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

F. L. Ludwig[1]

Stanford University, Stanford, California

John Horel

University of Utah, Salt Lake City, Utah

C. David Whiteman

Pacific Northwest National Laboratory, Richland, Washington

(submitted July 2 , 2003)

(Revised December , 2003)

[1] Corresponding Author: Francis L. Ludwig, Environmental Fluid Mechanics Laboratory, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, phone: 650-725-5948, Fax: 650-725-9720, E-mail: fludwig@stanford.edu

ABSTRACT

Empirical orthogonal functions (EOFs) have been determined for three wind data sets from stations in valleys south of the Great Salt Lake in Utah. Two of the data sets were for summer months, with individual days selected from the MesoWest archive to represent conditions conducive to well-developed thermally driven flows. The remaining data set 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 Salt Lake area. This experiment investigated stable atmospheric conditions in the complex urban terrain around Salt Lake City, Utah. In all three data sets, 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 to 58 percent of the variance in the wind component data sets (8 or 10 stations each). The second EOFs accounted for 13 to 18 percent of the variance. In the summer data sets, 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 data set. 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) approach to three such data sets 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 methodology also reduces the number of parameters necessary to describe conditions at several locations simultaneously. This is important because it is often desirable to stratify data sets so that the cases in each category can be related to other parameters. Stratifying a data set also can make it possible to select specific cases from the different strata to be modeled or subjected to detailed analysis. When a data set 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 2 classes each (say, positive and negative), the result will be 2^{16} (65536) possible categories. However, if the number of parameters required to describe the entire data set 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 methodology for doing this through the use of empirical orthogonal functions (EOF)* has been described by many authors (e. g. Hardy 1977; Horel 1981, 1984; Kutzbach 1967; Lorenz 1956; Ludwig and Byrd 1980, and Lumley 1981, von Storch and Zwiers 1999; Wilks 1995). These previous studies often had different objectives than ours, but the underlying methodology is much the same. There are two approaches to determine EOFs for vector data sets. 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 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 meso-scale models. Kaihatu et al. (1998) have also noted that the approach used here is better at preserving non-divergence.

Those wishing 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),

---

* EOF, eigenvector and principal component analysis are other terms often used for the same general approach

Lorenz (1955), von Storch and Zwiers (1999) and Wilks (1995) are all valuable for their descriptions of scalar EOF methodologies.

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 quite 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 Figure 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 Figure 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 SLC radiosonde data), and clear to partly cloudy skies. "Clear to partly cloudy" was defined using the criterion of Whiteman et al. (1999), i. e., 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) 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, i. e. 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 by mid-morning. 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 mid-evening 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 data sets 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), NWS, and Department of Defense (DOD) have been supplemented with data collected via 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 is a subset of the data set 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 less than an hour, 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 four years) was not included in the data set 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 importantly, 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). Very briefly, 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 non-divergence on the surfaces, thereby forcing flow around terrain obstacles Winds at selected WOCSS analysis grid points formed the data set. 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 data sets from all selected locations for every half hour of the ten VTMX intensive operating periods (IOP); no case was eliminated for want of an observation at one of the sites. This is important because the EOF analysis requires the product of the data matrix and its transpose. Direct calculation of that product requires the data set to be complete for each case used. Although methods do exist for estimating the matrix from incomplete data sets (e.g. Eslinger et al. 1989), they have some drawbacks which we avoided.

The desire for complete data sets affected the selection of stations from the summer data sets. MesoWest stations that had fewer than 2000 hours were removed from further consideration even if they met the selection criteria discussed earlier. Then, sets of eight stations each in the Rush Valley and the Salt Lake Valley 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 via the WOCSS analyses. The information in the Table relating to the two valley systems comes from the University of Utah's MesoWest web pages:

http://www.met.utah.edu/cgi-bin/database/stn_state.cgi?%20state= UT&state_name=Utah . The information concerning the PNL and NCAR sites comes from the VTMX 2000 data archive. The PNL data can be found at *http://etd.pnl.gov:2080/vtmx/surfmet.htm* , and the NCAR information 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 Figure 1. The different data sets are shown by different symbols on the map: Mesowest stations in the Tooele/Rush Valley by solid circles with the site number; those in the Utah/Salt Lake Valleys by Xs, and the VTMX locations by solid squares; the SLC and TOO locations were also included in the VTMX data set. There were 2281 hours with complete data sets from the Tooele/Rush Valleys, and 1399 from the Utah/Salt Lake Valleys. The approach taken to the VTMX data allowed use of ten stations without loss of cases due to missing observations. However, since a few IOPs were terminated after less than 24 hours, there were only 454 half-hourly cases from the ten VTMX IOPs available for EOF determination.

3. EOF analysis methodology

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 not very informative if there are more than 2 or 3 sites or categories. There are, however, some commonly used techniques for reducing the number of variables required to describe the patterns in a data set, 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 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 producing well defined flow patterns. The data we have chosen are ideally suited to this type of analysis because they represent cases with the maximum local influences from heating and topography.

Almost fifty 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 data sets. 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 in order 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. Specifically, for two-component wind data collected at N locations at time t, the data can be represented as follows:

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

The elements of the column vector on the left hand side (LHS) of Equation 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 data set, 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 quite 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 i^{th} EOF, $a_i(t)$, is given by the inner product of the i^{th} 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 equation 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 half or more of the variance of the data set. The correlation between estimated and observed values is proportional to the square root of the explained variance, 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 quite evident. Just as the LHS observations in Equation 1 can be plotted as vectors at the appropriate locations on a map, so can the pairs of components in the average and each of the EOFs, 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, $(u_{2,1}\ \ v_{2,1})$ 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 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 it 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).

4. Results

a. Averages

Figure 2 shows the averages for the three analyzed data sets. The Tooele/Rush Valley (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 cancelled in the average by the weaker land breezes. The results in the Salt Lake/Utah Valleys (Fig. 2 B) show net downslope and down-valley flow, both for the selected summer conditions and for the fall 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, free of larger scale effects. The analysis of hourly means by Stewart et al. (2002) showed that the down-slope and down-valley winds in this data set 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 half the July winds at the Salt Lake City 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 explaining the most variance for the Tooele/Rush Valley data set are shown in Figure 3. Nearly half, 47 percent, of the variance is accounted for by the first EOF (Fig 3A). Stated another way, the correlation between observed components (the LHS of equation 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, 16 percent, bringing the total to 63 percent for the two patterns shown.

The first EOF (Fig 3A) is well organized. All 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 when 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 negative and EOF 2 positive) or land breeze (EOF 1 positive and EOF 2 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 Figure 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/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 the VTMX 2000 experiment based on samples at ten locations and at half hour intervals for the ten 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 the more compact and homogeneous area represented. Not unexpectedly, the first EOF (Fig 5A) shows the same general pattern of along valley flow seen in the other two data sets.

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 where 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 (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 Figure 6. The plots were prepared with the aid of the software package, Data Desk 6.0 (Velleman 1997). The rectangle in each box plot spans the half of the values that are 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. Referring back to Figure 3A, it can be seen that this 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 Figure 7A that variability is much greater during the transition periods. Variability is also greater during the day than at night, when 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 EOF2 coefficients for the Tooele/Rush Valley cases. From about midnight until 1000, the median coefficients are near zero, indicating that the pattern represented by EOF 2 in Figure 3 is quite weak. There are then modestly positive values from about 1000 until 1400, indicating the onset of a lake breeze while there is still some drainage from the Rush Valley. The medians have moderately large negative values from about 1800 until midnight, reflecting the reverse transition from lake 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 Figure 7A. The onset of up-valley/upslope flows (negative EOF 1 coefficients) begins about noon judging by the medians and lasts until after 2000. As in the Rush Valley, this is also the period of greatest spread in the individual values. The 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 when the median coefficients for EOF 2 (Figure 8B) differ much from zero in the Utah/Salt Lake Valleys are those between about 1100 and 1700 LST, when they tend to be positive, and between 1800 and 2200, when they are negative. This is different from the other area, where the non-zero 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.

Comparing Figures 7 and 8 shows that the coefficients for the two EOFs become of opposite sign between 0800 and 0900 LST in the Tooele Valley (Figure 7). This marks the onset of the lake breeze. In the Salt Lake valley, the transition occurs about three hours 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 to 2100 LST. In the other system, that occurs about an hour 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 Figure 9 may help. It shows wind field estimates constructed from the median coefficients (Figures 7 and 8) for various hours of the day, using the mean plus first one or two EOF terms of Equation 1. The Tooele/Rush Valley averages and EOFs (Figures 2 and 3) were used for the winds on the west side of the figures, and the Utah/Salt Lake Valley averages and EOFs (Figures 2 and 4) 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 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 Figure 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 up-valley flows are evident with flows penetrating from the Utah Valley through the pass that separates it from the Rush Valley. Finally, both flows reverse during the evening.

The preceding discussion, and comparison of Figures 3A and 4A, and of 7A and 8A, suggests that the first EOFs derived from the data in the two valley systems represent very similar physical processes with similar diurnal evolution. This is supported by the scatter plot and regression line in Figure 10A. The regression line between the EOF 1 coefficients in the two valleys was calculated for those 765 hours when complete data sets 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 (Figure 10B). As mentioned earlier, the variance explained by the second EOF 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.

The diurnal cycle of coefficients for the last set of data, derived from the ten IOPS of VTMX 2000 are shown in Figure 11. The observations used for this figure were generally available twice per hour, usually 15 minutes before and 15 minutes 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 (Figure 5A) are much the same as for the other two data sets. Somewhat positive values corresponding to down-valley winds from the south-southeast persist from about 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 and1500. 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 data set, although there are significant differences in the EOFs themselves (Figures 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 when 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, 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 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 Figure 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-to-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. . ."

The passage of the cold front is well marked 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 Figure 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 reversal 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 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 Figure 4B). The observations made by the Doppler radar at Salt Lake International airport support this interpretation. There is a pronounced shear line through the region at about 0600 LST (1300 UT).

5. Discussion

There are several important conclusions to be drawn from the analyses presented here. 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 occasionally more esoteric features, such as larger scale eddies.

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

Finally, 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, while 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 further, the EOFs derived from observations can be used as filters for detecting the presence and magnitude of the same features in model output. Both 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.

Acknowledgements. 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 (NOAA), the University of Utah's Cooperative Institute for Regional Prediction, and the Salt Lake City National Weather Service (NWS) Forecast Office. We are grateful to the many VTMX participants who provided their data. They include those from Argonne, Livermore, Los Alamos and Pacific Northwest National Laboratories, from Arizona State University, the University of Utah, NCAR, NOAA and others. We also wish to 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**, 10884-10890.

Federal Climate Complex Asheville, 1992: *International Station Meteorological Climate Summary*, National Oceanic and Atmos. Admin, Asheville, NC, CD-ROM.

Hardy, D. M., 1977: Empirical eigenvector analysis of vector observations, *Geophys. Res. Letters*, **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*. 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.

Horel, J. D., 1984: Complex principal component analysis: theory and examples, *J. Appl. Meteor.*, **23**, 1660-1673.

Horel, J., M. Splitt, L. Dunn, J. Pechmann, B. White, C. Ciliberti, S. Lazarus, J. Slemmer, D. Zaff, and J. Burks, 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. and Oceanic Tech.*, **15**, 927-941

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

Lorenz, E. N., 1956: *Empirical Orthogonal Functions and Statistical Weather Prediction, Scientific Report 1, Statistical Forecasting Project*, Mass. Inst. Tech., Cambridge, Mass. (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.

Ludwig, F. L., 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.

Ludwig, F. L., R. Street, and Y. Chen, 2002: Detection and interpretation of patterns of motion in mesoscale atmospheric flows, *Amer. Geophys. Union Fall Meeting and Abstracts CD, A51C-0068*.

Lumley, J. L., 1981: *Coherent Structures in Turbulence, Transition and Turbulence (R. E. Meyer, ed.)*, Academic Press, New York, 215-242.

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

Mahrt, L., 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, Volume 1 (Second Ed.)*, Cambridge University Press, Cambridge, 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*," Center for Fluid Mechanics Report No. 90-212, Brown Univ., Providence, R. I., 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., Ithaca NY, 356 pp.

von Storch, H. and F. W. Zwiers, 1999: *Statistical analysis in climate research*, Cambridge University Press, Cambridge, 484 pp.

Whiteman, C. D., 2000: *Mountain meteorology, fundamentals and applications,* Oxford University Press, New York, 355 pp.

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

Whiteman, C. D., 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.

List of Tables

List of figures

FIGURE 8 Box plots of hourly EOF coefficient values derived from the Utah/Salt Lake Valley Data

A. EOF 1

B. EOF 2

FIGURE 9 Wind estimates at two-hour 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 indicated by arrows, magnitudes by straight lines.

FIGURE 10 Scatter plots and regression lines for EOF coefficients in the Tooele/Rush and Utah/Salt Lake Valleys

A. EOF 1 – Correlation coefficient, r= 0.68

B. EOF 2 – Correlation coefficient, r= 0.51

FIGURE 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

B. EOF 2

FIGURE 12 EOF coefficients during VTMX 2000 IOP 9, 20-21 October 2000

Solid line is EOF 1. Grey 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.

TABLE 1: Wind observing sites used

Site ID/ Map No. (Fig. 1)	Site Name	North Latitude degrees	West Longitude degrees	Elevation (m)	Organization[1]
Toole/Rush Valleys -- Figure 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 – Figure 1 symbol X:					
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 Muni. AWOS	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 Internat'l	40.78	111.97	1288	NWS/FAA
TPC / 16	Timpanogos Cave	40.4406	111.7063	2438	NWS/UAC
VTMX 2000 Sites – Figure 1 symbol ■ – This set also includes TOO / 8 (●) and SLC / 15 (X)					
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 Elem. 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

[1.] FAA = Federal Aviation Administration, NCAR = National Center for Atmospheric Research, NWS = National Weather Service, PNL = Pacific Northwest National Lab., UAC = Utah Avalanche Center, TCDEM = Tooele County Department. of Emergency Management, UDAQ = Utah Department of Air Quality, UDOT = Utah Department of Transportation

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

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