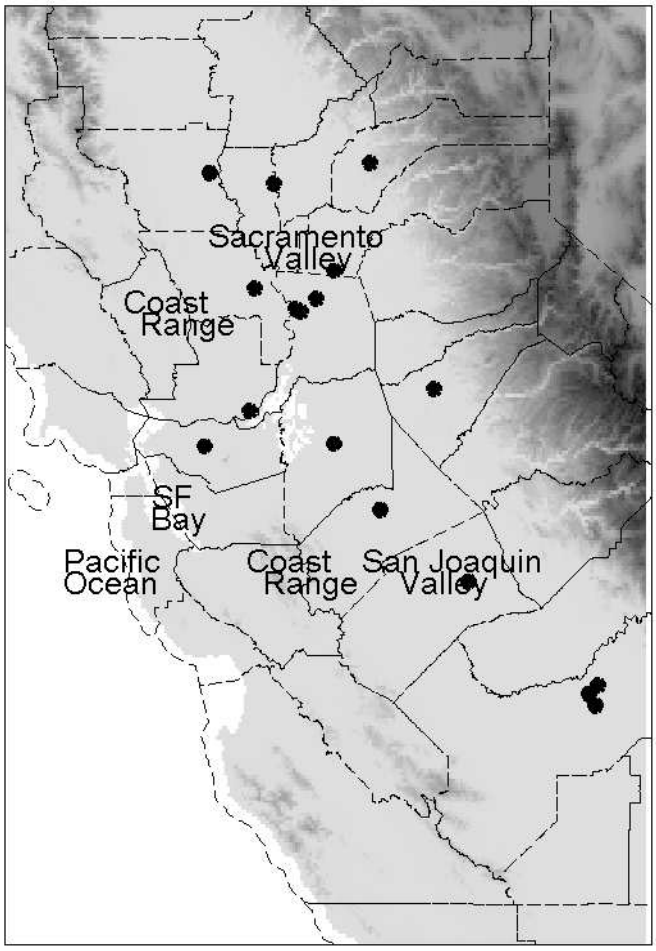

Figure 1: A map of the study domain displaying the spatial extent and position of the PM monitors.

a. *Wavelet Transforms*

The standard application of Fourier Transform decomposes a weather map into its frequency components, which can then be used to obtain a power spectral density (PSD) function. Since the PSD function is defined in the frequency domain, spatial information about the original reconstructed variables is lost. Though determination of all existing frequencies in the data is possible, there is no information about the exact spatial location of an event. Since the purpose of the current analysis is to recognize, locate, and measure the spatial extent of certain weather patterns and their associated anomalies, information about location, continuity, and transition of frequency components becomes critical.

2-D Wavelet transforms (WT) provide an excellent tool to analyze geospatial data by providing space/frequency localization at predetermined bands of frequencies crucial for identifying pat-

terns varying in spatial location and associated pressure gradients. Strong pressure gradients translate to high frequency (low scale) information and relatively weak patterns with widely spaced contours would be captured at lower frequency scales. The elaborate details of an atmospheric pattern are contained at multiple levels (frequencies or scales) of wavelet coefficients obtained after decomposition. The independence of wavelet coefficients across scales makes it possible to divide the task of analyzing unmanageable weather map data into smaller tractable subtasks and later aggregate the results.

Wavelet Transform (WT) allows a perfect resolution for both space and frequency, which is achieved by dilating and translating a finite wavelet function at different frequency ranges (Strang and Nguyen 1996). The main wavelet function is referred to as the mother wavelet and is used to obtain wavelet basis functions. Since the wavelet transform is not unique, there are different types of mother wavelet functions that can be used to obtain a WT. The basis functions are expressed as

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi(\frac{x-b}{a}) \tag{1}$$

where $\psi_{a,b}(x)$, the mother wavelet, is a space function with finite energy and fast decay, and a and b represent the dilation and translation parameters, respectively. The 1-D continuous wavelet transform (CWT) can be defined as

$$CWT(a,b) = \int_{-\infty}^{\infty} f(x)\psi^*_{a,b}(x) \tag{2}$$

The discrete wavelet transform (DWT) is a shortened version of the CWT designed to remove the redundancy by limiting the number of wavelet decomposition levels using a dyadic scale, i.e., $a = 2^j$ and $b = k2^j$ where j and k are positive integers.

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k) \tag{3}$$

$$DWT(j,k) = \int_{-\infty}^{\infty} f[x]\psi^*_{j,k}(x) \tag{4}$$

where $f[x]$ now represents a discrete function.

Weather maps are two-dimensional signals, with x and y representing the independent variables and z, the height of the pressure surface, as the dependent variable. The rows and columns of the matrix are treated as independent 1-D signals. A high-pass and a low-pass filter are applied to the map in the x-direction (zonal), and the results are downsampled by deleting every other column. This results in two matrices of approximately half the size of the original, one containing high-frequency components of the rows, and the other containing low-frequency components. These two matrices are then each filtered down the columns (meridional direction) using high-pass and low-pass filters and downsampling the results along the rows (deleting every other row). The resulting four matrices are approximately one-fourth the size of the original image and represent the smoothed approximation, the horizontal detail, the vertical detail, and the diagonal details , respectively. The detailed explanation of theory can be found in Carmichael et al. (2004).

b. *Principal Component Analysis*

Gridded weather data present copious spatial information and a large set of variables which makes it difficult to analyze using data analysis methods especially when target sample size is small. The large number of available measurements obscures the underlying relationships that exist in the data (X) and makes it difficult to draw conclusions based on any single variable. Principal Component Analysis (PCA) was proposed as a data reduction technique which accounted for the fact that the underlying variability of the data can be explained by relatively few unobservable variables (Johnson and Wichern 2002). These latent variables (Y_i) are linear combinations of the measured variables which maximize the variance explained in particular directions. The obtained subspace happens to be the same as that defined by the chosen eigen vectors corresponding to the largest eigen values of the covariance matrix of the data.

$$Y_i = e_i'X = e_{i1}X_1 + e_{i2}X_2 + ... + e_{ip}X_p \tag{5}$$

$$Var(Y_i) = e_i'\sum e_i = \lambda_i \tag{6}$$

$$Cov(Y_i, Y_k) = e_i'\sum e_k = 0, i \neq k \tag{7}$$

The original data matrix is written as a product of a scores matrix Y and the principal component loadings matrix P^T consisting of a chosen number of right eigen vectors of the covariance of the data matrix. The errors in approximating the original data matrix using a small number of principal components are a useful measure of deviations from the modeled subspace. T^2 statistic is then utilized to compare daily averaged weather maps to the target group if the final scores are within confidence bounds.

$$X = Y \times P' + E \tag{8}$$

$$T^2 = Y'\Lambda Y = \sum_1^r \frac{Y_i^2}{\lambda_i} \tag{9}$$

$$T^2 \sim \chi_r^2 \tag{10}$$

A simple 2-sample t-test was performed to compare the 500-hPa geopotential heights for sample days from the rest of the population. Figure 2 shows the latitude/longitude coordinates failing the test at $\alpha = 0.05$ clearly indicating specific regions and associated weather patterns (Aleutian Low, Onshore Ridge over western United States etc.) in the northern hemisphere. The spatial correlation structure and teleconnection patterns between these statistically significant features characterizes the uniqueness of the target sample set from days in the extended PM season. Due to lack of PM data it is impossible to identify all target sample days during the 1999–2007 period. A large number of winter days chosen during the observation period justifies the assumption of

normality of the population distribution. The coordinates failing the t-test on a relatively small target sample are envisioned as belonging to the extremes of the distribution. The proposed scheme does not utilize these coordinates or features directly because of the large dimensionality of geopotential height values at resulting significant coordinates.

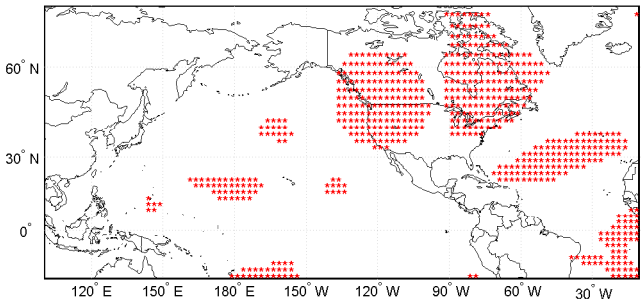

Figure 2: Two sample t-test results for individual latitude/longitude coordinates. The highlighted coordinates indicated in the figure pass the test at α = 0.01 confidence level.

Figure 3 shows the composite map of the 500-hPa geopotential heights for sample days in the expanded spatial domain (20° S to 75° N latitude and 100°E to 350°E longitude). The onshore ridge over west coast of United States is known to be associated with bulk of PM exceedances in Bay Area and adjoining Central Valley regions. Surface pressure gradients over Central California are relatively weak. Upper-level flows over Central California are also weak due to anticyclonic blocking. Relatively calm conditions develop throughout most of the Central Valley and east Bay Area, however, flow becomes organized and easterly surface winds are present. The strongly anticyclonic conditions are associated with limited vertical dispersion due to increased atmospheric stability and large-scale subsidence. Early morning low-level temperature inversions are also likely to develop due to nocturnal cooling under the clear-sky conditions. When compared to Figure 2 we observe 1) a trough south of Alaska over the Pacific Ocean, 2) the polar low north of Japan and located along east coast of Russia, 3) the diffuse high pressure pattern located to the north-northeast of Papua New Guinea, 4) the trough over Canada and eastern United Stated 5) high pressure located northeast of Caribbean in the north Atlantic.

The 500-hPA composite weather map is decomposed using a Daubechies-4 wavelet with 6 levels of decomposition and significant results are shown in Figure 4. Reconstructed weather maps using the sum of the horizontal, vertical and diagonal details at all scales capture the patterns observed in the sample composite shown earlier in Figure 3. Scale 1 details capture very high frequency details where as scale-6 captures gradients in geopotential as we move from the equator towards the poles and are not shown. Scales 2 through 5 capture the distinctive events discussed in one or multiple levels. The onshore ridge over western United States, predominantly appears in reconstructed details at scales 3, 4 and 5 as corroborated by the spatial location of significant anomalies. The trough located south of Alaska has its signature in the details at scales 4 and 5. A close look at scale 4 reveals almost all significant anomalies associated with synoptic weather patterns discussed earlier. Details at scale-5 capture high pressure cells over subtropical latitudes corresponding to large scale subsidence associated with the descending branch of Hadley cell and the jet-stream situated over midlatitudes. This roughly corresponds to synoptic weather patterns being identified at a lower scales and larger global circulation patterns being described at higher scales.

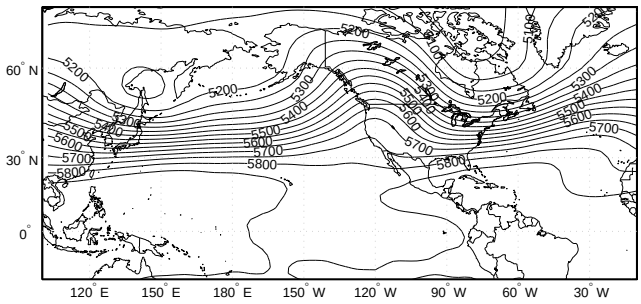

Figure 3: The 500-hPa geopotential heights composite map for specific target samples.

2-D wavelet decomposition is applied on 500-hPa geopotential height maps for all samples belonging to the target group. A two sample t-test is performed on all detailed wavelet coefficients (horizontal, vertical and diagonal) at every scale to identify a set of coefficients which significantly distinguish the target group from the entire population days belonging to the extended PM season at $\alpha = 0.01$ significance level. This procedure can be viewed as a thresholding technique to separate the coefficients which contribute the most on each scale. These statistically significant horizontal, vertical and diagonal detailed coefficients are then combined to construct a different PCA model at each scale. PCA is essential to identify the correlation structure between wavelet coefficients at all scales to characterize the teleconnections patterns amongst significant events at each scale. A confidence ellipsoid is constructed using a reduced order PCA model which is made to account for 95% of the covariance explained by the original set of wavelet coefficients. This step also avoids redundancies and singularity of the covariance matrix if an ordinary multivariate measure of distance were to be used.

The pattern recognition algorithm decomposes the 500-hPa geopotential height weather map using the same wavelet transform formulation to determine values of previously identified statistically significant wavelet coefficients. The values obtained are then projected onto the PCA-model subspace at each scale and the T^2 statistic is used to test if the new sample belongs to the population containing target samples. Only if the target sample lies within the bounds of the confidence ellipsoid at each scale, we accept that the particular day belongs to the population defined by the target samples. This procedure is repeated for all days belonging to the extended PM season from 2008 back to the

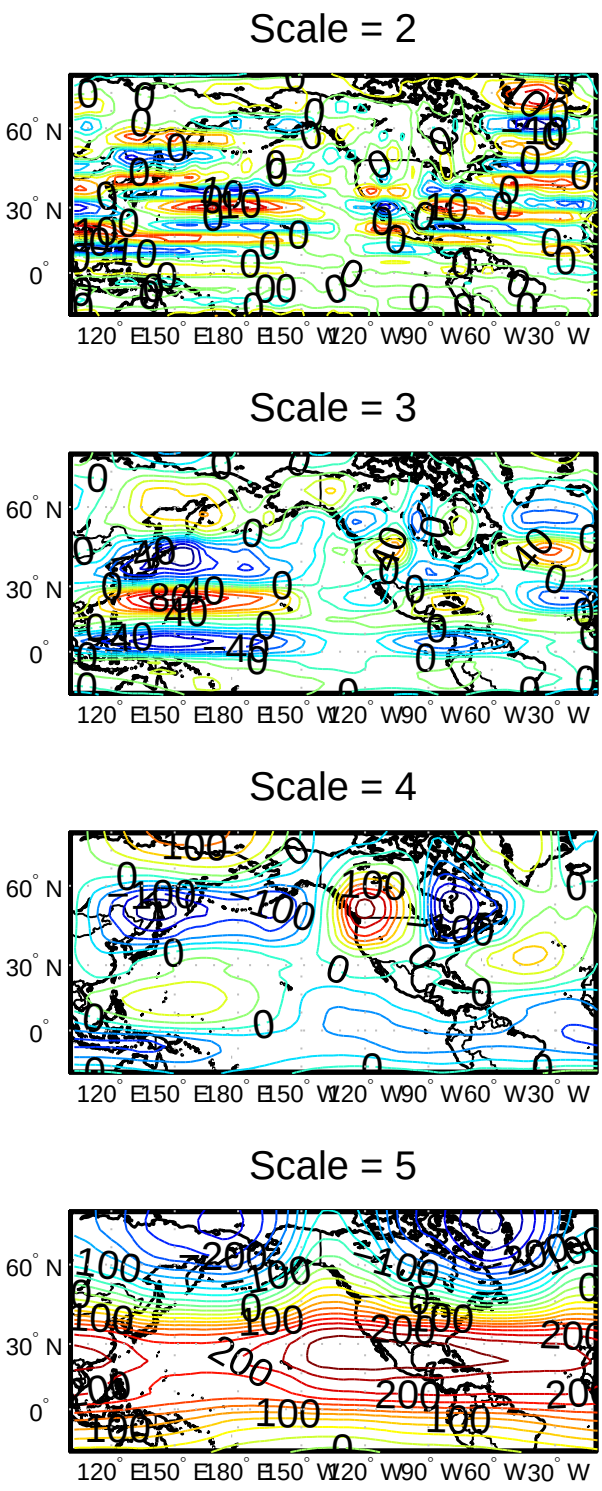

Figure 4: Reconstructed Maps using the sum of horizontal, vertical and diagonal detailed coefficients.

year 1948. The number of matches obtained each PM season corresponds to days when regional meteorological conditions are highly conducive to trigger widespread exceedances throughout the study domain. The actual occurrence of an exceedance day at any given site would additionally depend on numerous localized influences i.e., emissions, upwind conditions (incase of transport) and chemical transformation of PM.

3. Results and Discussion

The set of days identified by the algorithm in every PM season from 1950–51 through 2007–08 are shown in Figure 5. These matches correlate well with the actual number of exceedances occurring during each season obtained using PM measurements available from 1999–2007 ($\rho = 0.7066$). 73.5% of the days identified by the algorithm during 1999-2007 (including the target samples) are found to be exceedances at one or more sites in the study domain. Though meteorological conditions may be conducive to buildup of high $PM_{2.5}$ levels in the region, actual occurrence of an exceedance depends upon several other factors (local emissions and upwind conditions).

El Niño events characterized by anomalous heating/cooling of surface waters in eastern/central Pacific Ocean have a profound influence on winter weather patterns in California. El Niño, a localized event in the Pacific ocean basin, has a strong climatic response at remote locations on the globe due to its associated teleconnections patterns and explains a high proportion of climate variability on the sub-decadal scale (Yeh et al. 2009). Traditionally, canonical El Niño was characterized by warm anomalous sea surface temperatures off the coast of Peru. Recent Niño studies identify a more frequent and emerging El Niño pattern called the Central Pacific (CP) El Niño (also termed as El Niño Modoki or dateline El Niño) (Ashok and Yamagata 2009). Increased warming in the pacific as a result of climate change is projected to increase the frequency of occurrence of the CP-type El Niño) compared to EP-type El Niño) events (Yeh et al. 2009).

Using the classification of Larkin and Harrison (2005), we segregate the data into EP-El Niño, CP-El Niño and La Niña years and observe the following correlations. La Niña years present a high negative correlation (as high as -0.84) between the number of matches and the Niño3 and Niño1+2 region SST anomaly indices for all seasons. Thus, during a strong La Niña event with highly negative SST anomalies in central and eastern pacific, we expect Central California region to be more prone to PM exceedances than average winter conditions. For EP-type El Niño years, there also appears to be a strong negative correlation (-0.58) between identified matches and summer with spring (MAM) Niño3 and Niño3.4 SST anomalies. This presents a possibility of high SST anomalies in the Niño3 region (eastern Pacific) resulting in low number of PM exceedance days during the winter season. For CP-type El Niño events, there is a fairly strong negative correlation (-0.69) between identified matches and Winter (DJF) Niño-4 SST anomalies indicating decreased PM exceedances associated with the warming in central/western Pacific.

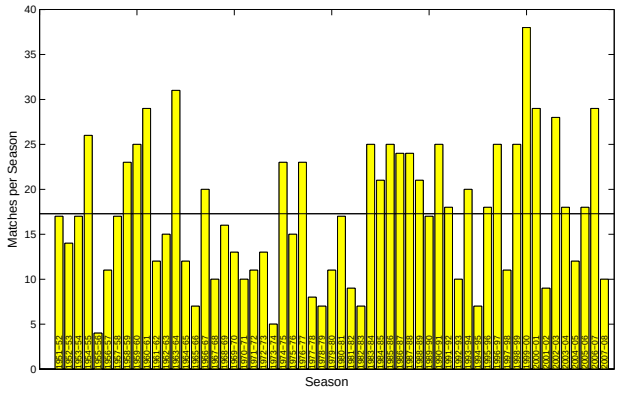

Figure 5: Barplot showing the number of matches identified each year from 1951–2008.

4. Conclusions

The proposed methodology is used to scan historical databasees of the daily averaged NCEP/NCAR reanalysis weather maps (1951–2008) for similar events conducive to poor air quality in the past. The matches returned by the algorithm have large scale weather conditions in the expanded domain conducive to widespread and in certain cases severe PM exceedances in California's Central Valley. The correlation patterns between sea surface temperature anomaly indices in the Pacific Basin and identified matches are used to understand the response wintertime air quality in California's Central valley to ENSO variability (a subdecadal scale global atmospheric pattern). The projected warming trends for sea surface temperatures are expected to influence PM response in Central Valley consistent with results obtained earlier.

References

Ashok, K. and T. Yamagata, 2009: Climate change: The el niño with a difference. *Nature*, **461**, 481–484.

Beaver, S. and A. Palazoglu, 2006: Cluster analysis of hourly wind measurements to reveal synoptic regimes affecting air quality. *Journal of Applied Meteorology and Climatology*, **45**, 1710–1726.

Carmichael, M., R. Vidu, A. Maksumov, A. Palazoglu, and P. Stroeve, 2004: Using wavelets to analyze afm images of thin films: Surface micelles and supported lipid bilayers. *Langmuir*, **20**, 11557–11568.

Jacob, D. and D. Winner, 2009: Effect of climate change on air quality. *Atmospheric Environment*, **43**, 51–63.

Johnson, R. A. and D. W. Wichern, 2002: *Applied Multivariate Statistical Analysis*. Prentice hall, Upper Saddle River, NJ 07458, fifth edition.

Larkin, N. K. and D. E. Harrison, 2005: On the definition of el niño and associated seasonal average u.s. weather anomalies. *Geophysical Research Letters*, **32**.

Strang, G. and T. Nguyen, 1996: *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA.

Vautard, R. and D. Hauglustaine, 2007: Impact of global climate change on regional air quality: Introduction to the thematic issue. *Comptes Rendus Geosciences*, **339**, 703–708.

Yeh, S. W., J. S. Kug, B. Dewitte, M. H. Kwon, B. P. Kurtman, and F. F. Jin, 2009: El niño in a changing climate. *Nature*, **461**, 511–514.