# A Statistical Framework to Identify the Influence of Large-Scale Weather Events on Regional Air Pollution

ANGADH SINGH AND AHMET PALAZOGLU

*University of California, Davis, Davis, California*

ABSTRACT

Regional air pollution episodes occur as a result of increased emissions and prevalence of conducive meteorological conditions. The frequency of occurrence of such favorable conditions on a regional scale may be influenced by large-scale climatic events like ENSO and the Pacific decadal oscillation (PDO). The scarcity of measurements of criteria pollutants, especially ozone and particulate matter (PM), prior to the last 10–15-yr period, limits the scope of observing the influence of climate variability during recent decades on regional pollution levels. The authors propose a novel statistical framework to utilize available measurements and characterize synoptic influences on regional PM pollution in California's Central Valley during 1998–2008. The identified target conditions are used to develop a classification scheme to scan historical climate datasets dating back to 1948. The procedure identifies exceedance-conducive days during 1950–98, when no $PM_{2.5}$ measurements were available. Temporal patterns in seasonal frequency of these identified exceedance-conducive days are investigated for temporal patterns driven by ENSO and PDO.

## 1. Introduction

Air pollution episodes in a region can be attributed to high emission levels and favorable meteorological conditions. Most air pollution control strategies are aimed at emissions reductions as a measure to keep pollutant levels in check. The outcome of such strategies depends upon relatively accurate understanding of chemical transformation and long- and short-range transport of primary–secondary pollutant species in the atmosphere. Meteorological variables like temperature, humidity, wind, and solar insolation heavily influence surface pollution levels at a sampling site (Seinfeld and Pandis 1998). Regional levels of ozone and fine particulate matter (PM) have been extensively studied in the past for their detrimental effects on human health (Pope and Dockery 2006). A significant proportion of variability in ozone and PM concentrations is attributed to meteorological variability (Beaver and Palazoglu 2006; Beaver et al. 2010; Wise and Comrie 2005; Tai et al. 2010). Though meteorological models for air quality simulations have been developed and remain an indispensable tool in predicting pollution levels, they are subject to uncertainties (Seaman

2000). Studies investigating the effects of climate variability and climate change (during the recent past and near future) on criteria pollutants like ozone and PM in the United States typically employ a combination of a general circulation model (GCM) to predict future climate and a chemical transport model (CTM) to obtain pollutant concentrations (Jacob and Winner 2009; Mickley et al. 2004; Patz et al. 2005; Leung and Gustafson 2005; Tagaris et al. 2007; Racherla and Adams 2006, 2008).

The U.S. Environmental Protection Agency's (EPA's) Air Quality System (AQS) records for PM measurements date back to 1998 and lack availability of regularly sampled PM measurements before the previous decade. PM with aerodynamic diameter $\leq 2.5$ $\mu$m ($PM_{2.5}$) measurements are sampled irregularly (1-, 3-, or 6-day sampling) and present a high probability of missing high-PM days (Watson and Chow 2002). The ability of statistical modeling studies quantifying meteorological effects on regional PM levels is limited because of a lack of actual measurements. Tai et al. (2010) use 11 yr of data (1998–2008) to investigate the sensitivity of $PM_{2.5}$ to meteorological variables in the United States and the implications of changing climate. Mahmud et al. (2010) consider a 7-yr analysis period (2000–06) to validate a GCM and air quality model output against actual observations. Though air quality measurements are limited, climatic datasets extend beyond 1998 and contain information about the prevalence

*Corresponding author address:* Ahmet Palazoglu, University of California, Davis, One Shields Avenue, Davis, CA 95616.
E-mail: anpalazoglu@ucdavis.edu

of meteorological conditions that would have been favorable for PM buildup in the past.

In this paper, we aim to develop a methodology to characterize large-scale weather conditions favoring PM buildup on a regional scale. Historical climatic datasets are then scanned to identify days with similar large-scale conditions that would be characterized by high PM levels in a current emissions scenario, which implies emissions during 1998–2008 under which a target set of high-PM days is identified. To demonstrate the applicability and potential of our method, we choose Central Valley in California as our study domain. Previous studies indicate a strong influence of synoptic weather conditions in driving surface meteorological conditions in the Central Valley (Beaver and Palazoglu 2006; Beaver et al. 2010). During summer, urban areas in the Central Valley are plagued by high ozone levels while PM is a major concern in the winter. The highest PM concentrations typically occur when stagnant wind conditions combined with strong temperature inversions can trap pollutants close to the surface (Kleeman et al. 2005). Beaver et al. (2010) find that an aloft anticyclonic ridge, located directly over the Central Valley, with weak surface pressure gradients and coolest overnight inland temperatures, was favorable for the buildup of PM levels in the region. The Central Valley receives almost all of its annual rainfall during winter under the influence of midlatitude storms (Mitchell and Blier 1997). Precipitation, strong winds, and vertical mixing result in low PM levels on the surface. The annual PM air quality in the region is tightly connected with the relative frequency of occurrence of these predominant weather patterns during winter. In a similar study, Leibensperger et al. (2008) show that a decrease in the summertime midlatitude cyclone frequency across eastern North America has led to increased stagnation days and poor ozone air quality.

Variability in the frequency and intensity of precipitation under midlatitude cyclones in the western United States with applications to California has been extensively studied (Cayan et al. 1998; Dettinger et al. 1998; Mitchell and Blier 1997). Large-scale climate variability associated with El Niño–Southern Oscillation (ENSO) and Pacific decadal oscillation (PDO) has a profound influence on precipitation events in various regions of United States (Schonher and Nicholson 1989; Cayan et al. 1999; Schubert et al. 2008; Zhang et al. 2010; Gershunov and Barnett 1998b; Brown and Comrie 2004; Higgins et al. 2007). Most studies focus on the frequency of extreme precipitation events (Cayan et al. 1999; Gershunov and Barnett 1998a; Gershunov and Cayan 2003; Schubert et al. 2008; Zhang et al. 2010). For air quality purposes, even an ordinary precipitation event would suffice for ventilating this region and leading to low pollutant levels

(Leibensperger et al. 2008). Thus, large-scale conditions favoring low PM levels have been extensively studied but little is known about variability of the frequency of PM-conducive weather.

In the current study, we utilize National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis-1 data (1998–2008) to identify large-scale weather patterns during widespread PM pollution events in the Central Valley. The procedure identifies statistically significant events on daily averaged gridded 500- and 1000-hPa geopotential height patterns. The correlation structure between statistically significant features is attributed to the teleconnection between relatively remote weather events. Next, we develop a classification scheme to identify days with statistically significant similarities and extrapolate the occurrence of such conditions to the period 1950–2008. The results obtained can be utilized to study the implications of climate variability in the past (1948–present) on regional air quality. The variability in season totals and comparison between the relative seasonal frequencies of favorable conditions and rain events would provide crucial insight into annual PM air quality in the region. Though centered on Central Valley in California, the methods developed are equally applicable to any other region with a strong synoptic influence on local weather. Seasonal frequencies of such conditions during El Niño–La Niña events and dependence on PDO are investigated to detect any effects of large-scale climatic events.

Section 2 describes the data and provides a comprehensive discussion of methods applied to the ensuing study. The application of the proposed framework to central California PM pollution is discussed in section 3. The results are divided into two parts: we first discuss the validity of the proposed algorithm's output and then study the influence of large-scale climatic variability, associated with ENSO and PDO, on the frequency of identified days. Section 4 provides a brief summary and concludes with future research directions and broader implications of the developed methodology.

## 2. Data and methods

### a. Data

The target group selection is performed using daily averaged $PM_{2.5}$ measurements collected during 1998–2008 based on EPA-designated federal reference method (FRM) samplers in central California. The $PM_{2.5}$ dataset, obtained from the California Air Resources Board (CARB), contained measurements from 39 sites. Figure 1 shows California's Central Valley bounded by the Coast Range to the west, Sierra Nevada to the east, Cascade Range to the north, and Tehachapi Mountains to the
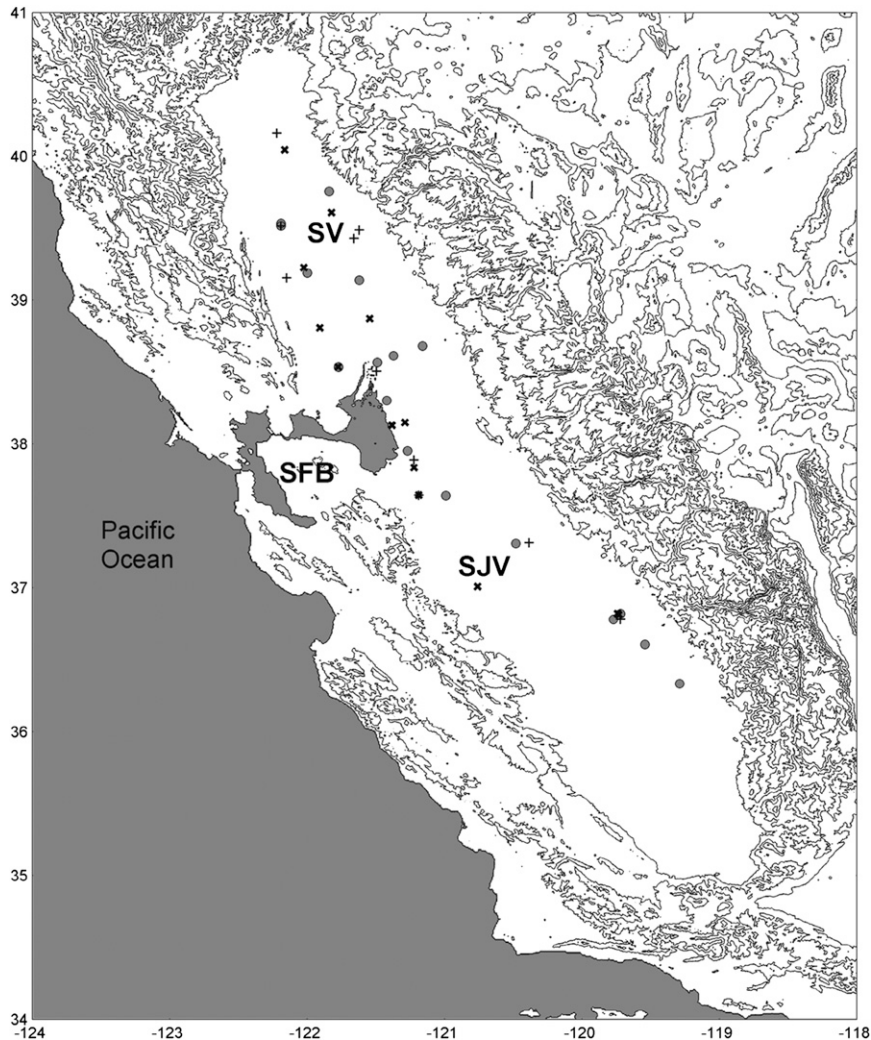
FIG. 1. A map of the study domain displaying the spatial extent, position of the PM monitors indicated by dots, and contour plots showing elevation drawn every 600 m, starting at 600 m. SFB = San Francisco Bay Area, SV = Sacramento Valley, and SJV = San Joaquin Valley. The black dots indicate $PM_{2.5}$ monitoring sites in the valley, the $\times$ and + signs represent the NCDC and CIMIS sites sampling precipitation, respectively.

south. Comprising Sacramento Valley to the north and San Joaquin Valley toward the south, the Valley is divided by the delta region of Sacramento and the San Joaquin Rivers. Around the delta region, the Valley is connected to the San Francisco Bay Area to its west by Carquinez Strait, a gap in the Coast Range. Identification of poor PM air quality days in the valley requires PM monitoring sites to accurately reflect valleywide PM levels (spatial coverage) and have adequate temporal continuity. PM monitors in close proximity to the northern and southern rims are omitted to avoid edge effects. Using these subjective criteria, 16 sites with relatively continuous measurements and adequate spatial representativeness were chosen to reflect regionwide fine particulate levels in the Central Valley.

The current National Ambient Air Quality Standards (NAAQS) impose a threshold on 24-h (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 with aerodynamic diameter $\leq 10$ $\mu$m ($PM_{10}$) levels. The NAAQS also impose a threshold of 15 $\mu g$ m$^{-3}$ for the annual average $PM_{2.5}$ concentration; no annual threshold exists for $PM_{10}$. Days with $PM_{2.5}$ levels exceeding the NAAQS threshold are termed ''exceedances.'' PM exceedances in central California occur almost exclusively in the winter during November–February (Watson and Chow 2002) and exceedances in the summer are considered atypical. During 1998–2008, the number of exceedances during November–February

was 680 as compared with 46 exceedances during June–September. The November–February period will be referred to as the PM season from now on.

Precipitation data for 1948–98 were obtained from National Climatic Data Center's hourly precipitation data (NCDC HPD). The hourly data are then used to calculate daily totals for 10 sites located along the valley floor (Fig. 1). Precipitation for 1998–2008 was obtained from CARB's Air Quality and Meteorological Information System (AQMIS). The CARB database contains meteorological data from a multiple agencies like the California Irrigation Management Information Systems (CIMIS), Remote Automatic Weather Stations (RAWS), and NCDC. Fourteen sites included in the analysis belonged to the CIMIS network. The NCEP–NCAR reanalysis-1 (NNR1) data were obtained online from the National Oceanic and Atmospheric Administration/Earth System Research Laboratory (NOAA/ESRL) Physical Sciences Division (http://www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml). The monthly atmospheric and oceanic time series were also downloaded from the NOAA/ESRL Internet site. The original source of the Niño-1+2, Niño-4, and Niño-3.4 SST indices was NOAA Climate Prediction Center (CPC). The monthly PDO values were obtained from the Joint Institute for the Study of Atmosphere and Ocean (http://jisao.washington.edu/pdo/PDO.latest). The monthly mean values for the indices were subsequently seasonally averaged as winter [December–February (DJF)], spring [March–May (MAM)], summer [June–August (JJA)], and autumn [September–November (SON)] averages.

### b. Methods—Theory and application

PM measurements are first utilized to identify days with high PM levels in the region. These days define a target group that provide a template for identifying the underlying atmospheric phenomena and help scan past climatic records for such occurrences. NNR1 geopotential heights (m) of the 500- and 1000-hPa pressure levels are utilized to sample the upper-air and surface patterns. Only a subset of the data belonging to a domain extending from 20°S to 75°N latitude and 100° to 350°E longitude are used.

The entire procedure can be broadly divided into two phases: feature extraction and classification. The feature extraction step involves identifying statistically significant features on geopotential height fields that characterize the target group. The significance test procedure uses the nonparametric bootstrapping resampling method described later (section 2c). The test is first performed grid point by grid point and regions that are statistically significant when compared with the nontarget group are identified. Because of the large number of comparisons

($31 \times 101$ in our case) and nonindependence of the tests, 2D wavelet transforms are then used for dimension reduction and obtaining independent metrics to gauge statistical significance (section 2d). Principal component analysis (PCA) is used to capture the correlations between significant features identified and to construct a multivariate distance metric for classification.

### c. Bootstrap significance testing

Developing a hypothesis testing procedure for a small sample size and a large number of atmospheric variables (geopotential height at every grid point for all pressure levels included) is convoluted because of multiple comparisons and nonindependence of gridpoint values. Generally, the target group sample mean (or median) is compared to the overall winter mean to determine grid points (regions) where one can reject the hypothesis supporting their equality at desired significance levels. A small sample size is not suited for comparison using two-sample $t$ test without making assumptions about normality of the target data. Also, since these comparisons are made at each grid point, the significance level of the hypothesis test procedure for an entire weather map is hard to define.

Grotjahn and Faure (2008) propose an empirical hypothesis testing methodology based on bootstrap resampling to identify statistically significant features on reanalysis weather maps. The testing involves comparing a test function of the state variable at a given gridpoint location with values of the test function obtained for bootstrap replicates constructed from days not belonging to the target group. For example, if geopotential heights were being tested for significance with sample mean as the test statistic, we would compare the sample mean to an empirical distribution derived from replicated blocks. If the target sample consists of $n$ days and $N$ represents the total number of days, $(N)$ days are subjected to random sampling with replacement to construct $n \times m$ windows with sample size equal to that of the target group ($n$ days). Here, $m$ controls the number of replicates being generated from the original data (typically, $m = 1000$). The empirical distribution of the replicated means is estimated. Departures from the overall mean are considered to be significant if the target group mean does not belong to an interval characterized by a significance level, $\alpha$. Thus, a coordinate failing the hypothesis test at a significance level of $\alpha = 0.01$ implies that the target group mean lies outside an interval defined by 0.5th and 99.5th percentiles as its lower and upper bounds, respectively.

The test can be easily extended to identify relevant features on weather composites using 2D wavelet coefficients instead of applying it directly to geopotential heights at the 500- and 1000-hPa levels. The target set is

decomposed using a discrete wavelet transform at a pre-determined set of frequency bands and the computed details (horizontal, vertical, and diagonal) are used as input to the hypothesis testing procedure. The new variables (coefficients) obtained are viewed as the result of applying a nonlinear transformation to the original reanalysis weather map data.

### d. Two-dimensional wavelet transforms

The goal of our classification scheme is to identify, locate, and measure the spatial extent of relevant synoptic weather patterns and associated anomalies. The information about location and spatial variability (frequency) are critical to the performance of the classification scheme. 2D wavelet transforms (WT) provide an excellent tool to overcome the limitations associated with Fourier transform (FT) in analyzing spatial data. Two-dimensional wavelet transforms have been used in improving upon point-to-point comparison methods to capture spatial dependencies in climatic data. Briggs and Levine (1997) explore the use of WT for developing multivariate field forecast verification measures using geopotential height fields. Casati et al. (2004) use WT for the verification of spatial precipitation forecasts for a short-range numerical weather prediction system. Livina et al. (2008) developed a wavelet-coefficient score to compare two-dimensional air temperature and specific humidity fields in reanalysis and modeled datasets. Wang and Lu (2010) provide an excellent demonstration of two-dimensional continuous wavelet transform (2D CWT) to meteorological data obtained from numerical model simulations.

WT provides space–frequency localization at predetermined bands of frequencies crucial for identifying patterns varying in spatial location and gradients. For the geopotential height data, strong pressure gradients with closely spaced isopleths translate to high-frequency (low scale) information and relatively weak patterns with widely spaced contours are captured at lower-frequency scales. The details of an atmospheric pattern are contained at multiple levels (frequency–scale) of wavelet coefficients obtained after decomposition. In addition to the space–frequency characterization, the wavelet coefficients are utilized to remove low-scale variability not relevant on a global scale. The independence of wavelet coefficients across scales also enables us to divide the task of analyzing voluminous weather maps into smaller tractable subtasks and later aggregate the results.

WT achieves resolution in both space and frequency by dilating and translating a finite wavelet function (basis) at different frequency ranges (Addison 2002). The main wavelet function is referred to as the mother wavelet and is used to obtain wavelet basis functions. Since WT is not unique, there are different types of mother wavelet functions that can be used to obtain a WT. Basis functions are expressed as

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

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 1D continuous wavelet transform (CWT) is defined as

$$T(a,b) = \int_{-\infty}^{\infty} f(x)\psi_{a,b}^{*}(x). \qquad (2)$$

The asterisk indicates the complex conjugate function of the original mother wavelet function. The convolution can be viewed as a cross correlation between the signal and derived basis functions at various scales. Discrete wavelet transform (DWT) is an abbreviated version of the CWT designed to remove the redundancy by limiting the number of wavelet decomposition levels using a dyadic scale; that is, $a = 2^{j}$ and $b = k2^{j}$ where $j$ and $k$ are positive integers:

$$\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k) \quad \text{and} \qquad (3)$$

$$\mathrm{DWT}(j,k) = \int_{-\infty}^{\infty} f(x)\psi_{j,k}^{*}(x), \qquad (4)$$

where $f(x)$ now represents a discrete function.

The NNR1 data are three-dimensional with two-dimensional meteorological fields available at 17 different pressure levels. The topography of geopotential height surfaces at 500- and 1000-hPa levels is characterized in the present study. Thus, WT concepts defined for one-dimensional signals are extended to include the two-dimensional pressure-level surface profile:

$$T(a,\mathbf{b}) = \frac{1}{a}\int_{-\infty}^{\infty} \psi\left(\frac{\mathbf{x} - \mathbf{b}}{a}\right)f(\mathbf{x})\,d\mathbf{x}, \qquad (5)$$

where $\mathbf{b}$ is the coordinate vector $(b_1, b_2)$ and $\mathbf{x}$ pertains to the zonal and meridional coordinates, respectively.

The rows and columns of the data matrix are treated as independent 1D signals. Alternating high- and low-pass filters (Fig. 2) are applied in the $x$ direction (zonal), and the results are downsampled by deleting every other column (dyadic up–downsampling). This results in two matrices of approximately half the size of the original: one containing high-frequency components of the rows and the second containing the low-frequency contributions. These matrices are then filtered down along the columns, that is, $y$ or meridional direction, using high- and low-pass
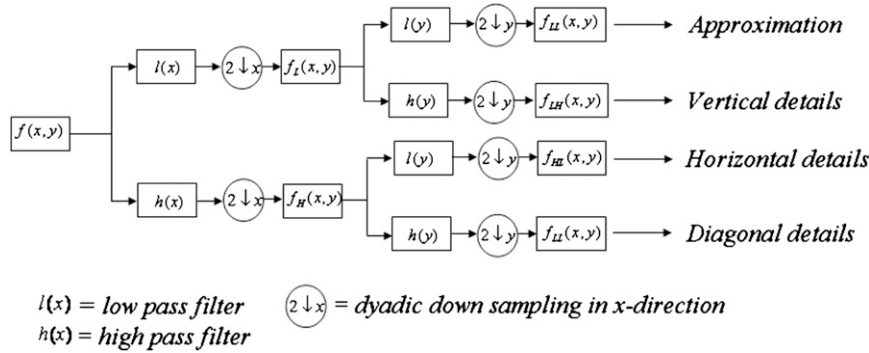
FIG. 2. A block diagram showing the chronology of wavelet decomposition of 2D data at a single level.

filters and subsequently downsampling the resulting matrices along rows (deleting every other row). The resulting four matrices are approximately one-fourth the size of the original image and represent the *smoothed approximation* and the *horizontal*, *vertical*, and *diagonal details*, respectively.

The detail coefficients obtained at lower levels of decomposition are high-frequency contributions representing small-spatial-scale variability and noise. Coefficients at higher levels of decomposition capture important features pertaining to varying spatial scales, anywhere between regional–mesoscale and global-scale phenomena. A summation of the detailed coefficients can be used to reconstruct the total detailed information at corresponding levels of decomposition.

The horizontal, vertical, and diagonal detail coefficients obtained by decomposing the target days are compared to the nontarget days using the bootstrap-resampling hypothesis testing method. A high level of significance ($\alpha = 0.01$) ensures statistical strength when the test is applied to multiple wavelet coefficients across multiple levels. Coefficients passing the test are retained and the rest discarded resulting in a considerable dimension reduction. The significant coefficients (capturing feature information) are concatenated into a single multivariate data vector at every level of decomposition (scale). The information captured in the detailed coefficients is mutually exclusive across scales, thus dividing the original 2D geopotential height field into different multivariate data vectors at each scale. Thus problem of comparing and classifying geopotential heights is effectively split into $s$ different problems, where $s$ pertains to levels of decomposition with distinct spectral variability.

### e. Principal component analysis

Gridded climatic data present copious amounts of spatial information and a large number of variables makes it difficult to use conventional statistical methods when the target sample size is small. A large number of variables obscures the underlying relationships that exist in the data ($X$) and makes it difficult to draw definitive conclusions based on any single variable. The univariate testing procedure establishes the significance of individual features on the weather map but fails to capture the underlying correlation structure (teleconnections) between these patterns. This structure can be explored using PCA, also known as empirical orthogonal functions (EOF) in the oceanic–atmospheric sciences literature (North and Wu 2001; Hannachi et al. 2007; Navarra and Simoncini 2010). PCA (EOF) analysis finds linear combinations of the original variables that explain variability contained in the covariance (correlation) matrix:

$$Y_i = \mathbf{e}_i' X = \mathbf{e}_{i1} X_1 + \mathbf{e}_{i2} X_2 + \cdots + \mathbf{e}_{ip} X_p, \quad (6)$$

where $Y_i$ is the $i$th principal component score and $\mathbf{e}_i$ is called the loading vector.

PCA also serves as a dimension reduction technique, which summarizes the underlying variability in the data by a relatively few unobservable variables (Johnson and Wichern 2002). These latent variables are combinations of original variables that maximize the variance explained in the particular direction and orthonormal to the others. The subspace containing these latent structures is defined by the chosen eigenvectors corresponding to relatively few significantly large eigenvalues of the covariance matrix. The original data can be written as a product of a scores matrix $\mathbf{Y}$ and a principal component loadings matrix $\mathbf{P}^{\mathrm{T}}$, consisting of selected right eigenvectors of the covariance matrix. The errors ($\mathbf{Q}$) in approximating the original data using a truncated principal components model are useful in quantifying the deviations of measurements from the modeled subspace:

$$\mathbf{X} = \mathbf{Y} \times \mathbf{P}' + \mathbf{Q}. \quad (7)$$

The $T^2$ statistic, defined as

$$T^2 = \mathbf{Y}'\mathbf{\Lambda}\mathbf{Y} = \sum_1^r \frac{Y_i^2}{\lambda_i}, \qquad (8)$$

is a nonlinear transformation that characterizes the variability of the target group in the modeled subspace. In this paper, PCA is performed on the wavelet coefficients resulting from multilevel decomposition of target geopotential height fields to derive a threshold. Bootstrap resampling is performed to obtain the average estimated 99th percentile for the $T^2$ statistic providing an empirical upper limit for similar fields.

## 3. Central California PM case study

$PM_{2.5}$ exceedances typically occur during the winter season in California's Central Valley. Highest PM levels are observed on days with strong anticyclonic influence resulting in stagnant conditions and reduced vertical mixing due to subsidence. Interestingly, the valley receives almost all of its annual rainfall during these winter months. Days with stormy conditions (induced by midlatitude cyclones) are characterized by high wind speeds and precipitation, which lead to increased vertical mixing and wet deposition of atmospheric particulates. The number of $PM_{2.5}$ exceedances in any given year then depends upon the relative frequency with which these two opposing processes occur during the core PM season.

$PM_{2.5}$ data are not suited for studying effects of climate variability on air pollution on interannual or decadal scales because of the lack of measurements prior to the 1998/99 winter season. An irregular sampling schedule (1, 3, and 6 day) of 24-h average $PM_{2.5}$ levels during early years of operation makes it difficult to identify all the exceedances occurring at a given site on a given day. In addition, local phenomena induce heterogeneity in weather conditions experienced at different locations in the valley. Thus, an exceedance at Fresno in the San Joaquin Valley may not be representative of conditions at sites in the Sacramento Valley. The 16 monitoring sites with good temporal continuity of data are chosen and shown in Fig. 1.

### a. Exceedance selection

Under strongly anticyclonic conditions, the study region experiences $PM_{2.5}$ exceedances at almost all the sites. To identify widespread PM exceedances and avoid local phenomena, the chosen sites are divided into three regional clusters belonging to Sacramento Valley (SV), northern San Joaquin Valley (NSJV), and central San Joaquin Valley (CSJV). We intend the selected subregions to capture regional dynamics of PM emanating from Sacramento, Stockton, and Fresno, California, metropolitan

areas, respectively. The available PM measurements from all sites belonging to a subregion are averaged to obtain a measure of the PM levels in the region. The averages calculated may not involve measurements from every site in the subregion because of the irregular sampling schedule. Days with no measurements in the entire subregion are omitted to avoid ambiguity and subjectivity. Imputation of missing values is not performed to avoid unintentionally biasing the results. Since almost all PM exceedances occur during the winter months (November–February), exceedances occurring outside this period are considered atypical of recurring meteorological scenarios contributing to high PM levels in the region. The days with averaged $PM_{2.5}$ levels within all three subregions exceeding the NAAQS threshold are chosen to represent the target sample. A total of 48 days are identified (from 1998/99 through 2008/09) using the aforementioned criterion to represent poor PM air quality conditions in the study domain. These days are referred to as "target days" in subsequent text.

The selection criterion aims to minimize the influence of measurement uncertainties, local emissions, and microscale meteorological effects that do not represent domain-wide conditions. For instance, there are 141 days where regionally averaged $PM_{2.5}$ levels exceed 35 $\mu g$ m$^{-3}$ in NSJV but only 48 days belonging to this set result in averaged PM levels exceeding the threshold in both NSJV and SV. Exclusive exceedances, occurring in one or two subregions, may or may not have large-scale weather signatures central to the current theme. The identified target samples do not represent the entire target population during the chosen 11 winter PM seasons because of the constraints imposed by data sparsity. The target days are characterized by calm conditions in the valley with a land breeze exiting into the Pacific Ocean through the gap in the Coast Range. Averaged surface winds at 1100 and 0000 Pacific standard time (PST) overnight temperature anomalies (in standard deviations) are shown in Fig. 3. A weak surface pressure gradient over central California results in calm conditions typical of anticyclonic blocking. Low overnight surface temperatures indicate that a low-level inversion is also likely to develop because of nocturnal cooling under clear skies. Reduced vertical mixing traps local emissions close to the valley surface resulting in PM buildup if such conditions persist for long periods of time (3–5 days). No precipitation is recorded at any of the selected sites on days belonging to the target sample.

### b. Feature extraction

The 500-hPa pressure level is assumed to represent aloft conditions capturing synoptic weather systems affecting central California. However, to ensure that the days identified capture surface conditions, we use 1000-hPa
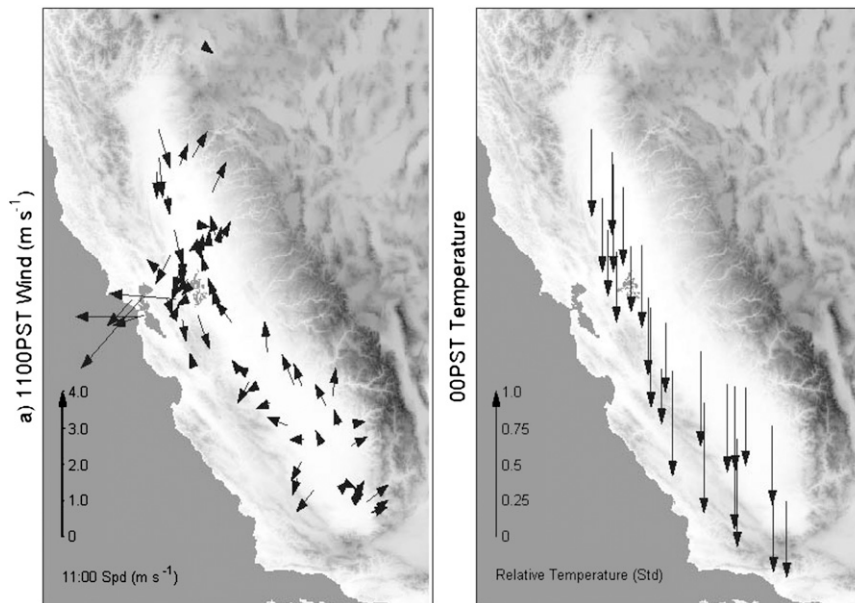
FIG. 3. Meteorological conditions under target samples characterized by (left) averaged wind flow patterns at 67 locations at 1100 PST and (right) standardized overnight temperatures (defined as number of standard deviations relative to the overall hourly mean temperatures at specific sites) at 24 sites for 0000 PST. Downward-pointing arrows indicate negative deviations from average values.

geopotential height weather maps and perform the tests independently at two pressure levels. Any event on a geopotential height field is uniquely characterized by characteristics such as varying spatial location, size, and associated gradients. The bootstrap test procedure is used to compare geopotential heights between the target sample and the nontarget winter days at each grid point on 500- and 1000-hPa levels. The statistically significant coordinates at $\alpha = 0.05$ level are shown in Fig. 4 (20°S–75°N latitude and 100°–350°E longitude). The significant regions represent events coexisting with local conditions in the valley.

The geopotential height composites for the target days at 500- and 1000-hPa levels are shown in Fig. 5. The 500-hPa composite shows a strong ridge along the west coast of United States and a trough extending over Canada and into the eastern United States. Less prominent features include what appears to be a short-wave trough north of the Aleutian Islands over the Bering Sea, a trough–low pressure pattern northwest of Japan, and a ridge off the northwest coast of Africa. A nexus between the aloft on-shore ridge and PM$_{2.5}$ exceedances in the San Francisco Bay Area has been identified in a previous study (Beaver et al. 2010). The 1000-hPa composite map shows the North Pacific high, a relatively strong Great Basin high, and a strong low over the Aleutian Islands. It should be noted again that some of the nontarget days are nonidentifiable exceedance days because of data sparsity and their identification would further strengthen the hypothesis test

results. The features identified are likely to be more prevalent than they might appear during the 11-yr period.

The geopotential height composites may delineate important synoptic weather events but are not readily amenable to 2D wavelet analysis. To enhance the efficacy of the procedure in separating the target group from the mean state, we use the long-term daily anomaly maps instead. Shown in Fig. 6, these maps offer more vivid contrasts between the target and nontarget days. The anomalies accentuate the less obvious trough east of Japan and relatively higher pressure in the central North Pacific at the 500-hPa level. Anomaly maps also differentiate statistically significant features from semipermanent patterns. Significant anomalies at concurrent geographical coordinates indicate increased intensity of such semipermanent features. The Aleutian low, observed at the 1000-hPa level, is considerably deeper during high-PM target days and is indicated by negative anomalies centered slightly north of the islands.

The study domain for 500-hPa geopotential height fields (encompassing 20°S–75°N latitude and 100°–350°E longitude) presents a large number of variables (i.e., a $37 \times 104$ matrix or 3848 points). The information in Fig. 6 contains synoptic-scale patterns that can be summarized by relatively few parameters. 2D wavelet decomposition was performed using Daubechies-3 (db3) wavelet function at five levels of decomposition. The selection of the wavelet function and the decomposition levels was
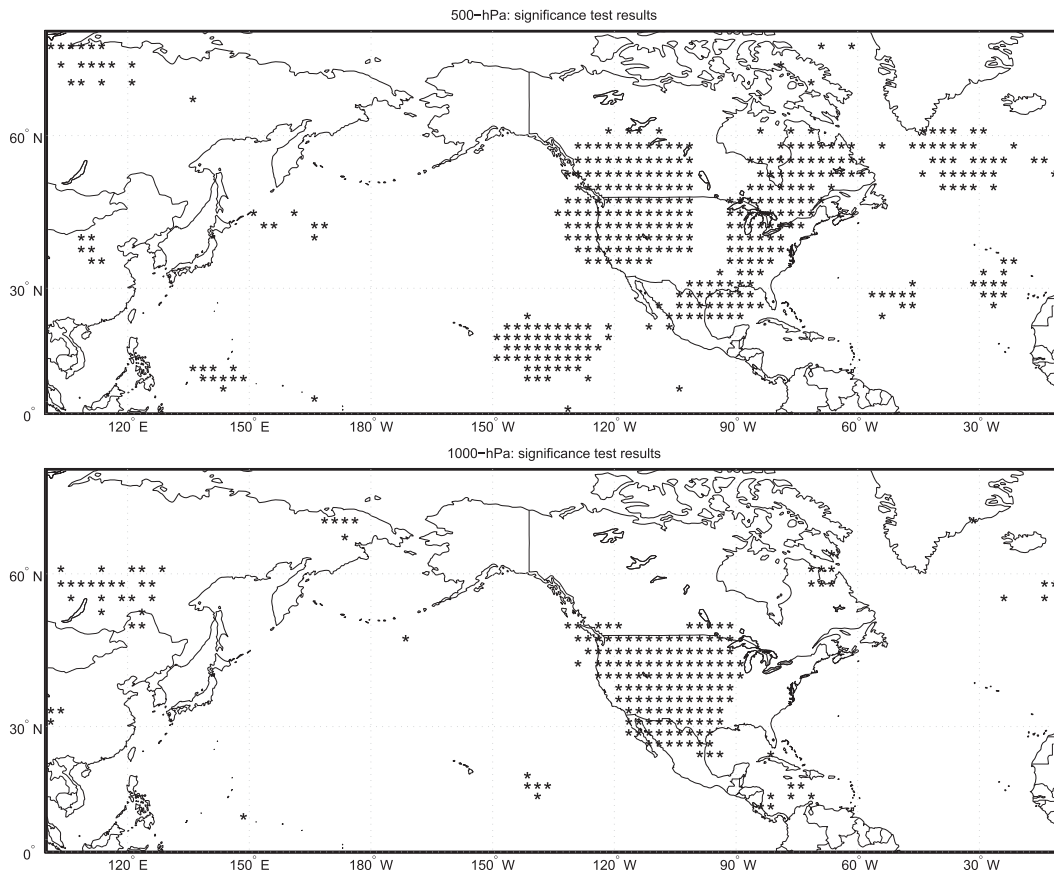
FIG. 4. Geographical coordinates at which geopotential heights at (top) 500- and (bottom) 1000-hPa levels for the target group of high PM levels are significantly different from the mean levels of nontarget days. The coordinates passing the test are shown as black asterisks.

performed by maximizing the log-energy entropy across 20 wavelet functions derived from Daubechies, Symmlet, and Coiflet families (Sun et al. 2009) and minimizing across scales. The high-frequency information contained in the anomaly maps is limited because of the 2.5° × 2.5° grid resolution. Thus, low zero crossings and relatively few scales for decomposition suffice for the current analysis.

Following the procedure outlined in Fig. 2, the first step reduces the original (43 × 104) grid to four (24 × 54) matrices. The vertical, horizontal, and diagonal details extracted in this step mostly contain high-frequency information or noise, which is usually filtered out. Subsequent decompositions yield 14 × 29, 9 × 17, 7 × 11, and 6 × 8 matrices summarizing frequency at decreasing frequency levels. All coefficients obtained at scales 2–5 are then subject to the resampling test to identify statistically significant coefficients at $\alpha = 0.01$ significance level. This procedure is analogous to thresholding used in denoising and filtering. Coefficients that are not significant are set to zero, and nonzero coefficients at each

level are used to reconstruct the details at scales 2–5 (Figs. 7 and 8). It is crucial to verify if the decomposition captures relevant information at each scale. Scale 2 captures relatively small regions of positive and negative anomalies centered around locations corresponding to the features identified on the anomaly composites (Fig. 6). Reconstruction at subsequent scales 3–5 show synoptic features characterized at decreasing levels of frequency. The nonzero region over western United States at scales 2, 3, and 4 corresponds to positive anomalies associated with a high pressure ridge observed in the geopotential height composites. Reconstructed details indicate shape and gradients associated with an event spread across multiple scales. The total number of statistically significant detail coefficients (horizontal, vertical, and diagonal) required to summarize the presented information is shown in Table 1.

The significant horizontal, vertical, and diagonal detail coefficients are concatenated to form a variable vector amenable to multivariate analysis. For a target set of 48 days, we would have a 48 × 6 data matrix for scale = 3

a) 500−hPa Geopotential Height Composite Map

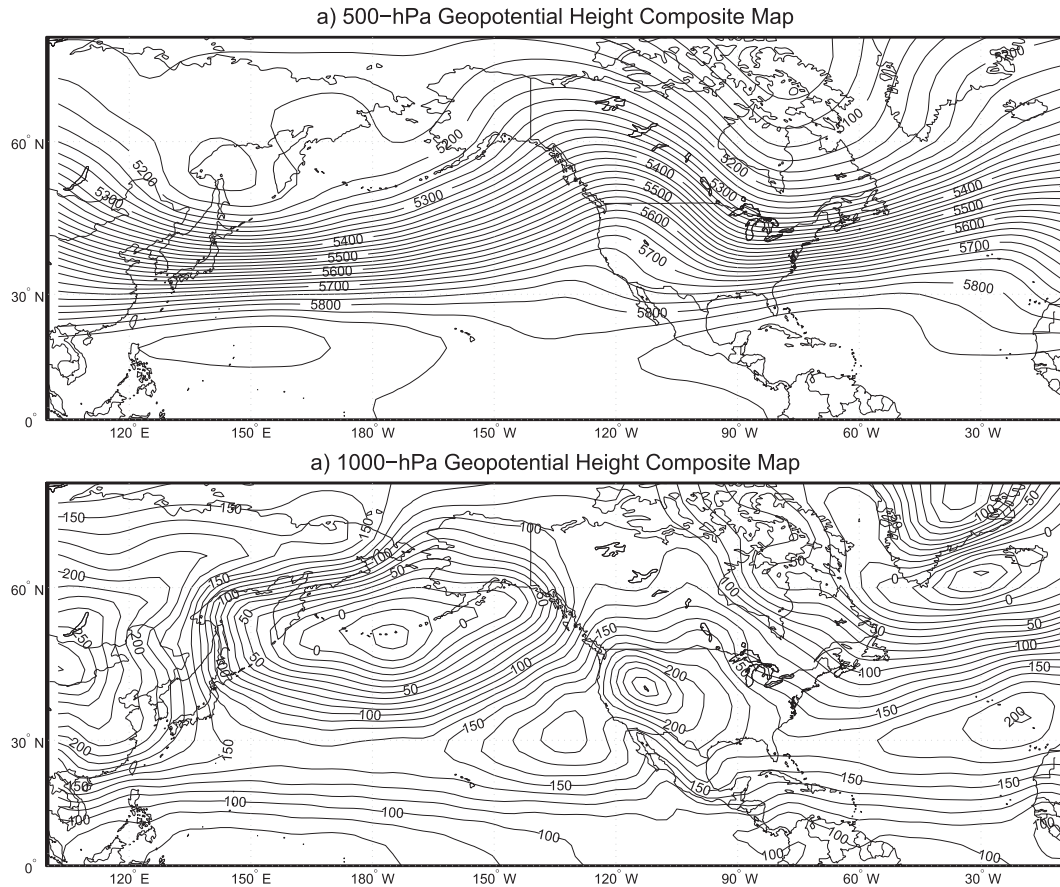a) 1000−hPa Geopotential Height Composite Map

FIG. 5. A composite map showing the geopotential height fields at (top) 500- and (bottom) 1000-hPa levels for the target group. Contours are displayed every 25 m and labeled at a 100-m interval for 500-hPa height fields and displayed every 50 m and labeled every 12.5 m for the 1000-hPa composites.

at the 500-hPa level 1. The chosen coefficients for target days are subject to PCA (EOF) analysis to capture the spatial correlation structure between detail coefficients corresponding to teleconnection patterns associated with significant events at any given scale. Reduced-order PCA models, accounting for 99% of the original variability, are estimated and used in subsequent analysis to avoid redundancy in data and ill conditioning of the covariance matrix. The $T^2$ statistic is calculated for the retained principal components at each scale and a bootstrap test is used to estimate the 99th percentiles. This statistic serves as a threshold during classification. The daily averaged geopotential height fields are decomposed using WT and projected coefficients are then used to determine if the field can be classified as an exceedance-conducive (EC) day. The 99th percentile ensures a relatively relaxed threshold to avoid overfitting.

*c. Classification results*

The constructed template is then used to identify occurrences with similar patterns during an extended period, 1951–2008. The test procedure decomposes every day of the annual PM season during the extended period using the same parameterization to identify days with similar (not identical) occurrences in the past. 2D wavelet decomposition is performed using the target group parameters and a check for similarity is performed at scales 2–5 for the chosen pressure levels using the $T^2$ statistic. Days passing the test at scales 2–5 and both pressure levels are accepted as matches. The algorithm is designed toward identifying relative patterns in the form of pressure anomalies pertaining to the shape and gradients of a weather event. Minor inconsistencies in magnitude are not expected to result in a serious error if the overall pattern associated with the feature is well represented.

Figure 9 shows the number of matches found every winter PM season (November–February) from 1951/52 to 2007/08. The number of EC days is highest (colored white) for the 1976/77 season with 48 matches and lowest (colored black) for the 1955/56 and 1982/83 PM seasons with 7 and 9 days, respectively. The bar plot

b) 500−hPa Geopotential Height Anomaly Composite Map



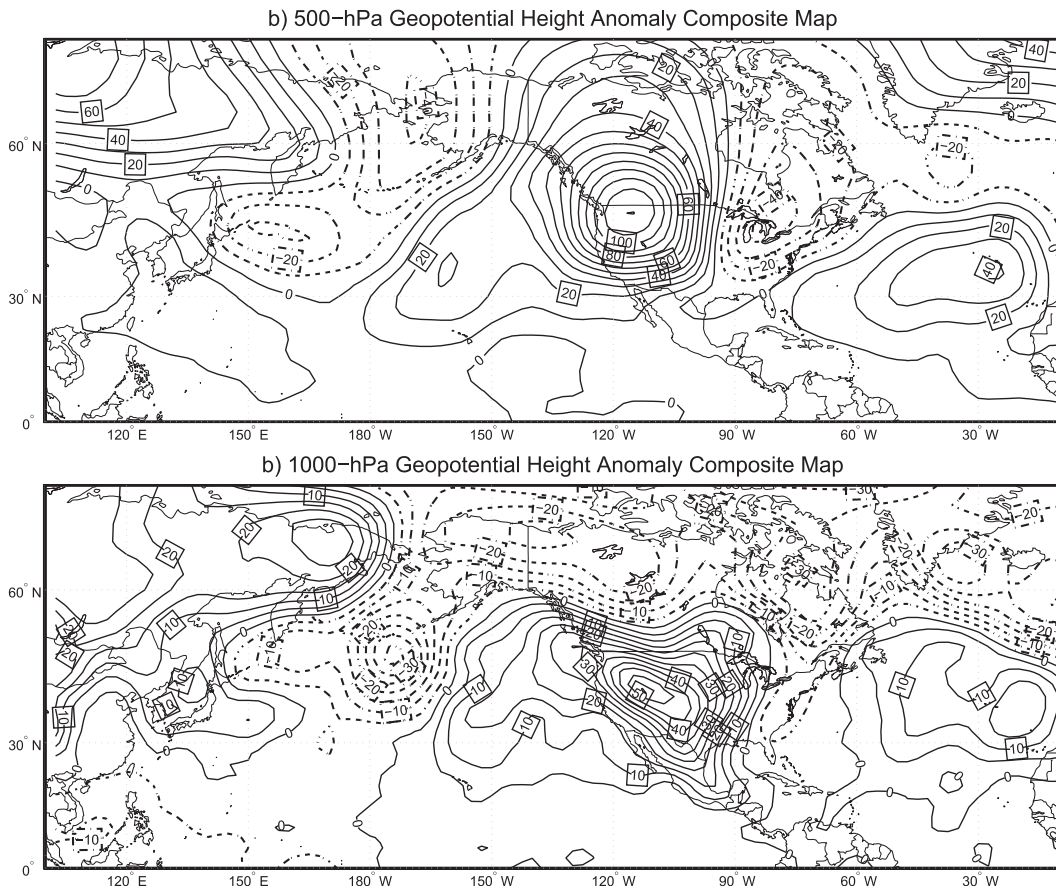b) 1000−hPa Geopotential Height Anomaly Composite Map



FIG. 6. A composite map showing the geopotential height anomaly composites at (top) 500- and (bottom) 1000-hPa levels for the target group. Contours are displayed every 10 m and labeled at a 20-m interval for 500-hPa anomaly composites and displayed every 5 m and labeled every 10 m for 1000-hPa composites. Contours representing negative anomalies are displayed as dash–dotted lines.

shows considerable interannual-scale variability in EC days count for the study period. The grayscale chosen to represent the bars indicates a color contrast before and after the 1982/83 season. Prior to 1982/83, the frequency of low-EC-days seasons appears higher, indicating variability at an interdecadal time scale.

Figure 10 shows composite maps of the geopotential height anomalies for all EC days found during the extended study period. The anomalies at 500-hPa level capture the shapes and locations of features contained in the target set very accurately (see Fig. 6). The anomalies are lower in magnitude because of a higher threshold set to avoid overfitting. The smooth low pressure anomaly east of Japan in Fig. 10 is reduced to a kink in the low pressure anomaly pattern over the Aleutian Islands. Similar results hold for the 1000-hPa level anomaly composite where the features are well represented but with lower magnitudes. The geopotential height composites for the identified EC days are almost identical at the 500-hPa level (Fig. 11). At the 1000-hPa level, the low pressure

center appears to be shifted slightly eastward over the Aleutians. A tighter threshold would yield a lower number of matches and retain the exact location of this low pressure pattern but possibly lead to overfitting. The procedure does reasonably well at identifying EC days during the study period capturing previously identified features with consistency.

The difference in $PM_{2.5}$ levels during EC days versus non-EC days can only be observed during 1998–2008 on days when the PM concentrations are sampled. Figure 12 compares the $PM_{2.5}$ levels recorded at Roseville, Modesto, and Clovis, California. These sites are displayed because of their proximity to the major urban centers like Sacramento, Stockton, and Fresno. The urban sites are avoided to discount the influence of localized emissions and microscale events exhibiting high PM during atypical conditions. Median $PM_{2.5}$ levels are approximately 15–20 $\mu g$ m$^{-3}$ higher for EC days at the sites shown. The difference between mean $PM_{2.5}$ levels is statistically significant ($p <$ 0.0001) at all monitoring sites included in the study. EC
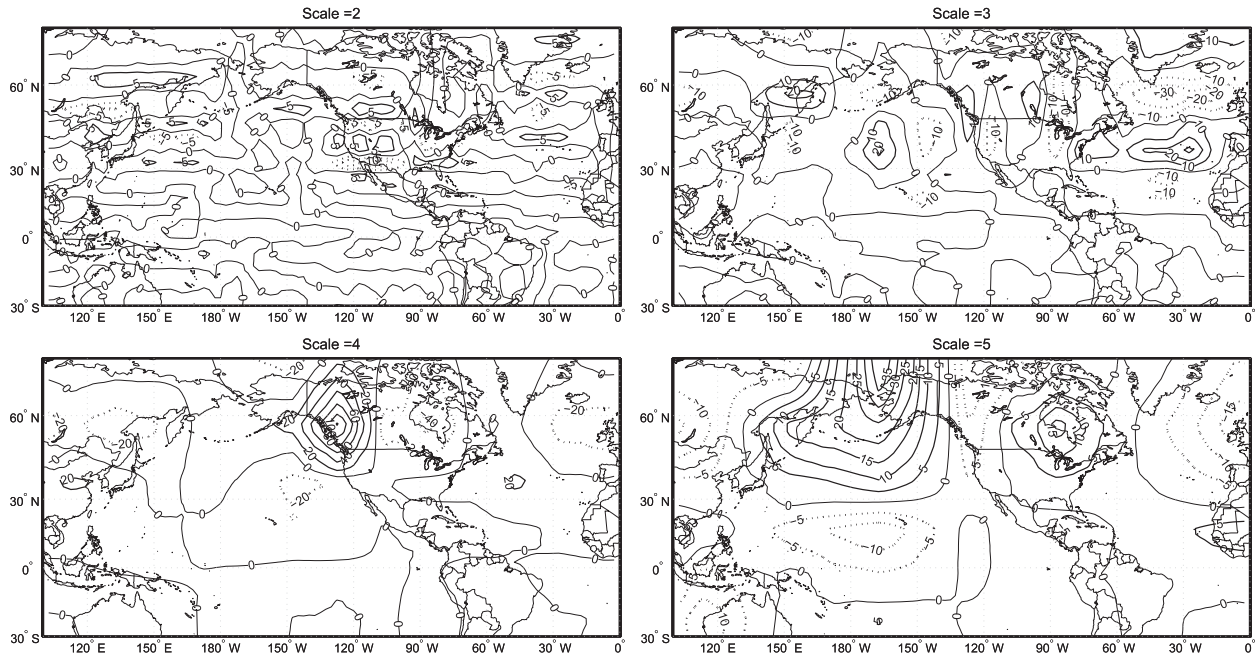
FIG. 7. Reconstructed details using the sum of horizontal, vertical, and diagonal detail coefficients obtained at scales 2–5 for the 500-hPa geopotential height anomaly maps. Contours representing negative anomalies are displayed as dash–dotted lines.

days represent favorable meteorological conditions but do not guarantee an exceedance. EC days were identified during the period 1998–2008, and 148 days were exceedances at one or more sites in the study domain. We do observe a few EC days with fairly low PM$_{2.5}$ levels. It should be noted from Fig. 12 that the highest PM$_{2.5}$ levels almost exclusively occur on days belonging to the EC

days category validating the performance of the proposed method.

Daily precipitation records extending back to 1951 are used to investigate any influence of precipitation frequency on the PM exceedance potential in the Central Valley. A "rain" day is defined when the averaged daily precipitation exceeds 0.05 inches at two or more subregions (SSV,
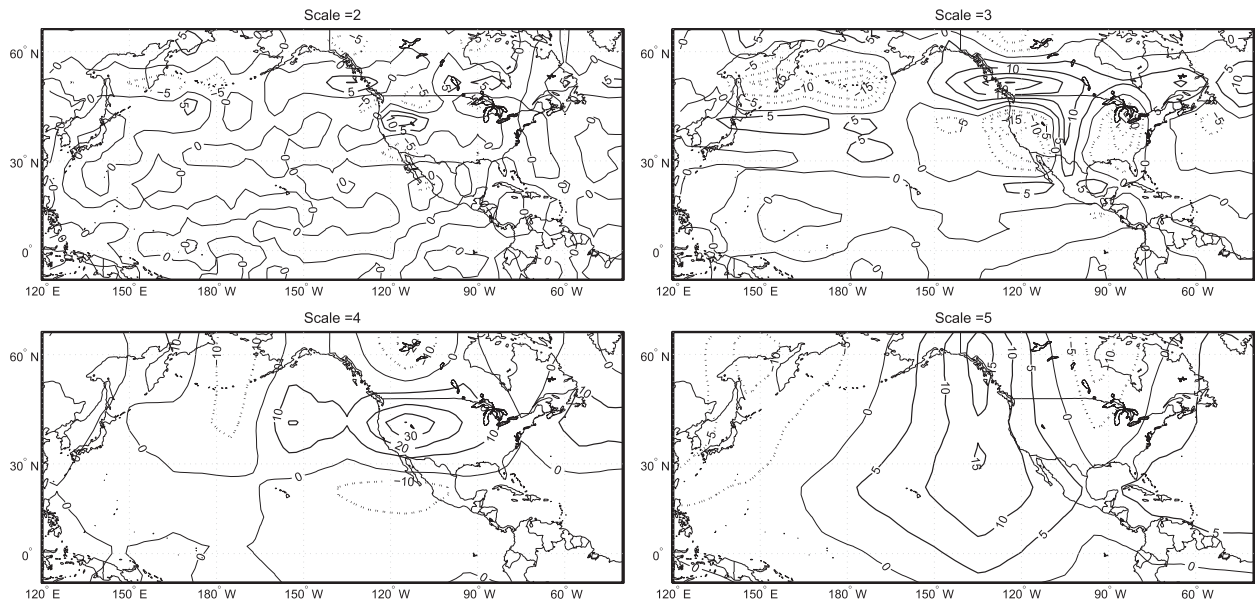


FIG. 8. As in Fig. 7 but for the 1000-hPa geopotential height anomaly maps.

TABLE 1. The number of statistically significant horizontal, vertical, and diagonal detail coefficients obtained at scales 1–5 at the 500- and 1000-hPa level.

| Scale | 500-hPa level | 1000-hPa level |
|-------|---------------|----------------|
| 1 | 44 | 51 |
| 2 | 18 | 16 |
| 3 | 6 | 10 |
| 4 | 8 | 7 |
| 5 | 5 | 3 |

NSJV, and CSJV) constituting the study domain. Reasonable precipitation results in wet deposition and vertical mixing and disqualifies the day from being conducive to a widespread PM exceedance. A straightforward test for correlation indicates significant anticorrelation between EC days and the number of rain days per season ($\rho = -0.4224$). A linear regression model is fit (shown in Fig. 13) and an $F$ test confirms a linear relation with a $p$ value = 0.001. Figure 13 is analogous to Fig. 9b in Leibensperger et al. (2008), where the authors compare the number of ozone pollution days with frequency of midlatitude cyclones for individual years in the 1980–2006 period. Frequency of days with poor air quality (ozone and PM) appears to be deeply connected to the frequency of stormy conditions in the specified region.

### d. Large-scale ocean–atmosphere connection

ENSO events characterized by anomalous heating/cooling of surface waters in the eastern/central Pacific Ocean have a profound influence on winter weather patterns in the western United States (Schonher and Nicholson 1989; Dettinger et al. 1998; Cayan et al. 1998, 1999; Schubert et al. 2008; Praskievicz and Chang 2009; Zhang et al. 2010). The ENSO variability has strong climatic responses at remote locations on the globe because of its associated teleconnections patterns and explains a high proportion of climate variability on the subdecadal scale (Yeh et al. 2009; Cayan et al. 1998).
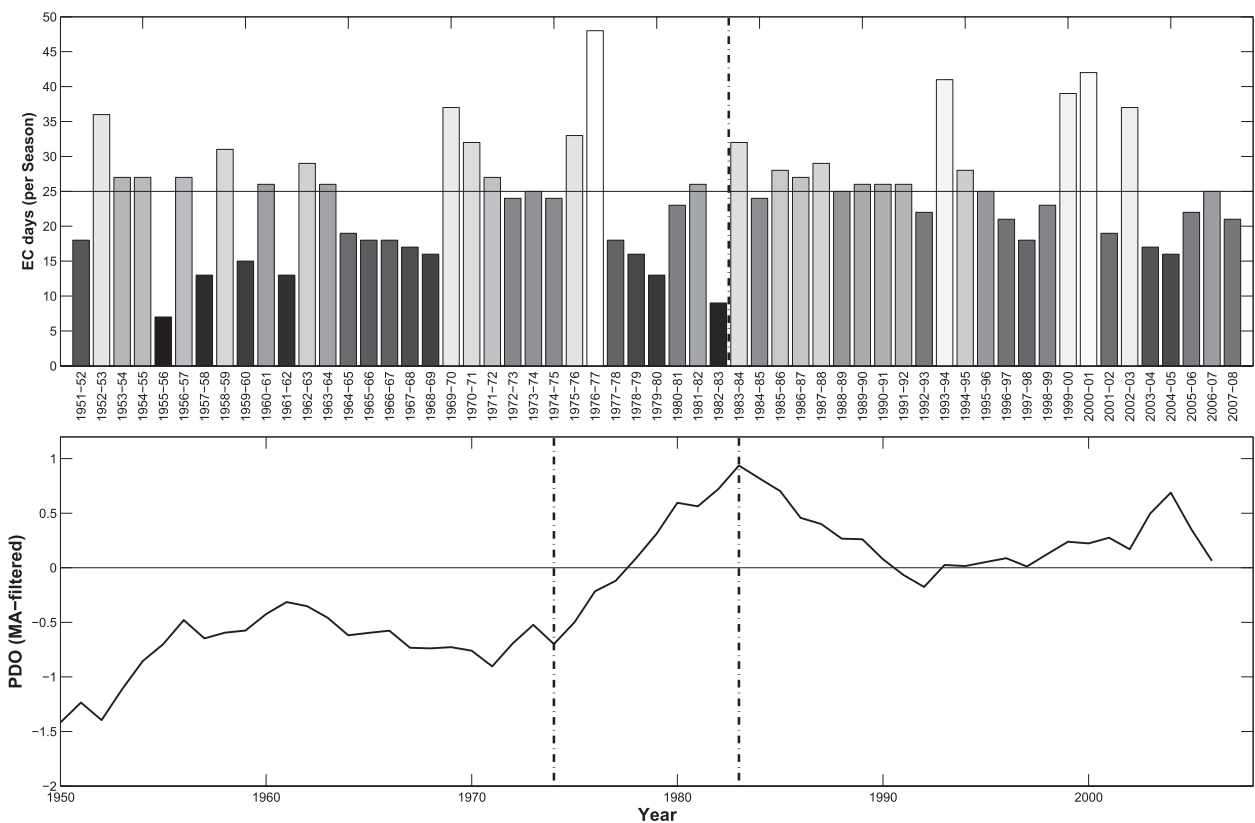


FIG. 9. (top) Bar plot showing the frequency of identified EC days every winter season during the study period, 1951–2008. The bars are colored in grayscale with black assigned to the lowest EC-day count and white assigned to the highest. The horizontal line indicates the median (=25) of EC days identified during each season. (bottom) Seasonally averaged winter (DJF) PDO index (Mantua et al. 1997) smoothed using a moving average filter with a period =10 yr shown from 1951 to 2008. The vertical line in the top panel corresponds to the year 1983 and the two lines in the bottom panel correspond to 1976 and 1983, respectively. The year 1976 marks a shift in the PDO from being largely negative to positive phase during the study period (Mantua and Hare 2002).

## a) 500−hPa Geopotential Height Composite Map
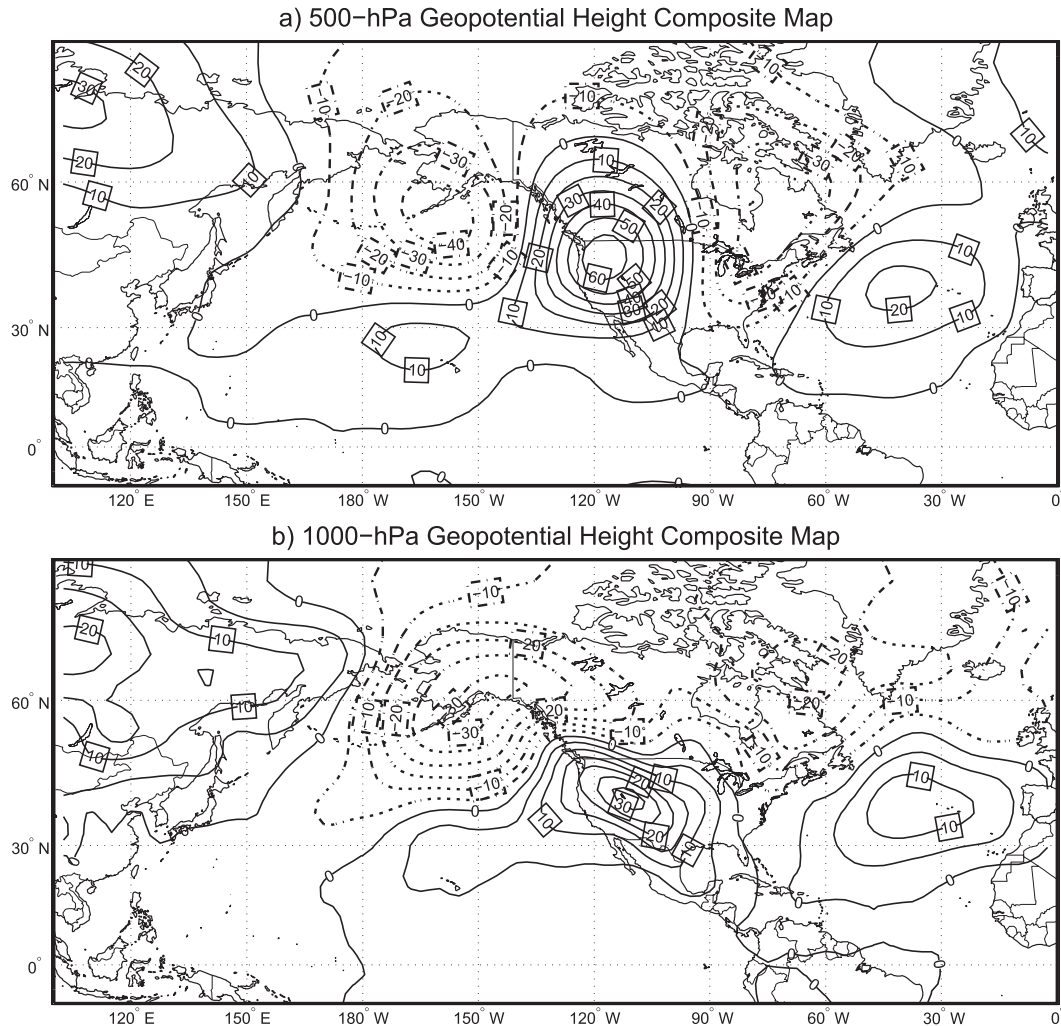
## b) 1000−hPa Geopotential Height Composite Map

FIG. 10. Geopotential height anomaly composites at the (top) 500- and (bottom) 1000-hPa levels for all matches returned by the classification algorithm during the period of 1951–2008. Contours are displayed every 5 m and labeled at a 10-m interval. Contours representing negative anomalies are displayed as dash–dotted lines.

Traditionally, canonical El Niño (warm ENSO event) was characterized by warm anomalous sea surface temperatures off the coast of Peru. Recent El Niño studies identify a more frequent pattern called the central Pacific (CP) El Niño [also termed as El Niño Modoki (Ashok and Yamagata 2009) or date line El Niño (Larkin and Harrison 2005)]. 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 relative to east Pacific El Niño events (referred to as EP–El Niño) (Yeh et al. 2009). Using the classification of Yeh et al. (2009), we define EP–El Niño, CP–El Niño, and La Niña years (seasons) and investigate the frequency of EC days during these winters.

Figure 14 shows relative frequencies of EC days against the rain days during a PM season under EP–El Niño (Fig.

14, top panel), CP–El Niño (Fig. 14, center), and La Niña (Fig. 14, bottom panel) events. The EC days response is inconsistent across EP–El Niño events but an anticorrelation between EC days and rain days is clearly delineated. The 1982/83 and 1997/98 EP–El Niño events are characterized by strong SST anomalies and have a relatively high frequency of rain days and a low exceedance potential. The 1976/77 PM season has the highest EC days identified during the entire study period. It also has the lowest rain days when compared with all winter seasons considered. The Department of Water Resources, California, declared the 1976/77 period as the driest in the modern era (Peters et al. 1976). The Central Valley PM response to CP–El Niño is relatively weak (Fig. 14, center) and hard to determine given the few events during the study period. The year 2002 has a high

## a) 500–hPa Geopotential Height Composite Map



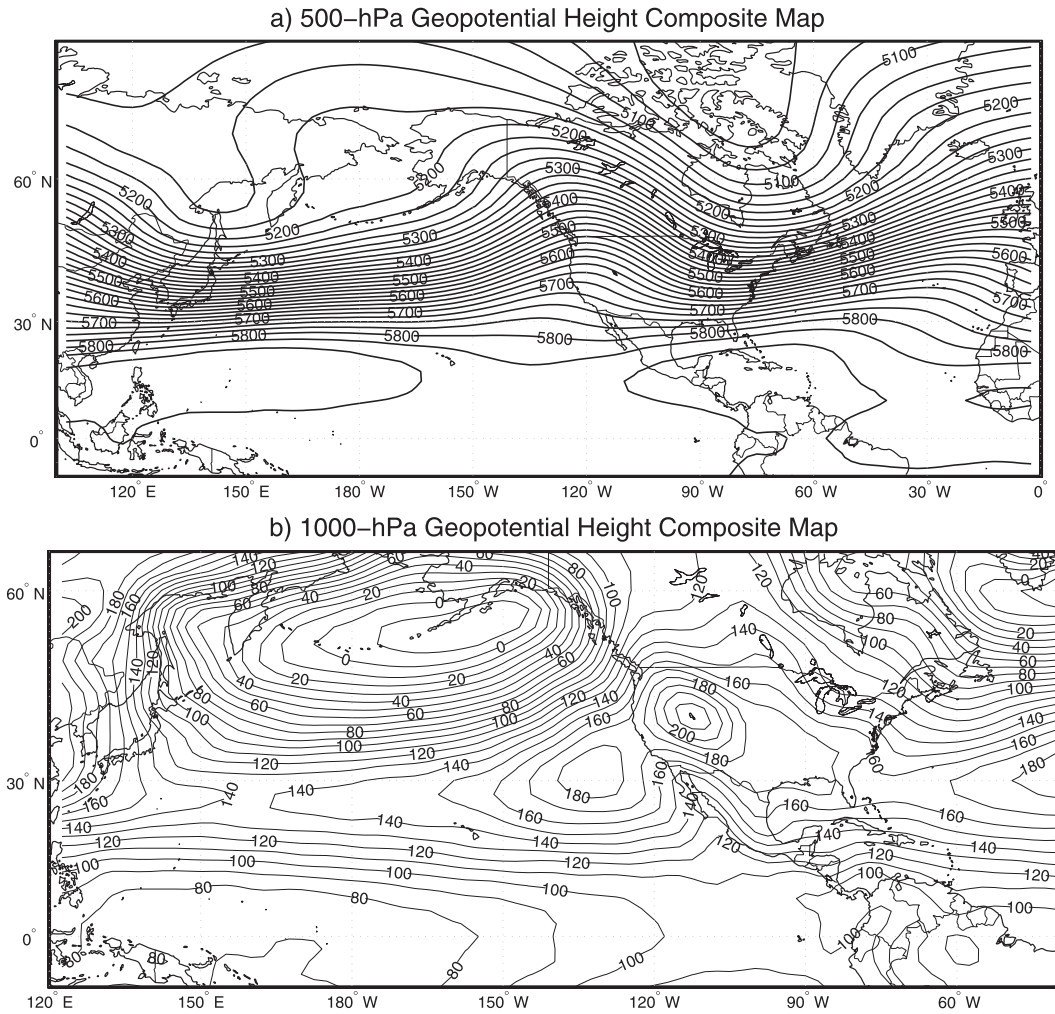## b) 1000–hPa Geopotential Height Composite Map



FIG. 11. As in Fig. 10, but for geopotential height. Contours are displayed every 25 m and labeled at a 100-m interval at the 500-hPa level and displayed every 10 m and labeled at 20-m intervals at the 1000-hPa level.

incidence of EC days but Larkin and Harrison (2005) classify 2002 as a conventional El Niño and not a date line–type event. Figure 14 (bottom panel) indicates the lack of consistent EC days response during La Niña

winters. The 1955/56 event has the lowest incidence of EC days during the study period.

Figure 9 compares seasonal EC days frequency (Fig. 9, top) with smoothed value of winter (DJF) averaged PDO
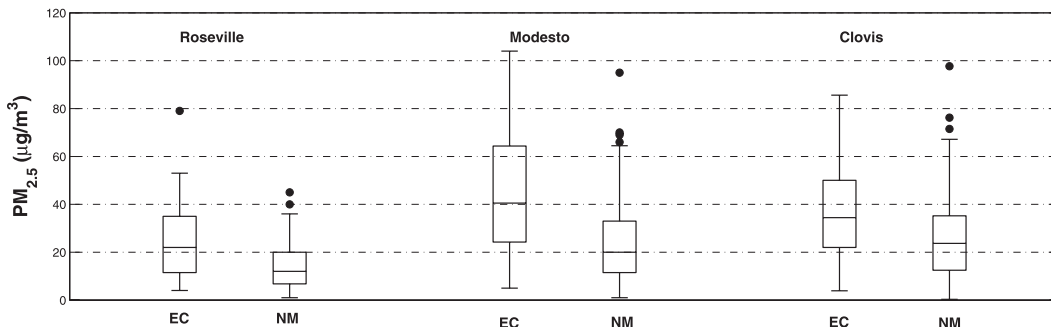


FIG. 12. Box plots showing the comparison between $PM_{2.5}$ levels sampled at Roseville, Modesto, and Clovis for EC days vs nonmatch days during 1999–2008.
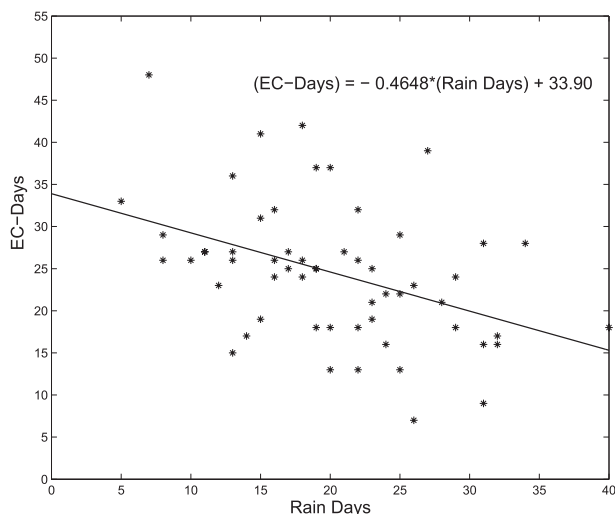
FIG. 13. Scatterplot showing the number of EC days vs rain days in a PM season. The straight line shows a linear regression fit demonstrating anticorrelation between the two quantities.

index (Fig. 9, bottom). A moving-average filter (period = 10 yr) is chosen for smoothing and best representing variability on decadal time scales. Mantua et al. (1997) observed a shift in the polarity of PDO index from negative to positive around 1977. The PDO index has been mostly positive ever since and could possibly explain the lower frequency of occurrence of EC days in the last 3 decades. A simple $F$ test for a one-way analysis of variance (ANOVA) comparing prior/post 1982/83 season frequencies yields a $p$ value of 0.135. A Mann–Whitney test confirms that there is no difference at $\alpha = 0.05$ level. It should be noted that these tests can be influenced by high-EC-day seasons during this period. It is observed that 9 out of the 10 lowest EC-day-count PM seasons have occurred prior to 1983 with 2004/05 CP–El Niño event being the only exception in the last 25 years.

High variability in the seasonal PM response to ENSO and PDO patterns may in part be attributed to the geographical location of the study domain. Dettinger et al. (1998) discuss the opposing precipitation patterns of the Pacific Northwest and the southwestern United States where the north is mostly dry and the south receives higher-than-normal rainfall during strong El Niño events. Wise (2010) investigated the hydroclimatic variability associated with this dipolar pattern and propose a narrow transition zone around 40°N, which passes through the northern half of our study domain. The exact location of the transition zone is variable and depends on the ENSO and PDO phases. Southern California or the Pacific Northwest is expected to show a stronger and more

consistent ENSO responses in comparison with the indeterminacy of the current study region.

## 4. Summary and conclusions

The paper proposes an objective statistical framework to identify large-scale weather conditions responsible for poor regional air quality. The lack of regularly sampled air quality measurements in the past provided the impetus to characterize such patterns during the period when data are available. The information obtained is used to extrapolate the occurrence of regional conditions leading to poor air quality during years when sampling was not performed. The extent to which the methodology can be used to extrapolate is limited by the availability of quality reanalysis data (1948 for the NCEP/NCAR Reanalysis-1 dataset). The identified occurrences can then be utilized to study interannual and decadal trends otherwise not possible.

Regional PM air pollution episodes in California's Central Valley were used to demonstrate applicability of the proposed framework. Forty-eight days were chosen as a target set to represent high $PM_{2.5}$ levels exceeding NAAQS $PM_{2.5}$ threshold in the chosen study domain. The 500- and 1000-hPa geopotential height anomaly composites for the target set show strongly anticyclonic conditions aloft and close to the valley surface in the study domain. The composite maps indicate other statistically significant weather events coexisting with local features and provide a template for large-scale conditions associated with elevated $PM_{2.5}$ concentrations in the valley. The classification procedure is then used to scan NNR-1 reanalysis datasets to extrapolate days with EC conditions in 60 winter PM seasons (November–February) during 1948–2008.

A total of 1397 days were identified with conditions matching the classification criteria during PM seasons starting 1951/52 through 2007/08. Geopotential height composites for identified EC days capture statistically significant events observed for the target group. Mean $PM_{2.5}$ levels during EC days were significantly higher than non-EC days during 1998–2008, when $PM_{2.5}$ measurements were available. Also, the right tail for distribution of $PM_{2.5}$ levels at all sites indicates that almost all of the highest concentrations occur during the EC days. The frequency of occurrence of EC days was found to have a significant anticorrelation to number of rainy days in a winter PM season. The relative frequency of occurrence of EC days and rain days in PM seasons during El Niño and La Niña events does not exhibit a consistent response across different events. The frequency of high-EC-days seasons seems to be consistently higher in the last 25 years of the study period (1983–2008), indicating an influence (not statistically significant) of the PDO index reverting to a mostly positive state in the period.
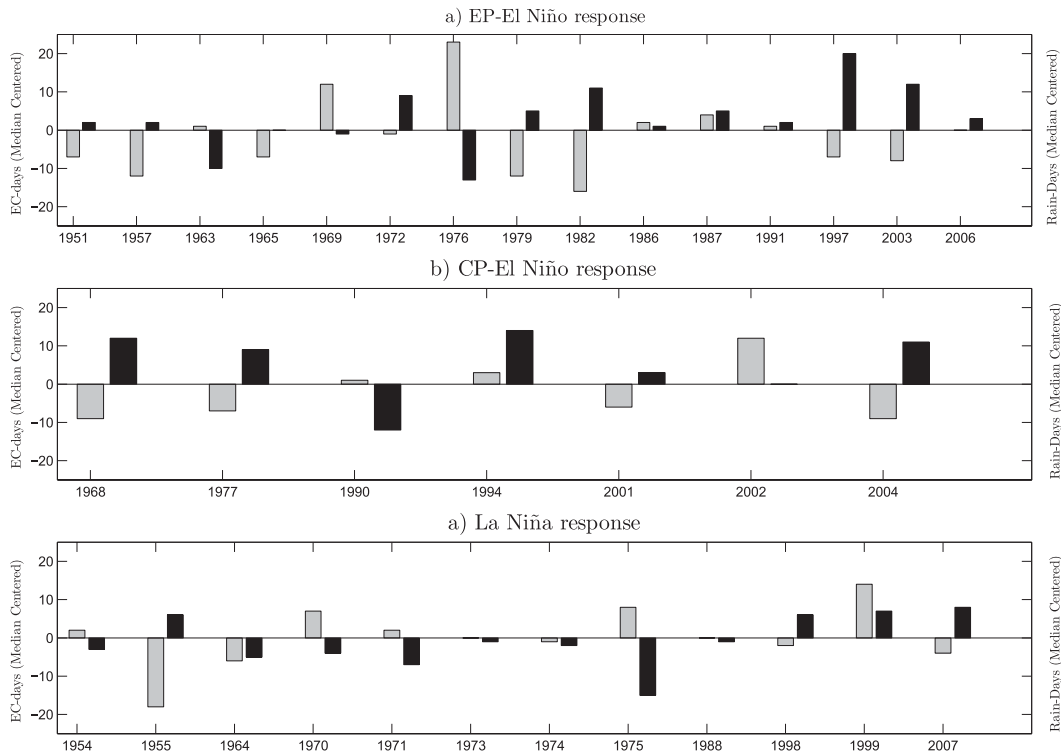
FIG. 14. Bar plots showing the number of median centered EC days (gray) and rain days (black) for every PM season during (top) EP–El Niño, (middle) CP–El Niño, and (bottom) La Niña events.

Future studies will focus on regional PM and ozone responses in various regions of United States designated as "nonattainment" zones by the EPA. The proposed method when applied to these nonattainment zones can provide vital information to regulatory agencies developing long-term implementation plans to achieve NAAQS in the near–distant future. The method does not account for the impact of emissions on pollution levels and is not capable of predicting pollution events driven by increased local emissions. The merit of the current approach lies in resolving variability of local meteorological conditions leading to high pollutant (ozone and PM) levels at interannual/interdecadal time scales. This will help to differentiate poor air quality resulting from varying local weather driven by natural climate variability on these time scales from anthropogenic emissions induced climate change. The proposed framework will determine, wherever possible, the drivers of local weather patterns at time scales associated with ENSO, PDO, and the North Atlantic Oscillation (NAO) toward this goal.

## REFERENCES

Addison, P., 2002: *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance.* Taylor and Francis, 400 pp.

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. *J. Appl. Meteor. Climatol.,* **45,** 1710–1726.

——, ——, A. Singh, S. Soong, and S. Tanrikulu, 2010: Identification of weather patterns impacting 24-h average fine particulate matter pollution. *Atmos. Environ.,* **44,** 1761–1771.

Briggs, W., and R. Levine, 1997: Wavelet and field forecast verification. *Mon. Wea. Rev.,* **125,** 1329–1341.

Brown, D., and A. Comrie, 2004: A winter precipitation 'dipole' in the western United States associated with multidecadal ENSO variability. *Geophys. Res. Lett.,* **31,** L09203, doi:10.1029/2003GL018726.

Casati, B., G. Ross, and D. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.,* **11,** 141–154.

Cayan, D., M. Dettinger, H. Diaz, and N. Graham, 1998: Decadal variability of precipitation over western North America. *J. Climate,* **11,** 3148–3166.

——, K. Redmond, and L. Riddle, 1999: ENSO and hydrologic extremes in the western United States. *J. Climate,* **12,** 2881–2893.

Dettinger, M., D. Cayan, H. Diaz, and D. M. Meko, 1998: North–south precipitation patterns in western North America on interannual-to-decadal timescales. *J. Climate,* **11,** 3095–3111.

Gershunov, A., and T. Barnett, 1998a: ENSO influence on intra-seasonal extreme rainfall and temperature frequencies in the contiguous United States: Observations and model results. *J. Climate,* **11,** 1575–1586.

——, and ——, 1998b: Interdecadal modulation of ENSO tele-connections. *Bull. Amer. Meteor. Soc.,* **79,** 2715–2725.

——, and D. Cayan, 2003: Heavy daily precipitation frequency over the contiguous United States: Sources of climatic vari-ability and seasonal predictability. *J. Climate,* **16,** 2752–2765.

Grotjahn, R., and G. Faure, 2008: Composite predictor maps of extraordinary weather events in the Sacramento, California, region. *Wea. Forecasting,* **23,** 313–335.

Hannachi, A., I. Jolliffe, and D. Stephenson, 2007: Empirical or-thogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol.,* **27,** 1119–1152.

Higgins, R. W., V. B. S. Silva, W. Shi, and J. Larson, 2007: Re-lationships between climate variability and fluctuations in daily precipitation over the United States. *J. Climate,* **20,** 3561–3579.

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

Johnson, R. A., and D. W. Wichern, 2002: *Applied Multivariate Statistical Analysis.* 5th ed. Prentice Hall, 767 pp.

Kleeman, M., Q. Ying, and A. Kaduwela, 2005: Control strategies for the reduction of airborne particulate nitrate in California's San Joaquin valley. *Atmos. Environ.,* **39,** 5325–5341.

Larkin, N. K., and D. E. Harrison, 2005: On the definition of El Niño and associated seasonal average U.S. weather anomalies. *Geophys. Res. Lett.,* **32,** L13705, doi:10.1029/2005GL022738.

Leibensperger, E., L. Mickley, and D. Jacob, 2008: Sensitivity of US air quality to mid-latitude cyclone frequency and implica-tions of 1980–2006 climate change. *Atmos. Chem. Phys.,* **8,** 7075–7086.

Leung, L., and W. Gustafson Jr., 2005: Potential regional climate change and implications to U.S. air quality. *Geophys. Res. Lett.,* **32,** L16711, doi:10.1029/2005GL022911.

Livina, V., N. Edwards, S. Goswami, and T. Lenton, 2008: A wavelet-coefficient score for comparison of two-dimensional climatic-data fields. *Quart. J. Roy. Meteor. Soc.,* **134,** 941–955.

Mahmud, A., M. Hixson, J. Hu, Z. Zhao, S. Chen, and M. Kleeman, 2010: Climate impact on airborne particulate matter concen-trations in California using seven year analysis periods. *Atmos. Chem. Phys.,* **10,** 11 097–11 114.

Mantua, N., and S. Hare, 2002: The Pacific decadal oscillation. *J. Oceanogr.,* **58,** 35–44.

——, ——, Y. Zhang, J. Wallace, and R. Francis, 1997: A Pacific interdecadal climate oscillation with impacts on salmon pro-duction. *Bull. Amer. Meteor. Soc.,* **78,** 2033–2039.

Mickley, L., D. Jacob, B. Field, and D. Rind, 2004: Effects of future climate change on regional air pollution episodes in the United States. *Geophys. Res. Lett.,* **31,** L24103, doi:10.1029/2004GL021216.

Mitchell, T., and W. Blier, 1997: The variability of wintertime pre-cipitation in the region of California. *J. Climate,* **10,** 2261–2276.

Navarra, A., and V. Simoncini, 2010: *A Guide to Empirical Or-thogonal Functions for Climate Data Analysis.* 1st ed. Springer, 200 pp.

North, G., and Q. Wu, 2001: Detecting climate signals using space–time EOFs. *J. Climate,* **14,** 1839–1863.

Patz, J., D. Campbell-Lendrum, T. Holloway, and J. Foley, 2005: Impact of regional climate change on human health. *Nature,* **438,** 310–317.

Peters, H., R. Allison, K. Lloyd, D. McKillop, L. Miller, W. Slater, and H. Stoddard, 1976: The California drought—1976. Tech. Rep., Department of Water Resources, California, 18 pp.

Pope, C. A., III, and D. W. Dockery, 2006: Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manage. Assoc.,* **56,** 709–742.

Praskievicz, S., and H. Chang, 2009: Winter precipitation intensity and ENSO/PDO variability in the Willamette Valley of Ore-gon. *Int. J. Climatol.,* **29,** 2033–2039.

Racherla, P., and P. Adams, 2006: Sensitivity of global tropo-spheric ozone and fine particulate matter concentrations to climate change. *J. Geophys. Res.,* **111,** D24103, doi:10.1029/2005JD006939.

——, and ——, 2008: The response of surface ozone to climate change over the eastern United States. *Atmos. Chem. Phys.,* **8,** 871–885.

Schonher, T., and S. Nicholson, 1989: The relationship between California rainfall and ENSO events. *J. Climate,* **2,** 1258–1269.

Schubert, S., Y. Chang, M. Suarez, and P. Pegion, 2008: ENSO and wintertime extreme precipitation events over the contiguous United States. *J. Climate,* **21,** 22–39.

Seaman, N., 2000: Meteorological modeling for air-quality assess-ments. *Atmos. Environ.,* **34,** 2231–2259.

Seinfeld, J. H., and S. N. Pandis, 1998: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change.* J. Wiley and Sons, 1326 pp.

Sun, W., R. Mukherjee, P. Stroeve, A. Palazoglu, and J. Romagnoli, 2009: A multi-resolution approach for line-edge roughness de-tection. *Microelectron. Eng.,* **86,** 340–351.

Tagaris, E., K. Manomaiphiboon, K.-J. Liao, L. R. Leung, J.-H. Woo, S. He, P. Amar, and A. G. Russell, 2007: Impacts of global climate change and emissions on regional ozone and fine particulate matter concentrations over the United States. *J. Geophys. Res.,* **112,** D14312, doi:10.1029/2006JD008262.

Tai, A., L. Mickley, and D. Jacob, 2010: Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: Implications for the sensitivity of $PM_{2.5}$ to cli-mate change. *Atmos. Environ.,* **44,** 3976–3984.

Wang, N., and C. Lu, 2010: Two-dimensional continuous wavelet analysis and its application to meteorological data. *J. Atmos. Oceanic Technol.,* **27,** 652–666.

Watson, J. G., and J. C. Chow, 2002: A wintertime $PM_{2.5}$ episode at the Fresno, CA, supersite. *Atmos. Environ.,* **36,** 465–475.

Wise, E. K., 2010: Spatiotemporal variability of the precipitation dipole transition zone in the western United States. *Geophys. Res. Lett.,* **37,** L07706, doi:10.1029/2009GL042193.

——, and A. C. Comrie, 2005: Meteorologically adjusted urban air quality trends in the southwestern United States. *Atmos. En-viron.,* **39,** 2969–2980.

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.

Zhang, X., J. Wang, F. Zwiers, and P. Groisman, 2010: The in-fluence of large-scale climate variability on winter maximum daily precipitation over North America. *J. Climate,* **23,** 2902–2915.