15.3 Statistical Modeling of Downslope Windstorms in Boulder, Colorado

Andrew E. Mercer
University of Oklahoma, School of Meteorology, Norman, Oklahoma

Michael B. Richman
University of Oklahoma, School of Meteorology, Norman, Oklahoma

Howard B. Bluestein
University of Oklahoma, School of Meteorology, Norman, Oklahoma

John M. Brown
Forecast Systems Laboratory, NOAA, Boulder, Colorado

Abstract

Downslope windstorms are of major concern to those living near the Boulder, Colorado area, often striking with little warning, bringing clear air wind gusts of 35-50 m/s or higher, and producing widespread damage across the city. Models used for forecasting these dangerous events are often not accurate. Hence, there is a need to apply different linear and non-linear statistical modeling techniques to a 10-year mountain-windstorm dataset.

A set of eighteen predictors, based on a decade of data, are used in this study. Linear regression, neural networks and support vector models are employed to relate the predictors to windstorm events. For the linear model, stepwise linear regression is applied. It is difficult to determine which predictor is the most important, although significance testing indicates 700 hPa flow is highly significant. The nonlinear techniques employed, support vector regression and a feedforward neural network did not filter out any predictors. The study indicates that there is a potential for improvement in peak wind forecasting using different methods and predictors.

The models are evaluated using RMSE and median residuals. The support vector regression model performed best. Stepwise linear regression yielded results that were accurate to within 8 m/s, whereas a neural network reduced errors of 6 to 7 m/s and support vector regression had errors of 4 to 6 m/s. 85% of these forecasts based on nonlinear techniques predicted maximum wind gusts with an RMSE of less than 6 m/s, and all forecasts predicted wind gusts with an RMSE of below 12 m/s. In comparison, a linear model forecast wind gusts better than 6 m/s 60% of the time, and better than 12 m/s 95% of the time. These results suggest that meaningful improvements to mountain wind forecasts are achieved by application of newer non-linear techniques, such as neural networks and support vector regression.

1. Introduction

Mountain windstorms occur often in cities that are located on the lee side of the Rocky Mountains, particularly Boulder, Colorado (Figure 1). According to recent data from the National Climate Data Center, winds of 45 ms^{-1} or more are commonly observed near Boulder, Colorado, about ten times each year. Such storms cause extensive tree and property damage. Despite this, no modeling technique is currently used operationally that predicts mountain wind events accurately. A previous study by Leptuch (2001) attempted to model wind speeds using a set of 18 predictors using linear models. That work established a baseline for which improvement can be assessed.

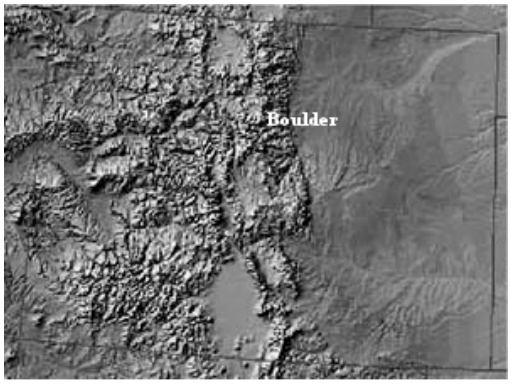

Figure 1. Topographical map of Colorado. Boulder is marked in the lee of the Rocky Mountains.

According to Brinkman (1973), two general classes of windstorms are observed in the Rocky Mountains. The first is the classic Chinook, which is referred to in this study as a "prefrontal" windstorm. Prefrontal windstorms occur when a surface low-pressure system moves into northern Colorado or just east of that location, leading to a positive pressure perturbation on the windward side of the Rockies and a negative pressure perturbation on the lee side of the Rockies. Such storms are common and are observed with strong warm advection over the Pacific Northwest, intense northwesterly flow over the Continental Divide, and lee cyclogenesis above the high plains. The second class of windstorm is known as the "cold bora" type (denoted as "postfrontal" in this study). This type of windstorm occurs often when a surface anticyclone intensifies west of the Continental Divide and/or moves into northern Colorado. Such positioning of the anticyclone leads to a rapid pressure rise west of the Continental Divide and a slower pressure rise east of the Divide.

A number of additional theories have been proposed to account for the formation and intensification of mountain windstorm events (Klemp and Lilly 1975, Peltier and Clark (1979), etc.). However, this paper considers the aforementioned situations that are based on lee-wave theory. By limiting the cases to these two, implementation of predictors is more straightforward. It is important to remember that owing to the randomness of wind gusts and the variability of wind gust speeds, no theory can account for all types of mountain windstorms.

Klemp and Lilly (1975) show that, when the wavelength of a gravity wave is long enough to permit the wave to propagate vertically, the wave may propagate upward into a region where the Scorer parameter (see section 2.2.13) decreases rapidly with height. Part of the energy from the wave is reflected downward, possibly resulting in wave resonance. In their study, multiple-layer models were analyzed, and peak wind gusts that were generated were approximately double what are actually observed in the atmosphere. Non-linearities in the true atmosphere, including diffusion, were attributed to this apparent discrepancy in the results. Klemp and Lilly's study was deemed useful for locating areas where lee waves should have a strong response, and these areas are expected to be useful for the prediction of mountain windstorms. Klemp and Lilly's theory is not valid for longer waves, such as those considered by Scorer and Kleiforth (1959).

In a pair of papers by Clark and Peltier (1977) and Peltier and Clark (1979), the development of finite-amplitude mountain waves was analyzed by observing flow over an isolated obstacle. These studies determined that the linear theory proposed by Klemp and Lilly (1975) produced wind gusts that were too strong in comparison with those observed during mountain windstorms. Hence, they developed a non-linear dynamic model that found that a critical layer induced by the mountain wave is responsible for producing mountain windstorms.

According to Smith (1985) and Durran and Klemp (1987), there is a mean-state critical layer when the phase speed of a mountain wave equals the wind speed at some level in the atmosphere with vertical wind shear. In the presence of a critical layer, any portion of the mountain wave

**Corresponding Author Address*: Andrew E. Mercer, School of Meteorology, University of Oklahoma, Norman, OK, 73072-7307; email: toine3212@ou.edu

that is above the critical layer will be advected in the opposite direction, leading to a steepening of the wave and possible wave breaking. The level at which the wind shifts is determined to be the critical layer (i.e., a wind shift from a negative u component to a positive u component -the level at which the flow goes to zero is defined as the critical layer). This situation, however, is rarely observed near the lee side of the Rocky Mountains; thus, other parameters were sought that would predict windstorms successfully upstream of the mountain ranges. Clearly, numerous ideas exist about the formation of significant mountain waves.

In section 2, the methods used in forming the statistical models are presented, while in section 3 the results from applying the statistical modeling to the datasets are shown. The key results are summarized in section 4.

2. Methods

2.1 Data

This study made use of National Weather Service (NWS) 0000 and 1200 UTC soundings from five sites: North Platte, Nebraska; Denver, Colorado; Grand Junction, Colorado; Lander, Wyoming; and Salt Lake City, Utah. The soundings were obtained between 1 January 1969, and 31 December 31 1978. However, owing to the climatology of windstorms, no soundings between 15 May and 15 September were used.

The verification wind-gust data were obtained from the (then) National Bureau of Standards (NBS), 325 Broadway, in Boulder. Stimulated by the extremely destructive storms of 7 January 1969 (Bergen and Murphy 1978) and 10-11 January 1972 (Doyle 2000), there was an intensive ongoing effort to study Boulder windstorms during this period. However, most of these data remained in analogue, strip-chart form and have never been processed. The NBS data collection was unique in that the Guard Force at the NBS site was charged with the responsibility to record by hand every 3 hours (at 0000, 0300, 0600, …, 2100 UTC) the peak gust during the previous 3-h period, if this peak gust equaled or exceeded 10 ms^{-1}. The NBS site is located in a residential area of Boulder, 1.5-3km to the east of the base of the Front-Range Foothills. The local terrain slopes gently upward to the SW –W before the foothills are abruptly encountered. The 3-cup anemometer was located atop the northeast end of the main building on the site, roughly 15m above ground level. This anemometer is of unknown type, but one of us (JMB) recalls that it was of a generic rugged design characterized by a distance constant rather larger than that typical of current designs.

The handwritten records were obtained from Wayne Sangster, formerly of the NWS Central Region, who used these data to develop his "Sangster Method", discussed in Section 2.2.18, and were subsequently put in electronic form. Since we are interested in relating sounding properties upstream (west) of the Continental Divide to windstorms occurring just downstream, we chose to correlate the sounding data to Boulder gust occurrence near or after the launch time of the balloons (typically about 1115 and 2315 UTC). Given the nature of the 3-h periods for which we had peak gusts, we chose to relate the soundings to peak gusts occurring in seven different time periods: 0-12, 0-6, 6-12, 0-3, 3-6, 6-9, and 9-12 h after nominal sounding times. These represent the time period over which the peak gust was observed. For example, in a 1200 UTC sounding, the 0-6 hour peak gust would be the gust observed between 1200 and 1800 UTC. With these different combinations, prediction could take place over different time periods spanning the 12h interval between soundings. Roughly 5% of daily wind data were missing and thus were not available for our study.

2.2 Parameters

The following is a list of the 18 predictors, each of which is followed by a description and brief definition. In the following discussions, low levels are classified as the layer between 3100 and 5600 m (roughly 700 mb to 500 mb), mid-high levels are the layer between 5600 and 7400 m (roughly 500 mb to 400 mb), and high levels are the layer between 5600 and 9400 m (roughly 500 mb to 300 mb). Parameters based out of the Grand Junction National Weather Service Office (GJT), the Lander National Weather Service Office (LND), and a mean between the two, led to three sets of parameters for each of the two types of windstorms.

2.2.1 Temperature advection

The first variable calculated was the 700 hPa geostrophic temperature advection. Assuming thermal wind balance, it was estimated using the 700 hPa geostrophic wind and the 500 – 700 hPa mean temperature gradient. Positive values represented warm advection, and negative values represented cold advection. The units of this parameter are K day^{-1}.

2.2.2 700 hPa geostrophic wind direction

This quantity is calculated by fitting the 700 hPa wind height fields from three different sites, GJT, LND, and LBF, to a plane. The resulting u and v wind components from this plane are used to determine the geostrophic wind using trigonometry. It is assumed that the data are approximately geostrophic. Data from DEN are neglected, since mesoscale terrain effects there sometimes contaminate the data. The 700 hPa wind direction is determined to be significant since it is approximately the average height of the mountains and is therefore used to determine the likelihood of the formation of lee waves. The units of this parameter are degrees.

2.2.3 700 hPa geostrophic wind speed [m s^{-1}]

The calculation of 700 hPa geostrophic wind speed is the same manner as for the 700 hPa geostrophic direction. Observational data indicates that there is a minimum value of wind speed at the ridge-top level below which windstorms do not occur, roughly 9 m/s.

2.2.4 $\overline{U}$ DEN [m s^{-1}]

This wind component is an approximation to the flow component normal to the terrain west of the Continental Divide west of Boulder. It is derived from the Denver rawinsonde. The direction of $\overline{U}$ backs linearly with height from 290° below 3100 m to 270° above 4200 m. Inspection of a topographic map (Fig. 10) reveals that the Front Range west of Boulder (crest approximately 3700 - 4200m) is oriented N-S. However, when considering the state as a whole, the 3350m (~11,000 ft) elevation contour west of the Continental Divide tends to be oriented NNE-SSW (Fig. 10). To account for this, we take the perpendicular used to compute $\overline{U}$ as backing with height in the elevation range indicated.

2.2.5 Local $\overline{U}$ [m s^{-1}]

This parameter is computed identically to $\overline{U}$ DEN, but for either GJT or LND, or an average of the two.

2.2.6 700 hPa – 500 hPa geostrophic shear direction

This parameter is defined as the difference between the direction of the 700 hPa geostrophic wind and the 500 hPa geostrophic wind. If it is positive, it corresponds to show warm air advection, and if it is negative, it is deemed to show cold air advection. This is significant in that the parameter is used to diagnose the synoptic type of storm system as either prefrontal or postfrontal. A postfrontal case is defined as one characterized by a sounding containing more than 15° of backing in this layer, while all other cases were deemed as prefrontal. The units of this quantity are in degrees.

2.2.7 Ratio of the 700 hPa wind speed by the 700 hPa geostrophic wind speed at Denver

This quantity diagnoses the degree of ageostrophy at DEN. The total wind is selected, as opposed as normal-to-barrier wind, so that the results give the departure from geostrophy in cases where the winds are both normal to the barrier and parallel to the barrier. This parameter is dimensionless.

2.2.8 Difference between 700 hPa wind direction and the 700 hPa geostrophic wind direction

This parameter is used to determine whether the DEN sounding site was contaminated by lee vortices. A filtering program is applied to determine if the DEN sounding site was not suitable to apply as one of the three soundings used to determine the 700 mb geostrophic wind. Negative values imply cross-height flow, likely due to boundary-layer friction. The units of this parameter are degrees.

2.2.9 Mountain top relative humidity

During the cool season, a downslope wind event induced by convective outflow can occur. Such an even may occur when evaporative cooling from precipitation falling from a foehn cloud intensifies the downslope wind, leading to strong surface wind gusts. This event is very difficult to forecast using objective methods, and relative humidity was chosen to account for these specific types of events. This parameter is dimensionless.

2.2.10 Cross mountain height difference

This quantity is the difference between the 850 hPa height (m) at GJT and that at DEN. It is used to determine the intensity of synoptic scale forcing that is present in the area. A large value implies strong synoptic flow and high winds, whereas a small value implies weak synoptic flow and lighter winds.

2.2.11 Static-stability ratio

This parameter is the ratio of the mean-layer Brunt Väisälä parameter, N^2, of low levels to that at high levels. This parameter is applied to

diagnose the change with height of static stability in the atmosphere. Decreasing static-stability with height is thought to be favorable for the development of windstorm formation. This parameter has no dimensions.

2.2.12 Froude height

This parameter is the height at which the Froude number equals unity. This quantity represents the height at which upslope winds have enough momentum to go across the top of the mountain, as opposed to being directed around the mountaintop. Froude height is calculated by:

$$Fr = \int_H \frac{N}{\overline{U}} dz \qquad (1)$$

where *N* is the Brunt-Väisälä frequency and $\overline{U}$ is the magnitude of the normal-to-mountain component of the wind. This quantity is integrated from ridge-top down to the surface; thus, H is the mountain height. This quantity is given in kilometers.

2.2.13 Integrated Scorer parameter

This parameter is designed to determine the phase shift ϕ_l of vertically propagating waves between the Froude height and the tropopause, and is given by the following:

$$\phi_l = \frac{1}{2\pi} \int l dz \qquad (2)$$

where l is the Scorer parameter (Scorer 1949) and represents the vertical wave number, which is defined as 2π /L_z, with L_z being the vertical wavelength in meters. A phase shift of nπ represents the optimum phasing for a reflecting wave. The original equation for the Scorer parameter, (equation 3), has the following two terms:

$$l^2 = \frac{N^2}{\overline{U}^2} - (\frac{1}{\overline{U}}) \frac{\partial^2 \overline{U}}{\partial z^2} \qquad (3)$$

where $\overline{U}$ is the magnitude of the normal-to-mountain component of the wind and *N* is the Brunt-Väisälä frequency. In calculating the Scorer parameter, it can be shown that including the second term leads to significant noise in a small data set; thus, this term was neglected. If the value of *l* is determined to be negative over a thin layer, it is set it to zero. Negative values of the Scorer parameter are an artifact of small superadiabatic layers that are created by the splines used to fill missing sounding data. In contrast to Klemp and Lilly (1975), the Froude height is selected as a lower limit instead of 700 hPa. This parameter is dimensionless.

2.2.14 Characteristic impedance ratio

This parameter, *(C.I.R),* is used to determine the joint vertical variation of cross-mountain flow and static stability. This parameter was first defined in Blumen (1985), and is given as:

$$C.I.R. = \left(\frac{36}{25}\right) \frac{\sum_{i=1}^{n} (N_i \overline{U}_i)_{low}}{\sum_{i=1}^{n} (N_i \overline{U}_i)_{high}} \qquad (4)$$

where $\overline{U}_i$ is the magnitude of the normal-to-mountain component of the wind, 36/25 is a reflection coefficient determined numerically by Blumen (1985), and N_i is the Brunt-Väisälä frequency. If the quantity $N_i \times \overline{U}_i$ tends not to increase with height, there is a greater likelihood of windstorms. The quantity *i* represents the height of the calculation of the *C.I.R.* This quantity is dimensionless.

2.2.15 Lowest Tropopause

This parameter shows the height in kilometers of the lowest of the tropopause levels at DEN, GJT, and LND. Lower values of this parameter are generally associated with synoptic scale low-pressure systems with strong vertical forcing, leading to strong winds, etc.

2.2.16 Local Tropopause

This parameter is the tropopause height in kilometers at the observing site. This parameter is used for comparisons with tropopause heights at LND and GJT. In the case in which one of the two locations had the lower tropopause, this value is used for both the lowest tropopause and the local tropopause.

2.2.17 Postfrontal Parameter

This parameter, *(PF)*, calculated using the following equation, is designed to measure forward wind shear in postfrontal situations.

$$Pf = \overline{U}_{mounta\,\mathrm{int}\,op} \left(\frac{\overline{U}_{mounta\,\mathrm{int}\,op}}{\overline{U}_{mid-high}}\right) \qquad (5)$$

where $\overline{U}$ is the magnitude of the normal-to-mountain component of the wind.

A commonly observed feature of windstorms is strong cross-mountain flow at the mountaintop associated with cyclogenesis in the Plains and weaker cross-mountain flow aloft with a near-zero thermal wind component. The dimensions of this parameter are meters per second.

2.2.18 Sangster Parameter

This parameter is closely related to the NW component of the geostrophic wind NW of Boulder, CO. According to Sangster (1977), this parameter is defined as:

$$Z_{850SLC} + Z_{850GJT} - 2Z_{850LND} \quad (6)$$

where Z_{850SLC}, Z_{850GJT}, *and* Z_{850LND} are the 850 hPa heights at SLC, GJT, and LND.
This parameter is currently used operationally in Boulder as a forecast tool for windstorm events. Units are in meters.

2.3 Models

2.3.1 Linear Regression Model

The linear regression model is the first of three types of models used to forecast peak wind gusts. Multiple linear regression was used due to the multivariate nature of the data set. The prediction equation used in multiple linear regression, according to Wilks (1995) is:

$$\hat{Y} = B_o + \sum_{i=1}^{k} B_k x_k \quad (7)$$

where x_k is each individual parameter and B is the coefficient for each predictor and is analogous with the slope. Each coefficient is calculated using the following:

$$B_i = \frac{n\sum_{j=1}^{n} x_{ij} y_{ij} - \sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} y_{ij}}{n\sum_{j=1}^{n} (x_{ij})^2 - (\sum_{j=1}^{n} x_{ij})^2} \quad (8)$$

where B_i is each coefficient, n is the number of observations, x_{ij} is each predictor from each row, and y_{ij} is the wind observation from each row.

An example of the solution obtained by multiple linear regression is given in Figure 2. In Figure 2, a prefrontal case from GJT at 0000 UTC was selected for analysis. This figure shows a plot of the Sangster parameter versus the peak wind gusts over 0-12 hours. A hypothetical regression line is plotted as well, to see how the linear regression would fit a line to the data.

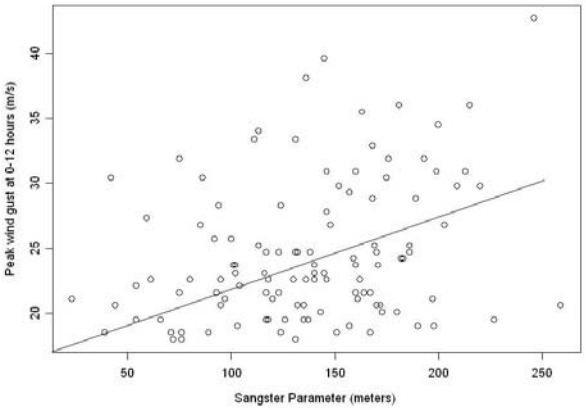

Figure 2. Scatter plot of Sangster parameter vs. Peak wind gusts for prefrontal GJT 0000 TC with hypothetical regression line.

2.3.2 Support Vector Regression Model

The support vector regression (SVR) model was selected since it has been shown to offer promising results in other studies (Richman et al. 2005) as well as its ability to model nonlinear input data. In SVR, the functional dependence of the dependent variable y is based on a set of independent variables x. Like other regression problems, the relationship between the independent and dependent variables is given by a deterministic function $f(x)$ plus the addition of some additive noise. The goal is to find a functional form for $f(x)$ that can correctly predict new cases that the SVR has not been presented with before. This can be achieved by training the SVR model on a sample set, i.e., training set, a process that involves the sequential optimization of an error function.

Consider a data set (x_i,y_i), where x is the input and y is the output. It is desired to find a function $f(x)$ that has ε deviation from y_i for all training data. For a basic example, consider a linear case, with the following as our function:

$$f(x) = < w, x > + b_i \quad (9)$$

The idea is to minimize the normalization of w, which is the weight, and this is found as:

$$\min(\frac{1}{2} \| w^2 \|) \textit{ such that}$$

$$y_i - < w, x_i > - b \le \varepsilon \textit{ and} \quad (10)$$

$$-y_i + < w, x_i > + b \le \varepsilon$$

For this case, it was assumed that there was a feasible result, so Lagrangian duality theory (Haykin 1999) was applied to obtain an equation for w, indicated by equation 11,

$$w = \sum_{i=1}^{n} (a_i - a_i^*) x_i \quad (11)$$

and substituting this result into equation 9:

$$f(x) = \sum (a_i - a_i^*) < x, x_i > + b_i \quad (12)$$

After obtaining equation 12, one can substitute a kernel function in for $<x,x_i>$, say $k(x,x_i)$. In this example, one can substitute this kernel function directly. For the case of nonlinear separability, each data point is mapped to a higher dimension using a feature map φ. This makes the dot product become $<\varphi(x),\varphi(x_i)>$, and a kernel function can be used to replace this dot product as well. There are three common kernel functions that are used with support vector regression, including:

1. polynomial $k(x, y) = (x^T y + 1)^p$ (13)

2. radial basis function $k(x,y) = e^{\frac{-1}{2\sigma^2}||x-y||^2}$ (14)

3. tangent hyperbolic

$$k(x, y) = \tanh(\beta_o x^T y + \beta_1) \quad (15)$$

For this study, the polynomial kernel function was chosen, as after several runs this kernel function was found to offer the most accurate predictions on the training data. After implementing the kernel function, a loss function is implemented. It is used to define the differences between f(x) and y.

$$C(f(x)-y)=(f(x)-y)^2 \quad (16)$$

By using this loss function, the problem solution is given by the following equation:

$$\max(w(a, a_i)) = \max[\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i - a_i^*)(a_j - a_j^*)k(x_i, x_j) + \sum_{i=1}^{n}(a_i - a_i^*)y_i - \frac{1}{2C}\sum_{i=1}^{n}(a_i^2 + a_i^{*2})]$$

subject to:

$$\sum_{i=1}^{n}(a_i - a_i^*) = 0 \quad (17)$$

By setting $\beta_i = (\alpha_i - \alpha_i^*)$, equation 17 becomes:

$$\max(w(a, a_i)) = \max[\sum_{i=1}^{n}\sum_{j=1}^{n}(\beta_i)(\beta_j)k(x_i, x_j) + \sum_{i=1}^{n}\beta_i y_i - (\frac{1}{2C}\sum_{i=1}^{n}(\beta_i^2)]$$

subject to

$$\sum_{i=1}^{n}\beta_i = 0 \quad (18)$$

This leads to the regression function being given by:

$$f(x) = \sum \bar{\beta}_i k(x_i, x) + \bar{b} \quad (19)$$

where $\bar{\beta}_i$ is the mean of β and $\bar{b}$ is:

$$\bar{b} = \overline{Y_i - \sum_{j=1}^{n}\beta_i k(x_i, x_j)}$$

Figure 3 shows a sample of support vector machines from Richman et al. (2005) in which a set of binary data are optimally separated into two classes in feature space, represented by 1 and -1. These numbers are applied to a support vector machine in an attempt to determine the optimal solution for the modeling function, represented by the dashed line in Figure 3. This figure is merely an illustration, as support vector regression does not classify data, but instead yields a modeled result that does not penalize small errors in the margin. In our case the forecast is a wind speed in meters per second.

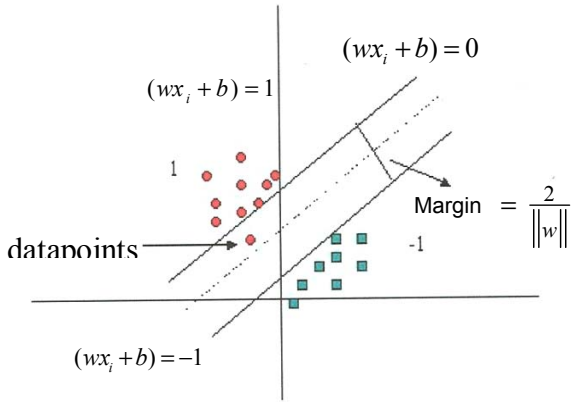

Figure 3. A geometric representation of Support Vector Regression. The dashed line is the optimal solution.

2.3.3 Neural Networks Model

The final form of model used in this study was the feed-forward neural network (Haykin 1999). The neural network uses a nonlinear technique to attempt to fit a curve to a dataset, as opposed to fitting a line. The neural network is named as it is thought to be similar to operation of the human brain. In the brain, data are entered into the body through nerve endings; the data run through a network of nerves and end up at the brain. The brain then processes the nerve input and yields an output to the senses. Similarly, the neural network first enters an input dataset through the input layer, then the data travels through the neurons to the hidden layer where the data are modeled, and the data finally exit through the output layer of the network as a modeled result.

The neural network operates by using an activation function to introduce nonlinearity into the system, therefore attempting to model peak winds in a nonlinear way. The primary goal of the neural network is to discover an optimal set of weighted neurons that model each predictor in the best way. The following shows the goal for the neural network in function form.

$$y_j(n) = \varphi_j(v_j(n)) \tag{20}$$

where y_i is the output of the neural network (in this case a predicted peak wind gust), φ_i is the activation function, and v_i is the input into the hidden layer, defined as:

$$v_i = \sum_{i=0}^{m} w_{ji}(n) x_j(n) \tag{21}$$

where m is the number of input data (in this case 18), x_j is the input data set, and w's the initial weights selected at random. The activation function selected for this study is the hyperbolic tangent activation function, which is defined as:

$$\varphi_j(x_j(n)) = \frac{1 - e^{-x_j}}{1 + e^{-x_j}} \tag{22}$$

The goal of the feed-forward neural network is to discover a value of w that will minimize the error:

$$\min(e_{ij} = \sqrt{(y_{ij} - d_{ij})^2}) \tag{23}$$

where e_{ij} is the root mean square error in the model and d_{ij} is the observed peak wind gust. According to Hayken (1999), in order to discover the optimal solution for the weights, an optimization technique must be used. In our work, this was the steepest descent method. This method converges on the optimal weights the fastest by calculating the gradient at each weight, thus determining optimal decrease or increase values. The following shows how this method works.

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \tag{24}$$

where $\Delta w_{ji}(n)$ is the change in the weights, η is a constant called the learning rate that determines how quickly the system converges, δ_j is the gradient, and y is the output from the network. To put this result in terms of the activation function, take the derivative and multiply it by the error. This leads to the following:

$$\Delta w_{ji}(n) = \eta e_{ij}(n) \varphi'(v_j(n)) y_i(n) \tag{25}$$

where $\varphi'(v_j(n))$ is the derivative of the activation function. After adjusting the weights with the determined result from the neural network, one epoch has been completed. Normally, several hundred epochs are required to determine an optimal solution.

2.4 Methodologies

One complication with using neural networks and support vector regression is that many of the variables used in each calculation are not set to specific values. Multiple experiments of each set

are required to determine the optimal solutions for each of the two models. For example, in this study, five hidden nodes were selected for the neural network, with a learning rate of 0.3, a hyperbolic tangent activation function, and 1000 epochs. This was found to be the most efficient for both computing resources and for modeling. In support vector regression, a cost function value C of 1 was selected, with a quadratic loss function and a polynomial kernel function. These values were determined after performing many experiments with varying values for the cost function and the quadratic loss function. One other concern is the variable nature of the neural network, as weights are determined randomly and several model runs are needed to determine a mean of modeled wind speeds, as well as to converge on an optimal solution. Figure 4 represents the neural network selected for modeling peak wind speeds.

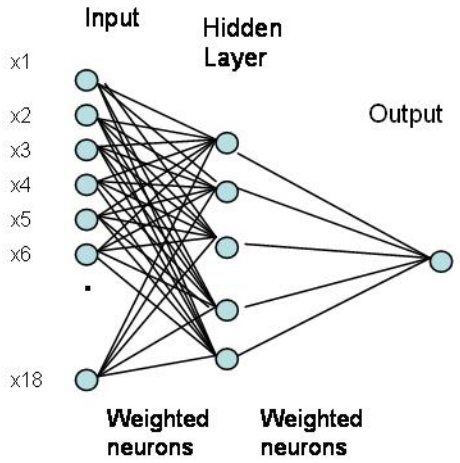

Figure 4: Architecture of neural network selected for this study. X values represent the 18 predictors; circles represent layers, which represent a run through the model; lines represent weighted neurons in which the data are optimally weighted.

Since raw sounding data were used to calculate the 18 predictors, it was common to find gaps in the sounding data. A Cardinal cubic spline was implemented to account for the holes in the soundings. A grid resolution of a 100 m resolution was used for the spline, with a spline tension coefficient of 0.5. All data were separated into prefrontal and postfrontal cases, as described previously. There was a further prescreening of data, as all data that had a 700 hPa wind direction that was not between 240° and 350° and all data that had a 700 hPa wind speed less than 8 ms^{-1} were removed (Brown 1986).

After obtaining data for each of the cases surviving the screening process, these were analyzed line by line in search for non-real values. These values were indicated as NA or Inf in the data matrix, and 99999 in the raw wind data. A large number of NA's were discovered in the integrated Scorer parameter and the static stability ratio because small superadiabatic layers were created by the cardinal splines. This screening device was designed to account for the fact that a superadiabatic layer leads to a Brunt-Väisälä frequency that is less than zero, and thus to a non-real integrated Scorer parameter and a non-real static stability ratio. This device would then find five levels that showed an increase in potential temperature with height, and set them to NA in the Scorer parameter and the static stability ratio. This device was considered unnecessary and removed, restoring many of the data points that were previously deleted. The remaining days with NA values were subsequently removed, leaving roughly 1000 cases for each prefrontal dataset and 200 test cases for each postfrontal dataset.

Another problem with the dataset was the abundance of 0 meter per second wind gusts. After investigating this problem, it was found that during data collection period all wind speeds that were below 10 ms^{-1} were recorded as 0 ms^{-1}. Since these data were all recorded 30 years ago, there was no access to the original wind data from these days. The lack of lower wind speed data proved troublesome when attempting to implement linear models on our data, since it led to a large bias in the results as well as poor residual values. A threshold was selected to remove all wind data that had peak gusts that were below 18 meters per second, classifying this threshold as a strong wind event. In doing this, we removed roughly 80% of our data, which made it more difficult to model. Table 1 shows how many data points were used in each model.

3. Results

The three models were tested on these data: a purely linear stepwise model using the "efroysom" (Insightful Corporation 2002) technique, a support vector regression model, and a neural network. Training and testing data were obtained by dividing the data sets into two halves. Each half was used in training, and the other half was used to test on the trained model. For example, the GJT prefrontal dataset for 0000 UTC had 230 data points. We divided the matrix of predictors for this case into two matrices of 115 points in each,

trained a model on the first half and tested on the second, then reversed the process. The training the 700 hPa cross-mountain wind in Denver. A complete table of the number of times each predictor appears in a model is included in Table 3 below. Other significant predictors included the Sangster parameter, the Froude height, 700 hPa temperature advection and the 500 hPa -700 hPa geostrophic wind shear.

Dataset name	0-12 h	0-6 h	6-12 h	0-3 h	3-6 h	6-9 h	9-12 h
prefrontal_gjt_0z	230	162	157	96	126	107	115
prefrontal_gjt_12z	222	137	170	97	104	128	116
postfrontal_gjt_0z	46	41	21	34	24	18	9
postfrontal_gjt_12z	58	39	41	21	30	38	26
prefrontal_lnd_0z	227	157	155	89	121	110	113
prefrontal_lnd_12z	223	137	171	92	59	126	117
postfrontal_lnd_0z	47	41	22	33	24	18	10
postfrontal_lnd_12z	59	39	42	22	30	39	28
prefrontal_lg_0z	232	162	159	95	59	36	115
prefrontal_lg_12z	230	144	176	101	109	131	122
postfrontal_lg_0z	48	42	23	34	25	19	10
postfrontal_lg_12z	57	38	40	21	29	37	25

Table 1: Number of data points used in each model

Training and testing data were selected at random using a random number program in Matlab (Mathworks, Inc. 2002). These data were used in each of the three statistical models listed previously. Table 2 lists the models used, as well as a number that corresponds to each number, which will be referenced in the results.

Model name	Number
GJT prefrontal 0000 UTC, first half	1
GJT prefrontal 1200 UTC, first half	2
GJT postfrontal 0000 UTC, first half	3
GJT postfrontal 1200 UTC, first half	4
LND prefrontal 0000 UTC, first half	5
LND prefrontal 1200 UTC, first half	6
LND postfrontal 0000 UTC, first half	7
LND postfrontal 1200 UTC, first half	8
GJT and LND prefrontal 0000 UTC, first half	9
GJT and LND prefrontal 1200 UTC, first half	10
GJT and LND postfrontal 0000 UTC, first half	11
GJT and LND postfrontal 1200 UTC, first half	12
GJT prefrontal 0000 UTC, second half	13
GJT prefrontal 1200 UTC, second half	14
GJT postfrontal 0000 UTC, second half	15
GJT postfrontal 1200 UTC, second half	16
LND prefrontal 0000 UTC, second half	17
LND prefrontal 1200 UTC, second half	18
LND postfrontal 0000 UTC, second half	19
LND postfrontal 1200 UTC, second half	20
GJT and LND prefrontal 0000 UTC, second half	21
GJT and LND prefrontal 1200 UTC, second half	22
GJT and LND postfrontal 0000 UTC, second half	23
GJT and LND postfrontal 1200 UTC, second half	24

Table 2. List of models and model numbers used in the figures.

3.1 Linear Model

S-Plus (Insightful Corporation 2002) was used to create the linear models that were designed to forecast peak wind gusts. After implementing a stepwise regression, often only a few predictors remained. This method showed that there were only two linear predictors that were present in many model runs, the 700 hPa wind speed and

Another method to determine the bias of each model is the median residual. Each residual was calculated as observed – predicted; thus, a negative value represents an over-forecast and a positive value represents an under-forecast. The median residual measures the bias of the model by showing the data point at the second quartile, which is a good measure of the central tendency of the data set. The median residual was selected for this study because the residual vectors contained numerous outliers. The mean would include these outliers, thus not providing a good interpretation of the central tendency of the models. Once a median was determined at each of the possible seven time increments, all increments were averaged to get a time-averaged statistic used for analysis of model performance. Figure 5 shows the time-averaged median residuals for these linear models. Each model number corresponds to Table 2, which shows the model name versus the model number. Although the models seemed to center around 0 ms, there was a slight bias towards the negative (in 18 out of the 24 models the bias was negative), implying that for linear regression the models tend to overforecast peak winds.

Predictor Name	Quantity
Temp advection	35
700 geos direction	17
700 geos magnitude	64
700 normal wind denver	56
700 normal wind local	29
500-700 geos shear	30
700 meas/700 geos ratio	20
700 meas - 700 geos	22
relative humidity	18
cross mountain height difference	25
static stability ratio	12
froude height	29
scorer parameter	15
characteristic impedence ratio	14
lowest tropopause level	22
local tropopause level	20
postfrontal parameter	29
sangster parameter	33

Table 3: Number of times each predictor appears in the linear regression model.

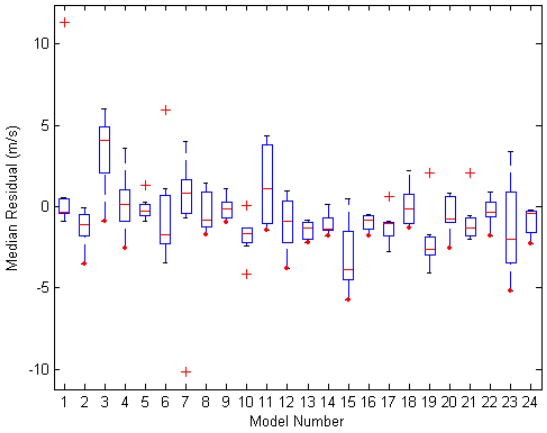

Figure 5. Time-Averaged boxplot of median residual in m/s vs. model number for the linear regression models. Numbers along the horizontal axis represent a particular model that is listed in table 3.

The parameter used to represent error was the root mean square error:

$$RMSE = \sqrt{(\hat{Y} - Y)^2} \qquad (26)$$

where $\hat{Y}$ is the observed peak wind gust and Y is the modeled, or where $\hat{Y} - Y$ is the residual winds in ms^{-1}. The models seemed fairly accurate, having a mean value of RMSE near 6.33 ms^{-1}. A plot of RMSE against the particular model number is given in Figure 6.

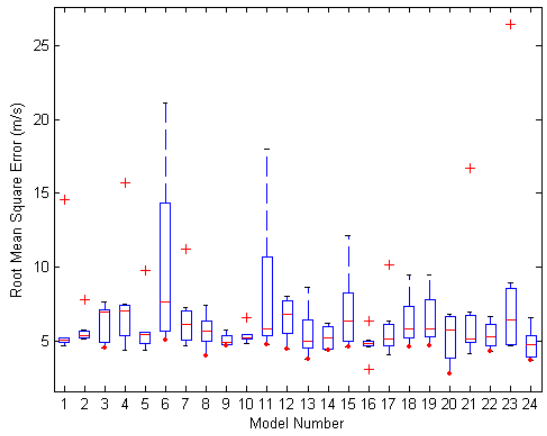

Figure 6. Same as Figure 5, but for RMSE instead of median residual.

Some problems with the linear regression model are observed when considering the RMSE values for a particular model. Some models had large outliers, as the LND value for a 6-12 hour prefrontal model run contained an RMSE value of 21.15 ms^{-1}. In part, this is attributable to high correlation between the predictors, as is shown in Table 4 below. Since values of correlation are desired to be below the absolute value of 0.5, the correlation between the 12th parameter and the 28th parameter was excessive (-0.71). These high correlations may have led to some degradation of results for the linear models, as these were improved upon using SVR.

Correlation of Coefficients	model 2 [12]	model 2[28]
model 2[12]		
model 2[28]	**-0.7083**	
model 2[20]	0.0831	-0.3344

Table 4: Sample correlation matrix. The model number comes from table 4, and the numbers in brackets represent the parameter selected. The parameters are numbered according to table 1.

3.2 Support Vector Regression Model

After applying linear models to the windstorm data, we implemented a Matlab (Mathworks, Inc. 2002) program to attempt to use SVR (Trafalis et al. 2003) to model the wind data. The code required training on half the data and testing on the other half.

Figure 7 indicates that these data, while showing no skewness, seem to center around 0 ms^{-1}, which implies that the model forecast with little bias. A very slight tendency towards overforecasting was observed, as more models seemed to report a negative value of median residual. However, this result is better than the result obtained from linear regression, as there are no large median residual values from using SVR.

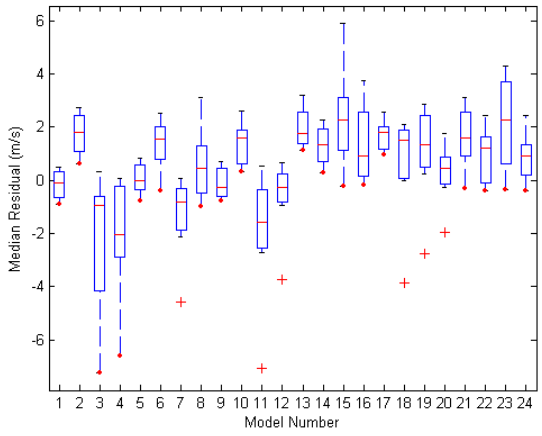

Figure 7. Same as Figure 5, but for support vector regression.

The mean value of RMSE for SVR from Figure 8 was 5.22 ms^{-1}, which is lower than the 6.33 ms^{-1} that was obtained from linear regression. Moreover, there are no significant outliers, as the largest value of RMSE is 11.69 ms^{-1}, and the minimum is also quite low at 3.42 ms^{-1}. Taken collectively, these results indicate that support vector regression was superior to multiple linear regression for modeling these data.

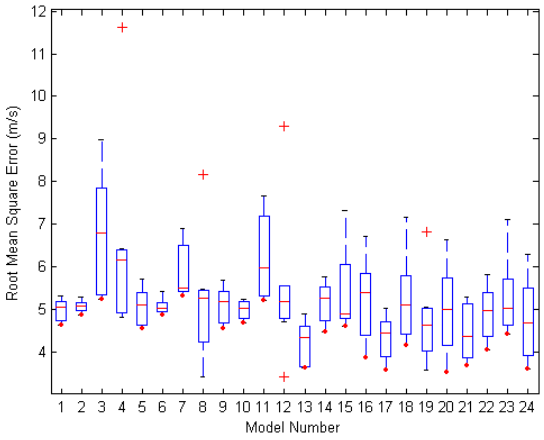

Figure 8. Same as Figure 6, but for support vector regression.

3.3 Neural Networks

A feedforward neural network (NN) (Haykin 1999) was applied to the windstorm data using the "newff" command in Matlab (Mathworks, Inc. 2002). Figure 9 shows a time-averaged median residual obtained from the artificial neural network from the 10 model runs, and Figure 10 illustrates the time-averaged RMSE results from the 10 runs.

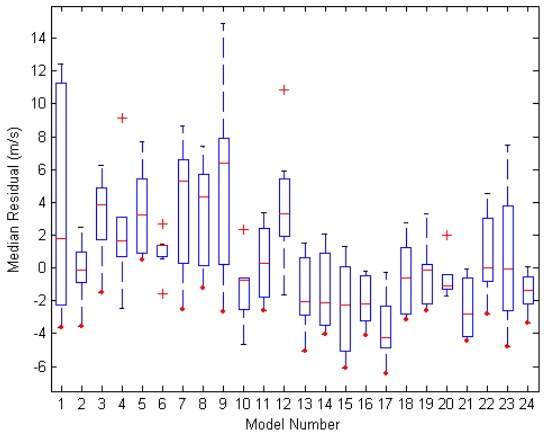

Figure 9. Same as Figure 5, but for neural networks.

Figure 10 indicates that the NN model overforecast peak winds when using the second half of the data set as a training set and the first half of the data set as a testing set. Ironically, it underforecast peak winds when using the first half of the dataset as a training set and the second half as a testing data set. The time-averaged RMSE values center near 7 ms^{-1}, which is comparable to linear regression for this dataset. Some extreme outliers were present in different NN model runs, including a 16.91 ms^{-1} RMSE value for the averaged dataset of LND and GJT at 0000 UTC for the 0-12 hour time step. One result of note is that the values of RMSE tended to be higher for a prefrontal case than for the postfrontal case.

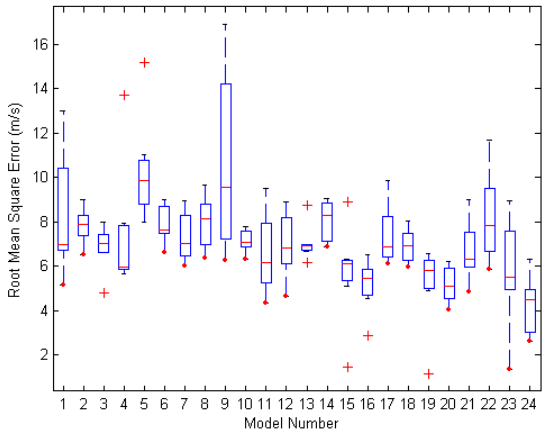

Figure 10. Same as Figure 6, but for neural networks.

4. Conclusions

The purpose of this study was to create models to be used to forecast peak wind gusts in the Boulder, Colorado area. These models were needed because we are currently unable to forecast downslope windstorm events well. The models created in this study will be given to the Boulder, Colorado National Weather Service office and be put into operation, in replacement of current forecasting techniques for downslope windstorms. Short range forecasting was also a result of this study, now peak wind speeds can be forecast only 3 hours in advance. Once a sounding is inserted into the model, peak wind speeds can be forecast with ease.

In the linear model, stepwise linear regression was used; using this model, significant predictors could have been ignored. It is difficult to see which predictors are the most

important, although Table 3 seems to indicate that 700 mb flow alone is most significant. Both support vector regression and the feedforward neural network did not filter out any predictors, but instead fit a non-linear function to all predictors, so that no important data were discarded. However, this study did not conclusively discover any set of these nineteen predictors that did a significantly better job of forecasting peak winds than any of the other predictors. Through this study, it was found that many different methods and predictors could be used to forecast peak winds successfully.

Through comparison of both RMSE and median residuals, it was found that our support vector regression model performed the best. 85% of our forecasts predicted maximum wind gusts with an RMSE of better than 6 ms^{-1}, and all of our forecasts predicted wind gusts with an RMSE of better than 12 ms^{-1}. Our linear model forecasted wind gusts better than 6 ms^{-1} 60% of the time, and better than 12 ms^{-1} 95% of the time. These results are a vast improvement to the techniques available to forecasters at this point. The result of this study is the knowledge that nonlinear modeling techniques such as support vector regression and neural networks are far superior to stepwise linear regression. These results can be implemented into an operational support vector regression model used by the Boulder National Weather Service office, an essential step resulting from this study.

Additionally, these results can be applied to different windstorm prone regions in the lee of the Rocky Mountains, determining the optimal set of predictors for a variety of regions and investigation of additional candidate predictors of peak wind gusts is important, so that models can be improved further.

Acknowledgements: Much of the work described herein was funded by COMET Grant # S-98-93860 through UCAR. We wish to thank Terry Simmons as well for providing data and assistance with this project.

REFERENCES

Bergen, W. R., and A. H. Murphy, 1978: Potential economic and social value of short range forecasts for Boulder windstorms. *Bull. Amer. Meteor. Soc.*, **59**, 29-44.

Blumen, W., 1985: Reflection of hydrostatic gravity waves in a stratified shear flow, Part 1: Theory. *J. Atmos. Sci.,* **42**, 2255-2263.

Brinkman, W. A. R., 1973: Strong downslope winds at Boulder, Colorado. *Mon. Wea. Rev.*, **102**, 592-602.

Brown, J. M., 1986: A decision tree for forecasting downslope windstorms in Colorado. Preprints, *11th Conference on Weather Forecasting and Analysis*, Kansas City, MO, Amer. Meteor. Soc., Boston, 83-88.

Clark, T. L. and W. R. Peltier, 1977: On the evolution and stability of finite amplitude mountain waves. *J. Atmos. Sci.*, **34**, 1715-1736.

Doyle, J. D., Durran, D. R., Chen, C., Colle, B. A., Georgelin, M., Grubisic, V., Hsu, W. R., Huang, C. Y., Landau, D., Lin, Y. L., Poulos, G. S., Sun, W. Y., Weber, D. B., Wurtele, M. G., Xue, M., 2000: An Intercomparison of Model-Predicted Wave Breaking for the 11 January 1972 Boulder Windstorm. *Mon. Wea. Rev.*, **128**, 901-914.

Durran, D. R. and J. B. Klemp, 1987: Another look at downslope windstorms. Part II: Nonlinear amplification beneath wave-overturning layers. *J. Atmos. Sci.*, **44**, 2341-2361.

Haykin, Simon, 1999: *Neural Networks: A Comprehensive Foundation.* Pearson Education, 842pp.

Insightful Corporation 2002: Splus. Version 6.1. University of Oklahoma.

Klemp, J. B. and D. K. Lilly, 1975: The dynamics of wave-induced downslope winds. *J. Atmos. Sci.*, **32**, 320-339.

Leptuch, P. A., 2001: Forecasting Downslope Windstorms in the Vicinity of Boulder, Colorado: an Empirical Statistical Technique. Masters Thesis, University of Oklahoma, 200 pp. [Available from Bizzell Library, University of Oklahoma, Norman, Oklahoma 73019.]

Mathworks, Inc. 2002: Matlab Version 6.5.0.180913a. University of Oklahoma.

Peltier, W. R. and T. L. Clark, 1979: The evolution of finite amplitude mountain waves. Part II: Surface wave drag and severe downslope windstorms. *J. Atmos. Sci.*, **36**, 1498-1529.

Richman, M.B., B. Santosa and T.B. Trafalis, 2005: Feature selection of radar-derived tornado attributes with support vector machines. *4th Conference on Artificial Intelligence Applications to Environmental Sciences*, San Diego, CA., Amer. Meteor. Soc., in press.

Sangster, W. E., 1977: An updated objective forecast technique for Colorado downslope winds. NOAA Technical Memorandum NWS CR-61, National Weather Service, Central Region, 24pp.

Scorer, R., 1949: Theory of waves in the lee of mountains. *Q. J. Roy. Meteor. Soc.*, **75**, 41-56.

Scorer. R., and H. Klieforth, 1959: Theory of mountain waves of large amplitude, *Q. J. Roy. Meteor. Soc.,* **85***,* 131-143.

Smith, R. B., 1985: On severe downslope winds. *J. Atmos. Sci.*, **42**, 2597-2603.

Trafalis, T. B., Santosa, B., and M. B. Richman, 2003: Prediction of Rainfall from WSR-88D Radar Using Kernel-based Methods. *International Journal of Smart Engineering System Design*, **5**, 429-438.

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