Using EOF Analysis to Identify Important Surface Wind Patterns in Mountain Valleys

F. L. LUDWIG

Stanford University, Stanford, California

JOHN HOREL

University of Utah, Salt Lake City, Utah

C. DAVID WHITEMAN

Pacific Northwest National Laboratory, Richland, Washington

(Manuscript received 2 July 2003, in final form 31 January 2004)

ABSTRACT

Empirical orthogonal functions (EOFs) have been determined for three wind datasets from stations in valleys south of the Great Salt Lake in Utah. Two of the datasets were for summer months, with individual days selected from the MesoWest archive to represent conditions conducive to well-developed thermally driven flows. The remaining dataset was for the month of October 2000 and was derived from a combination of MesoWest data and data collected during intensive observation periods of the Vertical Transport and Mixing Experiment (VTMX) conducted in the Great Salt Lake area. This experiment investigated stable atmospheric conditions in the complex urban terrain around Salt Lake City, Utah. In all three datasets, the primary EOFs represented flows that were directed predominantly along valley axes and were caused by channeled or thermally driven flow. Diurnal variations in EOF intensity showed that thermal effects were the most common causal mechanism. The along-valley EOFs accounted for 43%–58% of the variance in the wind component datasets (8 or 10 stations each). The second EOFs accounted for 13%–18% of the variance. In the summer datasets, the second EOF appeared to represent day–night transition periods; there was evidence of both side canyon flows and day–night transitional effects in the October dataset. The EOF approach has promise for classifying wind patterns and selecting representative cases for simulation or for further detailed analysis.

1. Introduction

It is easy to be overwhelmed by large amounts of data from many different locations collected on many different days. This paper describes the application of a well-established empirical orthogonal function (EOF)[1] approach to three such datasets to find patterns in the data and to quantify them objectively. The methods described here meet several common needs in trying to understand such data. They have revealed underlying patterns in the flow and their diurnal variations in such a way that the physical processes are apparent. The method also reduces the number of parameters necessary to describe conditions at several locations simultaneously, which is important because it is often desirable to stratify datasets so that the cases in each category can be related to other parameters. Stratifying a dataset also can make it possible to select specific cases from the different strata to be modeled or subjected to detailed analysis. When a dataset is described by 16 or more parameters, as with the cases described here, such categorization can be all but impossible. For example, if 16 variables are divided into two classes each (say, positive and negative), the result will be 2^{16} (65 536) possible categories. However, if the number of parameters required to describe the entire dataset can be reduced to one or two, then it becomes possible to develop manageable categories for the purposes mentioned above. Furthermore, the method allows a one- or two-parameter description of the simultaneous behavior at several locations.

The theory and methods for doing this procedure through the use of EOFs has been described by many authors (e.g., Hardy 1977; Horel 1981, 1984; Kutzbach 1967; Lorenz 1956; Ludwig and Byrd 1980; Lumley 1981; von Storch and Zwiers 1999; Wilks 1995). These previous studies often had different objectives than ours,

[1] EOF, eigenvector, and principal component analysis are other terms often used for the same general approach.

Corresponding author address: Francis L. Ludwig, Environmental Fluid Mechanics Laboratory, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020.
E-mail: fludwig@stanford.edu

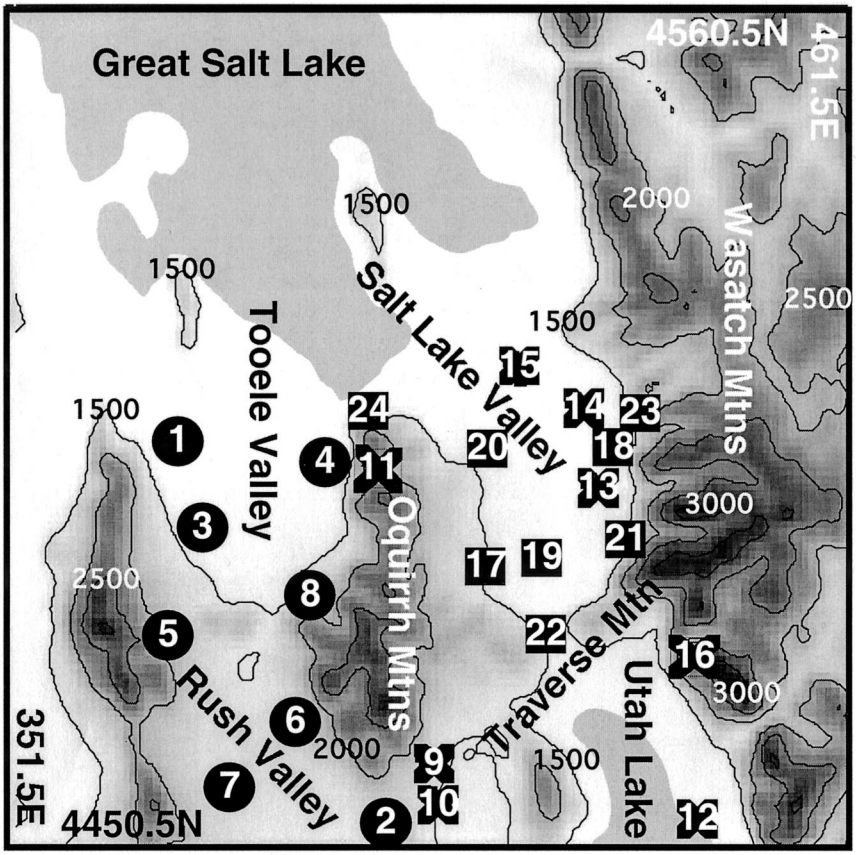

FIG. 1. Study region and station locations. See Table 1 (first column) for key to site numbers. Contour interval is 500 m.

but the underlying method is much the same. There are two approaches to determine EOFs for vector datasets. The individual winds can be represented by complex numbers (e.g., Hardy 1977; Lumley 1981) or the real components, as done here. Kaihatu et al. (1998) have examined the differences, advantages, and disadvantages of the two approaches. Here, we have chosen to use the real components to define physically significant flow patterns and their diurnal behavior and to obtain measures that can be used to classify and to select data for other types of studies. This choice is largely a matter of our preference for a simple approach that could be generalized to the three-dimensional flows provided by mesoscale models. Kaihatu et al. (1998) have also noted that the approach used here is better at preserving nondivergence.

Those interested in more information on the physical processes associated with the flow patterns revealed by the analyses presented here can find it in Whiteman's (2000) book on mountain meteorology. The paper by Kaihatu et al. (1998) provides a good discussion of vector EOF methods. Horel (1984) discusses the application of complex principal component analysis to the study of traveling atmospheric waves. Kutzbach (1967), Lorenz (1956), von Storch and Zwiers (1999), and Wilks (1995) are all valuable for their descriptions of scalar EOF methods.

2. Study area and data

a. The Salt Lake and Rush Valleys—Topography and known meteorological effects

Figure 1 shows the area encompassed by the study and the locations of some of the meteorological sites that were used. The sites are discussed in the next section. The topography is complex, with altitudes ranging from about 1270 m at the Great Salt Lake to over 3000 m in the Wasatch Mountains to the east. The Oquirrh Mountains also reach elevations of 3000 m and separate the Salt Lake and Utah Valleys from the Rush and Tooele Valleys to the west. The Utah Valley (marked by Utah Lake in Fig. 1) is separated from the Salt Lake Valley by a narrow cut (the Jordan Narrows) in the Traverse Mountains, through which the Jordan River passes. The higher-elevation Rush Valley is separated from the Tooele Valley by an east–west-oriented ridge, which is referred to as South Mountain (not labeled in Fig. 1).

The following meteorological discussion is based on

that given by Stewart et al. (2002). The two, nearly parallel, valley systems, combined with the presence of two large lakes and numerous side canyons exert marked influences on the air motions of the region. They lead to important diurnal cycles of lake–land breezes, slope flows, and valley flows, especially when synoptic influences are weak. The desired conditions for strong, thermally driven flows were defined by Stewart et al. (2002) as winds less than 7 m s^{-1} at the 700-hPa level [from the Salt Lake City International Airport (SLC) radiosonde data] and clear to partly cloudy skies. ''Clear to partly cloudy'' was defined using the criterion of Whiteman et al. (1999), that is, total daily solar radiation was more than 65% of the theoretical extraterrestrial solar radiation (from a solar radiation monitor at the University of Utah in Salt Lake City, Utah) for the day as determined from Whiteman and Allwine's (1986) solar model. Even when those conditions are not met, the local valleys will provide some constraints on flow patterns by channeling the wind so that it is parallel to their axes, that is, generally from the north-northwest or south-southeast.

At night, when the surface winds tend to be decoupled from the synoptic-scale flow, downslope and down-valley winds develop in the Utah–Salt Lake and Tooele–Rush Valley systems. Close to the Great Salt Lake, the down-valley winds can be reinforced by land breeze effects. Stewart et al. (2002) averaged the data according to the time of day and found that there are down-valley flows in both the Utah–Salt Lake and Rush–Tooele Valleys on undisturbed, fair-weather nights. At sunrise, the average down-valley winds weaken and change to up-valley winds by midmorning. During this transition, the lake breeze develops and penetrates the Tooele Valley first, and later, the Salt Lake Valley. In the afternoons, interactions take place among the upslope flows on the valley sides, the up-valley flows within the valleys, and the lake breeze.

Stewart et al. (2002) report little evidence of slope winds during the afternoon in the Rush Valley. The lake breeze penetrates southward from the Tooele Valley, across South Mountain, and into the Rush Valley. An evening transition begins around sunset, and within 3 h the average downslope and down-valley flows are reestablished in the Tooele and Salt Lake Valleys. The slope winds are strongest in midevening and taper off as the night progresses.

b. Data sources and selection

The analyses reported here were motivated by some analyses (Ludwig et al. 2002) of data specially collected during the Vertical Transport and Mixing Experiment (VTMX) in the Salt Lake Valley (Doran et al. 2002) and by the aforementioned work of Stewart et al. (2002), which was directed toward understanding thermally driven, summertime flows in valleys in the western United States. The basic datasets used for analysis came from the MesoWest data archive (Horel et al. 2002) at the University of Utah. For one of the examples discussed later, the MesoWest data were augmented with data from the VTMX 2000 campaign. The MesoWest archive consolidates information from numerous independently operated mesonets across the western United States. The conventional data from the Federal Aviation Administration (FAA), National Weather Service (NWS), and U.S. Department of Defense have been supplemented with data collected with phone modems, Internet connections, or radio transmissions from sites in regions that are not otherwise well sampled. Data in the MesoWest archive have undergone automated quality-control procedures that remove most spurious values.

The MesoWest summer data used here are a subset of the dataset used by Stewart et al. (2002) to study thermally driven circulations during the summer months from 1997 through 2000 in four study regions located throughout the western United States. Here the study is limited to the area encompassing the Tooele–Rush and Utah–Salt Lake Valleys along the Wasatch Front in Utah. Summer was chosen for its frequently weak large-scale flows that allow thermal effects to dominate local circulations. Fair-weather periods with weak winds aloft and clear to partly cloudy skies were identified in 12-h blocks centered on rawinsonde observation times, nominally 1100 and 2300 UTC. If data were reported for averaging periods of less than 1 h, the shorter-period values were vector averaged for the full hour. Any station with fewer than 50 observations for any hour of the day (over the 4 yr) was not included in the dataset that was analyzed. This latter criterion eliminated locations such as some smaller airports that closed at night.

The VTMX data were used somewhat differently from the MesoWest archived data. The VTMX program augmented the usual MesoWest stations with wind observations from about 15 locations in the Salt Lake Valley, temperatures at more than 50 others, and, very important, upper-air observations at several locations in the valley. As a consequence, it was possible to generate objective wind analyses using the winds on critical streamline surfaces (WOCSS) methods described by Ludwig et al. (1991). In brief, the WOCSS analysis defines surfaces on which the flow should take place, given that there is a maximum height to which the kinetic energy of the wind can lift a parcel of air in a stably stratified atmosphere. Winds are interpolated to those surfaces, which may intersect the terrain when the atmosphere is very stable. Then the interpolated winds are iteratively adjusted toward two-dimensional nondivergence on the surfaces, thereby forcing flow around terrain obstacles. Winds at selected WOCSS analysis grid points formed the dataset. WOCSS grid points near observation sites were chosen, because analyzed values at such points agree well with the nearby observations. The objective analyses provided complete datasets from all of the selected locations for every half hour of the

TABLE 1. Wind-observing sites used (ID gives site label, UAC = Utah Avalanche Center, TCDEM = Tooele County Department of Emergency Management, UDAQ = Utah Department of Air Quality, and UDOT = Utah Department of Transportation).

Site ID/Map No. (Fig. 1)	Site name	North lat (°)	West lon (°)	Elev (m)	Organization
	Toole–Rush Valleys (Fig. 1 symbol: ●)				
FLU/1	Flux	40.7045	112.5297	1286	TCDEM
FMP/2	Five Mile Pass	40.2316	112.1772	1640	TCDEM
GRS/3	Grantsville	40.5873	112.4769	1361	TCDEM
LAK/4	Lake Point	40.6800	112.2688	1298	TCDEM
MTB/5	Mormon Trail Bar	40.4502	112.4756	1646	TCDEM
OPH/6	Ophir Station	40.3519	112.3056	1695	TCDEM
TES/7	Tead South	40.3299	112.4080	1567	TCDEM
TOO/8	Tooele City	40.5144	112.3126	1565	TCDEM
	Utah–Salt Lake Valleys (Fig. 1 symbol: ×)				
CFO/9	Cedar Fort	40.3092	112.1013	1585	TCDEM
FFD/10	Fairfield	40.2627	112.0947	1494	TCDEM
FWP/11	Farnsworth Peak	40.659	112.202	2797	NWS
PVU/12	Provo Municipal Automated Weather Observation System	40.2240	111.7253	1369	NWS
QCW/13	Cottonwood (Holladay)	40.6445	111.8497	1323	UDAQ
QHW/14	Hawthorne	40.7344	111.8720	1311	UDAQ
SLC/15	Salt Lake City International Airport	40.78	111.97	1288	NWS/FAA
TPC/16	Timpanogos Cave	40.4406	111.7063	2438	NWS/UAC
	VTMX 2000 sites (Fig. 1 symbol: ■—This set also includes TOO/8 (●) and SLC/15 (×)				
CDW/17	Kennecott Copper slope	40.5394	112.0235	1485	PNL
HO1/18	Horseshoe Bend	41.134	111.783	1652	UDOT
M01/19	South Jordan City Hall	40.5517	111.9367	1349	PNL
M05/20	Hunter High School	40.6807	112.0258	1360	PNL
M06/21	Granite Elementary School	40.5731	111.8067	1564	PNL
NCAR/22	Jordan Narrows	40.4633	111.9317	1369	NCAR
UT5/23	Parleys Canyon	40.7122	111.8019	1498	UDOT
UT9/24	Lake Point	40.693	112.265	1311	UDOT

10 VTMX intensive operating periods (IOP); no case was eliminated for want of an observation at one of the sites. This fact is important because the EOF analysis requires the product of the data matrix and its transpose. Direct calculation of that product requires the dataset to be complete for each case used. Although methods do exist for estimating the matrix from incomplete datasets (e.g., Eslinger et al. 1989), they have some drawbacks that we avoided.

The desire for complete datasets affected the selection of stations from the summer datasets. MesoWest stations that had fewer than 2000 h were removed from further consideration even if they met the selection criteria discussed earlier. Then sets of eight stations each in the Rush and Salt Lake Valleys were selected. The choices were intended to represent flows along the axes and on the side slopes of the valleys, as well as at side canyon entrances where possible. The number of hours with complete data (all stations reporting) was determined for each set of eight stations that had been tentatively chosen. There was some trial and error required to maximize the numbers of available cases in each valley while still having locations that would represent important flow features.

Table 1 gives relevant station information for the locations used either directly (Tooele–Rush and Utah–Salt Lake Valleys) or indirectly through the WOCSS analyses. The information in the table relating to the two valley systems came from the University of Utah's MesoWest Web pages (available online at http://www.met.utah.edu/cgi-bin/database/stn_state.cgi?%20state=UT&state_name=Utah). The information concerning the Pacific Northwest National Laboratory (PNL) and National Center for Atmospheric Research (NCAR) sites came from the VTMX 2000 data archive. (The PNL data could be found online at http://etd.pnl.gov:2080/vtmx/surfmet.htm, and the NCAR information was at http://etd.pnl.gov:2080/vtmx/data/rawin/brown/readme_sondes.txt.) The MesoWest information includes photographs of many of the sites and descriptions of instruments.

The first column of Table 1 assigns a number to each site, so that they can be found in Fig. 1. The different datasets are shown by different symbols on the map: MesoWest stations in the Tooele–Rush Valleys are shown by solid circles with the site number, those in the Utah–Salt Lake Valleys are shown by Xs, and the VTMX locations are shown by solid squares; the SLC and Tooele City (TOO) locations were also included in the VTMX dataset. There were 2281 h with complete datasets from the Tooele–Rush Valleys, and 1399 from the Utah–Salt Lake Valleys. The approach taken to the VTMX data allowed use of 10 stations without loss of cases because of missing observations. However, because a few IOPs were terminated after less than 24 h,

there were only 454 half-hourly cases from the 10 VTMX IOPs available for EOF determination.

3. EOF analysis method

There are many possible ways to categorize atmospheric flow patterns represented by wind observations at a set of stations. They can be averaged by hour of the day as was done by Stewart et al. (2002). That approach uses categories (the hours) that do not necessarily relate to the features of the flow. As noted in the introduction, even simple categorization schemes are impractical and are not very informative if there are more than two or three sites or categories. There are, however, some commonly used techniques for reducing the number of variables required to describe the patterns in a dataset while losing only a small part of the information, and these techniques often reveal physically important connections in the data. Furthermore, they identify statistically independent patterns of flow that can often be associated with different physical processes. This property allows the temporal and other variations of the patterns to be examined separately, which simple averaging does not. The method does not always work well when there are not well-defined physical processes that produce well-defined flow patterns. The data we have chosen are ideally suited to this kind of analysis because they represent cases with the maximum local influences from heating and topography.

Almost 50 years ago, Lorenz (1956) described an approach using what he called ''empirical orthogonal functions'' (EOF) for representing pressure and temperature fields over the United States in a way that reduced the number of predictors necessary for a statistical forecasting scheme. The EOF approach has been widely used because it has many desirable properties, such as the fact that the resulting features are based on the characteristics of the data themselves, they are linearly independent of one another, and a measure of their relative importance is provided. The earlier analyses of Stewart et al. (2002) suggested that the flows in the Rush–Tooele and Salt Lake–Utah Valleys were sufficiently organized that the EOF analysis technique would be applicable.

Most EOF applications in meteorology have focused on scalar features or the development of a set of basis functions for numerical calculations that would be more efficient than the commonly used Fourier decomposition. Here, wind vectors are to be represented for purposes of data classification and analysis. Early applications by Lumley (1981) suggested that a similar kind of analysis could be used to extract coherent structures from turbulent flows. Hardy (1977) suggested a similar approach for classifying wind datasets. Ludwig and Byrd (1980) also applied the concept to vector fields, identifying patterns of variability in the inputs used for a linear diagnostic wind model to simplify the resulting calculations. Others using similar techniques have been Sirovich (1988) for analysis of turbulent flows, and Mahrt (1991) and Mahrt and Frank (1988) for analyses of wind component time series and vector variability along an aircraft flight path.

The determination of EOFs based on vectors is a relatively standard technique and, thus, will not be described here in detail. The following discussion is derived in large part from the classic report by Lorenz (1956), which provides a more complete mathematical derivation and discussion of the technique as applied to scalars; see also the more easily obtained works of von Storch and Zwiers (1999) or Wilks (1995). The objective in all cases is to reduce the number of variables that it takes to describe the data while losing as little information as possible. For two-component wind data collected at N locations at time t, the data can be represented as follows:

$$\begin{bmatrix} u_1(t) \\ v_1(t) \\ u_2(t) \\ v_2(t) \\ \vdots \\ u_N(t) \\ v_N(t) \end{bmatrix} = \begin{pmatrix} \overline{u}_1 \\ \overline{v}_1 \\ \overline{u}_2 \\ \overline{v}_2 \\ \vdots \\ \overline{u}_N \\ \overline{v}_N \end{pmatrix} + \sum_{i=1}^{2N} a_i(t) \begin{pmatrix} u_{1,i} \\ v_{1,i} \\ u_{2,i} \\ v_{2,i} \\ \vdots \\ u_{N,i} \\ v_{N,i} \end{pmatrix}. \quad (1)$$

The elements of the column vector on the left-hand side (LHS) of Eq. (1) represent the u and v components at N sites for time t; the different sites are denoted by the subscripts. The first right-hand side (RHS) vector has the mean u and v components for the N sites as determined from some complete dataset, such as the observations over several years. Each of the remaining terms on the right has two parts. The scalar terms $a_i(t)$ are functions of time and will be different for each set of observed data. As we shall see, a few of these scalars will usually describe the data very well. The column vectors that are multiplied by the scalar terms are the EOFs. They have the following properties: 1) They are unit vectors obtained by normalizing the eigenvectors of a matrix related to the covariance matrix. 2) They are arranged in decreasing order of explained variance. 3) They are orthogonal and hence uncorrelated. 4) At any time t, the coefficient of the ith EOF, $a_i(t)$, is given by the inner product of the ith EOF and the observation vector (LHS) minus the mean.

Because the terms on the RHS are arranged in descending order of importance (where importance is defined as the amount of variance explained by each EOF), the first few terms generally provide good estimates of the wind field. The agreement between observations and the estimates provided by Eq. (1) will improve as terms are added, but after the first two or three terms the effect of each additional term will usually be small. For well-correlated data, (e.g., winds from closely spaced stations, or winds governed by well-defined physical processes), the first term beyond the mean can explain one-half or more of the variance of the dataset. The corre-

lation between estimated and observed values is proportional to the square root of the explained variance, and so one or two terms will often produce estimates whose correlation with observed wind components is greater than 0.7.

Before proceeding, there are several other things to be said about EOFs. They can be displayed graphically so that the patterns are evident. Just as the LHS observations in Eq. (1) can be plotted as vectors at the appropriate locations on a map, so too can the pairs of components in the average and each of the EOFs be plotted. For example, with EOF 1, the vector pair ($u_{1,1}$ $v_{1,1}$) is plotted at (x_1 y_1), the location of the first observation site, ($u_{2,1}$ $v_{2,1}$) is plotted at (x_2 y_2), and so forth.

The EOFs do not necessarily reflect the results of a specific physical process. However, if there are one or two processes that are particularly strong, they are very likely to be reflected in one of the first few EOFs. When this is true, the single coefficients $a_1(t)$ and $a_2(t)$ will represent the intensity of certain patterns and their associated physical processes at time t. Another possibility is that two separate processes are similarly important statistically so that they become ''mixed together'' in the parameter subspace represented by the first two EOFs (Horel 1981). This has not been the case in the analyses below, but it can sometimes be resolved by linearly combining the EOFs that are in the subspace (Horel 1981; Richman 1981; von Storch and Zwiers 1999). The final warning about EOFs is that one should not invest much effort in the interpretation of those that explain little variance. They are subject to considerable uncertainty from sampling and round-off errors in the formation of a covariance matrix and its diagonalization.

The calculations required for EOF analysis are not prohibitive. As before, the determination of the coefficients $a_i(t)$ for each time only requires the determination of the inner product of the data vector at that time (minus the mean) and the normalized EOF vector. The most extensive calculations are the multiplication of the original data matrix by its transpose (giving the matrix closely related to the covariance matrix) and the finding of the eigenvectors of that matrix. The required operations are all standard and can easily be performed with available subroutines, such as those found in *Numerical Recipes* (Press et al. 1997).

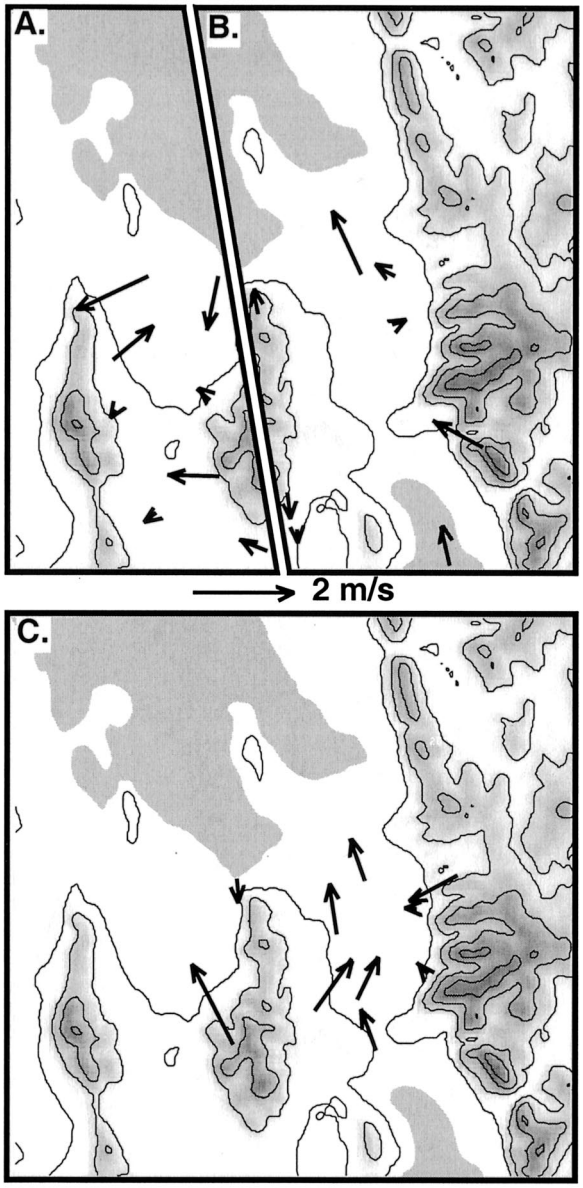

FIG. 2. Wind averages for (a) clear, light wind conditions in the Rush and Tooele Valleys, (b) clear, light wind conditions in the Salt Lake and Utah Lake Valleys, and (c) VTMX IOPs.

4. Results

a. Averages

Figure 2 shows the averages for the three analyzed datasets. The Tooele–Rush Valleys (Fig. 2a) shows similar downslope averages at the six southernmost stations, but the two northern sites show the effects of frequent, well-developed lake breezes, not canceled in the average by the weaker land breezes. The results in the Salt Lake–Utah Valleys (Fig. 2b) show net downslope and downvalley flow, both for the selected summer conditions and for the autumn days represented in the VTMX 2000 data (Fig. 2c). At all sites the average speeds are low, less than 2 m s^{-1}. This suggests that the climatological patterns in these data are weak.

It should be remembered that these are not unbiased means. The selection process for the summer data (i.e., clear, light wind conditions), and the criteria for conducting IOPs (i.e., conditions favorable to the development of slope flows) are likely responsible for the small synoptic effects. The data are strongly biased toward conditions that are conducive to thermally driven flows that are free of larger-scale effects. The analysis of hourly means by Stewart et al. (2002) showed that

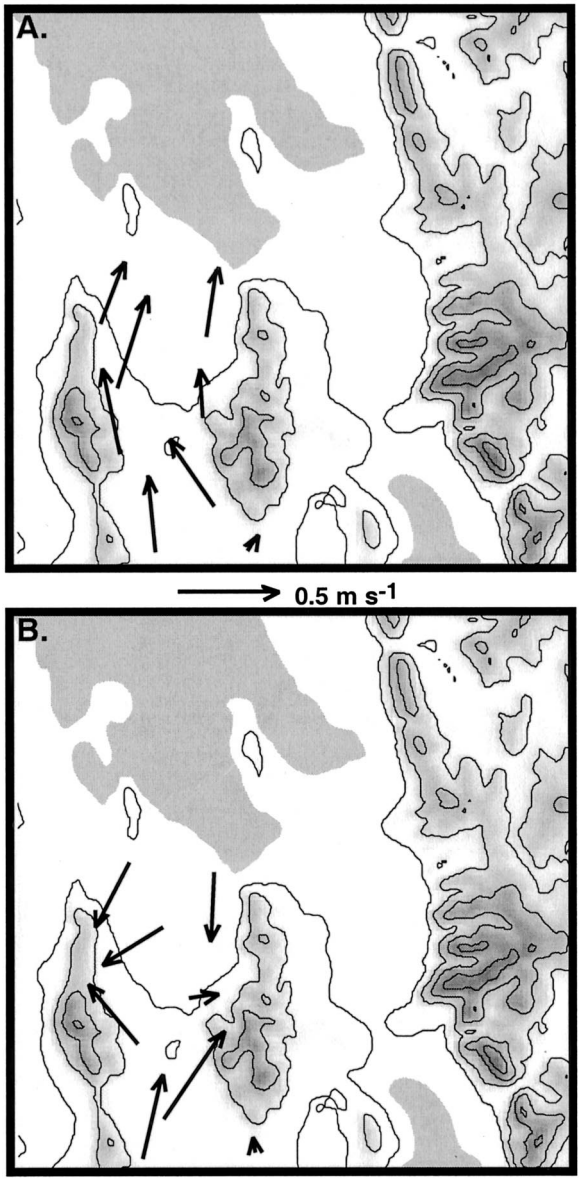

FIG. 3. Tooele–Rush Valley EOFs: (a) EOF 1 explains 47% of the variance, and (b) EOF 2 explains 16% of the variance.

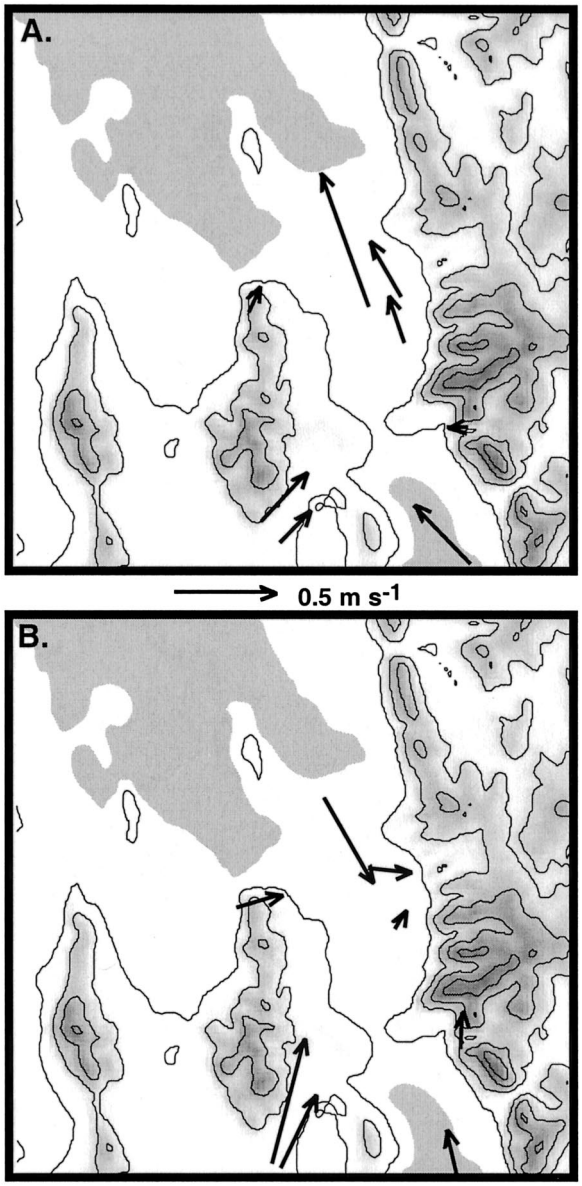

FIG. 4. Utah–Salt Lake Valley EOFs: (a) EOF 1 explains 43% of the variance, and (b) EOF 2 explains 13% of the variance.

the downslope and down-valley winds in this dataset were more consistent and lasted somewhat longer than the up-valley winds. Hawkes (1947), in contrast, found that the velocity of the up-valley wind is greater than that of the down-valley wind in many Austrian valleys. It appears that there are climatic effects that produce the observed average southerly components. This is consistent with the long-term averages for summer in the Salt Lake area. Between 1948 and 1990, more than one-half of the July winds at the Salt Lake City National Weather Service Forecast Office were from the directions south through southeast, and their mean wind speed was about 0.8 m s^{-1} greater than for winds with northerly components (Federal Climate Complex, Asheville 1992).

b. EOFs explaining the most variance

The two EOFs that explain the most variance for the Tooele–Rush Valleys dataset are shown in Fig. 3. Nearly one-half, 47%, of the variance is accounted for by the first EOF (Fig. 3a). Stated another way, the correlation between observed components [the LHS of Eq. (1)] and values estimated from just the first two terms of the RHS (the averages and the first EOF) would be nearly 0.7 (the square root of the explained variance, 0.47). The second EOF (Fig. 3b) explains much less variance,

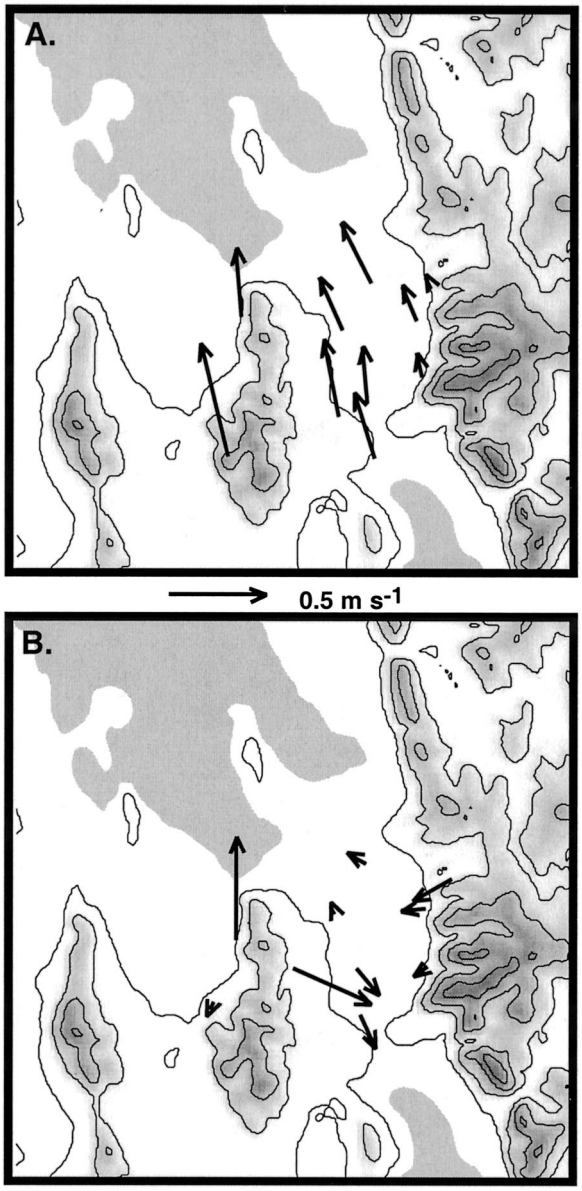

FIG. 5. EOFs derived from 10 VTMX 2000 IOPs: (a) EOF 1 explains 58% of the variance, and (b) EOF 2 explains 18% of the variance.

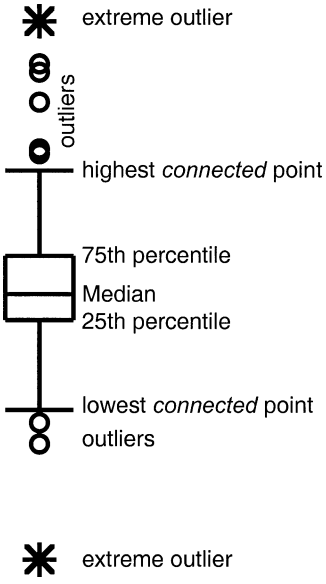

FIG. 6. Definition of box plot symbols.

16%, bringing the total to 63% for the two patterns shown.

The first EOF (Fig. 3a) is well organized. All of the directions tend to be aligned with the valley axes, indicating that the dominant physical processes are channeling and diurnal thermally driven along-valley flows. The diurnal cycles discussed later indicate that the latter is the more important contributor. When the coefficients are positive (negative), this EOF describes the down-valley (up valley) flow.

The second EOF is constrained to be spatially orthogonal to the first EOF, so it is not surprising that the second EOF defines an out-of-phase relationship between winds in the Toole and Rush Valleys (Fig. 3b). This kind of effect might be expected during times for which the land breeze has already changed to a lake breeze but the nighttime downslope flow has yet to quit. When the coefficients of the two EOFs are of opposite sign, the northern part of the valley will be dominated by a lake breeze (EOF 1 is negative and EOF 2 is positive) or land breeze (EOF 1 is positive and EOF 2 is negative).

Figure 4 presents the first two EOFs derived from the Utah–Salt Lake Valley MesoWest stations. The first EOF here (Fig. 4a), like the first EOF in the other valleys, reflects the importance of channeling and thermal flows parallel to the main valley axes. The second EOF is characterized by flows in opposite directions in the northern and southern parts of the domain (Fig. 4b). When the EOF-2 coefficient is positive (i.e., when its contributions will be in the same directions shown in Fig. 4), the two northern locations have lake-breeze–upslope flows while the southern sites have greater downslope components. As already discussed for the Tooele Valley, opposite signs of the coefficients for the first two EOFs characterize the lake-breeze–land-breeze circulation. Flow through the pass separating the Rush Valley from the Utah Valley is evident in both EOF 1 and EOF 2. Later, it will be shown that this is evidence for strong, frequently occurring flow from the Utah Valley to the Rush Valley in the late afternoon.

Figure 5 shows the first two EOFs for VTMX 2000 based on samples at 10 locations and at half-hour intervals for the 10 IOPs. Here, the sites are concentrated in the Salt Lake Valley, with none in the Utah Valley, but one site east of the Oquirrh Mountains (near TOO; see Fig. 1) has been included. The explained variance is appreciably greater for these data, perhaps reflecting

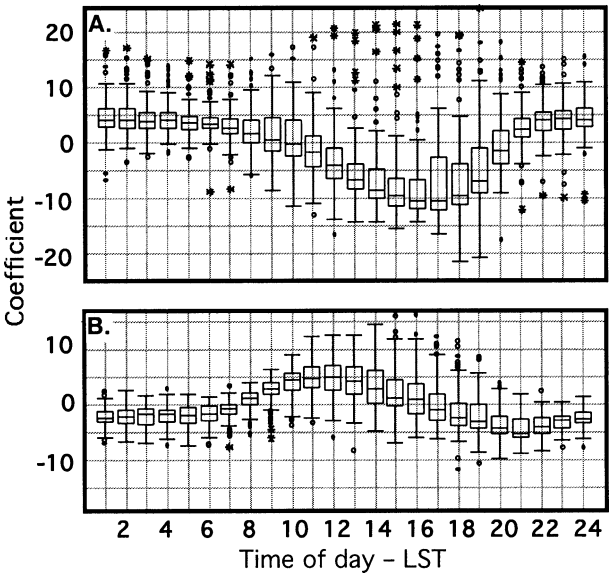

FIG. 7. Box plots of hourly EOF coefficients derived from the Tooele–Rush Valley data: (a) EOF 1 and (b) EOF 2.

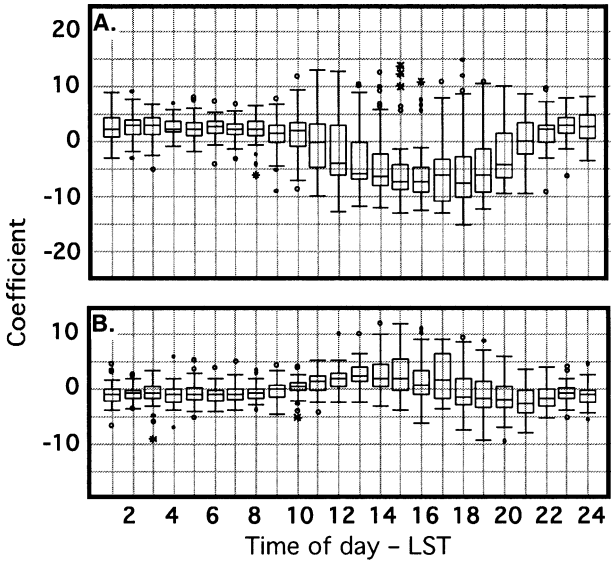

FIG. 8. Box plots of hourly EOF coefficient values derived from the Utah–Salt Lake Valley data: (a) EOF 1 and (b) EOF 2.

the more compact and homogeneous area represented. Not unexpected is that the first EOF (Fig. 5a) shows the same general pattern of along-valley flow that is seen in the other two datasets.

Figure 5b shows the second EOF (constrained to be spatially orthogonal to the first EOF). It is more complex than those discussed to this point, and it illustrates the importance of having information available from locations at which flows of interest can be found. In this case, the VTMX 2000 field campaign made it possible to monitor winds near the mouths of some of the canyons and along the broad west slope of the Salt Lake Valley. As a consequence, this EOF shows more evidence of canyon drainage and slope flow than was possible in the other cases. The slope flow at the Kennecott Copper slope site (site CDW) is particularly evident, which suggests that it is distinctly out of phase with the dominant north–south flow of the first EOF. The second EOF also has some of the characteristics seen in the other cases, in that there is a confluence (for positive coefficients) of the side canyon and slope flows throughout the length of the main valley.

c. Temporal variability

The discussions of the EOFs have made reference to flows that appear to be thermally driven. If indeed the EOFs reflect these kinds of flows, which are driven by diurnal heating cycles, then the coefficients of the EOFs should have pronounced diurnal cycles. Box plots for each hour have been used to show the diurnal tendencies in coefficient values. The box plot symbols that have been used are explained in Fig. 6. The plots were prepared with the aid of the commercial software package, Data Desk 6.0 (Velleman 1997). The rectangle in each box plot spans the half of the values that is between the lower and the upper quartile. The horizontal line within the rectangle marks the median. According to Velleman (1997), ''The whiskers extend from the top and the bottom of the box to depict the extent of the main body of the data.'' The small circles and asterisks at the extremes mark individual outlier values beyond that main body of data. Each box plot in subsequent figures represents about 95 values for the Tooele–Rush Valley cases, about 60 for the Utah–Salt Lake Valley hours, and about 20 for the VTMX 2000 IOPs. The number of cases was not exactly the same for each hour.

Figure 7 shows box plots of the coefficients found in the Tooele–Rush Valley data for each hour of the day. The diurnal trends are obvious. For EOF 1 (Fig. 7a), the median coefficients are positive between about 2100 and 0900 LST and are negative for most of the daytime hours. When referring back to Fig. 3a, it can be seen that this pattern corresponds to the expected cycle for a thermally driven flow, with increased down-valley winds at night and up-valley winds during the day. It is also evident from Fig. 7a that variability is much greater during the transition periods. Variability is also greater during the day than at night, during which time the nighttime stability at ground level tends to decouple the surface wind patterns from the effects of synoptic-scale features and their associated channeled flow.

Figure 7b shows the distribution of EOF-2 coefficients for the Tooele–Rush Valley cases. From about midnight until 1000 LST, the median coefficients are near zero, indicating that the pattern represented by EOF 2 in Fig. 3 is weak. There are then modestly positive values from about 1000 until 1400 LST, indicating the onset of a lake breeze while there is still some drainage from the Rush Valley. The medians have moderately

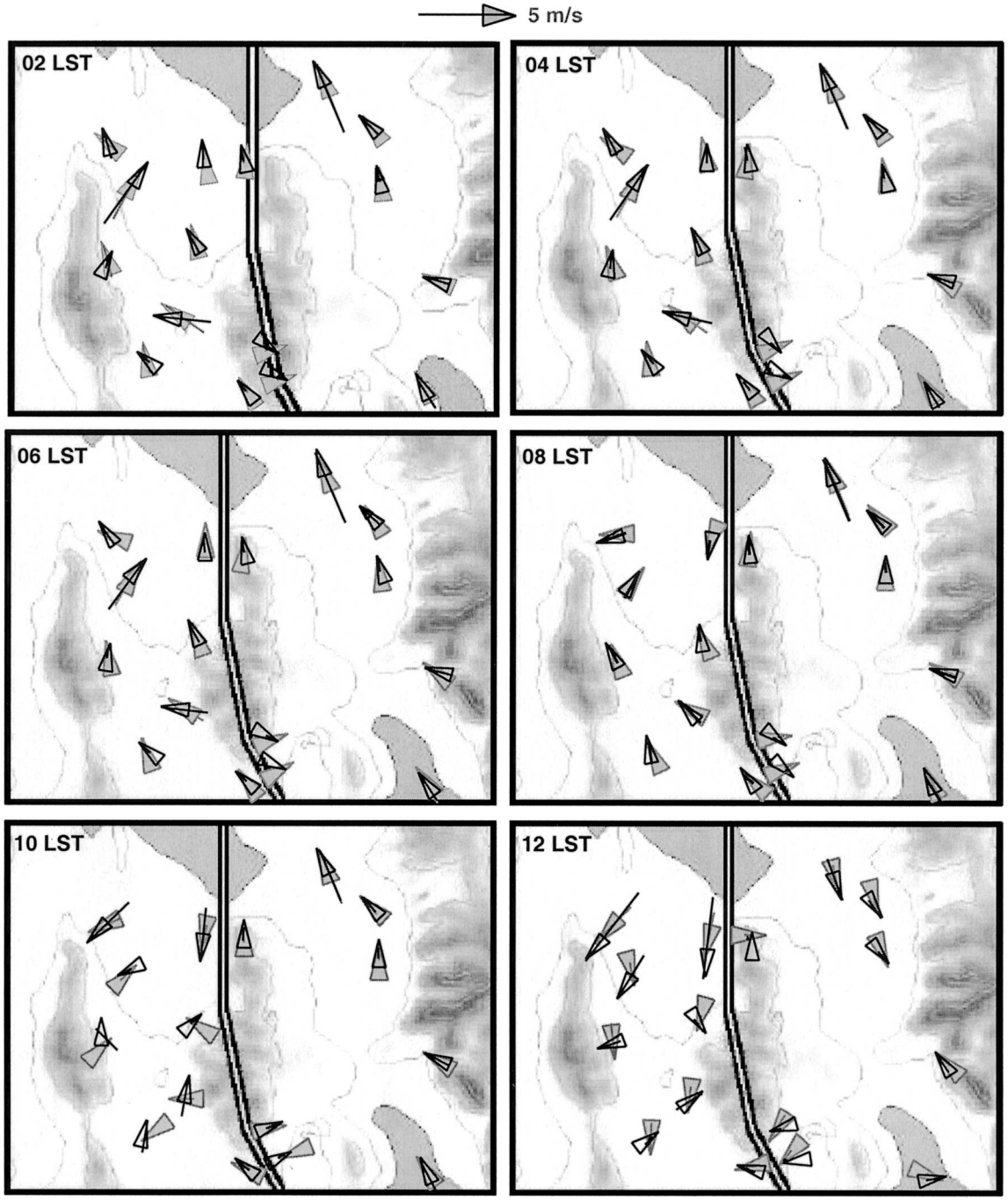

FIG. 9. Wind estimates at 2-h intervals reconstructed from the mean and first two EOFs using hourly median coefficient values. Gray arrows show results using first EOF only. Black outline arrows show results for first two EOFs. Directions are indicated by arrows; magnitudes are shown by straight lines.

large negative values from about 1800 LST until midnight, reflecting the reverse transition from lake breeze to land breeze. As with EOF 1, the greatest spread of observed values occurs during the convective part of the day.

Figure 8a shows that the diurnal pattern for the first EOF in the Utah–Salt Lake Valleys is similar to that for the other valley system in Fig. 7a. The onset of upvalley/upslope flows (negative EOF-1 coefficients) begins about noon, judging by the medians, and lasts until after 2000 LST. As in the Rush Valley, this is also the period of greatest spread in the individual values. The

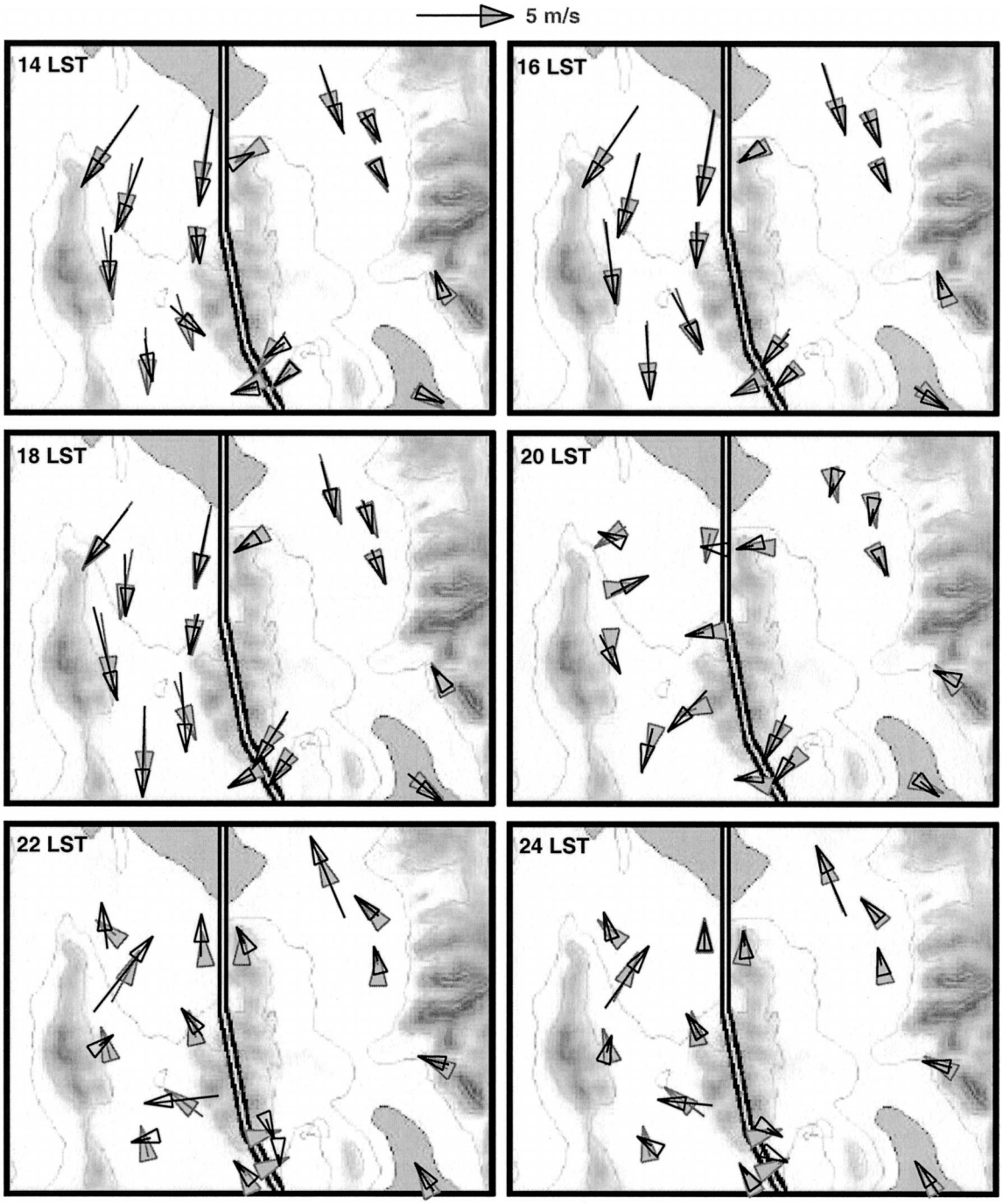

FIG. 9. (*Continued*)

positive coefficients marking the reverse flow at night are smaller but more consistent than was the case in the other valleys.

The only parts of the day for which the median coefficients for EOF 2 (Fig. 8b) differ much from zero in the Utah–Salt Lake Valleys are those between about 1100 and 1700 LST, during which they tend to be positive, and between 1800 and 2200, during which they are negative. This behavior is different from that of the other area, in which the nonzero periods corresponded more closely to morning and evening transitions. The second EOF in the Utah–Salt Lake Valleys predominantly reflects the lake/land breeze in the north. It appears that the north-northeasterly flow through the pass occurs with the onset of the lake breeze in the late morning, and they both reverse shortly after sunset.

A comparison of Figs. 7 and 8 shows that the coefficients for the two EOFs become of opposite sign between 0800 and 0900 LST in the Tooele Valley (Fig. 7). This switch marks the onset of the lake breeze. In the Salt Lake Valley, the transition occurs about 3 h later. A smaller lag is evident in the onset of the land breeze. In the western valley system, the two EOFs become of opposite sign at about 2000–2100 LST. In the other system, that occurs about 1 h later.

It is not easy to visualize how the wind fields would look from just looking at the averages, the EOFs, and the coefficients, but Fig. 9 may help. It shows wind field estimates constructed from the median coefficients (Figs. 7 and 8) for various hours of the day, using the mean plus the first one or two EOF terms of Eq. (1). The Tooele–Rush Valley averages and EOFs (Figs. 2 and 3) were used for the winds on the west side of the figures, and the Utah–Salt Lake Valley averages and EOFs (Figs. 2 and 4) were used for the others. The winds derived from just the average and first EOF are shown by gray arrows; those from the average and first two EOFs are shown by arrows outlined with black lines. The two sets of arrows are often very much alike, because the second EOFs usually contribute little (i.e., they have near-zero coefficients), except during transitions between day and night conditions. The first EOF is the major descriptor of the diurnal flow cycle. Because we have used median values for the coefficients, the wind fields shown in Fig. 9 can be considered as a picture of the wind patterns on a ''typical,'' albeit nonexistent, day.

In the Tooele–Rush Valleys there is a well-defined downslope flow during the night hours, which persists well after sunrise in the Rush Valley. However, as noted above, the lake breeze begins much earlier than in the Utah–Salt Lake Valleys and grows in strength as it is augmented by the heated upslope flow later in the day. Similar flows are seen in the Utah–Salt Lake Valleys, but the lake breeze begins later. Through the afternoon and early evening, the combined lake breeze and upvalley flows are evident, with flows penetrating from the Utah Valley through the pass that separates it from the Rush Valley. Both flows reverse during the evening.

The preceding discussion and comparison of Figs. 3a and 4a and of Figs. 7a and 8a suggest that the first EOFs derived from the data in the two valley systems represent very similar physical processes with similar diurnal evolution. This suggestion is supported by the scatterplot and regression line in Fig. 10a. The regression line between the EOF-1 coefficients in the two valleys was calculated for those 765 h for which complete datasets were available in both areas. The two are highly correlated (correlation coefficient $r = 0.91$). The same is not true of the EOF-2 coefficients (Fig. 10b). As mentioned earlier, the variance explained is much less for the second EOF than for the first. In addition, the physical processes represented by the second EOFs appear to be different for the two valleys. For these reasons, it is not to be expected that the coefficients would rise and fall together.

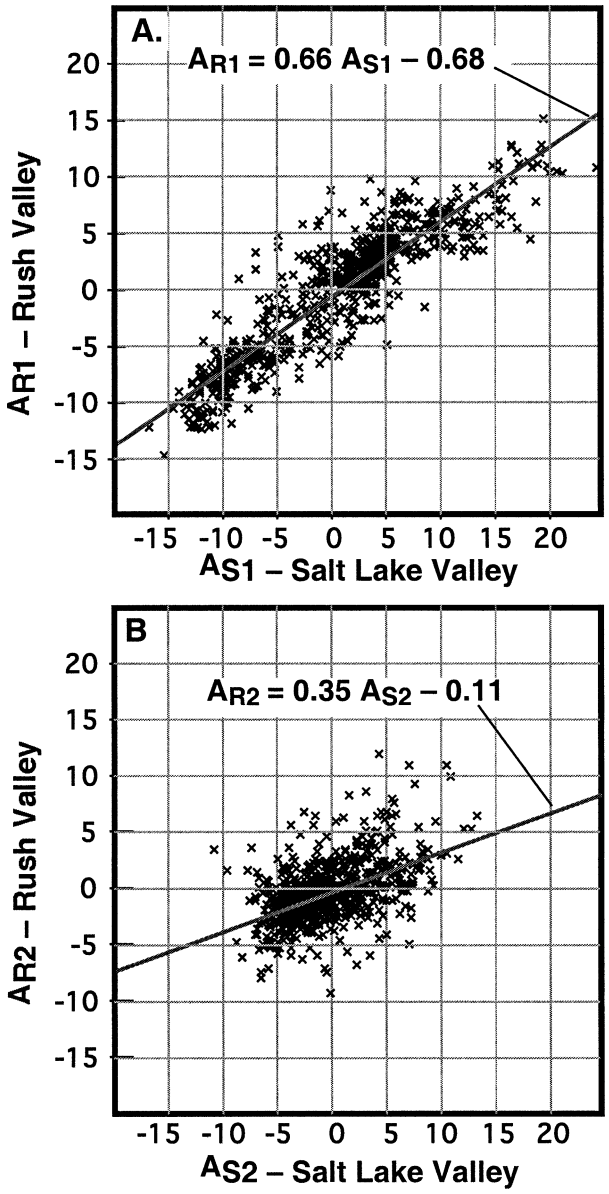

FIG. 10. Scatterplots and regression lines for EOF coefficients in the Tooele–Rush and Utah–Salt Lake Valleys: (a) EOF 1 ($r = 0.68$) and (b) EOF 2 ($r = 0.51$).

The diurnal cycle of coefficients for the last set of data, derived from the 10 IOPs of VTMX 2000, is shown in Fig. 11. The observations used for this figure were generally available 2 times per hour, usually 15 min before and 15 min after the hour. For purposes of display, the coefficients derived from the observations on either side of the hour have been plotted at the hour between them. The temporal changes for the first EOF (Fig. 5a) are much the same as for the other two datasets. Somewhat positive values corresponding to down-valley winds from the south-southeast persist from about

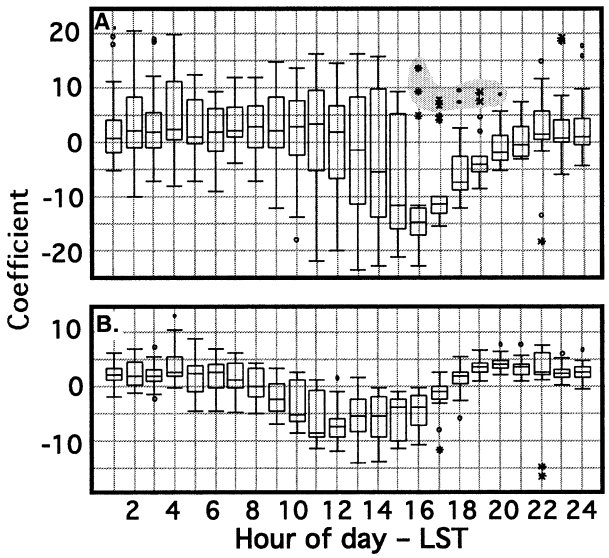

FIG. 11. Box plots of hourly coefficients derived from the VTMX 2000 field data: (a) EOF 1 (extreme values in shaded area are from IOP 9) and (b) EOF 2.

midnight through noon. The up-valley winds with negative coefficients usually begin just after noon and continue until about 2000 LST. Also, the greatest day-to-day variability is between about 0900 and 1500 LST. The sudden reduction in the day-to-day variability between 1500 and 1600 LST indicates that the up-valley flow was in almost all cases well established at 1600 LST, with the exception of the outliers, which all occurred during IOP 9. Their origin will be discussed later.

The EOF-2 coefficients' diurnal variability is also similar to the cycles for the other Salt Lake Valley dataset, although there are significant differences in the EOFs themselves (Figs. 4b and 5b). The median coefficients are slightly positive from about sunset until sunrise. Then, they are negative throughout the day, which suggests that EOF 2 is dominated by up-slope flows on both flanks of the Salt Lake Valley during that time. In addition, the opposite flows represented by EOF 2 in the northern and southern core of the valley (Fig. 5b) suggest a southward propagation of the lake breeze during the afternoon. A contingency table (Table 2) shows that there are a significant number of cases in which the two EOFs are of the same sign and either the lake breeze or the land breeze is reinforced, depending on the sign. By the nature of their determination, the coefficients are uncorrelated, and so the fact that there are similar numbers of cases in each of the boxes of the contingency table is to be expected.

d. Other influences

A considerable amount of corollary meteorological information was collected and archived during the VTMX 2000 campaign. This fact allows us to relate both the typical cases and the outliers to other conditions. This in turn, gives the opportunity to see what factors lead to coefficients of unusually large magnitude. For example, the outliers shown earlier in Fig. 11a were identified as occurring during IOP 9. Figure 12 shows the changes of coefficients for EOFs 1 and 2 during this IOP. The vertical bar in the figure marks the approximate time of passage of a cold front. The NWS analysis for 0500 LST 21 October 2000 showed a southwest–northeast-aligned cold front about 50-km southeast of Salt Lake City. The University of Utah weather log for this IOP describes conditions as follows: ''A short wave trough was approaching rapidly from the west, with strong 500 mb vorticity advection occurring ahead of the system. Associated vertical motions were strong ahead of this negatively tilted feature. Little temperature advection occurred over much of Utah at this time with the baroclinic zone to the west. However, 700 mb winds were still relatively strong and from the southwest. . . . Skies were cloudy throughout most of this short IOP. Local circulations were interrupted by cold frontal passage in the valley (~12 GMT). The clouds and winds prevented a morning radiational inversion from setting up. . .''

TABLE 2. Contingency table relating joint occurrences of positive and negative coefficients for the first two VTMX 2000 EOFs.

	Sign of EOF 2	
Sign of EOF 1	<0	≥ 0
<0	74	128
≥ 0	105	146

The passage of the cold front is marked well by the sudden change in EOF-1 coefficients from large positive values to large negative values. Large positive coefficients are indicative of strong southerly, down-valley winds (see Fig. 4a). In this case, the positive values are caused by the southerly flow ahead of the front, and the subsequent negative values are caused by the flow re-

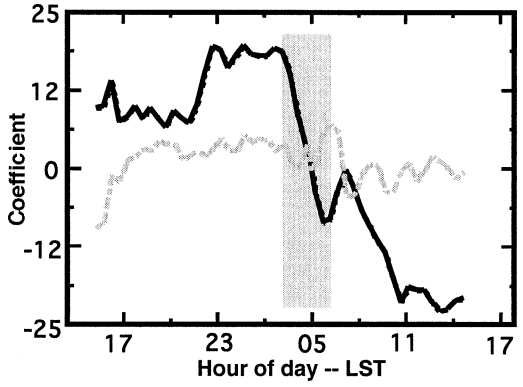

FIG. 12. EOF coefficients during VTMX 2000 IOP 9 (20–21 Oct 2000). Solid line is EOF 1. Gray line is EOF 2. Positive values increase down-valley (EOF 1) and down-canyon (EOF 2) flow components. Approximate time of frontal passage is shown by the vertical bar.

versal to northerly after the frontal passage. The contribution of the second EOF is small throughout most of the experimental period. The small spike around 0600 LST may mark the passage of the front through the middle of the area, with northerly winds behind it to the north and southerlies ahead in the south (see Fig. 4b). The observations made by the Doppler radar at Salt Lake City International Airport support this interpretation. There is a pronounced shear line through the region at about 0600 LST (1300 UTC).

5. Discussion

There are several important conclusions to be drawn from the analyses presented here. The first is the fact that this kind of analysis will identify a region's recurring patterns of motion. That done, it is often easy to deduce the underlying physical processes governing those motions, either from direct examination of the EOFs or from the diurnal variation of their intensity, as measured by the coefficient of the EOF. Although it was not done here, it is assumed that annual variations in the diurnal cycles would also be informative.

In the cases studied here, the dominant physical process has been the diurnal cycle of thermally induced flows. It was also noted that the thermal effects can be overwhelmed by synoptic events that mimic those patterns. These other mechanisms will, as was the case here, often occur at the ''wrong'' time of day, so that they are very evident as anomalous behavior. Other commonly occurring patterns that can be detected by EOF analysis include the sea-breeze cycle and, on occasion, more esoteric features such as larger-scale eddies.

Perhaps the EOFs are most useful as a succinct and objective method for characterizing datasets. In this case, we have limited the analysis to flow patterns, but mixed-parameter datasets can be used. Analysis of mixed parameter datasets must be done with great care to ensure that units are chosen so that the fluctuations in value are of comparable magnitude for all of the parameters. Often, standard deviations are used to make the parameters nondimensional and of comparable magnitude.

As shown by these analyses, typical patterns can be identified, or the EOF coefficients can be used to stratify the data in meaningful ways. Table 2 is a rudimentary example of a stratification system with four categories. One of the motivations for this work has been to develop ways to categorize data, so that certain cases can be selected for special analysis. In this way, examples of different types of atmospheric behavior can be identified for further modeling or analysis.

The EOF approach provides a means for objectively finding examples of the effects associated with different, identifiable physical processes, so long as they can be resolved by available data. There are many applications for this ability. A common one would be for selecting representative cases for modeling and developing air pollution abatement plans. Our objective has been to identify a few cases for modeling and analysis to identify when and where conditions occur that might lead to mixing in the nighttime stable atmosphere. Resources are too limited to pursue the modeling of very many days, and so we must classify conditions such that we ensure that we examine the major possibilities.

Last, the kind of analysis presented here has the potential for solving a long-standing problem in model performance evaluation. Simulation model output represents spatially filtered or averaged conditions, whereas observations are usually for a single point, often subject to small-scale effects that have not been modeled. Thus, when the two are compared, the causes of differences can be uncertain. A more valid comparison might compare EOFs derived from model output with those from observations. Going one step farther, the EOFs derived from observations can be used as filters for detecting the presence and magnitude of the same features in model output. Both of these approaches will provide quantitative measures of agreement, and both evaluate the performance over the whole domain, rather than on a point-by-point basis.

Acknowledgments. This work was supported by the U.S. Department of Energy, under the auspices of the Atmospheric Sciences Program of the Office of Biological and Environmental Research. The organizations responsible for the mesonet data archive that we used are the National Oceanic and Atmospheric Administration, the University of Utah's Cooperative Institute for Regional Prediction, and the Salt Lake City National Weather Service Forecast Office. We are grateful to the many VTMX participants who provided their data. They include those from the Argonne, Livermore, Los Alamos, and Pacific Northwest National Laboratories and from Arizona State University, the University of Utah, NCAR, NOAA, and others. We also acknowledge Jebb Stewart and Xindi Bian for their efforts in gathering and classifying the MesoWest data that have been used.

REFERENCES

Doran, J. C., J. D. Fast, and J. Horel, 2002: The VTMX 2000 campaign. *Bull. Amer. Meteor. Soc.*, **83,** 537–551.

Eslinger, D. L., J. J. O'Brien, and R. L. Iverson, 1989: Empirical orthogonal function analysis of cloud-containing coastal zone color scanner images of northeastern North American coastal waters. *J. Geophys. Res.*, **94,** 10 884–10 890.

Federal Climate Complex, Asheville, 1992: *International Station Meteorological Climate Summary.* Version 2.0, National Oceanic and Atmospheric Administration, CD-ROM.

Hardy, D. M., 1977: Empirical eigenvector analysis of vector observations. *Geophys. Res. Lett.*, **4,** 319–320.

Hawkes, H. B., 1947: Mountain and valley winds with special reference to the diurnal mountain winds of the Great Salt Lake region. Ph.D. dissertation, Ohio State University, 312 pp.

Horel, J. D., 1981: A rotated principal component analysis of the interannual variability of the Northern Hemisphere 500 mb height field. *Mon. Wea. Rev.*, **109,** 2080–2092.

——, 1984: Complex principal component analysis: Theory and examples. *J. Climate Appl. Meteor.,* **23,** 1660–1673.

——, and Coauthors, 2002: MesoWest: Cooperative mesonets in the western United States. *Bull. Amer. Meteor. Soc.,* **83,** 211–226.

Kaihatu, J. M., R. A. Handler, G. O. Marmorino, and L. K. Shay, 1998: Empirical orthogonal function analysis of ocean surface currents using complex and real-vector methods. *J. Atmos. Oceanic Technol.,* **15,** 927–941.

Kutzbach, J. E., 1967: Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *J. Appl. Meteor.,* **6,** 791–802.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction, Scientific Report 1, Statistical Forecasting Project. Massachusetts Institute of Technology Defense Doc. Center No. 110268, 49 pp.

Ludwig, F. L., and G. Byrd, 1980: A very efficient method for deriving mass consistent flow fields from wind observations in rough terrain. *Atmos. Environ.,* **14,** 585–587.

——, J. M. Livingston, and R. M. Endlich, 1991: Use of mass conservation and dividing streamline concepts for efficient objective analysis of winds in complex terrain. *J. Appl. Meteor.,* **30,** 1490–1499.

——, R. Street, and Y. Chen, 2002: Detection and interpretation of patterns of motion in mesoscale atmospheric flows. *Eos, Trans. Amer. Geophys. Union,* **83** (Fall Meeting Suppl.), Abstract A51C-0068.

Lumley, J. L., 1981: Coherent structures in turbulence. *Transition and Turbulence,* R. E. Meyer, Ed., Academic Press, 215–242.

Mahrt, L., 1991: Eddy asymmetry in the sheared heated boundary layer. *J. Atmos. Sci.,* **48,** 472–492.

——, and H. Frank, 1988: Eigenstructure of eddy microfronts. *Tellus,* **40A,** 107–119.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1997: *Numerical Recipes in Fortran 77: The Art of Scientific Computing.* Vol. 1, 2d ed. Cambridge University Press, 934 pp.

Richman, M. B., 1981: Obliquely rotated principal components—An improved meteorological map typing technique. *J. Appl. Meteor.,* **20,** 1145–1159.

Sirovich, L., 1988: Analysis of turbulent flows by means of empirical eigenfunctions. Brown University Center for Fluid Mechanics Rep. 90-212, 40 pp.

Stewart, J. Q., C. D. Whiteman, W. J. Steenburgh, and X. Bian, 2002: A climatological study of thermally driven wind systems of the U.S. Intermountain West. *Bull. Amer. Meteor. Soc.,* **83,** 699–708.

Velleman, P. F., 1997: *Data Desk Version 6.0 Handbook 2.* Data Description, Inc., 356 pp.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

Whiteman, C. D., 2000: *Mountain Meteorology, Fundamentals and Applications.* Oxford University Press, 355 pp.

——, and K. J. Allwine, 1986: Extraterrestrial solar radiation on inclined surfaces. *Environ. Software,* **1,** 164–169.

——, X. Bian, and J. L. Sutherland, 1999: Wintertime surface wind patterns in the Colorado River valley. *J. Appl. Meteor.,* **38,** 1118–1130.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 467 pp.