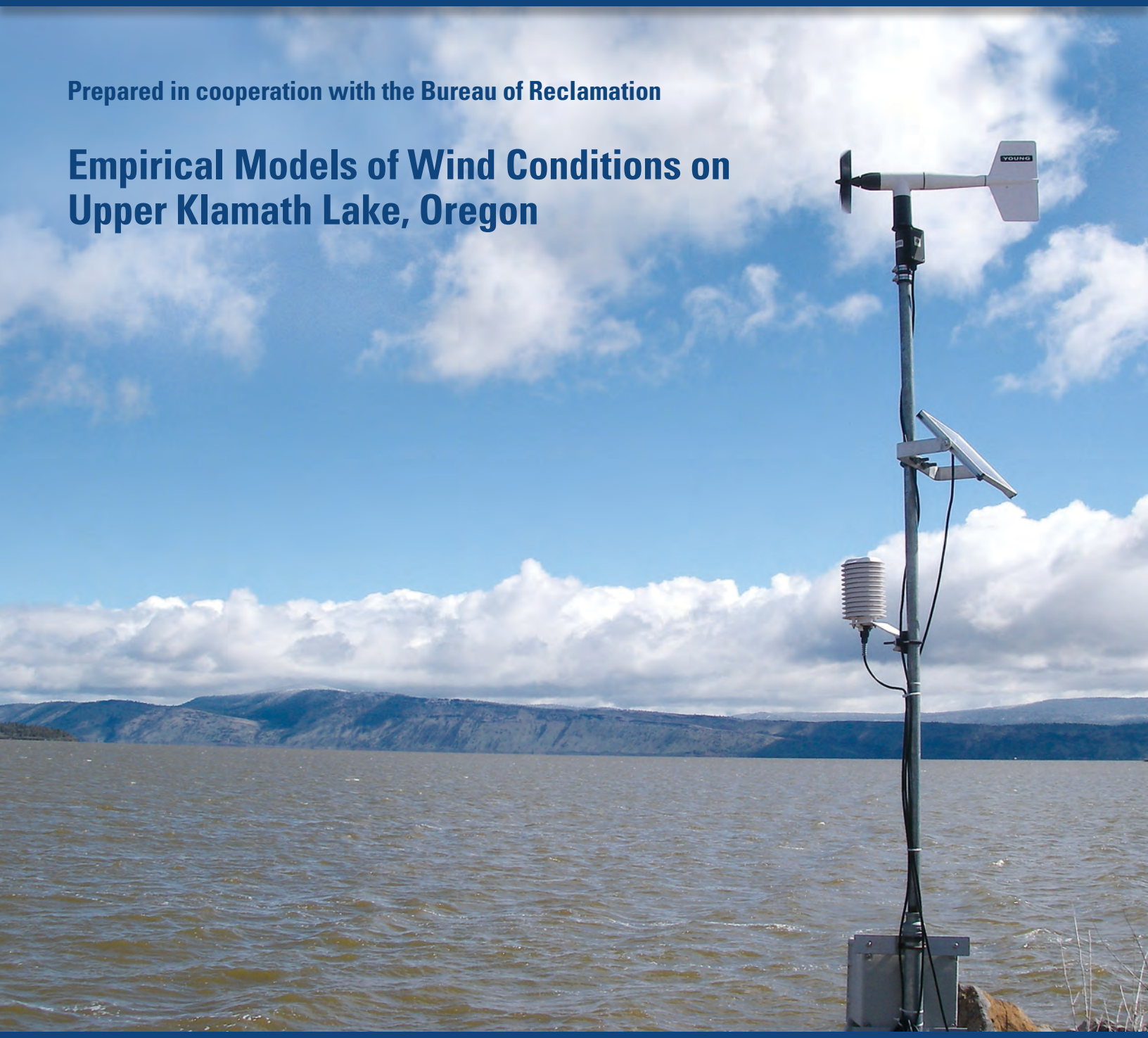
USGS
science for a changing world

Prepared in cooperation with the Bureau of Reclamation

# Empirical Models of Wind Conditions on Upper Klamath Lake, Oregon




Scientific Investigations Report 2010–5201

U.S. Department of the Interior
U.S. Geological Survey

**Cover:** Photograph of U.S. Geological Survey meteorologic station near Howard Bay in Upper Klamath Lake. (Photograph taken by Kristofor Kannarr, U.S. Geological Survey, 2010.)

# Empirical Models of Wind Conditions on Upper Klamath Lake, Oregon

By Norman L. Buccola and Tamara M. Wood

Prepared in cooperation with the Bureau of Reclamation

Scientific-Investigations Report 2010–5201

**U.S. Department of the Interior**
**U.S. Geological Survey**

**U.S. Department of the Interior**
KEN SALAZAR, Secretary

**U.S. Geological Survey**
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2010



Suggested citation:
Buccola, N.L., and Wood, T.M., 2010, Empirical models of wind conditions on Upper Klamath Lake, Oregon: U.S. Geological Survey Scientific-Investigations Report 2010–5201, 26 p.

# Contents



## Figures

# Tables



# Conversion Factors, Datums, and Abbreviations and Acronyms

## Conversion Factors

### Inch/Pound to SI

| Multiply | By | To obtain |
|---|---|---|
| | Length | |
| centimeter (cm) | 0.3937 | inch (in.) |
| meter (m) | 33.28 | foot (ft) |
| kilometer (km) | 0.621 | mile (mi) |
| | Velocity | |
| centimeter per second (cm/s) | 0.3937 | inch per second (in/s) |
| meter per second (m/s) | 33.28 | foot per second (ft/s) |
| meter per second (m/s) | 2.2369 | mile per hour (mi/h) |

## Datums

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Elevation, as used in this report, refers to distance above the Upper Klamath Lake Vertical Datum (UKLVD), which is used by the Bureau of Reclamation for reporting the elevation of Upper Klamath Lake.

# Conversion Factors, Datums, and Abbreviations and Acronyms—Continued

## Abbreviations and Acronyms

| | |
|---|---|
| ADCP | acoustic Doppler current profiler |
| AgriMet | Pacific Northwest Cooperative Agricultural Weather Network |
| ANN | Artificial Neural Networks |
| BIAS | Bias or Mean error |
| HIST | Historical Model |
| MARS | Multivariate Adaptive Regressive Splines |
| NASH | Nash-Sutcliffe error |
| NCDC | National Climatic Data Center |
| NOPPM | Un-filtered data |
| PPM1 | Pre-processing (filtering) method 1 |
| PPM2 | Pre-processing (filtering) method 2 |
| USGS | U.S. Geological Survey |

This page intentionally left blank.

# Empirical Models of Wind Conditions on Upper Klamath Lake, Oregon

By Norman L. Buccola and Tamara M. Wood

## Abstract

Upper Klamath Lake is a large (230 square kilometers), shallow (mean depth 2.8 meters at full pool) lake in southern Oregon. Lake circulation patterns are driven largely by wind, and the resulting currents affect the water quality and ecology of the lake. To support hydrodynamic modeling of the lake and statistical investigations of the relation between wind and lake water-quality measurements, the U.S. Geological Survey has monitored wind conditions along the lakeshore and at floating raft sites in the middle of the lake since 2005. In order to make the existing wind archive more useful, this report summarizes the development of empirical wind models that serve two purposes: (1) to fill short (on the order of hours or days) wind data gaps at raft sites in the middle of the lake, and (2) to reconstruct, on a daily basis, over periods of months to years, historical wind conditions at U.S. Geological Survey sites prior to 2005. Empirical wind models based on Artificial Neural Network (ANN) and Multivariate-Adaptive Regressive Splines (MARS) algorithms were compared. ANNs were better suited to simulating the 10-minute wind data that are the dependent variables of the gap-filling models, but the simpler MARS algorithm may be adequate to accurately simulate the daily wind data that are the dependent variables of the historical wind models. To further test the accuracy of the gap-filling models, the resulting simulated winds were used to force the hydrodynamic model of the lake, and the resulting simulated currents were compared to measurements from an acoustic Doppler current profiler. The error statistics indicated that the simulation of currents was degraded as compared to when the model was forced with observed winds, but probably is adequate for short gaps in the data of a few days or less. Transport seems to be less affected by the use of the simulated winds in place of observed winds. The simulated tracer concentration was similar between model results when simulated winds were used to force the model, and when observed winds were used to force the model, and differences between the two results did not accumulate over time.

## Introduction

The importance of wind in determining circulation patterns and water quality in Upper Klamath Lake has been well established (Laenen and LeTourneau, 1996; Kann and Welch, 2005; Wood and others, 2006; Holiman and others, 2008; Wood and others, 2008). The collection of wind data at or near the lake is, therefore, a critical contribution to studies of the lake's ecosystem and water quality. These studies address problems as diverse as (1) the occurrence of low dissolved-oxygen events (Kann and Welch, 2005; Wood and others, 2006), (2) the dependence of cyanobacterial vertical distribution on water-column stratification (J.W. Gartner, U.S Geological Survey, unpub. data, 2010), (3) the transport through the Williamson River Delta as a function of lake elevation, and wind speed and direction (T.M. Wood, U.S. Geological Survey, unpub. data, 2010), and (4) the transport of young larval suckers through the Williamson River Delta (Wood, 2009).

In 2005, as part of a cooperative study with the Bureau of Reclamation to determine water circulation patterns and heat transport in Upper Klamath Lake, several meteorological sites were established on the shoreline of Upper Klamath Lake and at two sites on rafts on the lake (fig. 1), to accurately describe the spatial variation of the wind over the surface of the lake. Wind data collected during the summers of 2005 and 2006 were used successfully as a forcing function in a hydrodynamic model of Upper Klamath Lake developed by Wood and others (2008). However, the collection of wind data since 2005 has been limited by the fact that the raft sites are in place only during the late spring through early fall months, May through September.

The ability to predict or, in the case of looking back in time, to "reconstruct" wind speed and direction at a site, would be useful for two reasons. First, the data needed to drive the hydrodynamic model are unavailable outside of May through September, which limits some applications of the model.

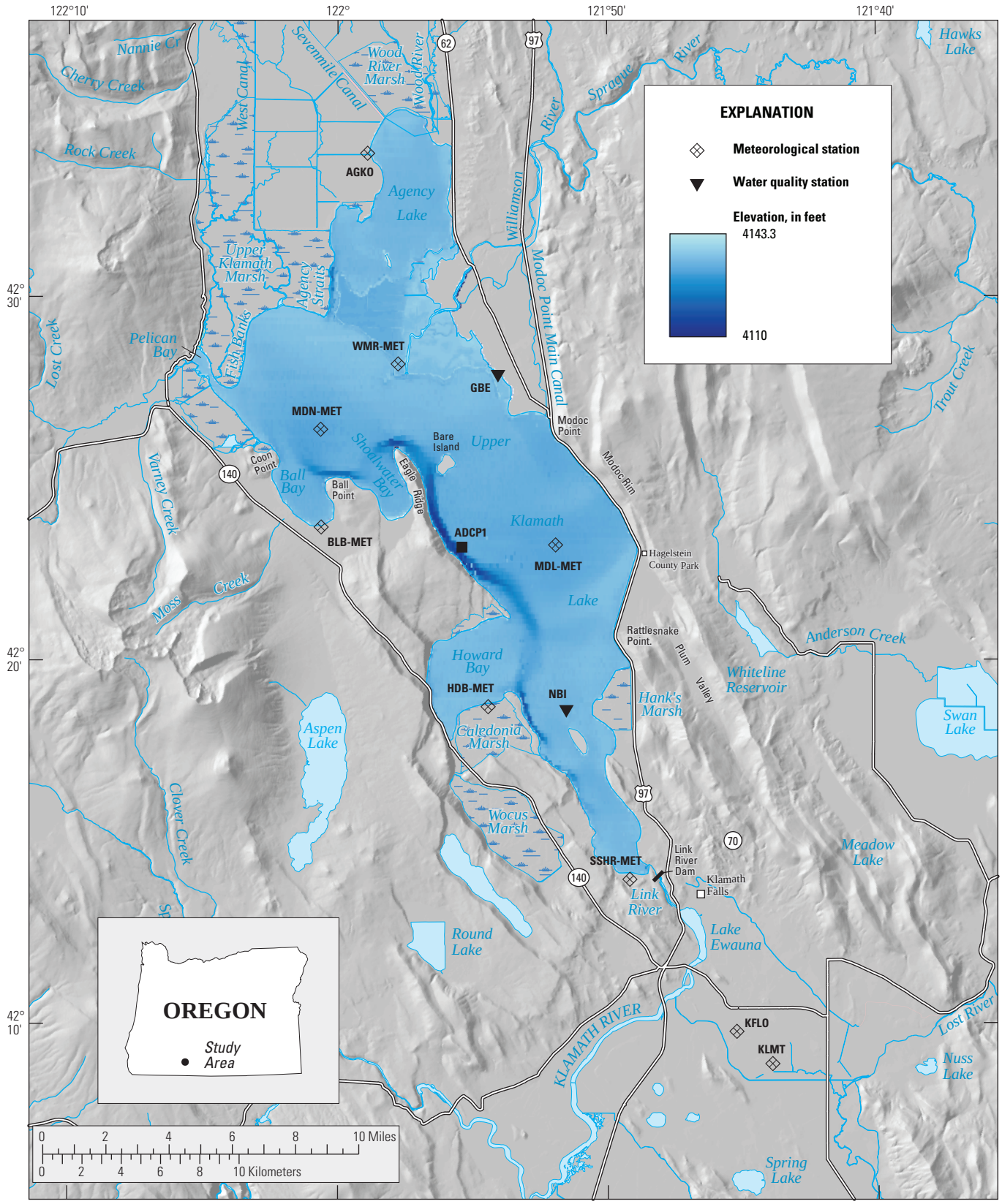



**Figure 1.** Upper Klamath Lake, Oregon, showing the location of sites used in this report.

For example, one application is to predict the passive transport of larval suckers through the Williamson River Delta and the lake. Larval suckers drift down the Williamson River in early spring, in some years prior to the availability of wind data at the two raft sites on the lake. Therefore, the ability to simulate early spring conditions at those sites a few days or weeks prior to the deployment of the rafts would make the hydrodynamic model more useful. In addition, there are occasional gaps in the wind records from May through October due to equipment failures that are too long to fill by simple interpolation. In these cases, reconstruction of the wind data is needed over a relatively short period of days or a few weeks, but at a high temporal resolution (at least hourly) to provide an accurate forcing function for the hydrodynamic model. The second reason that reconstructing the wind data at a site close to the lake would be useful is that it has the potential to extend the data record backward in time, prior to 2005, based on longer datasets collected elsewhere. This could provide several more years of overlap between wind data and water quality or other datasets for statistical analysis (for example, the Klamath Tribes water-quality monitoring began in 1986).

This study, done in cooperation with the Bureau of Reclamation, evaluated the accuracy of two multivariate, nonlinear, empirical methods for reconstructing the wind conditions at the raft sites on Upper Klamath Lake based on meteorological data collected at sites on the shoreline of the lake and at other locations in the basin. These two methods were Multivariate Adaptive Regressive Splines (MARS), a variation of linear regression, and Artificial Neural Networks (ANNs). MARS is a nonparametric technique that describes nonlinear dependencies on independent variables with piecewise linear segments of differing slope (Friedman, 1991). The change in slope occurs at a "knot" value that is part of the solution to the problem. Because the form of individual terms is linear, the dependencies of MARS models are simpler to understand and visualize than are those of the ANN models. The MARS models are built in two passes. On the forward pass, terms are added until the maximum specified number of terms is reached. On the backward pass, all existing terms are considered at each step and the least important term is removed, based on a generalized cross-validation comparison, which is a means of weighing goodness-of-fit against additional model complexity (Milborrow, 2009). With this recursive procedure, MARS selects the independent variables that constitute the best model.

Artificial Neural Networks have been used successfully in environmental studies to predict wind (Kretzschmar and others, 2004; Kulkarni and others, 2008), storm runoff (Shamseldin, 1997), and water-quality parameters, such as dissolved oxygen (Rounds, 2002), water temperature (Risley and others 2003), salinity (Conrads and Roehl, 1999; Conrads and others, 2006; Conrads and Roehl, 2007), sediment concentration (Rajaee and others, 2009), and pH (Cannon and Whitfield, 2001). ANNs rely on discerning and then "learning" the relations among variables based on many realizations of past events covering a large range in conditions, and therefore benefit from the large amount of data provided by the long records of hourly measurements collected at meteorological sites. Both ANN and MARS methods, being empirical in nature, allow data of varying types (for example, wind, air temperature, relative humidity, or solar radiation) to be tested as important explanatory variables without knowledge of a specific deterministic equation to describe the relation between dependent and independent variables, and do not require that the relations be linear. As such, these models are well suited to the current problem of reconstructing the wind data at a single site based on wind and other meteorological data collected nearby. The logical assumption is that the datasets are related, but the exact nature and physical description of that relation are unavailable.

Two sets of models were built for this study, and each set included the MARS and the ANN approach to the problem. The first set of models, used for reconstructing short periods of 10-minute wind data at the raft sites, is designated as the "gap-filling" models. The second set of models, used to reconstruct longer periods of daily wind data, is designated as the "historical" models.

## Purpose and Scope

This report presents the results of efforts to demonstrate the feasibility of models based on MARS and ANN to simulate the wind at two raft sites on Upper Klamath Lake and document their accuracy, using as input the meteorological variables measured at six other sites, four U.S. Geological Survey (USGS) sites on the shoreline of the lake, and two AgriMet sites located away from the shoreline. These models simulate the wind over periods of a few days to a week or more on a 10-minute basis to match the temporal resolution of the observations and are appropriate only for filling gaps in the data since 2005, when intensive collection of meteorological data around the shoreline of the lake began.

Second, this report presents the results of efforts to assess the feasibility and document the accuracy of models, based on MARS and ANN, to simulate the historical wind record at a single site on the lake shoreline on a daily basis. These models could be used to reconstruct the wind record at any lake site since 2000, using data from the long-term Pacific Northwest Cooperative Agricultural Weather Network (AgriMet) and the National Climatic Data Center (NCDC) sites located from 8.0 to 11.5 km from the shoreline of Upper Klamath Lake.

## Data Description

Wind measurements used in the gap-filling model were collected at six USGS meteorological sites—MDL, MDN, BLB, HDB, SSHR, WMR—as well as two nearby AgriMet sites (Bureau of Reclamation, 1997)—AGKO and KFLO (fig. 1). Meteorological instrumentation at sites MDL and MDN was mounted on floating rafts on the lake during the summer (approximately May through September); data were collected at the other sites continuously through the year. Construction associated with the breaching of levees around the Williamson River Delta dictated that the site near the mouth of the Williamson River be relocated twice during the period covered by this report. Site WMR was relocated to WME in May 2007, and WME was in turn relocated to WRW in April 2008. Data from WMR, WME, and WRW were used interchangeably in this report due to their close proximity; references to WMR wind data should be interpreted to mean data from the Williamson River Delta site, which was at one of these three locations, depending on the time period. During the periods used for calibration and validation of the gap-filling model, USGS data were measured at 10-minute intervals. AgriMet data were collected at 15-minute intervals, and for the purposes of the gap-filling wind model, were interpolated linearly (after a conversion to north-south and east-west components) to match the 10-minute intervals of the USGS data. The air temperature used in the gap-filling model was measured at site WMR and was collected at 10-minute intervals.

Datasets used in the historical wind model included the wind measured at the AgriMet sites AGKO and KFLO (daily mean speed and direction), as well as global solar radiation (cumulative daily), relative humidity (daily mean), and air temperature (daily mean) measured at AGKO. These variables were downloaded from the AgriMet website as daily values (Bureau of Reclamation, 1997). The daily mean wind speed and direction was then converted to east-west and north-south components, when appropriate, and depending on the model type. An additional dataset used in the historical model is the data collected since 1959 at the Klamath Falls Airport (KLMT) by the National Oceanic and Atmospheric Administration and distributed via the Internet by the NCDC (National Oceanic and Atmospheric Administration, no date). The KLMT site (fig. 1) is only about 2 km from the KFLO site, but the sensors at the AgriMet sites are at 2 m above the ground, whereas the sensors at the KLMT sites are at 10 m above the ground. Daily means of the wind components, air temperature, dew point temperature, sky cover, and altimetric pressure were calculated from the original NCDC data, which were collected at irregular intervals over the day, but generally at hourly temporal resolution. Daily means of sky cover were calculated after converting categorical values to numerical values as follows (adapted from ASOS user's guide [National Weather Service, 1998]): clear sky = 0, scattered clouds = 0.31, broken clouds = 0.75, overcast sky = 0.83, partial obscuration = 0.92, and obscuration = 1.0.

## Statistical Methods and Model Algorithms

Two statistical modeling algorithms were used and compared in both the gap-filling and historical models: Artificial Neural Networks (ANN) and Multivariate Adaptive Regressive Splines (MARS). ANNs provide a flexible method of relating input and output variables through an interconnected mesh of nonlinear transfer functions. The R-package *nnet* provided the feed-forward, single-hidden-layer ANN that was used in this work (Venables and Ripley, 2002). Nondefault parameters used in the *nnet* algorithm were *size* (number of units in the hidden layer) and *decay* (weight decay). The number of units in the hidden layer (*size* parameter) generally determines the complexity of the neural-network model and is relative to the complexity of the dataset that is being used to train or calibrate the model. The weight decay (*decay* parameter) is dependent on the *size* parameter, such that the former will not have much effect until the latter is of a sufficient value. Therefore, the minimum number of hidden-layer units (*size*) was determined first, as it primarily determined the complexity of the ANN. After *size* was determined, the optimum *decay* value was determined while holding *size* constant. The best model results were achieved through an optimization process in which *size* and *decay* were adjusted by trial-and-error iteratively until model-fit statistics from the calibration and validation time periods matched as closely as possible. The goal of this optimization process was to obtain parameter values, which resulted in models that could accurately simulate a wide range of wind conditions.

MARS is a multiple-regression technique that allows for nonlinearity between the dependent and independent variables described by a discrete change in slope between linear segments. MARS can achieve a closer fit to measured data using a larger number of input variables than least-squares regression while still utilizing recursive linear-regression techniques (Friedman, 1991). The R-package *earth* incorporates the MARS algorithms and has the ability to sort through a large initial set of input variables by removing input variables stepwise from this initial model framework until the most statistically effective subset of variables is left, resulting in a model that is not overfit (Milborrow, 2009). The nondefault parameter in the *earth* package that was manipulated for this work was *nk*, which is one of the criteria used to limit the number of input variables that are used to predict the dependent variable in the MARS model construction process. The value of *nk* was set to the smallest possible value (thereby limiting the complexity of the model) that did not show a substantial degradation in performance over a more complex model (one with a higher *nk* value). This iterative process was guided by minimizing the difference between fit statistics calculated over the calibration and validation time periods.

## Gap-Filling Wind Models

The dependent variables of the gap-filling models numbered four: the east-west and north-south components of the wind at two sites, MDN and MDL. The independent (explanatory) variables were the east-west and north-south components of the wind at six sites (HDB, BLB, WMR, SSHR, AGKO, KFLO), and air temperature at one site (WMR, fig. 1). Each time series was passed to both ANN and MARS models in three ways—with no preprocessing (NOPPM) and with two types of preprocessing with different degrees of smoothing. These two methods are denoted as preprocessing method 1 (PPM1), in which the original data were decomposed into low- and high-frequency components, and preprocessing method 2 (PPM2), in which an eigenvector filter was used to smooth the data (fig. 2).

The first preprocessing method (PPM1) consisted of decomposing each measured wind time series into low- and high-frequency components. The low-frequency component was a 24-hour moving average of the original data. The high-frequency component was obtained by subtracting the low-frequency component from the original data. This high-frequency component was smoothed using a 3-hour moving average. The low- and high-frequency components were passed to separate models that were used to simulate the low- and high-frequency components of the dependent variables with both ANN and MARS. Low-frequency model output was added to the list of possible inputs that were used in the high-frequency models. The final output time series was obtained by adding the two simulated components together (fig. 3).

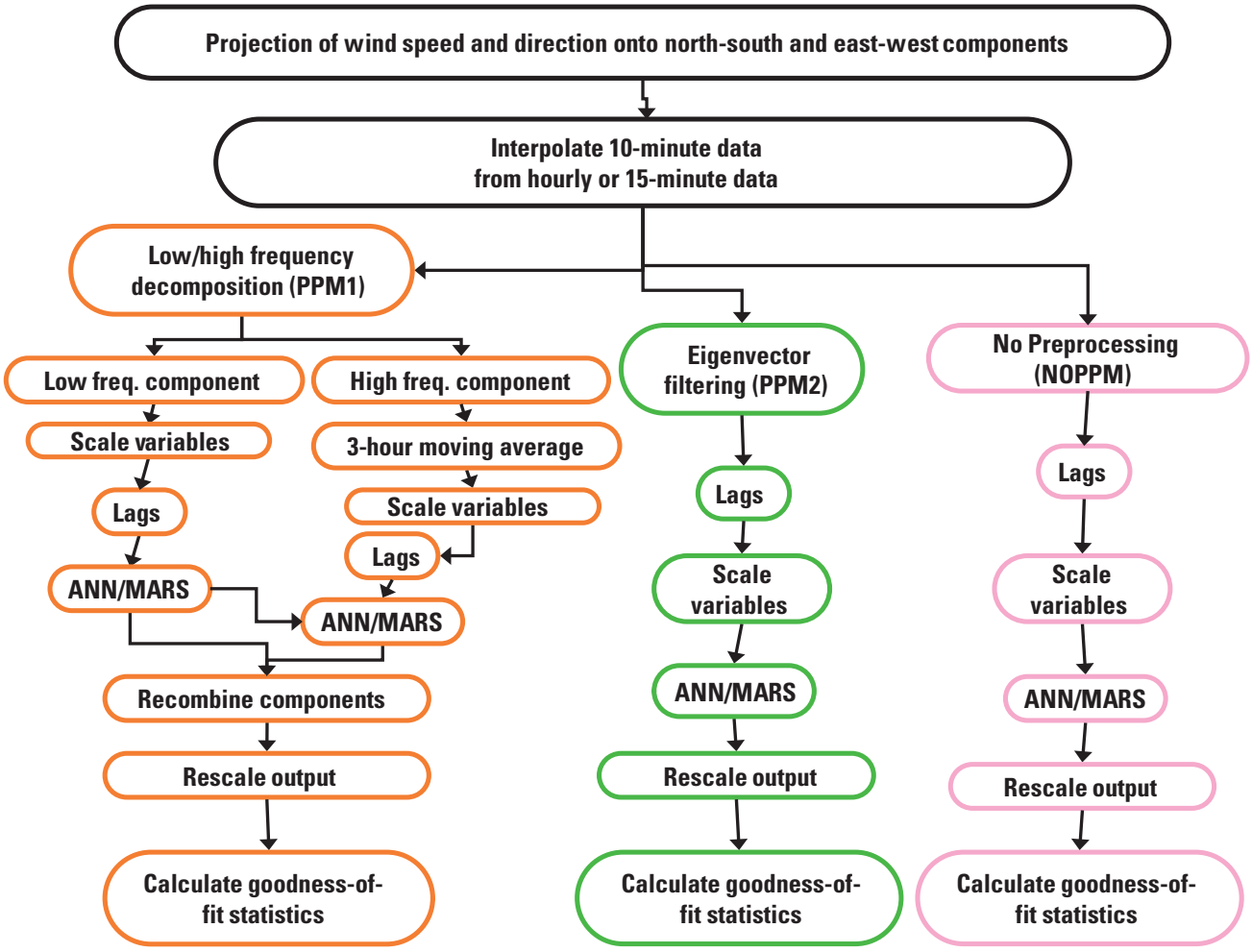



**Figure 2.** Steps in the construction of gap-filling models for wind time series on Upper Klamath Lake, Oregon.

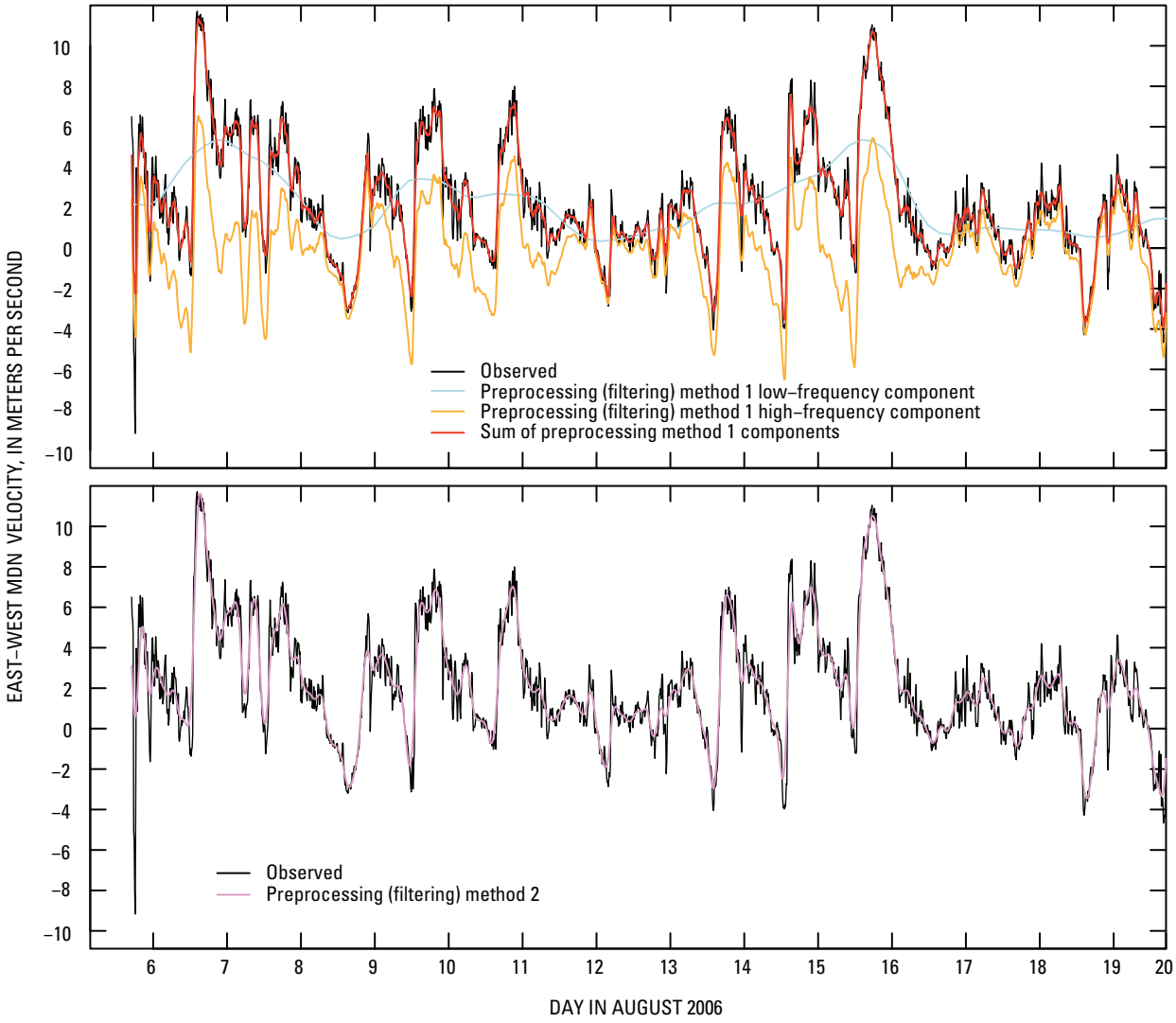



**Figure 3.** Examples of the original east-west components of the wind vector at site MDN on Upper Klamath Lake, Oregon, with the low- and high-frequency components obtained with preprocessing method 1 and the eigenvector smoothed version obtained with preprocessing method 2.

The second method of preprocessing the input data made use of an eigenvector filter, essentially performing a principal component analysis on the original time series and lagged copies of the time series. The function *decevf* in the R-package *pastecs* was used to perform this calculation (Ibañez and others, 2009). The parameter *lag*, in this filtering algorithm, was set to 4 hours after a trial-and-error optimization process. The goal of this filter optimization was to smooth each time series while minimizing the loss of important high-frequency fluctuations. The time series was then reconstructed using only the two most important eigenvectors, resulting in a smoothed version of the original time series. The preprocessed time series was then passed to both ANN and MARS models (fig. 2).

In order to incorporate the large-scale spatial features of the wind over the lake, time series of estimates of the two-dimensional divergence (the amount of spreading of the wind vectors) and curl (the amount of rotation of the wind vectors) of the wind were calculated and used as two additional inputs to the wind models. Based on a two-dimensional representation (at ground level) of the wind at sites WME (the northernmost lake sites) and SSHR (the southernmost lake sites), the divergence (*div*) and curl (*curl*) of the wind field $W(x,y)$ were calculated at each observation time as follows:

$$div(W) = \nabla \bullet W \approx \frac{(u_{WME} - u_{SSHR})}{\Delta x} + \frac{(v_{WME} - v_{SSHR})}{\Delta y}, \quad (1)$$

and

$$curl(W) = \nabla \times W \approx \frac{(u_{WME} - u_{SSHR})}{\Delta y} - \frac{(v_{WME} - v_{SSHR})}{\Delta x}, \quad (2)$$

where

$u_{WME}$ and
$v_{WME}$ are the east-west and north-south components, respectively, of the wind vector at WME,
$u_{SSHR}$ and
$v_{SSHR}$ are the east-west and north-south components, respectively, of the wind vector at SSHR,
$\Delta x$ is the longitudinal distance between the two sites,
$\Delta y$ is the latitudinal distance between the two sites,
• is the dot product, and
× is the cross-product of the two vectors shown.

The final preprocessing step was to scale the magnitude of each time series to the interval [0,1]. This assured that the range of all model inputs was the same, and in the case of the ANN model, that the range matched the range of the internal ANN output units (Venables and Ripley, 2002). Model inputs were scaled linearly using a method adopted from Rajee and others (2009):

$$x_s = \frac{(x - \min(x))}{(\max(x) - \min(x))}, \quad (3)$$

where

$x$ is the original time series,
$x_s$ is the scaled time series,
$\max(x)$ is the maximum of the time series, and
$\min(x)$ is the minimum of the time series.

This scaling was repeated for each time series; thus, the values of $\min(x)$ and $\max(x)$ were unique to each time series, and furthermore, were calculated for the calibration dataset and used unchanged in the validation dataset. The output of all models was postprocessed using equation 4:

$$x_o = x_s\left(\max(x) - \min(x)\right) + \min(x), \quad (4)$$

where

$x_o$ is the output time series.

In addition to preprocessing, lags were applied to each input variable to create additional model inputs that were mathematically independent. All variables were lagged by -6, +6, +12, or +24 hours, where a positive lag indicates that the time series was shifted forward in time and a negative lag indicates that it was shifted backward in time. All gap-filling models were calibrated using data from May 12, 2007, to September 29, 2007, and validated with data from May 12, 2006, to September 30, 2006. These dates bracketed the period when the rafts were located at sites MDN and MDL.

## Historical Wind Models

Two methods were used to simulate a daily wind record over a long period of time. The first method, denoted HIST1, used as a dependent variable the daily mean wind speed at WMR (fig. 1). For HIST1, the daily mean wind direction at the Klamath Falls Airport (KLMT), with a constant rotation, was used as a proxy for the wind direction at WMR. (For this purpose, the daily mean wind direction is defined as the direction of the vector whose orthogonal components are the daily mean of the east-west and north-south wind components.) The rotation of +5 degrees was determined from the distribution of the difference between the daily mean wind direction at WMR and KLMT over the calibration time period (fig. 4). The independent (explanatory) variables for this method were the daily mean wind speed, relative humidity, air temperature, and daily cumulative solar radiation measured at AGKO and KFLO, as well as the daily mean wind speed, sky cover, dew point temperature, air temperature, and altimetric pressure at KLMT.

The second method, denoted HIST2, used as dependent variables the daily mean of the east-west and north-south components of the wind at WMR and simulated each component separately (fig. 5). The independent variables for HIST2 were the daily mean east-west and north-south components of the wind, daily cumulative solar radiation, daily mean relative humidity, and daily mean air temperature measured at AGKO and KFLO, as well as, the daily mean east-west and north-south components of the wind, sky cover, dew point temperature, air temperature, and altimetric pressure at KLMT.

Both historical wind models were calibrated using data from 2006 through 2007 and validated using data from 2008 through 2009. Preprocessing for HIST1 and HIST2 methods consisted of scaling each input variable to [0,1], as was done for the gap-filling models. The scaled time series were then passed to both MARS and ANN models (eq. 3). Postprocessing consisted of rescaling the output using equation 4.

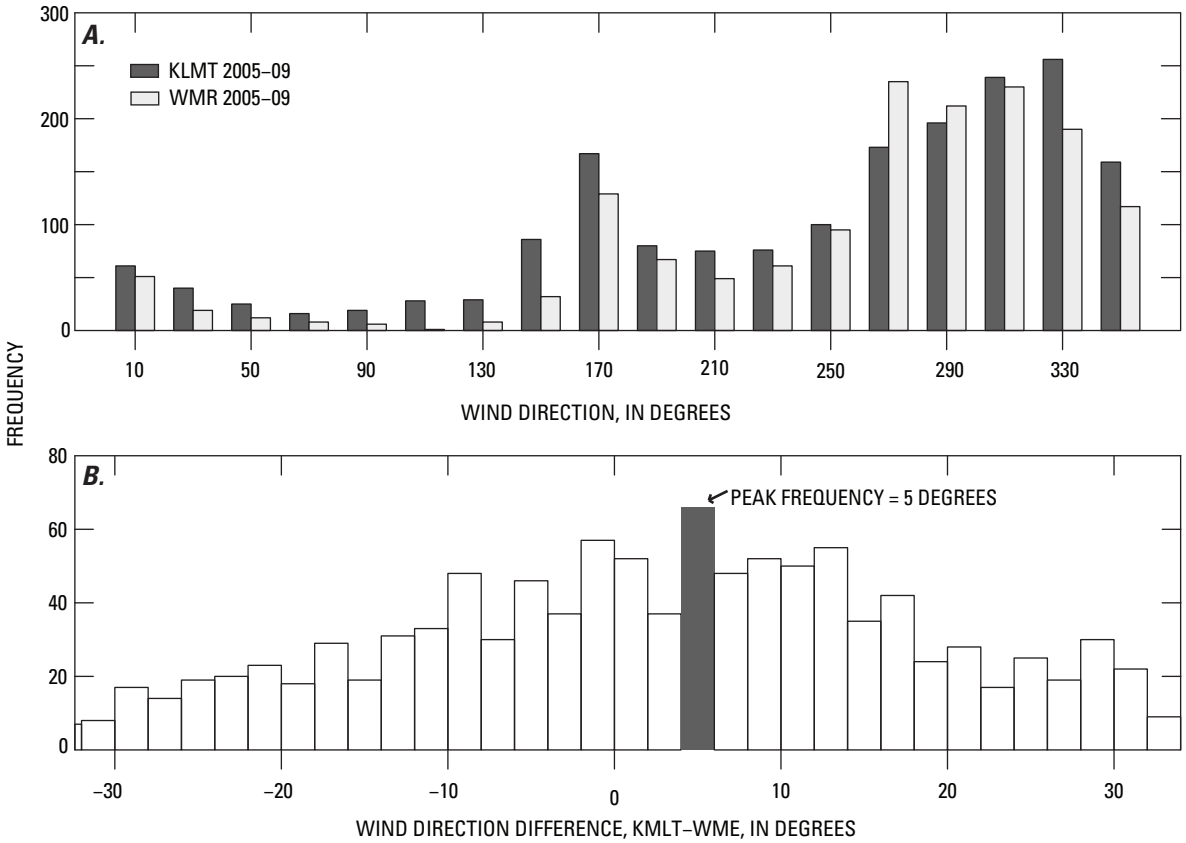



**Figure 4.** Wind direction at the mouth of the Williamson River (WMR) and Klamath Falls Airport (KLMT) near Upper Klamath Lake, Oregon (see fig. 1 for site locations): (*A*) daily mean wind direction at sites WMR and KLMT during 2005–09; (*B*) difference between the daily mean wind direction at KLMT and WMR during 2005–09. The daily mean wind direction is defined as the direction of the vector whose orthogonal components are the daily mean of the east-west and north-south wind components.

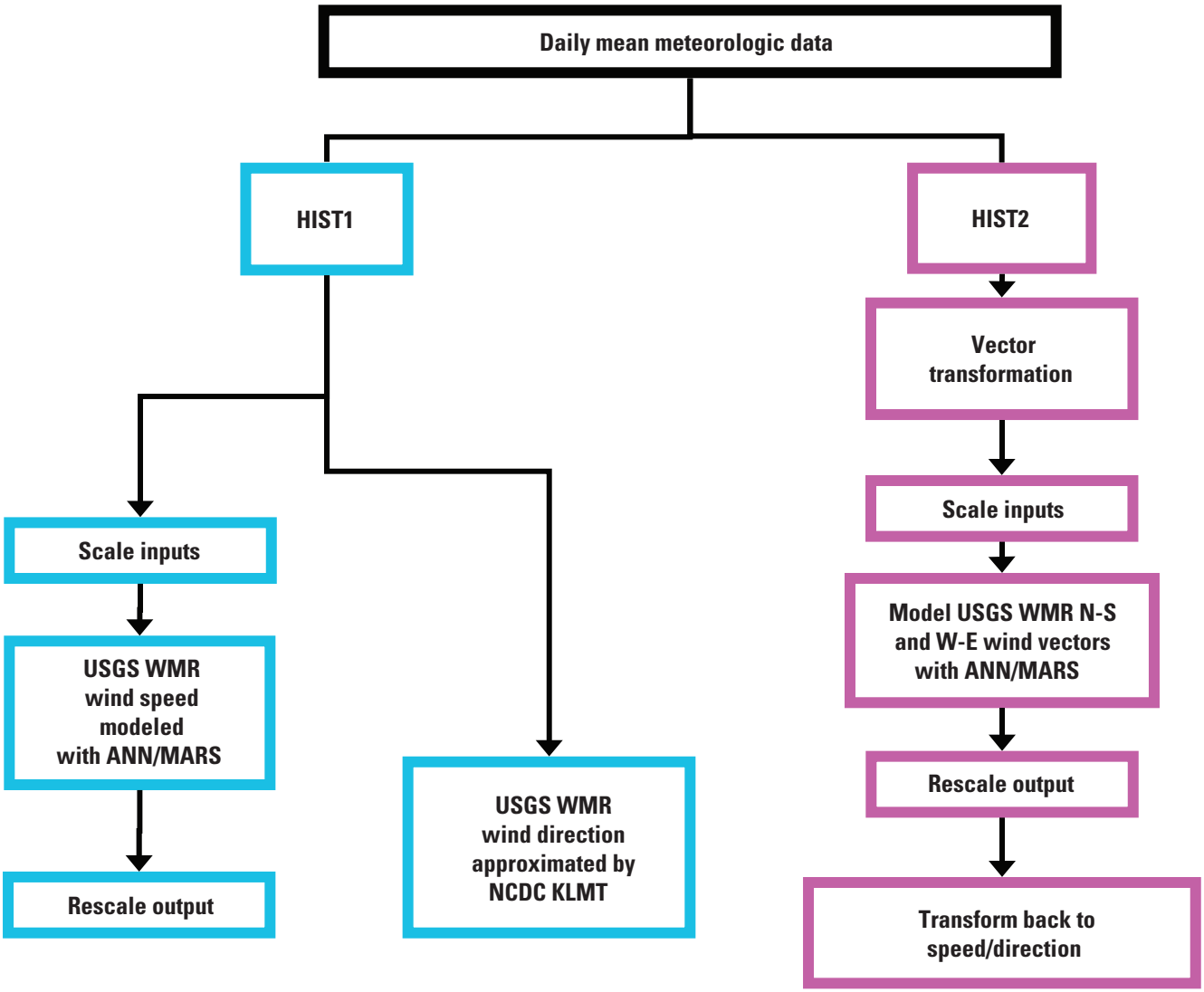



**Figure 5.** Steps in the construction of the historical wind models for Upper Klamath Lake, Oregon.

## Results of Gap-Filling Wind Models

The MARS algorithm was used to determine which explanatory variables were the most significant (based on optimized values of *nk*, table 1) and should be kept in each model. These selected variables were then used as inputs to both the ANN and MARS models. The optimized, nondefault values of parameters *size* and *decay* for the ANN models are provided in table 1.

A different set of the most significant explanatory variables was determined for each dependent variable and for each of the preprocessing methods used (table 2). Some features of the different sets are consistent with the geographic context. For example, site BLB is in the sets of explanatory variables for MDN models more than in the sets for MDL models. Similarly, site SSHR is in the sets of variables for MDL models more than in the sets for MDN models. These results reflect the closer proximity of BLB to MDN and SSHR to MDL. In general, the sets of variables in table 2 are notable for the number and variety of variables that contribute significantly to the models. The set for each dependent variable contains data from all sites around the lake, as well as the east-west and north-south components.

Goodness-of-fit statistics for the MARS models (MARS_PPM1, MARS_PPM2, and MARS_NOPPM) and the neural network models (ANN_PPM1, ANN_PPM2, and ANN_NOPPM) are shown in table 3. The Nash-Sutcliffe error (Weglarczyk, 1998), defined as $\mathrm{NASH} = 1 - \mathrm{MSE} / s_o^2$, where MSE is the mean-squared-error of the computed values, an $s_o^2$ is the variance of the observed values. NASH is a measure of the error between the simulated and observed time series normalized by the variability in the original time series, and approaches 1 as the mean-squared-error approaches zero. In general, NASH values greater than zero indicate a model fit that is more valuable than simply using the mean of the measured data. Likewise, NASH values less than zero indicate that a more accurate model could be achieved by using the mean of the measured data. The mean error, defined as $\mathrm{BIAS} = m_c - m_o$ (where $m_c$ = mean of simulated values and $m_o$ = mean of observed values), was calculated as a measure of the relative skewness of model results.

During calibration, ANN and MARS models resulted in similar fit statistics, whereas the validation period resulted in better statistical fits (indicated by higher NASH values and lower absolute BIAS values) for ANN models than for MARS models (table 3). The difference between validation and calibration of NASH values ranged from -0.07 to 0.05 for all ANN models, and from -0.26 to 0.06 for MARS models, indicating that ANN models resulted in a more consistent fit between validation and calibration periods than MARS models. This trend was most evident in the north-south MDL component MARS models during validation, where all three MARS models resulted in higher BIAS values than the north-south MDL ANN models. Overall, NASH values were higher (indicating more accurate fits) for models of the MDL north-south component than for the MDL east-west component, but the opposite was true for the models of the MDN north-south and east-west components. At site MDL, the north-south component was larger in magnitude than the east-west component, and at site MDN, the east-west component was larger in magnitude than the north-south component (figs. 6 and 7). Thus, the models generally do a better job of fitting the dominant component of the wind vector at both sites. Among all models, peak winds generally were underpredicted. Comparisons of the three most accurate models are shown in figures 6 and 7. Of these three models, peak winds tended to be simulated more accurately by the ANN_NOPPM model at both MDN and MDL, which captures the high-frequency signal better than the other models. Possibly indicative of this ability to model the high-frequency signal, the NOPPM models had a greater number of significant inputs and greater input variable diversity than the PPM1 and PPM2 models (table 2).

**Table 1.** Nondefault parameter values used in ANN and MARS wind models for Upper Klamath Lake, Oregon.

[**Abbreviations:** PPM1, preprocessing (filtering) method 1; PPM2, preprocessing (filtering) method 2; NOPPM, unfiltered data; n-s, north-south wind component; e-w, east-west wind component; ws, wind speed]

| Model algorithm | *R* function name | Parameter name | Gap-filling model preprocessing method | | | | | | | | | | | | Historic model method | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | PPM1 | | | | PPM2 | | | | NOPPM | | | | HIST1 | HIST2 | |
| | | | MDN | | MDL | | MDN | | MDL | | MDN | | MDL | | | | |
| | | | n-s | e-w | n-s | e-w | n-s | e-w | n-s | e-w | n-s | e-w | n-s | e-w | ws | n-s | e-w |
| MARS | *earth* | *nk* | 24 | | | | | | | | 32 | | | | 8 | 16 | 16 |
| ANN | *nnet* | *size* | 1 | 1 | 1 | 1 | 2 | 1 | 4 | 1 | 3 | 3 | 5 | 1 | 2 | 2 | 1 |
| | | *decay* | 1.0E-02 | | | | 1.0E-03 | | | | 1.0E-02 | 1.0E-04 | 1.0E-04 | 1.0E-03 | 1.0E-03 | 1.0E-02 | 1.0E-02 |

**Table 2.** Time series that were used as input to the gap-filling models for wind component vectors at sites MDN and MDL on Upper Klamath Lake, Oregon.

[**Abbreviations:** PPM1, preprocessing (filtering) method 1; PPM2, preprocessing (filtering) method 2; NOPPM, unfiltered data; n-s, north-south wind component; e-w, east-west wind component; air, air temperature; div, wind divergence; curl, wind curl; lfmod, low frequency model output]

| Order of importance as determined by MARS | north-south wind component at MDN | | | east-west wind component at MDN | | | north-south wind component at MDL | | | east-west wind component at MDL | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Site | Type of data | Time shift (hours) | Site | Type of data | Time shift (hours) | Site | Type of data | Time shift (hours) | Site | Type of data | Time shift (hours) |
| Dependent variables chosen for PPM1 low-frequency models | | | | | | | | | | | | |
| 1 | WMR | n-s | | WMR | e-w | | HDB | n-s | | HDB | n-s | |
| 2 | HDB | e-w | +6 | BLB | e-w | | AGKO | n-s | | WMR | e-w | |
| 3 | BLB | n-s | | HDB | n-s | | BLB | n-s | +6 | KFLO | e-w | |
| 4 | BLB | e-w | | WMR,SSHR | div | -6 | WMR | air | +24 | HDB | e-w | +6 |
| 5 | AGKO | e-w | +24 | WMR,SSHR | curl | | HDB | n-s | +6 | HDB | e-w | +24 |
| 6 | WMR | air | -6 | WMR,SSHR | curl | +6 | WMR | n-s | -6 | WMR | air | +24 |
| 7 | WMR,SSHR | div | -6 | BLB | n-s | -6 | WMR,SSHR | div | | AGKO | n-s | |
| 8 | WMR,SSHR | curl | | WMR | n-s | +24 | WMR | e-w | +24 | SSHR | n-s | +12 |
| 9 | HDB | n-s | | AGKO | n-s | +12 | WMR,SSHR | curl | +24 | HDB | n-s | +12 |
| 10 | SSHR | n-s | +12 | AGKO | e-w | +24 | SSHR | n-s | -6 | WMR,SSHR | curl | -6 |
| 11 | | | | HDB | e-w | +6 | | | | BLB | n-s | -6 |
| Dependent variables chosen for PPM1 high-frequency models | | | | | | | | | | | | |
| 1 | WMR | n-s | | WMR | e-w | | AGKO | n-s | | WMR | e-w | |
| 2 | BLB | n-s | | WMR | air | -6 | WMR | air | +6 | HDB | e-w | |
| 3 | BLB | e-w | | WMR | air | | HDB | n-s | | HDB | n-s | |
| 4 | WMR | e-w | +6 | BLB | e-w | | WMR | e-w | | KFLO | e-w | -6 |
| 5 | WMR | air | +24 | WMR | n-s | | WMR | air | | WMR | air | +6 |
| 6 | HDB | n-s | | BLB | e-w | +6 | WMR | n-s | | WMR | air | |
| 7 | KFLO | n-s | | BLB | n-s | | BLB | e-w | +12 | AGKO | e-w | |
| 8 | KFLO | e-w | | KFLO | e-w | | BLB | n-s | +6 | AGKO | n-s | +6 |
| 9 | WMR | e-w | | WMR,SSHR | div | +24 | SSHR | e-w | +12 | MDL-lfmod | n-s | +6 |
| 10 | WMR | e-w | -6 | KFLO | e-w | +6 | BLB | n-s | -6 | KFLO | e-w | +6 |
| 11 | | | | HDB | n-s | | | | | KFLO | n-s | +6 |

**Table 2.** Time series that were used as input to the gap-filling models for wind component vectors at sites MDN and MDL on Upper Klamath Lake, Oregon.—Continued

[**Abbreviations:** PPM1, preprocessing (filtering) method 1; PPM2, preprocessing (filtering) method 2; NOPPM, unfiltered data; n-s, north-south wind component; e-w, east-west wind component; air, air temperature; div, wind divergence; curl, wind curl; lfmod, low frequency model output]

| Order of importance as determined by MARS | north-south wind component at MDN | | | east-west wind component at MDN | | | north-south wind component at MDL | | | east-west wind component at MDL | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Site | Type of data | Time shift (hours) | Site | Type of data | Time shift (hours) | Site | Type of data | Time shift (hours) | Site | Type of data | Time shift (hours) |
| Dependent variables chosen for PPM2 models | | | | | | | | | | | | |
| 1 | WMR | n-s | | WMR | e-w | | AGKO | n-s | | HDB | e-w | |
| 2 | BLB | n-s | | WMR | air | | HDB | n-s | | HDB | n-s | |
| 3 | BLB | e-w | | WMR | air | +6 | BLB | n-s | +6 | WMR | e-w | |
| 4 | WMR | e-w | | BLB | e-w | | WMR | n-s | | KFLO | e-w | +6 |
| 5 | HDB | n-s | -6 | BLB | n-s | | WMR,SSHR | div | | BLB | n-s | +12 |
| 6 | WMR | e-w | +6 | WMR | n-s | | AGKO | n-s | -6 | SSHR | n-s | -6 |
| 7 | KFLO | e-w | | AGKO | e-w | +6 | AGKO | e-w | | AGKO | e-w | |
| 8 | KFLO | n-s | | AGKO | e-w | | WMR | e-w | | WMR | air | -6 |
| 9 | HDB | n-s | | WMR,SSHR | div | +24 | AGKO | n-s | +12 | KFLO | n-s | +12 |
| 10 | WMR | e-w | -6 | WMR | air | -6 | | | | AGKO | e-w | +6 |
| 11 | HDB | e-w | | KFLO | e-w | +12 | | | | AGKO | n-s | +24 |
| Dependent variables chosen for NOPPM models | | | | | | | | | | | | |
| 1 | WMR | n-s | | WMR | e-w | | HDB | n-s | | HDB | e-w | |
| 2 | BLB | n-s | | BLB | e-w | | AGKO | n-s | | WMR | e-w | |
| 3 | BLB | e-w | | WMR | air | | WMR | n-s | | HDB | n-s | |
| 4 | HDB | n-s | -6 | WMR | air | +6 | WMR | air | +6 | KFLO | e-w | +6 |
| 5 | WMR | e-w | | AGKO | e-w | | WMR,SSHR | div | | AGKO | e-w | |
| 6 | WMR | e-w | +6 | WMR | n-s | | WMR | air | | WMR | air | |
| 7 | HDB | n-s | | BLB | n-s | | HDB | n-s | +6 | WMR | air | +6 |
| 8 | WMR | e-w | -6 | AGKO | e-w | +6 | AGKO | n-s | +6 | SSHR | e-w | |
| 9 | KFLO | e-w | | SSHR | n-s | | SSHR | e-w | | AGKO | n-s | +24 |
| 10 | KFLO | n-s | | AGKO | n-s | | SSHR | n-s | -6 | KFLO | n-s | +12 |
| 11 | HDB | e-w | | WMR | air | -6 | AGKO | n-s | +12 | AGKO | e-w | +6 |
| 12 | AGKO | e-w | | WMR,SSHR | div | +24 | WMR | e-w | | SSHR | n-s | -6 |
| 13 | WMR | air | +24 | KFLO | e-w | | KFLO | n-s | +12 | BLB | e-w | |
| 14 | AGKO | n-s | -6 | WMR | air | +12 | BLB | e-w | +6 | KFLO | e-w | +12 |
| 15 | HDB | n-s | +12 | | | | | | | KFLO | e-w | |

**Table 3.** Goodness-of-fit statistics for the Upper Klamath Lake, Oregon, gap-filling wind models during the calibration (May 12–September 29, 2007) and validation periods (May 12–September 30, 2006).

[NASH and BIAS are the Nash-Sutcliffe error and bias error respectively, between the simulated and observed time series. NASH values are unit-less, BIAS values are in meters per second. **Abbreviations:** PPM1, preprocessing (filtering) method 1; PPM2, preprocessing (filtering) method 2; NOPPM, unfiltered data; n-s, north-south wind component; e-w, east-west wind component]

| Model | NASH | | | | BIAS | | | |
|---|---|---|---|---|---|---|---|---|
| | MDN | | MDL | | MDN | | MDL | |
| | n-s | e-w | n-s | e-w | n-s | e-w | n-s | e-w |
| Calibration | | | | | | | | |
| ANN_PPM1 | 0.44 | 0.83 | 0.70 | 0.68 | 0.00 | 0.00 | 0.00 | 0.00 |
| ANN_PPM2 | .48 | .83 | .73 | .67 | .00 | .00 | .01 | .01 |
| ANN_NOPPM | .49 | .84 | .75 | .67 | .00 | .00 | .01 | .01 |
| MARS_PPM1 | .51 | .84 | .74 | .69 | .00 | .00 | .00 | .00 |
| MARS_PPM2 | .49 | .84 | .72 | .68 | .00 | .00 | .00 | .01 |
| MARS_NOPPM | .47 | .82 | .70 | .67 | .00 | .00 | .00 | .01 |
| Validation | | | | | | | | |
| ANN_PPM1 | 0.49 | 0.78 | 0.75 | 0.62 | 0.17 | 0.09 | -0.01 | -0.12 |
| ANN_PPM2 | .53 | .79 | .70 | .60 | .25 | .04 | .04 | -.06 |
| ANN_NOPPM | .53 | .78 | .76 | .61 | .16 | .09 | .02 | -.04 |
| MARS_PPM1 | .25 | .76 | .70 | .58 | .04 | -.10 | .27 | -.02 |
| MARS_PPM2 | .51 | .78 | .67 | .61 | .17 | -.01 | .28 | -.07 |
| MARS_NOPPM | .53 | .76 | .73 | .51 | .17 | .07 | .13 | -.12 |
| Validation-calibration | | | | | | | | |
| ANN_PPM1 | 0.05 | -0.05 | 0.05 | -0.06 | 0.18 | 0.09 | -0.02 | -0.12 |
| ANN_PPM2 | .05 | -.04 | -.03 | -.07 | .24 | .04 | .03 | -.07 |
| ANN_NOPPM | .04 | -.06 | .00 | -.06 | .16 | .09 | .01 | -.05 |
| MARS_PPM1 | -.26 | -.08 | -.04 | -.12 | .04 | -.10 | .27 | -.03 |
| MARS_PPM2 | .01 | -.06 | -.05 | -.07 | .17 | -.01 | .27 | -.08 |
| MARS_NOPPM | .06 | -.06 | .03 | -.16 | .17 | .07 | .13 | -.13 |

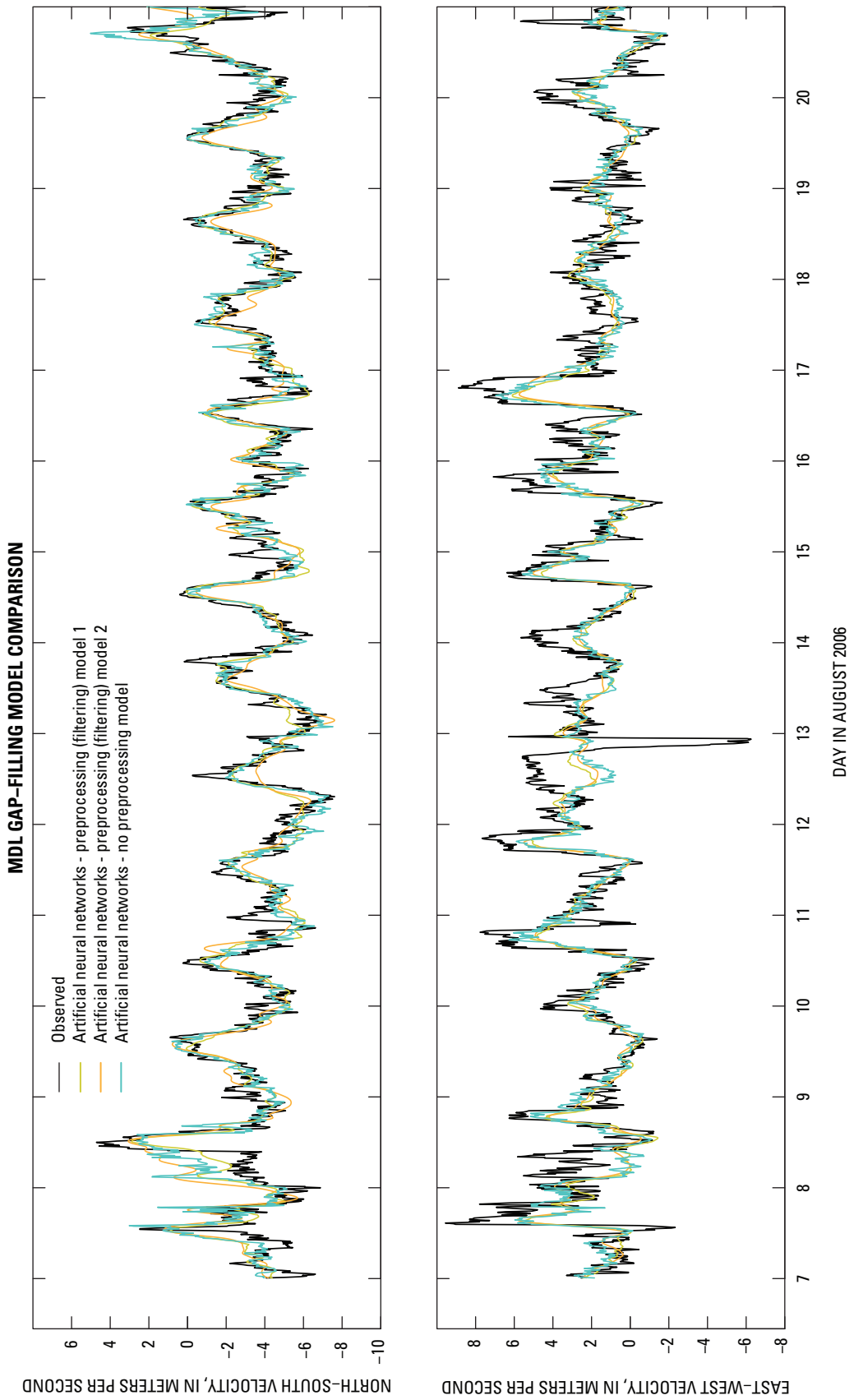



**Figure 6.** Simulated and observed east-west and north-south components of the wind vector at site MDL on Upper Klamath Lake, Oregon.

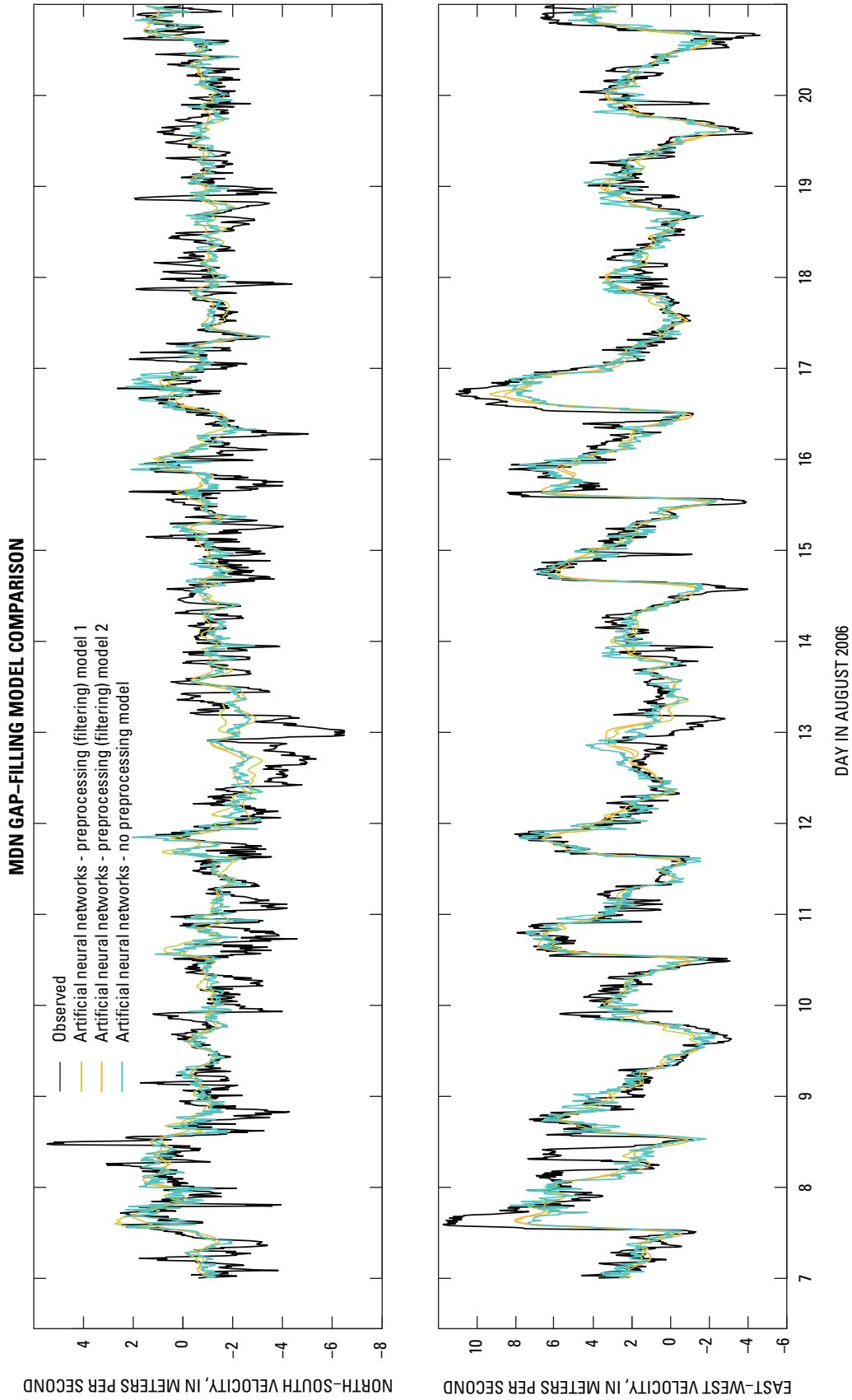



**Figure 7.** Simulated and observed east-west and north-south components of the wind vector at site MDN on Upper Klamath Lake, Oregon.

One objective of the gap-filling models is to generate wind data at the raft sites on the lake for use in a spatial interpolation that incorporates data from these sites as well as sites on the shoreline into a spatially variable wind field to drive the hydrodynamic model. A measure of how well the gap-filling models meet this objective is how well the hydrodynamic model, using the wind forcing created using the simulated winds over the lake, is able to simulate observed water currents. The most accurate gap-filling models (ANN models) were used to simulate the winds at sites MDN and MDL from mid-July through August 2006. These simulated winds were then used in combination with the measured winds at the other four sites around the lake to create a spatially variable wind. The three-dimensional UnTRIM hydrodynamic model of the lake, which is described in Wood and others (2008), was then used to simulate water currents and water temperature with this new wind forcing. The period July 26–August 31, 2006, was selected because this was the period selected previously for model validation (Wood and others, 2008). Goodness-of-fit statistics for currents simulated using only observed winds and observed winds at shoreline sites in combination with winds simulated with the gap-filling models were compared to site ADCP1 (table 4, fig. 1), where measurements from an Acoustic Doppler Current Profiler (ADCP) were available (Gartner and others, 2007).

The root-mean-square error and particularly the bias as measured by the mean error increased substantially with the use of the simulated winds at MDN and MDL. From the comparison of the simulated currents (fig. 8), it seems that the increased error largely is attributable to weaker peak currents resulting from the model forced with simulated winds. This is true on a daily basis, but particularly is noticeable during times of high velocities. Simulated currents derived from observed winds also underpredicted measured currents, but the error is greater when the simulated winds are used to force the hydrodynamic model. The increase in the mean error (bias) when the best ANN wind model is used to simulate the winds is 0.7 cm/s, and the increase in the root-mean-squared error is about 1.2 cm/s. In general, the gap-filling models seem to be adequate to fill data gaps over short periods of time (a few days or less). Given the increase in the error statistics, the use of the gap-filling models over periods of many days should be evaluated in the context of the problem being addressed.

**Table 4.** Upper Klamath Lake UnTRIM model goodness-of-fit velocity statistics at site ADCP1, Upper Klamath Lake, Oregon.

[The time period for the calculation is July 26–August 31, 2006. Data are in centimeters per second. **Abbreviations:** NOPPM, unfiltered data; PPM1, preprocessing (filtering) method 1; PPM2, preprocessing (filtering) method 2]

| Type of wind | Mean error | Root-mean-square error |
|---|---|---|
| Winds as observed | 2.3 | 3.9 |
| ANN_NOPPM | 3.0 | 5.1 |
| ANN_PPM1 | 3.2 | 5.3 |
| ANN_PPM2 | 3.4 | 5.5 |

The accurate simulation of the water currents in the lake is important in part because the transport of water-quality constituents and passively drifting larval fish is determined by the water currents. Errors in the simulation of currents will propagate into errors in the simulation of transport. A numerical tracer was used to assess the difference in simulated transport resulting from the use of the simulated winds instead of the observed winds at the two raft sites. The numerical tracer was initialized to zero at the beginning of the simulation. The only source of the tracer during the simulation was the Williamson River, where the tracer was put into the lake at a concentration of 10 (units are arbitrary). The time evolution of the concentration of the tracer at two sites in the lake, in Goose Bay (GBE) and north of Buck Island (NBI; fig. 1), is shown in figure 9. Overall, the tracers derived from simulated winds closely matched those derived from observed winds. Specifically, the mean absolute difference over the 47-day simulation between the concentration of the tracer as simulated using only observed winds and the concentration of the tracer as simulated using simulated winds at the raft sites was 0.052, 0.031, and 0.047 (units are arbitrary) at site GBE for ANN_NOPPM, ANN_PPM1, and ANN_PPM2, respectively. At site NBI, the mean absolute difference was 0.022, 0.023, and 0.011 for ANN_NOPPM, ANN_PPM1, and ANN_PPM2, respectively. Equally as important as the mean absolute difference is the lack of error propagation in the simulated tracers through time. That is, small differences occurring earlier in the tracer concentration time series do not compound and result in much larger differences later in the simulation.

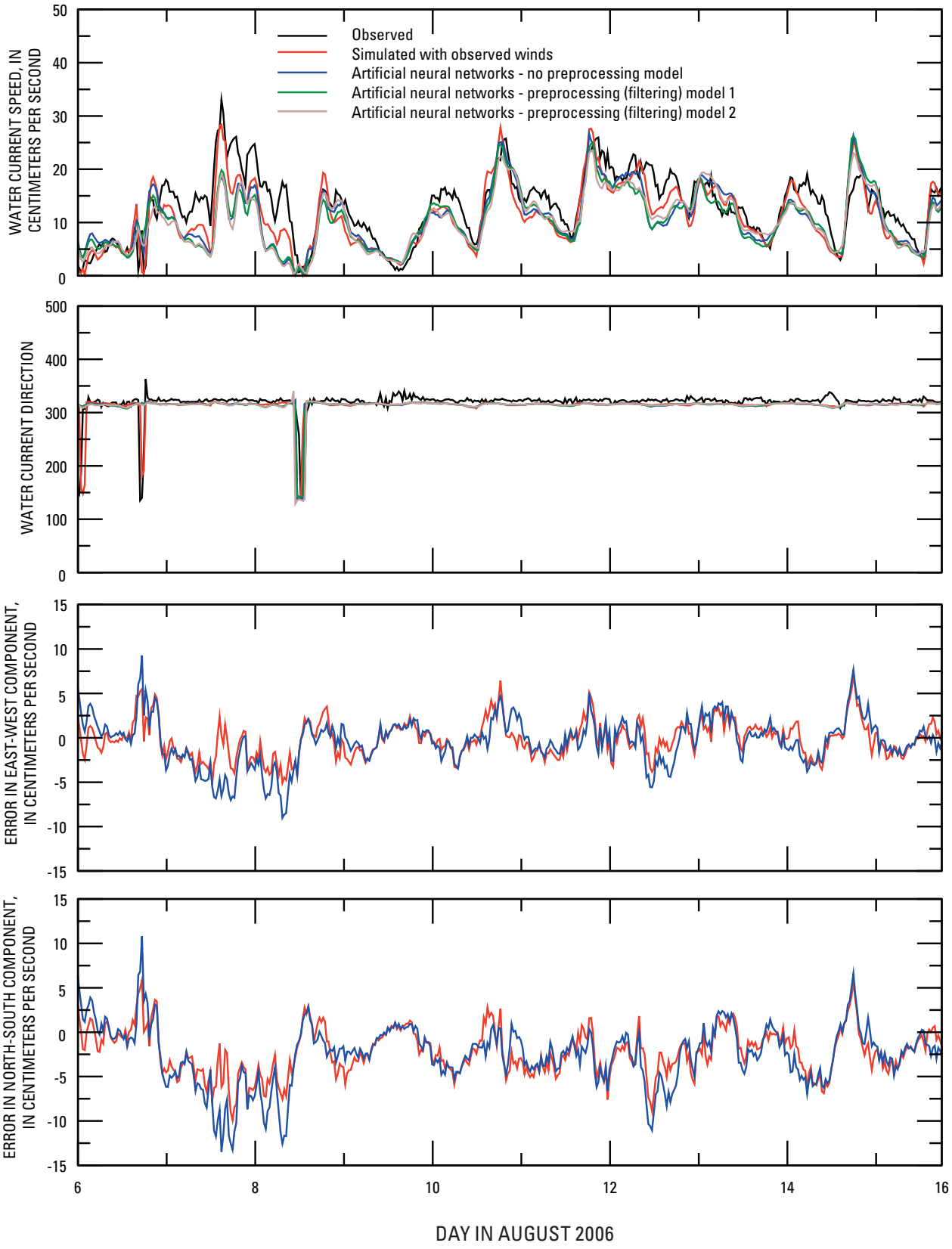



**Figure 8.** Observed and simulated currents at site ADCP1, Upper Klamath Lake, Oregon, August 2006.

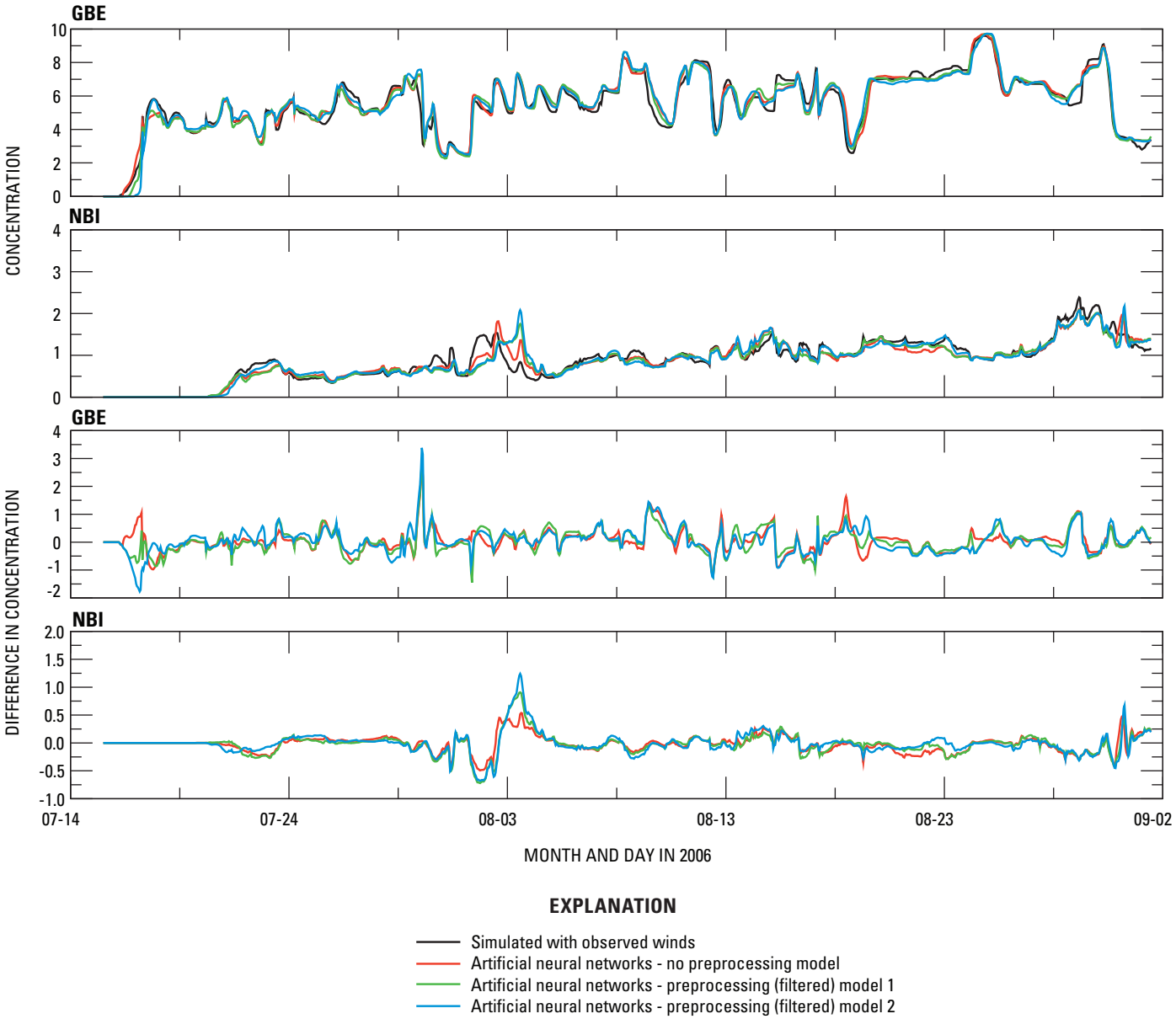



**Figure 9.** Simulated tracer concentration and residuals at sites GBE and NBI, Upper Klamath Lake, August 2006.

# Results of Historical Wind Models

Similar to the gap-filling models, the MARS algorithm was used to select the most significant explanatory variables for the ANN and the MARS historical wind models. These selected variables are shown in table 5. The model of the north-south wind component at WMR selected a greater diversity and larger number of input variables than that selected for the model of the east-west component—notably, relative humidity at AGKO and sky cover at KLMT were added. Solar radiation was selected by the MARS algorithm as a significant variable for all three dependent variables in the historical models (table 5).

Using the same statistical measures as above, the historical wind models are compared in table 6. For all models, NASH values ranged from 0.63 to 0.87 while BIAS values ranged from -0.24 to 0.11 m/s. HIST2 models had higher NASH values and lower BIAS values than HIST1 values, indicating a better statistical fit to measured data. Thus, the historical wind models that use the individual components of the wind as dependent variables (both ANN and MARS models) perform better than the historical wind models that use wind speed only as the dependent variable. Figure 10 displays the close fit to measured data that simulated HIST2 models obtained. As was the case for the gap-filling models, peak winds are the most difficult for the models to simulate consistently, and the models for the dominant (north-south) component of the wind perform better.

Winds reconstructed with the ANN_HIST2 model and measured wind at WMR during the calibration and validation period were aggregated on a weekly basis in order to visualize wind patterns for multiple years simultaneously (fig. 11). To create these "vector-bar" plots, the mean daily north-south wind components and east-west wind components were first converted to wind speed and direction. Next, the daily mean wind direction occurring each day was compartmentalized into eight directional bins. The maximum wind speed occurring in the most populated direction bin was used to determine the height of each bar. The mean of the wind-direction values associated with maximum wind-speed values determined the direction of the arrows. The maximum wind speeds and associated wind direction were used because they are assumed to be most environmentally significant for Upper Klamath Lake. The seasonal reconstructed winds generally matched the seasonal change in measured winds at WMR (red and black "vector-bars"). A closer agreement generally was observed during summer, when the winds are weaker. In cases where simulated wind direction differed from measured wind direction, the difference was infrequently more than 45 degrees.

**Table 5.** Daily time series used in the historical wind models of Upper Klamath Lake, Oregon.

[**Abbreviations:** n-s, north-south wind component; e-w, east-west wind component; ws, wind speed; rh, relative humidity; sc, sky cover; sr, solar radiation]

| Order of importance as determined by MARS | HIST1 | | HIST2 | | | |
|---|---|---|---|---|---|---|
| | Wind speed at WMR | | North-south wind component at WMR | | East-west wind component at WMR | |
| | Site | Type of data | Site | Type of data | Site | Type of data |
| 1 | AGKO | ws | KLMT | n-s | KLMT | e-w |
| 2 | KLMT | ws | AGKO | n-s | AGKO | e-w |
| 3 | AGKO | sr | AGKO | sr | KLMT | n-s |
| 4 | | | KFLO | n-s | AGKO | n-s |
| 5 | | | AGKO | rh | AGKO | sr |
| 6 | | | KLMT | sc | | |
| 7 | | | KLMT | e-w | | |

**Table 6.** Goodness-of-fit statistics for the historical wind models of site WMR during the calibration and validation period.

[NASH and BIAS are the Nash-Sutcliffe error and mean error, respectively, between the simulated and observed time series. NASH values are unit-less. BIAS values are in meters per second. **Abbreviations:** n-s, north-south wind component; e-w, east-west wind component]

| Model | NASH | | | | BIAS | | | |
|---|---|---|---|---|---|---|---|---|
| | Calibration | | Validation | | Calibration | | Validation | |
| | n-s | e-w | n-s | e-w | n-s | e-w | n-s | e-w |
| MARS_HIST1 | 0.84 | 0.63 | 0.81 | 0.70 | 0.10 | -0.16 | 0.06 | -0.24 |
| MARS_HIST2 | .87 | .78 | .87 | .73 | .00 | .00 | .05 | -.17 |
| ANN_HIST1 | .84 | .63 | .82 | .69 | .11 | -.15 | .07 | -.23 |
| ANN_HIST2 | .87 | .73 | .87 | .74 | .00 | .00 | .07 | .16 |

To demonstrate the use of the ANN_HIST2 model to reconstruct a long-term record, the model was used to reconstruct the daily wind record at WMR from January 1, 2004, through June 12, 2005, prior to the time when data were collected at this site (fig. 12). This approach has the potential to expand the length of the wind record at WMR or other sites around the lake backward in time to at least 2000 (when collection of the AgriMet data began) for the purpose of statistical investigation of the relation between wind and water quality in the lake.

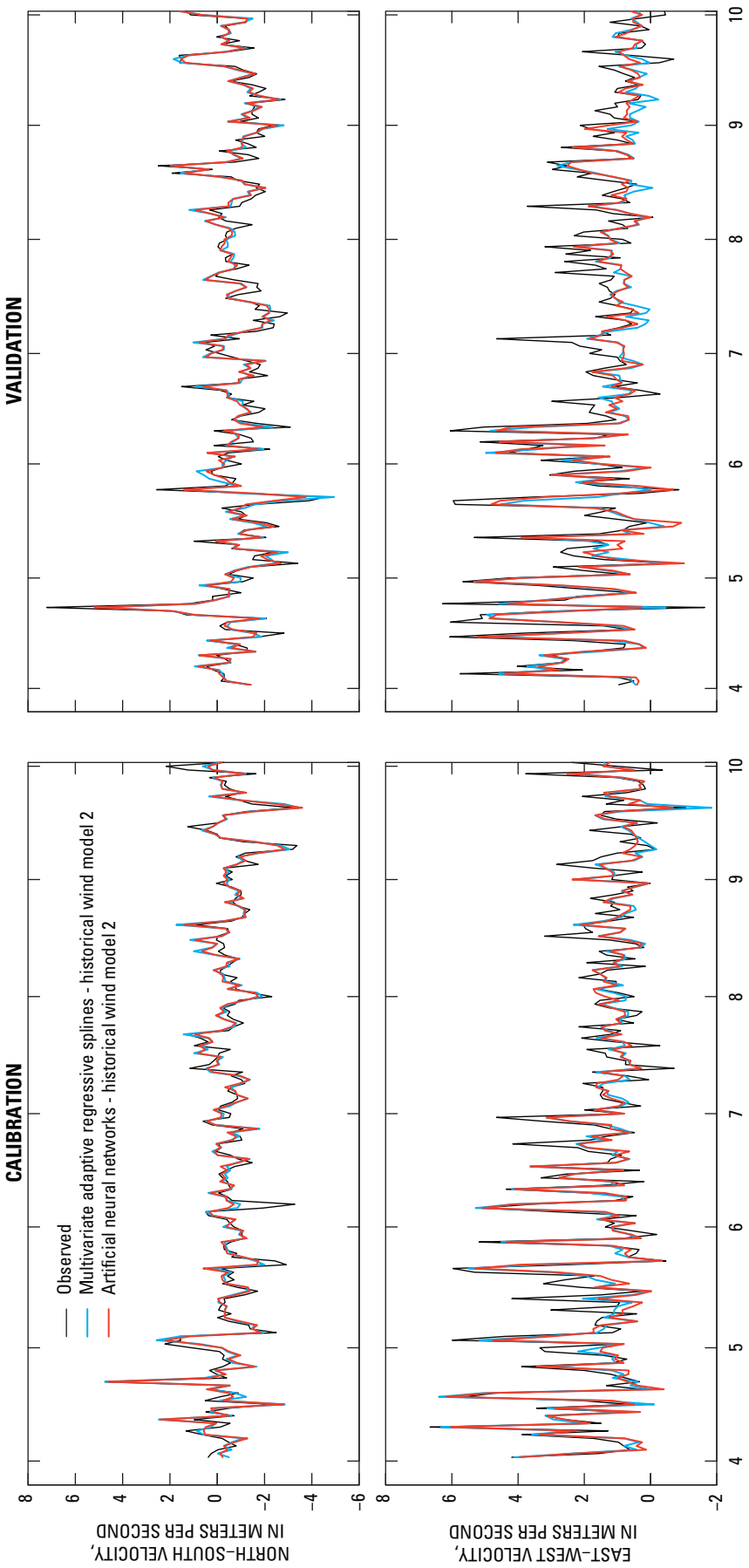



**Figure 10.** Comparison of calibration and validation years for MARS_HIST2 and ANN _HIST2 wind models, Upper Klamath Lake, Oregon.

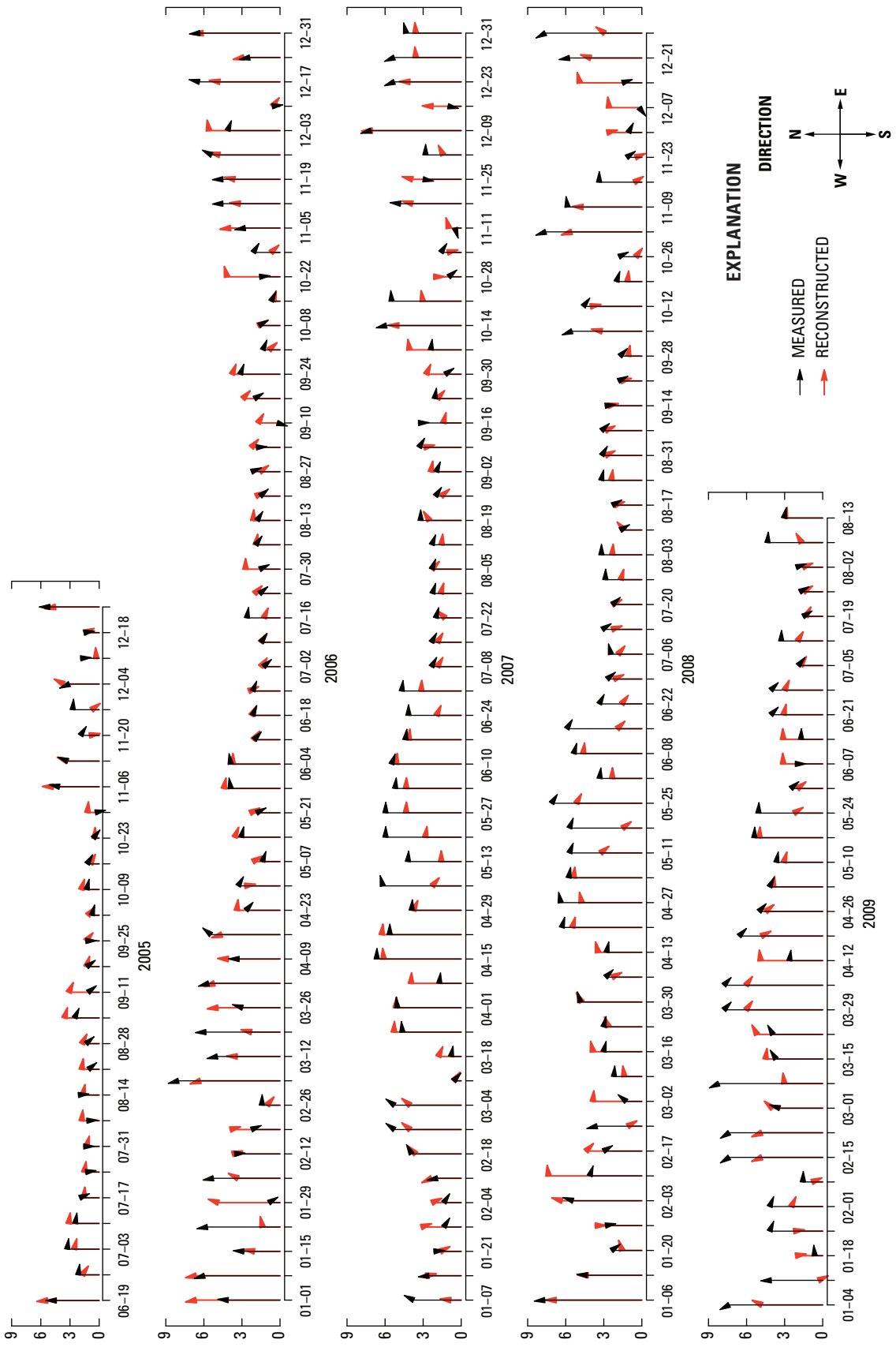



**Figure 11.** Observed and reconstructed wind data at WMR on Upper Klamath Lake, Oregon, aggregated on a weekly basis, 2005–09. Wind data were reconstructed using ANN_HIST2. Arrowheads indicate wind direction.

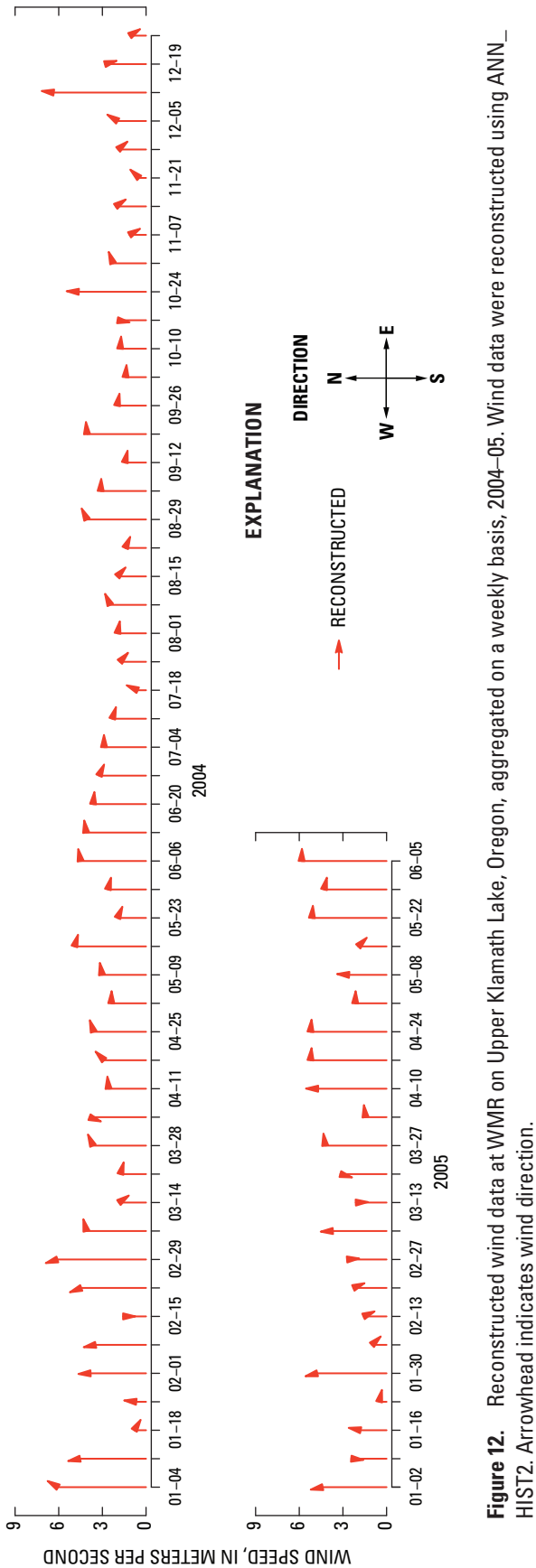



**Figure 12.** Reconstructed wind data at WMR on Upper Klamath Lake, Oregon, aggregated on a weekly basis, 2004–05. Wind data were reconstructed using ANN_HIST2. Arrowhead indicates wind direction.

## Summary and Conclusions

Nonlinear empirical models were developed using artificial neural networks (ANN) and multivariate adaptive regressive splines (MARS) to simulate wind data on Upper Klamath Lake. Both types of models were able to successfully simulate the wind on Upper Klamath Lake based on wind and other meteorological variables measured nearby. By using the MARS method of selecting model input variables prior to using the ANN, some insight into the dependence of wind on various meteorological variables at other land-based sites around the lake was gained. The most successful models were dependent on a diverse set of input variables measured at several sites around the lake.

MARS and ANN were used successfully to simulate wind data at two floating raft sites on Upper Klamath Lake. Both methods were able to capture most of the variability in the 10-minute wind time series. Two different preprocessing methods to smooth the input variables were investigated, but the preprocessing did not improve the performance of the models greatly, and the extra step of smoothing the data may not be necessary. Gap-filling models based on ANN performed moderately better than those based on MARS. Although the ANN approach does not directly reveal the nature of dependencies between input and output variables as well as the MARS approach, the greater accuracy achieved with the ANN approach may justify the added opacity. The set of explanatory variables that contributed to the "best" models as determined by the MARS algorithm included wind from all sites around the lake, as well as wind divergence and curl, and air temperature. The best models for the raft sites included both east-west and north-south wind components from other land-based sites and comprised from 11 to 15 independent variables.

The winds simulated with the ANN gap-filling models were used in the spatially variable wind forcing for the hydrodynamic model of Upper Klamath Lake. The error statistics for the simulated currents, when the simulated winds were used to force the model, were comparable to those for the currents that were simulated when only observed winds were used to force the model. It can be concluded that this technique is adequate for filling gaps of several days in the wind data collected from the rafts on the lake. A numerical tracer experiment indicated that the errors in transport that result when simulated winds are used to drive the hydrodynamic model probably are small, particularly when averaged over periods of several days or more.

The ANN and MARS were used to accurately reconstruct wind data at the Williamson River Delta site WMR on a daily timestep over several years. Both ANN and MARS historical models performed similarly. The datasets long enough to be useful as input to these models were more limited than those available for input to the gap-filling model. Significant explanatory variables selected by the MARS algorithm included wind, solar radiation, and relative humidity at AgriMet sites (AGKO and KFLO), as well as wind and sky cover at the Klamath Falls Airport site (KMLT). The most successful historical wind model simulated the north-south and east-west components separately. The accuracy was higher for this historical wind model than for the gap-filling model, and indicates the potential to simulate the wind on a daily timestep going backward in time at least to the year 2000, when the AgriMet data were first collected. Given the long data record at KMLT, there is a potential to develop an historical wind model based only on data collected at that site to reconstruct the wind at sites around UKL back to 1959, but that is beyond the scope of the current work and the accuracy of such a model has not been investigated for this report.

## Acknowledgments

This work was funded by the Bureau of Reclamation (Interagency Agreement 08AA200098) and the U.S. Geological Survey.

## References Cited

Bureau of Reclamation, 1997, AgriMet Weather Data—Pacific Northwest cooperative agricultural weather network: Bureau of Reclamation, accessed July 2, 2010, at http://www.usbr.gov/pn/agrimet/wxdata.html.

Cannon, A.J., and Whitfield, P.H., 2001, Modeling transient pH depressions in coastal streams of British Columbia using neural networks: Journal of the American Water Resources Association, v. 37, issue 1, p. 73–89.

Conrads, P.A., and Roehl, E.A., 1999, Comparing physics-based and neural network models for simulating salinity, temperature, and dissolved oxygen in a complex, tidally affected river basin: Proceedings of the 1999 South Carolina Environmental Conference, Myrtle Beach, S.C., March 15–16, 1999. (Also available at http://smig.usgs.gov/features_0399/nnm1.html/.)

Conrads, P.A., Roehl, E.A., Daamen, R.C., and Kitchens, W.M., 2006, Simulation of water levels and salinity in the rivers and tidal marshes in the vicinity of the Savannah National Wildlife Refuge, Coastal South Carolina and Georgia: U.S. Geological Survey, Scientific Investigations Report 2006-5187, 134 p. (Also available at http://pubs.usgs.gov/sir/2006/5187/.)

Conrads, P.A., and Roehl, E.A., Jr., 2007, Analysis of salinity intrusion in the Waccamaw River and the Atlantic Intracoastal Waterway near Myrtle Beach, South Carolina, 1995–2002: U.S. Geological Survey Scientific Investigations Report 2007-5110, 41 p. (Also available at http://pubs.usgs.gov/sir/2007/5110/.)

Friedman, J.H., 1991, Multivariate Adaptive Regressive Splines: Annals of Statistics, v. 19, no. 1, p. 1-67, accessed September 7, 2010, at http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.aos/1176347963/.

Gartner, J.W., Wellman, R.E., Wood, T.M., and Cheng, R.T., 2007, Water velocity and suspended solids measurements by in-situ instruments in Upper Klamath Lake, Oregon: U.S. Geological Survey Open-File Report 2007-1279, 136 p. (Also available at http://pubs.usgs.gov/of/2007/1279/.)

Hoilman, G.R., Lindenberg, M.K., and Wood, T.M., 2008, Water quality conditions in Upper Klamath and Agency Lakes, Oregon, 2005: U.S. Geological Survey Scientific Investigations Report 2008–5026, 44 p. (Also available at http://pubs.usgs.gov/sir/2008/5026/.)

Ibanez, Frederic, Grosjean, Philippe, and Etienne, Michele, 2009, Pastecs—Package for analysis of space-time ecological series, R package version 1.3–10, accessed May 13, 2010, at http://CRAN.R-project.org/package=pastecs.

Kann, Jacob, and Welch, E.B., 2005, Wind control on water quality in shallow, hypereutrophic Upper Klamath Lake, Oregon: Lake and Reservoir Management, v. 21, no. 2, p. 149–158.

Kretzschmar, Ralf, Eckert, Pierre, Cattani, Daniel, and Eggiman, Fritz, 2004, Neural network classifiers for local wind prediction: Journal of Applied Meteorology, v. 43, p. 727–738.

Kulkarni, M.A., Patil, Sunil, Rama, G.V., and Sen, P.N., 2008, Wind speed prediction using statistical regression and neural network: Journal of Earth System Science, v. 117, no. 4, p. 457–463.

Laenen, Antonius, and LeTourneau, A.P., 1996, Upper Klamath basin nutrient-loading study—Estimate of wind-induced resuspension of bed sediment during periods of low lake elevation: U.S. Geological Survey Open-File Report 95-414, 11 p.

Lindenberg, M.K., Hoilman, G.R., and Wood, T.M., 2009, Water quality conditions in Upper Klamath and Agency Lakes, Oregon, 2006: U.S. Geological Survey Scientific Investigations Report 2008-5201, 54 p. (Also available at http://pubs.usgs.gov/sir/2008/5201/.)

Milborrow, Stephen, 2009, "Earth—Multivariate adaptive regression spline models" R package version 2.3-5 (derived from mda:mars by Trevor Hastie and Robert Tibshirani), accessed August 24, 2009, at http://cran.r-project.org/web/packages/earth/index.html.

National Oceanic and Atmospheric Administration, [n.d.], NNDC climate data online: National Oceanic and Atmospheric Administration, accessed April 16, 2010, at http://www7.ncdc.noaa.gov/CDO/cdo.

National Weather Service, 1998, Automated surface observing system (ASOS), user's guide: Silver Springs, Md., National Oceanic and Atmospheric Administration, National Weather Service, 56 p., accessed August 2, 2010, at http://www.nws.noaa.gov/asos/aum-toc.pdf.

Rajaee, Taher, Mirbagheri, S.A., Zounemat-Kermani, Mohammad, Nourani, Vahid, 2009, Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models: Science of the Total Environment, v. 407, issue 17, p. 4916–4927.

Risley, J.C., Roehl, E.A., and Conrads, P.A., 2003, Estimating water temperatures in small streams in western Oregon using neural network models: U.S. Geological Survey Water-Resources Investigations Report 02-4218, 59 p. (Available at http://pubs.usgs.gov/wri/wri024218/.)

Rounds, S.A., 2002, Development of a neural network model for dissolved oxygen in the Tualatin River, Oregon, in Proceedings of the Second Federal Interagency Hydrologic Modeling Conference, Las Vegas, Nevada, July 29–August 1, 2002: Subcommittee on Hydrology of the Interagency Advisory Committee on Water Information, 13 p., accessed July 6, 2010, at http://or.water.usgs.gov/tualatin/ann_proceedings.pdf.

Shamseldin, A.Y., 1997, Application of a neural network technique to rainfall-runoff modeling: Journal of Hydrology, v. 199, p. 272–294.

Venables, W.N., and Ripley, B.D., 2002, Modern applied statistics with S, 4th ed.: Springer-Verlag, New York, 495 p.

Weglarczyk, Stanislaw, 1998, The interdependence and applicability of some statistical quality measures for hydrological models: Journal of Hydrology, v. 206, p. 98–103.

Wood, T.M., Hoilman, G.R., and Lindenberg, M.K., 2006, Water-quality conditions in Upper Klamath Lake, Oregon, 2002–04: U.S. Geological Survey Scientific Investigations Report 2006-5209, 52 p. (Also available at http://pubs.usgs.gov/sir/2006/5209/.)

Wood, T.M., 2009, Preliminary study of the effect of the proposed Long Lake Valley Project operation on the transport of larval suckers in Upper Klamath Lake, Oregon: U.S. Geological Survey Open-File Report 2009-1060, 24 p. (Also available at http://pubs.usgs.gov/of/2009/1060/.)

Wood, T.M., Cheng, R.T., Gartner, J.W., Hoilman, G.R., Lindenberg, M.K., and Wellman, R.E., 2008, Modeling hydrodynamics and heat transport in Upper Klamath Lake, Oregon, and implications for water quality: U.S. Geological Survey Scientific Investigations Report 2008–5076, 48 p. (Also available at http://pubs.usgs.gov/sir/2008/5076/.)

This page intentionally left blank.

Publishing support provided by the U.S. Geological Survey
Publishing Network, Tacoma Publishing Service Center

For more information concerning the research in this report, contact the

Director, Oregon Water Science Center
U.S. Geological Survey
2130 SW 5th Avenue
Portland, Oregon 97201
http://or.water.usgs.gov