Vol. 11: 173–190, 1999 | CLIMATE RESEARCH Clim Res | Published April 28

Empirical-statistical reconstruction of surface marine winds along the western coast of Canada

Manon Faucher[1,*], William R. Burrows[2], Lionel Pandolfo[3]

[1]Département des sciences de la terre, Université du Québec à Montréal, CP 8888, Succ. 'Centre-Ville', Montréal (Québec) H3C 3P8, Canada

[2]Numerical Prediction Research Division, Meteorological Research Branch, Atmospheric Environment Service, Downsview, Ontario M3H 5T4, Canada

[3]Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

ABSTRACT: CANFIS, an empirical-statistical technique, is used to reconstruct continuous daily surface marine winds at 6-hourly intervals at 13 Canadian buoy sites along the western coast of Canada for the 40 yr period 1958–1997. CANFIS combines Classification and Regression Trees (CART) and the Neuro-Fuzzy Inference System (NFIS) in a 2-step procedure. CART is a tree-based algorithm used to optimize the process of selecting relevant predictors from a large pool of potential predictors. Using the selected predictors, NFIS builds a model for continuous output of the predictand. In this project we used CANFIS to link large-scale atmospheric predictors with regional wind observations during a learning phase from 1990 to 1995 in order to generate empirical-statistical relationships between the predictors and buoy winds. The large-scale predictors are derived from the NCAR/NCEP 40 yr reanalysis project while the buoy winds come from the Canadian Atmospheric Environment Service buoy network. Validation results with independent buoy wind data show a good performance of CANFIS. The CANFIS winds reproduce the independent buoy winds with greater accuracy than winds reconstructed with a stepwise multivariate linear regression technique. In addition, they are better than the NCEP reanalyzed winds interpolated to the buoy locations. The reconstructed statistical winds recover more than 60% of the observed wind variance during an independent verification period. In particular, correlation coefficients between independent buoy wind time series and CANFIS wind time series vary between 0.61 and 0.98. Our results suggest that CANFIS is a successful downscaling method. It is able to recover a substantial fraction of the variation of surface marine winds, especially along coastal regions where ageostrophic effects are relatively important.

KEY WORDS: Marine wind modelling · Statistical downscaling · Classification and regression trees · Neuro-fuzzy inference system

1. INTRODUCTION

The objective of this study is to produce a historical wind data set for the west coast of Canada that will be reliable and without gaps in dates available. Continuous and reliable surface winds are needed to investigate climate change and its impact on British Columbia (BC) coastal fisheries because marine ecosystems are relatively sensitive to surface wind variability. Currently available data sets have 1 of 2 problems: they have either missing time periods or coarse spatial resolution. The latter problem is important in coastal areas where meso- or smaller scale phenomena are often generated by the rugged coastal topography. In general, coarse-resolution data sets fail to capture these small-scale events because the modification of the surface pressure field occurring over small distances near mountains and islands is beyond their reach. This causes the winds in low-resolution data sets to misrepresent the actual winds. This was shown to be the case with synoptic-scale wind data computed by the U.S. National Marine Fisheries Service (Thomson 1983) and with the surface wind data from the NCAR/NCEP (National Center for Atmospheric Research/National Centers for Environmental Protection) reanalysis project (Faucher & Pandolfo unpubl.).

Due to the absence of strong current systems along the west coast of Canada, surface winds have a stronger impact on water movements. In particular, they affect upwelling (a mechanism that brings deeper,

*E-mail: faucher@maia.sca.uqam.ca

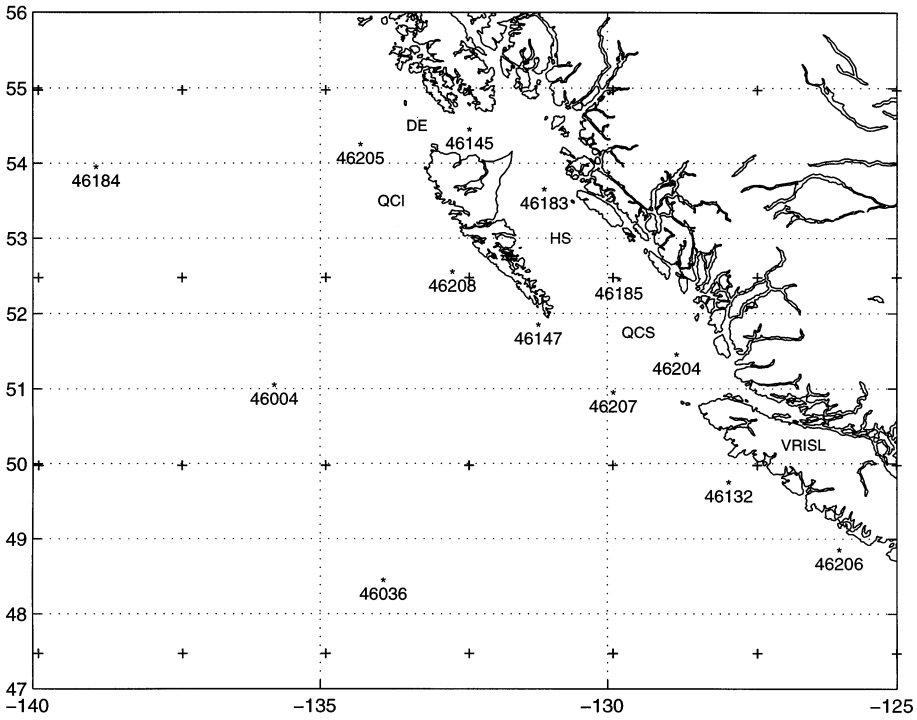


Fig. 1. Location of the 13 Atmospheric Environment Service (AES) weather buoys in the northeast Pacific used in this study. Buoy sites are indicated by a series of numbers of the form '46xxx'. (+) Locations of the grid points for the NCAR/NCEP reanalyzed data. VRISL: Vancouver Island; QCI: Queen Charlotte Islands; QCS: Queen Charlotte Sound; HS: Hecate Strait; and DE: Dixon Entrance

nutrient-rich waters to the surface), the depth of the mixed layer and the physical properties of the surface layer of the ocean (sea surface temperature, salinity and density gradient). If one assumes a bottom-up forcing, changes in the physical structure of the upper ocean will affect the primary production of phytoplankton by altering the light and nutrient concentration. This will then be reflected in the upper trophic levels via food consumption, predation, different swimming pattern and variations in growth and survival.

A statistical technique named CANFIS (Burrows 1998, Burrows et al. 1998) is used to estimate 6-hourly surface marine winds at 13 Canadian buoy sites along the coast of BC (Fig. 1) for the 40 yr period 1958–1997. Relationships between large-scale atmospheric variables and regional wind observations were built through an approach commonly called downscaling (Enke & Spekat 1997). The large-scale atmospheric variables come from the NCAR/NCEP 40 yr reanalysis project data (Kalnay et al. 1996) while the regional wind observations are from the buoy network of the Canadian Atmospheric Environment Service. Our downscaling approach combines 2 recent statistical techniques: Classification and Regression Trees (CART) (Brieman et al. 1984), and the Neuro-Fuzzy Inference System (NFIS) (Chiu 1994). CART is a tree-based algorithm used to optimize the selection of relevant predictors from a large pool of potential predictors. Our predictors correspond to various large-scale atmospheric variables derived from the NCEP reanalyzed data. NFIS is used to build final models of the predictand data. Our predictands are the horizontal wind components at each buoy location. The models are used to hindcast 6-hourly winds for the periods prior to and after the learning period (1990–1995). In terms of computer time, this is a relatively inexpensive yet accurate method compared with running a purely dynamical climate model, and is therefore very attractive (Kidson & Thompson 1998). Significant advantages of the CANFIS method are: (1) the tails of a predictand data distribution are modelled better than forward stepwise multivariate linear regression, and (2) output from the final NFIS model is continuous in time. Comparison between CANFIS and other methods indicates this hybrid method is beneficial for the construction of statistical models for a variety of predictands (Burrows et al. 1998).

The plan of the paper is as follows. The downscaling procedure is presented in Section 2, including a brief summary of CANFIS (see Burrows et al. 1998 for a complete description). The data are described in Section 3. Section 4 presents how the potential predictors are determined and how CART refines them to obtain a final set of relevant predictors. Then, after NFIS has been applied to generate continuous series of statistical winds, the CANFIS method is evaluated and compared with a stepwise multivariate linear regression technique (MLR) in Section 5. This is done by comparing the statistical wind data with independent wind observations. Finally, a conclusion is presented in Section 6.

2. DOWNSCALING PROCEDURE

2.1. Overview of the procedure

Large-scale atmospheric variables on a 2.5° × 2.5° grid from the 40 yr NCAR/NCEP reanalysis project (Kalnay et al. 1996) were used as inputs in the downscaling procedure. The data (listed in Section 3.2) were extracted at a series of grid points in an area from the coast of BC to approximately 400 km offshore (Fig. 1). These variables were used to compute a series of potential predictors determined from the atmospheric momentum equations (see Section 4), for coastal surface marine winds. The potential predictors are computed at the grid points with the series of definitions listed in Table 1 and interpolated to 13 buoy sites (Fig. 1). The potential predictors were matched with available wind observations at each buoy during the years 1990 to 1995, and the CANFIS procedure invoked to build models for horizontal components (zonal, 'CANFIS *u*', and meridional, 'CANFIS *v*'). Data were not available for all the years 1990 to 1995 at every buoy, and there were short gaps in time coverage at each buoy due to instrument down-time. The models were applied to reconstruct a 6-hourly wind data set without gaps at the buoy sites for 1958 to 1997.

2.2. CANFIS method

CANFIS is a 2-stage procedure. CART regression (Brieman et al. 1984) is used to select relevant predictors from a pool of potential predictors, and NFIS (Chiu 1994) is used to build an output model using those relevant predictors. Due to space limitation, only an abbreviated description of CART, NFIS, and CANFIS can be given here. The CART algorithm is complex. For theoretical details the reader should consult Brieman et al. (1984). Much practical information about the CART algorithm can be found in the software manual (Steinberg & Colla 1995). Abbreviated discussions about CART regression applied to other data-modeling problems can be found in Burrows (1997) and Burrows et al. (1995). More detail about NFIS appears in Chiu (1994) and Burrows et al. (1998). CART is a tree-structured algorithm for non-parametric data analysis. CART regression develops a decision-tree data partitioning structure which minimizes residual variance of the predictand, essentially clustering the data into a set of 'terminal nodes'. The procedure begins at the 'root node', where a data set comprised of cases of a predictand matched with predictors resides. The data set in the root node is divided into 2 descendent subsets (nodes) by a 'partition function'. A search algorithm tries the range of values of all single potential predictors and many linear combinations of potential predictors to find a partition function and a threshold value of it that minimizes the weighted average of the residual variances of the predictand in the 2 descendent nodes. Each descendent node (or data subset) is further partitioned until all variance has been explained, then the search for the 'best' tree structure begins. This very large tree structure is 'pruned' upwards incrementally from the bottom. Independent verification data is dropped down the tree at each pruning stage to calculate residual error. (In the current study, the original data was divided into 2 parts having approximately the same distributions in a histogram shape sense. 70% was kept aside for training and 30% as independent data for checking.) (For small data sets the error is estimated by cross-validation.) As the number of terminal nodes diminishes in the pruning process, the residual error decreases to a minimum then increases again as the number of terminal nodes steadily decreases until 1 is left (the root node). The tree structure with minimal error is deemed to be the 'best' tree. The predicted value at each node can be either the mean or median of the predictand data falling into the node. (The mean is used in this study.) Thus the regression response is composed of a finite set of discrete values. An example is given in Fig. 2, which is the tree derived for wind speed *(S)* at Buoy 46204 in Fig. 1. Table 2 shows the partition functions and terminal node values for the tree in Fig. 2. This tree has 18 internal nodes and 19 terminal nodes, and has a residual error of 0.263 for the training data and 0.297 for the independent data.

The ability of CART to model predictand data obviously depends on a judicious choice of potential predictors. These should have a known or suspected physical relation to the predictand. If such predictors are not offered, the ability of CART to build a tree is seriously impaired, and it is possible that no tree will be built. The predictors that appear in splitting decisions along the path to a terminal node nearly always make physical sense. The predictors used here are described in Section 4.

CART ranks the importance of each predictor on a scale of 0 to 100, as determined from its appearance as a primary or surrogate predictor in internal node-splitting decisions. The importance value of a CART predictor is somewhat ad hoc, as Brieman et al. (1984) point out. However a non-zero variable importance does indicate that a predictor has a role in modelling the predictand data. In the CANFIS procedure relevant predictors are identified as those with greater than zero importance.

Sometimes there are groups of highly correlated predictors that still survive the CART run because they were each deemed to have greater than zero importance. Predictors correlated with one another with

Table 1. List of the potential predictors computed at the NCAR/NCEP reanalysis project (REAN) data points and interpolated to the 13 buoy sites in the downloading procedure

Potential predictor	Definition	Name
REAN wind components u, v, and speed (S) at X = 0 (first sigma level) and X = 1000 and 925 hPa (pressure field values)	$u, v, S = \sqrt{u^2 + v^2}$	SU0, SV0, SS0 UX, VX, SX
Geostrophic winds at mean sea level (MSL)	See (a) terms in Eqs. (4) & (5)	UGMSL, VGMSL, SGMSL
Geostrophic wind at X = 925, 850, 700, 500 and 300 hPa (f = Coriolis parameter)	$u_g = -\left(\frac{1}{f}\right)\left(\frac{\partial \Phi}{\partial y}\right)$ $v_g = \left(\frac{1}{f}\right)\left(\frac{\partial \Phi}{\partial x}\right)$ $S_g = \sqrt{u_g^2 + v_g^2}$	UGX, VGX, SGX
Isallobaric ageostrophic wind at MSL	See (b) terms in Eqs. (4) & (5)	UIMSL, VIMSL
Advective ageostrophic wind at MSL	See (c) and (d) terms in Eqs. (4) & (5)	UAA, VAA
Convective ageostrophic wind at X = 925 hPa	See (e) terms in Eqs. (4) & (5)	UCX, VCX
Curvature of the MSL pressure field (X = 'MSL') and of X = 925 hPa height field (tangential $[K_i]$ and orthogonal $[K_n]$ components, where p [pressure] is replaced by ϕ [geopotential height] for 925 hPa)	$K_i = \frac{\partial^2 p / \partial x^2}{\partial p / \partial y}$ $K_n = \frac{\partial^2 p / \partial x \partial y}{\partial p / \partial y}$	KIX, KNX
Temperature tendency near the surface (2 m above) at 925 and 850 hPa	$\partial T / \partial t$	DTX
Temperature gradients between 2 m and 925 hPa, 2 m and 850 hPa and 925 and 850 hPa (X = 'TOP': top of mixed layer)	$\partial T / \partial z \Big\|_{x_1}^{x_2}$	DTPX
'Thermal wind' (baroclinicity) near the surface (2 m above) and at X = 925, 850 and 700 hPa (g = acceleration due to gravity)	$\frac{\partial u_g}{\partial z} = \frac{-g}{f}\left(\frac{\partial \ln T}{\partial y}\right)$ $\frac{\partial v_g}{\partial z} = \frac{g}{f}\left(\frac{\partial \ln T}{\partial x}\right)$ $\frac{\partial \sqrt{(u_g^2 + v_g^2)}}{\partial z}$	DUGX, DVGX, DSGX
Wind shear between the first sigma level and 925 or 850 hPa, and between 925 and 850 hPa (X = 'TOP')	$\partial u / \partial z$ and $\partial v / \partial z$	DUZX, DVZX
Modified vertical velocity at 925 hPa (h_m is mountain height, l is across-width scale, θ_0 is a referenced temperature, V is the horizontal wind, p_s is a pressure reference [taken at the MSL], R is the universal gas constant, and c_p is the specific heat at constant pressure)	$w = -\frac{R_l C^{-1} \frac{d\theta}{dt}}{(1 + R_l F \theta)}$ with $R_l = \frac{V}{fl}$, $C = \frac{N^2 h_m^2}{f^2 l^2}$, $F = \frac{f^2 l^2}{g h_m}$ and $N^2 = \frac{g}{\theta_0}\left(\frac{\partial \ln \theta}{\partial z}\right)$ where $\theta = T\left(\frac{p_s}{p}\right)^{R/c_p}$	W925
Temperature near the surface (2 m above) and at X = 1000, 925, 850 and 700 hPa	T	T0, TX
Mean sea level pressure	p	PMSL1013
Geopotential height at X = 1000, 850, 700, 500 and 300 hPa	Φ	PHIX
Vertical velocity at X = 1000, 925 and 850 hPa	ω	OMEGAX
Froude number and Rossby radius at 925 hPa	$F_N = V/(h_m N)$ and $R_l = V/fl$	FNX, ROX
Julian day	–	JULDAY

more than a threshold value 0.9 are gathered, and the one correlated highest with the predictand is retained. The value of 0.9 is arbitrary and was based on trial and error. If a significantly higher value is used, a group will be small and few predictors are eliminated, and, if a significantly lower value than 0.9 is used, a group will be large and many predictors are eliminated.

CART regression is capable of modeling a predictand with better accuracy and fewer predictors than forward stepwise multivariate linear regression (Burrows 1997) except when the predictand-predictor relationship is highly linear. We found the same result here (see Section 5). The piecewise-continuous output of CART by itself is sufficient for some problems. However, a smooth output is much more attractive for the purpose of wind modeling. Therefore, once CART has identified the relevant predictors, NFIS (Chiu 1994) is used to construct wind models and obtain continuous predicted wind values. NFIS is a computationally efficient algorithm that gives a highly optimized model in a single pass through the data. Its model can be tuned further with the Adaptive Neuro-Fuzzy Inference System (ANFIS) developed by Jang (1993), but this is time consuming and we found insignificant improvement because the NFIS model is already highly optimized. A brief description of the NFIS procedure follows.

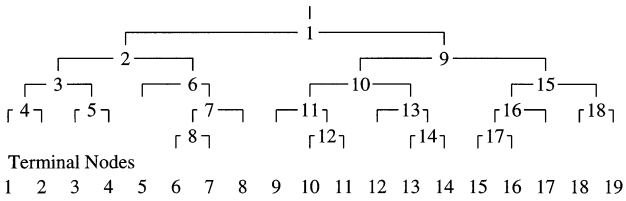


Fig. 2. CART tree derived for wind speed *(S)* at Buoy 46204. There are 18 internal nodes in the upper part of the diagram and 19 terminal nodes in the bottom row. See Table 2 for values in each node

Given training data with many cases [data points $\mathbf{x} = (\mathbf{y}, z)$] of a predictand, z, matched with N predictors, $\mathbf{y}$, a set of c multivariate (N + 1 variables) cluster centers is found by the subtractive clustering method. A 'potential' is calculated for each point based on its Euclidean distance from all other data points. The data point with the highest potential is chosen as the first cluster. The potential of each remaining data point is reduced by an amount determined by its distance from the first cluster, and the one with the highest adjusted

Table 2. (a) Partition functions in internal nodes for CART tree shown in Fig. 2 and (b) terminal node values for the predictand. Node numbers are integer values in top row of each section. Below node numbers is a node-splitting decision in (a) and a node value in (b). The predictand is wind speed *(S)* at buoy 46204 in Fig. 1. Predictor acronyms are explained in Table 1. Units for wind components are m s^{-1}, geopotential height is in dam (decametres)

(a) Partition functions in internal nodes

1	2	3	4
SS0 < 8.1	SS0 < 4.9	SS0 < 3.2	PHI1000 < 336
5	6	7	8
UIMSL < 1.8	SS0 < 6.2	SU0 < 3.7	DUG0 < 0.0
9	10	11	12
SS0 < 12.9	SS0 < 10.6	SS0 < 8.7	SS0 < 4.6
13	14	15	16
VAA < –2.4	DVG925 < 0	SS0 < 16.8	SU0 < 1.0
17	18		
PHI500 < 5284	U925 < –1.2		

(b) Terminal node values for predictand

1	2	3	4	5	6	7	8	9	10
4.9	2.6	3.7	8.6	4.8	5.7	6.8	5.3	7.0	8.4
11	12	13	14	15	16	17	18	19	
7.4	13.6	10.0	8.8	13.8	11.8	10.3	15.9	12.9	

potential is chosen as the second cluster if it passes acceptance tests. The potentials of remaining data points are reduced again by an amount determined by their distance from the second cluster, and the one with the highest newly adjusted potential is chosen as the third cluster if it passes acceptance tests, and so on. The total number of clusters found depends on 4 parameters for which optimal values must be found: the 'radius of influence' ($\mathbf{r}_a$), a vector defining a neighborhood around a data point; the 'squash factor' (φ), a parameter for avoiding closely spaced clusters; $\overline{\varepsilon}$, a parameter to determine if a candidate cluster should be accepted; and $\underline{\varepsilon}$, a parameter to determine if a candidate cluster should be rejected. Typically, though not necessarily, $\overline{\varepsilon}$ and $\underline{\varepsilon}$ are held fixed while a few runs are done with the training data varying $\mathbf{r}_a$ and φ values to find a combination that gives minimal residual error for independent data. That combination is used to build a final NFIS model using the complete data set.

The set of cluster centers serves as the basis for a fuzzy rule set that describes the behaviour of the system. Gaussian membership functions allow for predictor membership in every rule. For each rule i, a 'degree of fulfillment', μ_i, is determined from the Euclidean distance between a predictor vector $\mathbf{y}$ and its value at the cluster center $\mathbf{y}_i^*$ as follows

$$\mu_i = e^{-\alpha\|\mathbf{y}-\mathbf{y}_i^*\|^2} \quad (1)$$

where $\alpha = \frac{4}{\mathbf{r}_a^2}$ and $\mathbf{r}_a$ is defined above. The output of rule (cluster center) i is its predictand value multiplied by the degree of fulfillment for rule i. The final output vector Z is a weighted average of the c outputs, z_i^* from the cluster centers

$$Z = \frac{\sum_{i=1}^{c} \mu_i z_i^*}{\sum_{i=1}^{c} \mu_i} \quad (2)$$

We want a model to fit the predictand z with the predictors $\mathbf{y}$. Write the predictand output of each cluster center i as a linear combination of the N predictors

$$z_i^* = \mathbf{G}_i\mathbf{y} + b_i \quad (3)$$

where $\mathbf{G}_i$ is a row vector of coefficients for rule i and b_i is a constant. The coefficients $\mathbf{G}_i$ and b_i are found by solving a fast-executing recursive least-squares estimation procedure suitable for large data sets (see Chiu 1994).

We used the CANFIS procedure with data for the period 1990–1995 to (1) identify relevant predictors for u and v wind components, (2) build response models that give continuous output, and (3) use them to reconstruct a 6-hourly wind data set without gaps at the buoy sites for the period 1958–1997.

3. DATA

3.1. Buoy wind data

A 6 yr set (1990 to 1995) of 6-hourly buoy data served to fit the CANFIS models at selected buoy sites along the BC coast. Hourly wind observations are provided by ten 3 m discus buoys and three 6 m NOMAD (Navy Oceanographic Meteorological Automated Devices) buoys from the Canadian Atmospheric Environment Service network. Buoy locations are shown in Fig. 1. The NOMAD buoys are the 3 farthest from the coast. R. M. Young anemometers are located 3.7 and 4.7 m above the discus buoy platform, and 4.4 and 5.2 m above the NOMAD buoy platform. The accuracy of the anemometers for measuring wind directions and speeds are ±5° and 0.6 m s^{-1}. More information on the buoys and their instruments are in Ocean Data Acquisition System Buoy Service Reports published each year by Environment Canada (Environment Canada 1995).

Hourly wind observations were processed at the Institute of Ocean Sciences (Sidney, BC) for quality control and missing data. Details on buoy data processing are given in Cherniawsky & Crawford (1996). In this study, a 6-hourly moving average is applied to contiguous hourly data to eliminate some irregularities and high-frequency fluctuations that are not relevant to our study. We used an odd-length filter and assigned the result to the values corresponding to 00:00, 06:00, 12:00 and 18:00 h for each day. Missing data are identified with a special code and not treated. Four observations per day are retained to match with reanalyzed data presented below.

3.2. Reanalyzed data

A set of 6-hourly meteorological data without gaps was obtained from the NCEP/NCAR 40 yr reanalysis project (Kalnay et al. 1996) at a series of grid points over the northeastern Pacific. We refer to this data as the REAN data. The following REAN data were extracted for the period 1958–1997 to build the CANFIS models and to conduct the hindcast experiment:

- u and v components of the wind at the first sigma level (σ_1 = 0.995 ≈ 42 m) and at 1000, 925 and 850 hPa;
- pressure at mean sea level (MSL);
- geopotential height at 1000, 925, 850, 700, 500 and 300 hPa;
- temperature at 2 m and at 1000, 925, 850 and 700 hPa;
- vertical velocity at 1000, 925 and 850 hPa;
- potential temperature at the first sigma level (σ_1 = 0.995).

These data are available on a latitude-longitude grid with a 2.5° × 2.5° resolution. All data are interpolated at the 13 Canadian buoy locations along the British Columbia coast (Fig. 1) with a bicubic spline method.

4. PREDICTORS

4.1. Potential predictors

The REAN wind components themselves are a large-scale analysis of the surface wind and so are expected to explain a large portion of the variance in the wind at the buoys. A series of dynamical predictors for the u and v surface wind components (Table 1) is determined from the balance of forces governing the motion in the planetary boundary layer (PBL) in mid-latitudes. These predictors are computed at the grid points of the NCAR/NCEP data set and interpolated to the buoy sites. The wind components in the PBL can be decomposed dynamically in the following way (Saucier 1955):

$$u = -\left(\frac{1}{\rho f}\right)\left(\frac{\partial p}{\partial y}\right) - \left(\frac{1}{f^2\rho}\right)\frac{\partial}{\partial x}\left(\frac{\partial p}{\partial t}\right) - \frac{u}{f^2\rho}\frac{\partial^2 p}{\partial x^2} - \frac{v}{f^2\rho}\frac{\partial}{\partial y}\left(\frac{\partial p}{\partial x}\right) - \frac{g w}{f^2 T}\frac{\partial T}{\partial x} - \frac{C_d|\vec{V}|\bar{v}}{fh} \quad (4)$$

$$v = +\left(\frac{1}{\rho f}\right)\left(\frac{\partial p}{\partial x}\right) - \left(\frac{1}{f^2\rho}\right)\frac{\partial}{\partial y}\left(\frac{\partial p}{\partial t}\right) - \frac{u}{f^2\rho}\frac{\partial}{\partial x}\left(\frac{\partial p}{\partial y}\right) - \frac{v}{f^2\rho}\frac{\partial^2 p}{\partial y^2} - \frac{g w}{f^2 T}\frac{\partial T}{\partial y} - \frac{C_d|\vec{V}|\bar{u}}{fh} \quad (5)$$

(a) (b) (c) (d) (e) (f)

where ρ is the air density, f the Coriolis parameter, p the pressure, u and v the wind components, g the acceleration of gravity, w the vertical velocity, T the temperature, C_d the drag coefficient and h the height of the PBL.

These equations form the basis for the dynamical predictors computed from the large-scale REAN data. Terms (a) are the components of the geostrophic wind. The geostrophic balance usually dominates and captures the mean surface flow relatively well offshore over the ocean where friction is limited. This component is well represented by most large-scale data sets. However, it fails to represent correctly the surface winds in areas where the pressure field changes rapidly in time and/or space. This is often the case near the coast of BC, where changing pressures and friction may cause large ageostrophic accelerations. Therefore, the coastal wind dynamics usually involves relatively large ageostrophic winds. These winds, represented by term (b) to (f) in Eqs. (4) & (5), are caused by (b) local changes of surface pressures due to moving and/or developing pressure systems (isallobaric wind), (c) and (d) horizontal variations of the pressure field (advective wind), (e) vertical displacements from one pressure pattern to another (convective wind) and (f) the friction in the boundary layer (antitriptic wind). Ageostrophic winds can be relatively small and negligible in some cases. However, some of them may be as large as the geostrophic component especially in baroclinic conditions. In addition, the rugged west coast topography and islands play an important role in the distribution of regional airmasses. These often generate meso- or smaller scale systems that can create departures from geostrophy. There is further discussion on this subject in Bluestein (1992).

The above predictors, together with additional ones, are listed in Table 1 with the definitions and abbreviated names. This table includes the surface wind components and wind speed from the REAN data set and terms (a) to (e) from Eqs. (4) & (5). Term (f) is represented by 3 series of predictors: (1) temperature gradients and wind shear for stability parameters and (2) thermal wind for baroclinic effects. In addition, there are expressions for the vertical motion, the Rossby radius and the Froude number to account for the adjustment of the momentum equations in channels and for blocking effects along the coast (Overland 1984, Overland & Bond 1995). Finally, the Julian day is also included. Some predictors might be redundant but they were all kept for the tree building process.

4.2. Final predictors

From the CART ranking (not included), the most important predictors by far are the surface wind components from the REAN data set. This is not surprising because the REAN data set contains buoy wind information. The remaining predictors have much smaller rank numbers. However, they are considered important to capture some of the variance associated with meso- and smaller scale effects, especially in coastal waters. In general, 71 different predictors were used at least once in different NFIS wind models. However, only 12 of them appear in more than 20 % of the models (Fig. 3). From the most to the least frequent, these predictors are:

(1) u component of the surface wind (SU0),
(2) v component of the surface wind (SV0),
(3) magnitude of the surface wind (SS0),
(4) Julian day (JULDAY),
(5) v component of the geostrophic wind at mean sea level (VGMSL),
(6) u component of the surface advective ageostrophic wind (UAA),
(7) temperature gradient below 850 hPa (DTP0850),
(8) u and v components of the thermal wind near the surface (DUG0, DVG0),
(9) u component of the geostrophic wind at 850 hPa (UG850),
(10) u component of the thermal wind at 850 hPa (DUG850) and
(11) geopotential height at 300 hPa (PHI300).

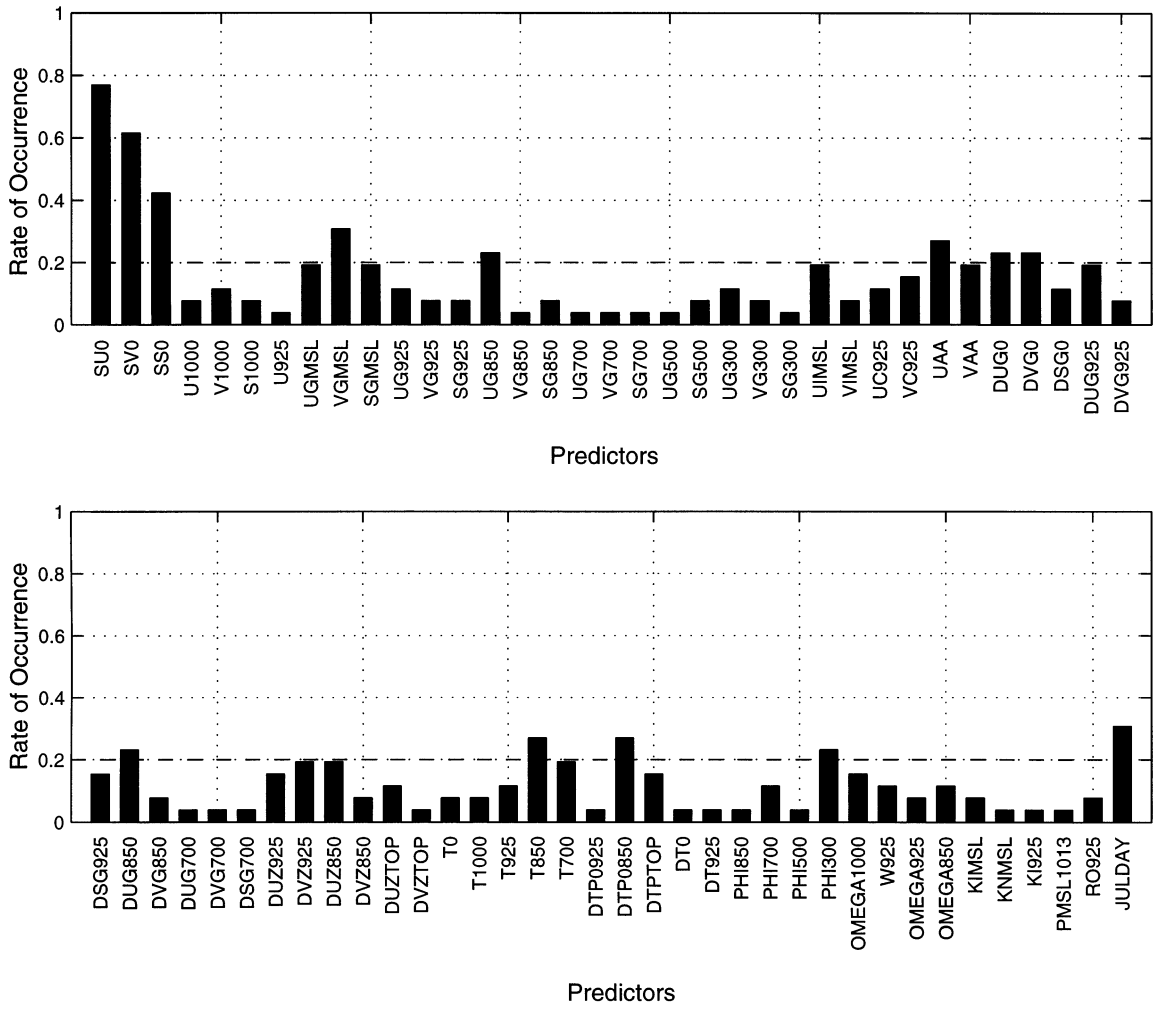


Fig. 3. Rate of occurrence of each predictor that enters the wind models. Dashed line represents a rate of occurrence of 20%. See Table 1 for predictor names

Predictors 1, 2, 3, 5, 9 and 11 represent the large-scale pressure field, predictor 4 represents climatology, predictor 7 is a stability parameter, predictor 6 is the advective effect and predictors 8 and 10 represent the baroclinicity in the lower atmosphere. In retrospect, the climatology predictor may have been better represented as the sine or cosine of the Julian day. However, analysis of MLR results showed Julian day is a minor predictor in terms of variance explained. Even though it was picked in about two-thirds of the models, it appeared only after several other predictors had already been chosen. In spite of their physical similarity, the REAN surface winds and geostrophic MSL winds are sufficiently different that they survived the CART predictor selection process as separate predictors. Analysis of MLR results showed the MSL geostrophic winds are also minor predictors since they always appeared only after the REAN surface winds and several other predictors had already been picked.

Except for the fact that REAN wind components are excellent predictors, no general consensus emerged for the number and type of predictors necessary to model buoy winds in different areas. However, in most cases, the predictands at the coastal sites require more predictors than those at the offshore sites. This can be seen in Fig. 4, where the number of predictors is shown for each predictand and buoy site. Faucher & Pandolfo (1998) showed that the buoy network can be divided into 3 main groups according to the influence of the coastal topography on the atmospheric surface flows. Group 1 is composed of Buoys 46132, 46145, 46183, 46185, 46204 and 46206, where the wind is primarily bi-directional due to important channeling and steering effects. In addition, meso- and smaller scale eddies often develop in Hecate Strait and Queen Charlotte Sound during the summer season, especially behind cold fronts. Group 2 is formed of Buoys 46147, 46205, 46207 and 46208, where the predominant wind direc-

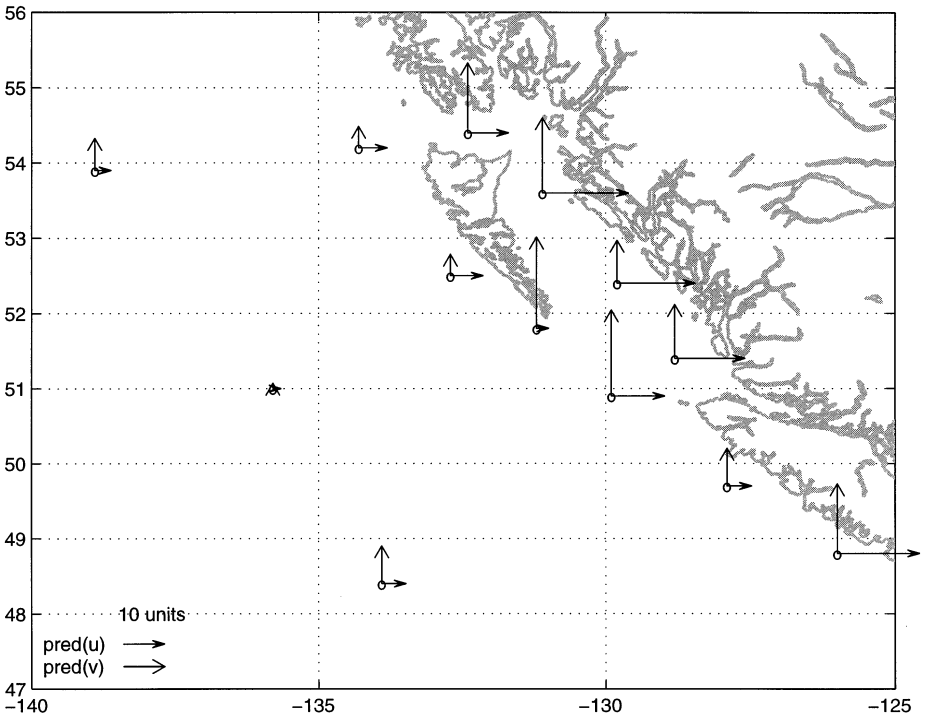


Fig. 4. Number of relevant predictors for each predictand (*u* and *v*) at each buoy site. Vector length along *u* and *v* axes is proportional to the number of predictors required to model the given wind component as determined by CART

tions span a relatively wide range of angles due to limited influence from the coastal boundaries. Group 3 includes Buoys 46004, 46036 and 46184, where the wind distribution is more representative of the large-scale synoptic atmospheric circulation. Despite the lack of consistency in the type of predictors in each group, there is some agreement in the number of predictors required to model each predictand. More than 10 predictors are required to model the wind components at most sites in Group 1. Ageostrophic influences are important in this group and they are partially lost at the scale of the REAN surface winds. Therefore, the REAN surface winds alone are insufficient to reproduce the observed winds at coastal buoy sites. On the other hand, less than 7 predictors enter the wind models for buoys in Group 3. In particular, the *v* component of Buoy 46004 is entirely represented by the REAN *v* component at the first sigma level. Finally, the number of predictors for Group 2 varies between 3 and 17.

5. VERIFICATION OF THE DOWNSCALING PROCEDURE

5.1. Comparison with forward stepwise regression

The fit of each CANFIS model for the learning data was compared with the fit by MLR in which forward stepwise predictor selection was continued until an equal number of predictors were selected. Results for all buoys are shown in Fig. 5. The order of buoys plotted in Fig. 5a–f is the same, and was found by sorting increasing root-mean-square-error (RMSE) values for CANFIS *u* in Fig. 5a. RMSE comparisons for all wind magnitudes are given in Fig. 5a,b. For all buoys the RMSE for CANFIS *u* and *v* models is less than for MLR models. At 2 buoys (46183 and 46147), the difference in RMSEs is more than 10% of the MLR model's RMSE for either *u* or *v* or both, at 7 buoys the difference is more than 5%. The least error is for the outer buoys and the greatest error is for buoys relatively close to the rugged coast. The improvement of CANFIS models extended into the 0–10 and 90–100 percentile ranges of the observed *u* and *v* wind distributions for all buoys. Fig. 5c,d show RMSEs for combined data sets segregated by the 0–10 and 90–100 percentile values of observed data, where the boundary values used for segregating model data are those used for the observed data. The CANFIS models at all buoys have less RMSE than the MLR models. At 7 buoys (46207, 46206, 46204, 46183, 46147, 46145 and 46132) the difference in RMSEs is more than 10% of the MLR model's RMSE for either *u* or *v* or both, and at 11 buoys the difference is more than 5%. The number of buoys with these improvements is much greater than those mentioned above for the non-segregated wind data. Fig. 5e,f

shows a 'normalized percentile range population score' for combined data sets segregated by the 0–10 and 90–100 percentile values of observed data. The score for a data segregation bin is defined as the population of model data in the bin minus the population of observed data in the bin, divided by the population of observed data in the bin, where the boundary values used for segregating model data are those used for the observed data. The bounds of the score are –1 to 9, with 0 being a perfect score. A positive score means the observed percentile range is over-represented by the model data, and a negative value means it is under-represented. Fig. 5e,f shows the combined 0–10 and 90–100 percentile range of observed *u* and *v* data is under-represented to a lesser degree by CANFIS models than by MLR models at nearly all the buoys. The best scores by both models were at the 3 buoys far from shore, the worst fits at buoys in the vicinity of Queen Charlotte Sound, Dixon Entrance, and off the south end of Vancouver Island. The greatest discrepancy between CANFIS and MLR *u* and *v* population scores was near the coast at Buoy 46183; however, 2 of the 3 outer buoys showed relatively large discrepancies.

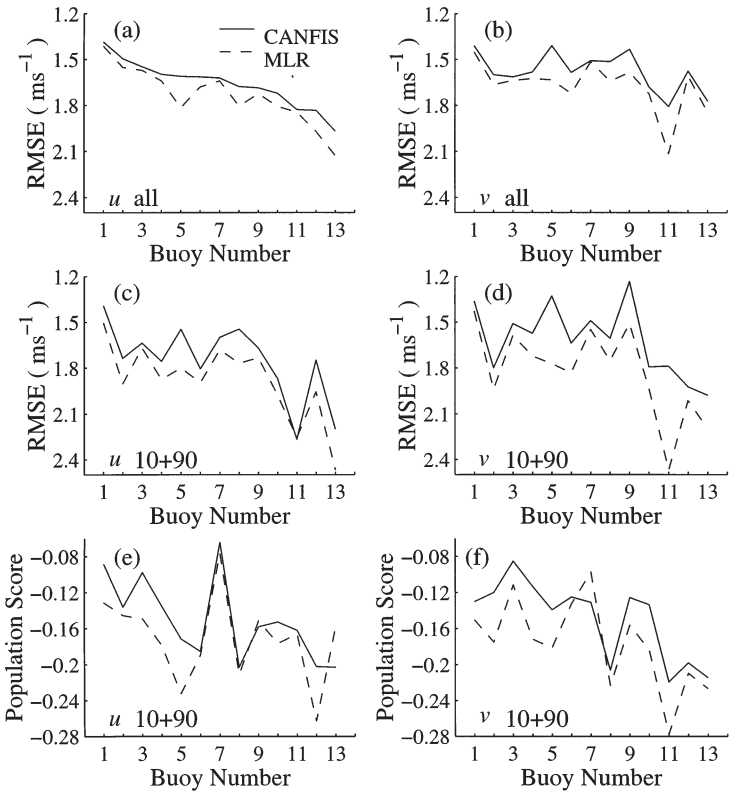


Fig. 5. (a,b) Root-mean-square-error (RMSE) for CANFIS and MLR models of *u* and *v* at each buoy; (c,d) RMSE for CANFIS and MLR model data segregated by boundary values obtained for the combined 0–10 and 90–100 percentile ranges of observed data; (e,f) 'normalized percentile range population score' for the combined 0–10 and 90–100 percentile segregation values of observed data. The score is described in Section 5.1. The order of buoy numbers plotted in (a–f) is the same and was found by sorting increasing RMSE values for CANFIS *u* in (a). Numbers 1 to 13 on *x*-axis correspond to Buoys 46036, 46208, 46004, 46205, 46183, 46207, 46184, 46145, 46132, 46185, 46147, 46206 and 46204

5.2. Comparison with independent data

Construction and training of the models were done using buoy and REAN data for the period 1990–1995. Then, REAN data covering the last 40 yr were fed to the models to generate a continuous daily series of surface winds at the locations of the buoys for 1958–1997. In this section, these CANFIS statistical winds are evaluated by comparing them with independent buoy-measured winds. A comparison of MLR versus buoy winds is also included. Since the training period is much smaller than the period for which CANFIS winds are generated (6 yr versus 40 yr) all available buoy data were used to build and train the models. However, some buoy observations exist for 1988 and 1989. Because these consist of only a sparse set of observations they were not used in training the models but are used in this section to validate the CANFIS winds.

Buoy wind observations used for the validation are provided by the following 7 buoys: 46004, 46036, 46184, 46204, 46205, 46206 and 46207. The data were split into 10 different cases (see Table 3 for specific dates). In the first part of the verification, 2-dimensional histograms and time series are presented to compare the regional wind distributions and variability. In the second part, statistical tests are applied to evaluate the accuracy of the CANFIS winds.

The REAN surface winds are included in the verification as a reference and also for comparison because they are generally the dominant predictor. However, the REAN surface wind corresponds to the first sigma level ($\sigma_1 = 0.995 \approx 42$ m) and requires adjustment to represent winds at the height of buoy anemometers (near 5 m). A logarithmic wind profile is used to convert the data:

$$u_{5m} = \frac{u_{REAN} \ln(z/z_0)}{\ln\left[\frac{-H}{z_0}\ln\left(\frac{-0.005P_{msl}}{Hg\rho_0}+1\right)\right]} \qquad (6)$$

where u_{REAN} (m s^{-1}) is the REAN wind, z_0 (m) is the roughness length taken from the REAN data set, z (m) is the approximate height of the anemometer, g (= 9.8 m s^{-2}) is the gravitational constant, H (= 6.9) is the scale height of the atmosphere, P_{msl} (Pa) is the pressure at

Table 3. Buoy sites used and the dates for which data was available for comparison between CANFIS and independent observed winds

Buoy site	Period		Length of period (no. of days)	Missing data points
	From	To		
46004	2 Sep 1988, 18:00 h	20 Jan1989, 06:00 h	139.75	12
46036	2 Sep 1988, 18:00 h	17 Jul 1989, 18:00 h	318.25	53
46184	2 Sep 1988, 18:00 h	17 Jul 1989, 18:00 h	318.25	55
46204	8 Sep 1989, 00:00 h	2 Dec 1989, 06:00 h	85.50	2
46205 (a)	23 Nov 1988, 00:00 h	20 Mar 1989, 18:00 h	117.75	22
46205 (b)	8 Sep 1989, 00:00 h	14 Nov 1989, 12:00 h	77.75	0
46206 (a)	23 Nov 1988, 06:00 h	20 Jan 1989, 06:00 h	58.25	0
46206 (b)	6 Feb 1989, 00:00 h	4 May 1989, 00:00 h	87.25	9
46206 (c)	1 Sep 1989, 18:00 h	8 Nov 1989, 00:00 h	67.50	0
46207	18 Oct 1989, 18:00 h	2 Dec 1989, 06:00 h	44.75	0

the mean sea level and is taken from the REAN data set and ρ_0 (= 1.225 kg m^{-3}) is the approximate surface air density (Stull 1988). Eq. (6) is applied to each component (*u* and *v*). Also, the REAN wind data are interpolated to the buoy sites using a bicubic spline method.

5.2.1. Graphical representation and scores

The regional wind distributions are presented in a series of 2-dimensional histograms for representative buoy sites of Groups 1, 2 and 3 (Fig. 6a–c). Arrows have been drawn on top of histograms to show the predominant wind directions. These directions are determined from a fuzzy K-mean (FKM) clustering method (Dillon & Goldstein 1984).

In Group 1, the stronger winds are primarily bidirectional due to channeling and steering effects of the coastal topography on surface flows associated with synoptic low and high pressure systems. This can be seen from the ellipsoidal shape (skewness) of the histograms for observed data at Buoy 46206, located near the west coast of Vancouver Island (Fig. 6a). Both the CANFIS and MLR data capture the skewness of the observed winds distribution relatively well. However, the CANFIS and MLR distributions have higher kurtosis (i.e. they are more peaked) than the observed one, with the CANFIS histogram being slightly closer to the observed. In contrast, the large-scale REAN data misrepresent the distribution of observed winds and do not properly resolve the predominant wind directions. Also, the REAN data indicate a third predominant direction which is not present in the observed data for Buoy 46206. In Group 2, the influence from the west coast topography is limited, with less steering effect. This is shown by a more circular shape of the observed histogram at Buoy 46205 (Fig. 6b). This wind distribution is represented relatively well by the 3 data sets CANFIS, MLR and REAN. However, the CANFIS data resolve the northerly predominant directions slightly better. Finally, the observed wind distribution of Group 3 is presented in Fig. 6c. Its shape is well reproduced by the 3 data sets since it is largely due to the passage of large-scale synoptic weather systems. However, the CANFIS distribution does slightly better in representing the most probable winds. Its kurtosis is closer to the observed while that of the MLR and REAN distributions is respectively higher and lower than the observed.

In addition to histograms, time series of 6-hourly CANFIS, MLR and observed data (*u* and *v* components) have been analyzed for each buoy site. As an example, the time series for Buoy 46206 are presented in Fig. 7. Correlation coefficients (CANFIS versus observations) for the *u* and *v* components are 0.89 and 0.87 respectively. These indicate a good match between observations and CANFIS winds. Differences occur mostly at high frequencies. The MLR winds time series are very similar to the CANFIS ones and often overlap with them. For all time series, the correlation coefficients between CANFIS or MLR and observed wind components fall between 0.85 and 0.93 for sites in Groups 1 and 2 while those between REAN and observed winds fall between 0.6 and 0.9 (Fig. 8). Discrepancies between CANFIS (or MLR or REAN) and observed winds increase for weaker winds. Weak winds tend to be more important in coastal areas due to frictional effects and are partly responsible for the lower correlation at coastal sites. Fig. 8 indicates that on average the correlation coefficients between CANFIS and observed winds slightly (largely) surpass those between MLR (REAN) and observed winds for buoys near the coastline. Offshore, the correspondence between CANFIS, MLR or REAN winds and observed ones is almost perfect for both directions (zonal and meridional) with correlation coefficients above 0.95.

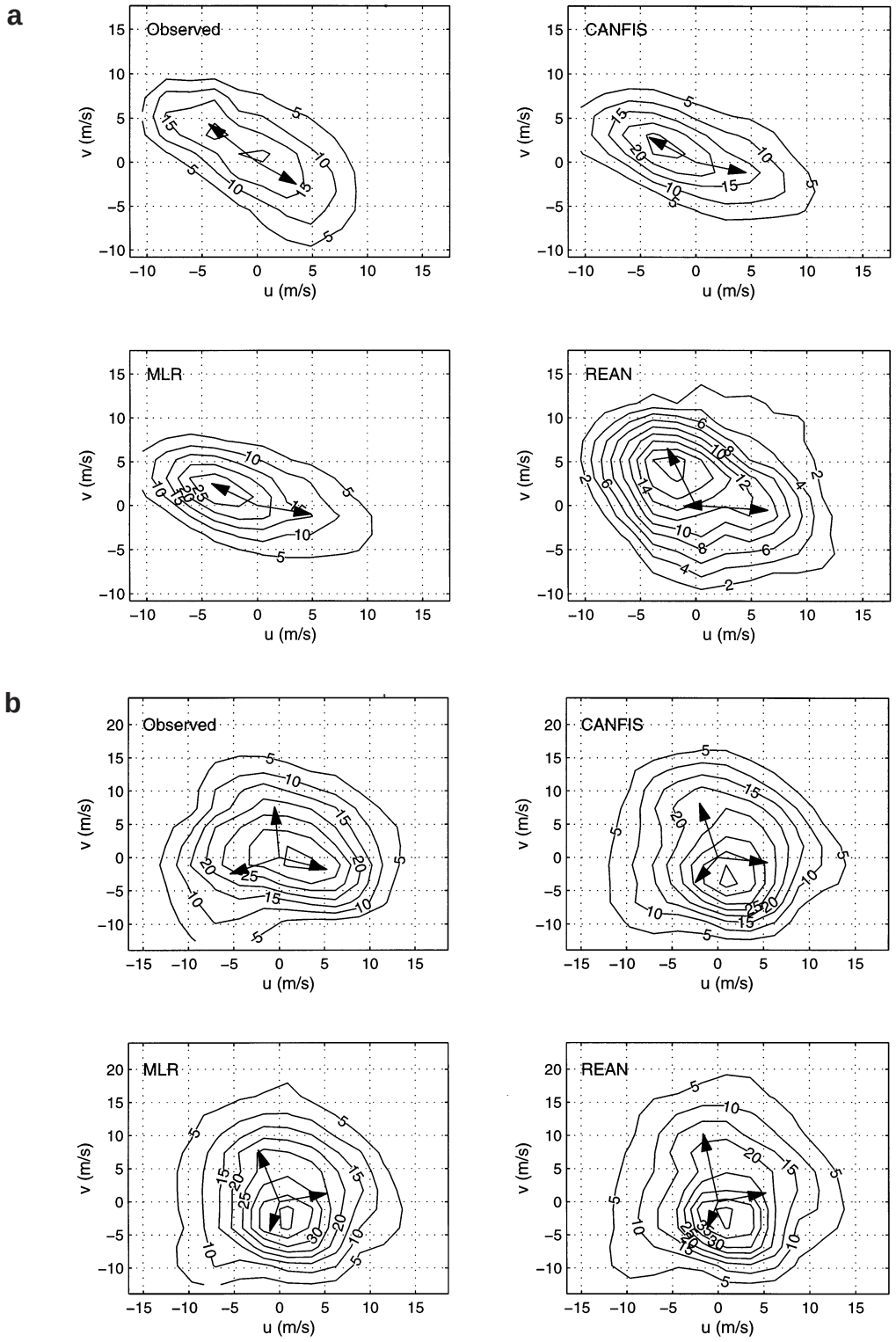
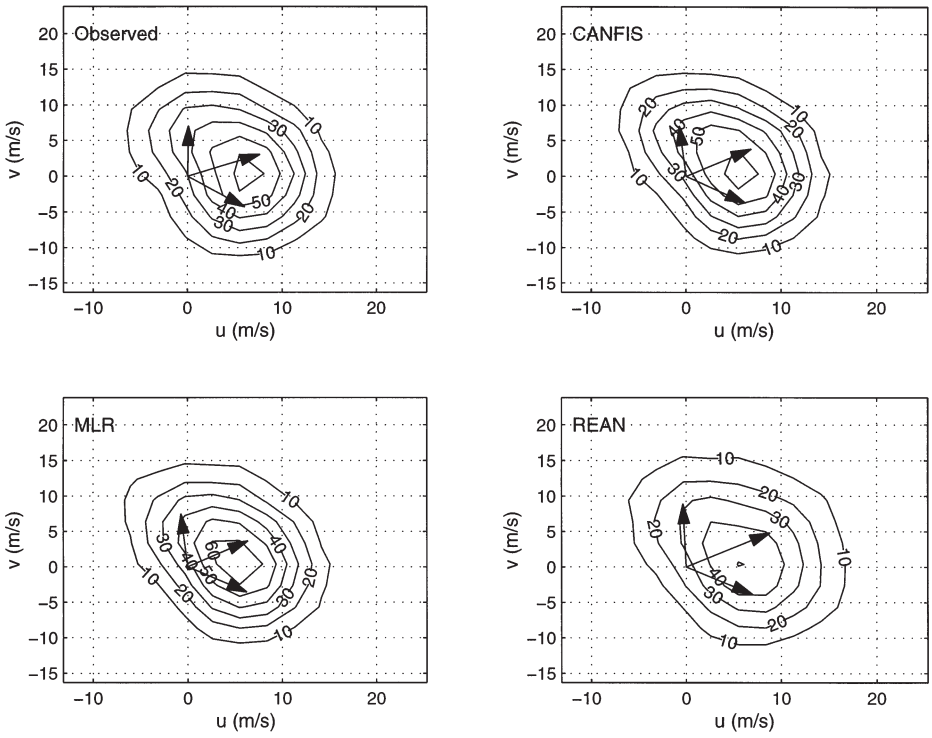


Fig. 6 (above and facing page). Two-dimensional histograms for 3 buoy sites (1 buoy for each group): (a) Buoy 46206 case (a) for Group 1, (b) Buoy 46205 case (a) for Group 2, and (c) Buoy 46004 for Group 3. The histograms are constructed from 6-hourly zonal *(u)* and meridional *(v)* components of buoy, CANFIS, MLR and reanalyzed NCEP winds. The *x*- and *y*-axes represent wind speeds in m s^{-1}, while contours represent the number of counts per bin at 5 counts bin^{-1} intervals for (a) and (b) and 10 counts bin^{-1} intervals for (c). Arrows: predominant wind directions determined from fuzzy K-mean clustering

c



The fraction of variance explained by model-generated winds and REAN winds is presented in Fig. 9. It is clear that CANFIS and MLR winds simulate the temporal variability of buoy winds better than REAN winds. The latter always overestimate the variance of the meridional winds as well as that of the offshore zonal winds. On the other hand, the variance of CANFIS and MLR winds is nearly equal to that of buoy winds, except for sites close to the coast where the meridional wind variance can be badly underestimated or overestimated.

To give a quantitative measure of the quality of agreement between predicted (CANFIS and MLR) or REAN winds and the independent buoy observations, the mean relative and mean absolute errors (Bias and MAE) are computed for each case:

$$\mathrm{Bias}_{\mathrm{mo}} = \frac{1}{\mathrm{n}}\sum_{i=1}^{\mathrm{n}}(S_{\mathrm{m}}^{i} - S_{\mathrm{o}}^{i}) \tag{7}$$

$$\mathrm{MAE}_{\mathrm{mo}} = \frac{1}{\mathrm{n}}\sum_{i=1}^{\mathrm{n}}|S_{\mathrm{m}}^{i} - S_{\mathrm{o}}^{i}| \tag{8}$$

where S^i represents wind components or wind speed for time i, subscript m represents CANFIS, MLR or reanalyzed data and subscript o indicates observations. In addition, the amount of explained variance in each data set is estimated via the reduction of variance (RV) score:

$$\mathrm{RV} = 1 - \frac{\overline{\sum(S_{\mathrm{m}} - S_{\mathrm{o}})^2}}{\sum(\overline{S}_{\mathrm{o}} - S_{\mathrm{o}})^2} \tag{9}$$

where $\overline{S}_{\mathrm{o}}$ is the mean of the observed values.

Bias values shown in Fig. 10 indicate that the magnitudes of CANFIS and MLR winds, are generally too weak, while those of REAN winds are generally too strong. On average, biases are slightly smaller for CANFIS than for MLR winds and both types of model-generated winds show biases much smaller than those for REAN winds. The MAE scores in Fig. 11a indicate that the CANFIS and MLR errors are relatively small. In general, MAEs for predicted winds are one-fifth to one-half the size of the observed standard deviations, which is quite acceptable. The MAEs of the REAN data are also smaller than observed standard deviations in all cases and are also considered acceptable. However, they are always equal to or larger than the CANFIS or MLR errors. The RV scores (Fig. 11b) indicate that a large fraction of the wind variability is reproduced by CANFIS and MLR winds at each buoy site. The RV index is larger than 60 % for all cases. Differences between CANFIS and MLR RV values are very small. In contrast, REAN data do not fare as well. In particular, the reduction of variance at Buoy 46206 (buoy site 9) is just below 50 % for the zonal component and as low as –30 % for the meridional component. A negative number indicates that a climatological value

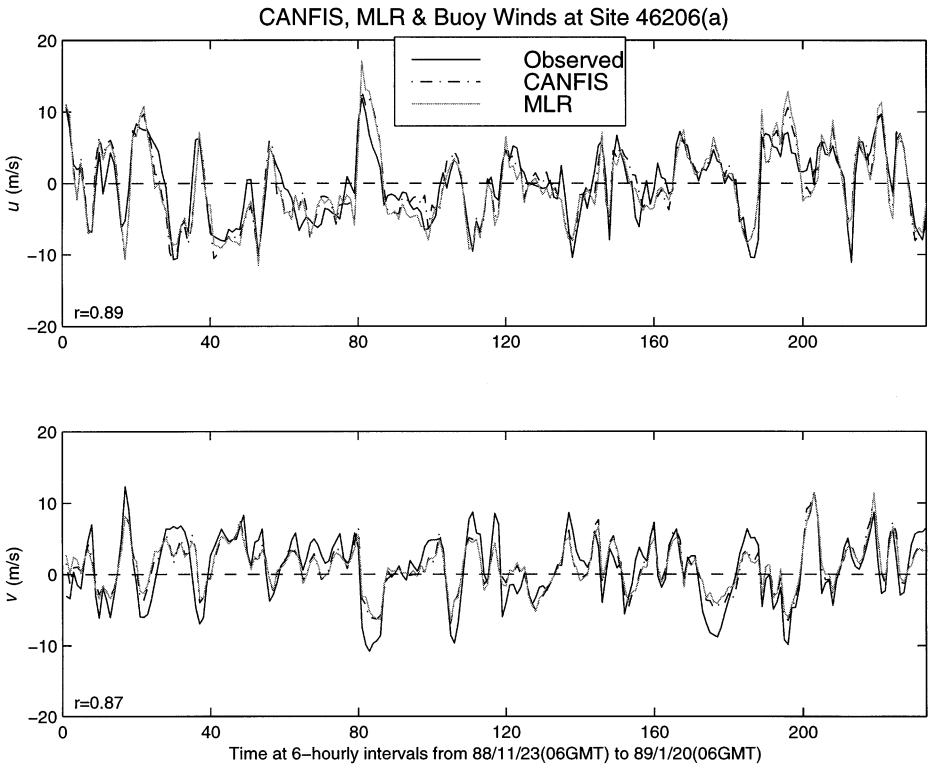


Fig. 7. Time series of 6-hourly buoy, CANFIS and MLR *u* and *v* components for Buoy 46206 case (a). The correlation coefficient between buoy and CANFIS data is indicated in the lower left corner for each time series. Values are in m s^{-1}. Dashed line is a reference corresponding to a speed of 0 m s^{-1}

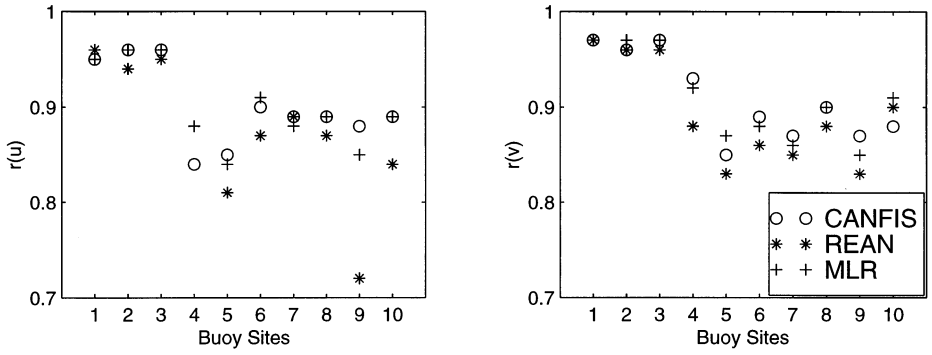


Fig. 8. Correlation coefficients between CANFIS, MLR and REAN and buoy data for *u* and *v* components for the buoy sites listed in Table 3. Numbers 1 to 10 on *x*-axis correspond to Buoys 46004, 46036, 46184, 46204, 46205 case (a), 46205 case (b), 46206 case (a), 46206 case (b), 46206 case (c) and 46207. Correlation coefficient r*(u)* for REAN versus buoy winds for buoy site 4 is 0.61

would give a better representation of the true wind in that case.

5.2.2. Statistical tests

In this section, statistical tests are performed on the wind data to evaluate objectively the accuracy of model-generated winds. Comparisons are also made with REAN data.

A univariate statistical *t*-test at the 5% significance level is applied on wind components *u* and *v* to determine if the means of 2 samples (CANFIS or MLR versus observed winds and REAN versus observed winds) differ significantly from each other. Quantile plots (not included) revealed that the winds

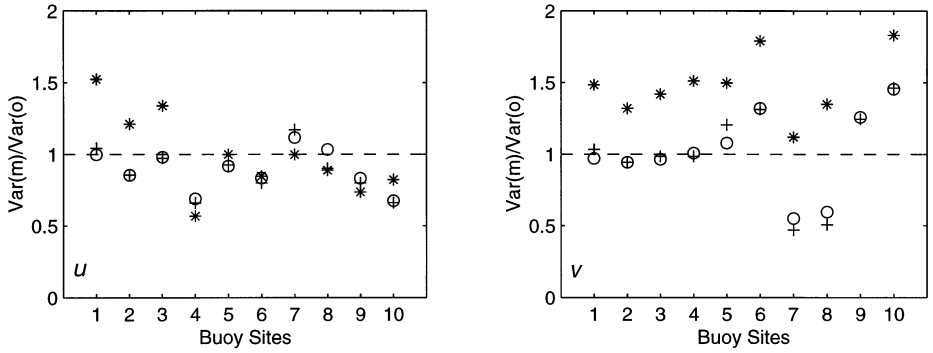


Fig. 9. Fraction of variance explained by the model results presented as the ratio of model wind variance, var(m), to buoy wind variance, var(o), for *u* and *v* components for the buoy sites listed in Table 3. REAN winds are also included for comparison. Var(REAN)/var(o) = 3.25 for *v* at site 9. (○) CANFIS, (+) MLR; and (∗) REAN. Dashed line is a reference corresponding to a perfect ratio of 1. Buoy site numbers as in Fig. 8

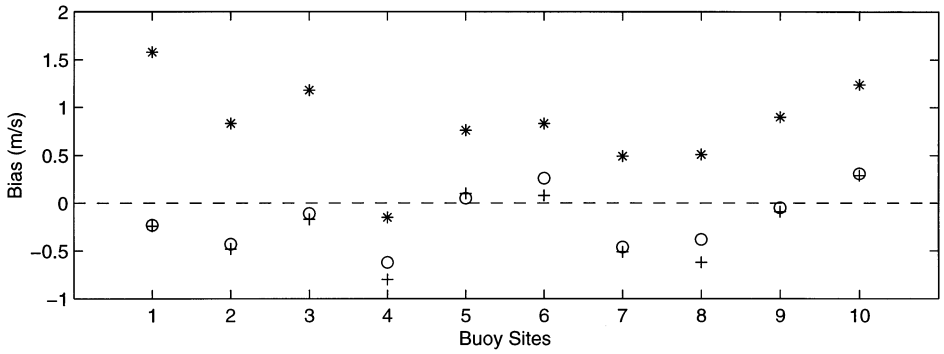


Fig. 10. Bias of wind speed for buoy sites listed in Table 3 for CANFIS (○), MLR (+) and REAN (∗) data. Values are in m s^{-1}. Dashed line is a reference corresponding to a zero bias. Buoy site numbers as in Fig. 8

distributions are close enough to normal distributions that they qualify for a *t*-test. An equivalent sample size equal to the actual sample size divided by 6 was used to eliminate the serial correlation and reduce the chance of falsely rejecting the null hypothesis. Six data points correspond to 36 h which approximates the temporal correlation in wind data coming from the passage of weather systems along the BC coast. Firstly, a series of *t*-tests are applied to the zonal and meridional wind components. Secondly, the zonal and meridional winds are further split into 4 components (northerly, southerly, easterly and westerly) and another series of *t*-tests are applied to each of them.

The results of *t*-tests performed on the zonal and meridional components are presented in Fig. 12. Acceptance of the null hypothesis occurs in the area between the dashed lines in Fig. 12. The figure indicates there is no significant difference in mean wind components between CANFIS or MLR and observed data except for buoy site 1. By comparison, significant differences between REAN and observed mean winds appear for many cases, especially for the zonal component.

In relation to coastal wind dynamics, we were interested in determining if one particular wind direction is better reproduced by the models. Hence, *t*-tests on the mean, conditional upon various wind directions, were conducted. Fig. 13 shows that CANFIS or MLR resolve southerly winds with greater accuracy than any other direction. When disagreement occurs the models tend to overestimate southerly winds. The lowest agreement is seen for the easterly component. In general, CANFIS or MLR tend to underestimate easterly winds. In contrast, REAN winds seem to resolve northerlies and westerlies with greater accuracy than CANFIS or MLR. When disagreement occurs REAN winds tend to overestimate buoy westerly and southerly winds while they tend to underestimate buoy easterly winds. (Positive *t*-statistics for easterly winds indicate underestimation due to the sign convention in meteorology: negative values are used for easterly as well as northerly winds.)

Statistical *t*-tests were also conducted at the 5% significance level to compare the variance of CANFIS,

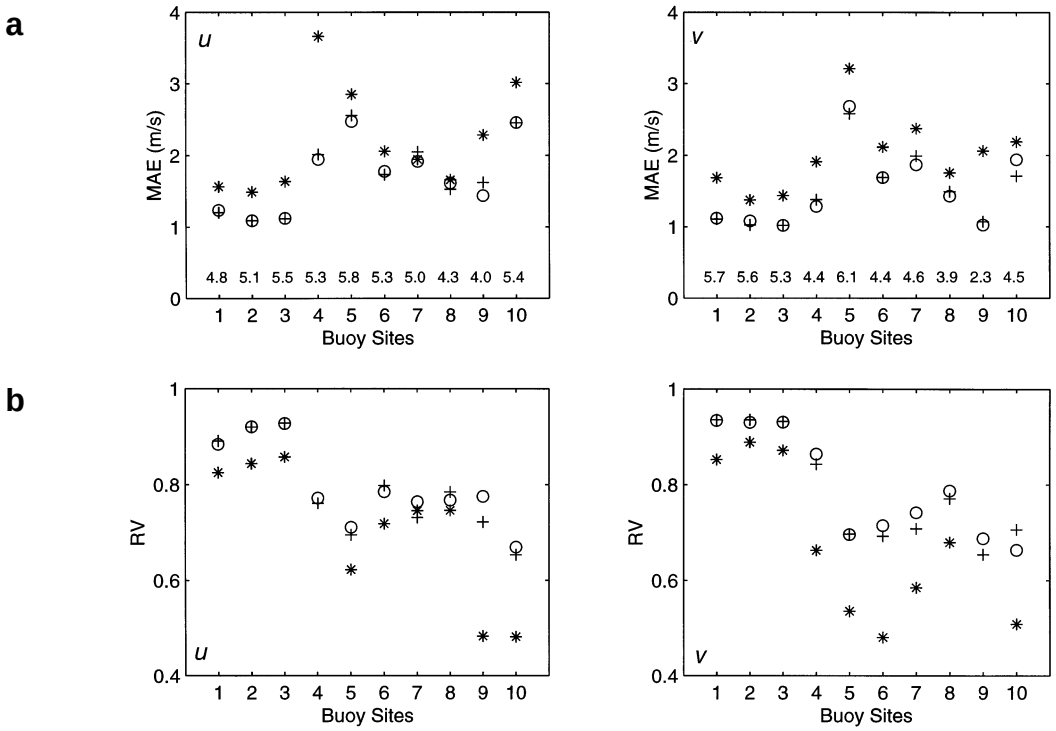


Fig. 11. (a) Mean absolute error (MAE) of wind components u and v for buoy sites listed in Table 3 for CANFIS (○), MLR (+) and REAN (∗) data. Standard deviations of each wind component at each buoy site are indicated at the bottom of the plots. Values are in m s^{-1}. (b) Reduction of variance (RV) on wind components u and v for the same buoy sites as in (a) for CANFIS, MLR and REAN data. Values are fractions. RV(u, REAN) = 0.2 at buoy site 4 and RV(v, REAN) = –0.3 at buoy site 9. Buoy site numbers as in Fig. 8

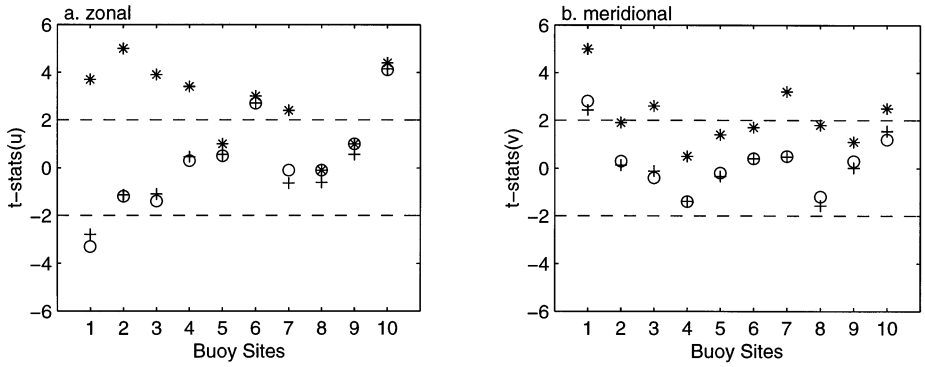


Fig. 12. Values of t-statistics (t-stats) for the (a) zonal and (b) meridional components. (○) Observations versus CANFIS; (+) Observations versus MLR; (∗) observations versus REAN. Dashed lines correspond to a 5% significance level of the tests for 28 to 202 equivalent degrees of freedom. Buoy site numbers as in Fig. 8

MLR and REAN data to that of buoy winds. Results (not included) showed that the null hypothesis could not be rejected for CANFIS, MLR or REAN winds. Thus, all data sets capture the variance well at that significance level. However, Figs. 9 & 11b indicate that the variance of buoy winds is generally better represented by the model-generated data.

In summary, model-generated winds resolve mean wind patterns near the west coast of British Columbia relatively well. In particular, they capture the southerly component well over much of the area. However, CANFIS or MLR easterly winds do not fare as well. This indicates the effect of the coastline on wind dynamics. CANFIS or MLR data are clearly more accu-

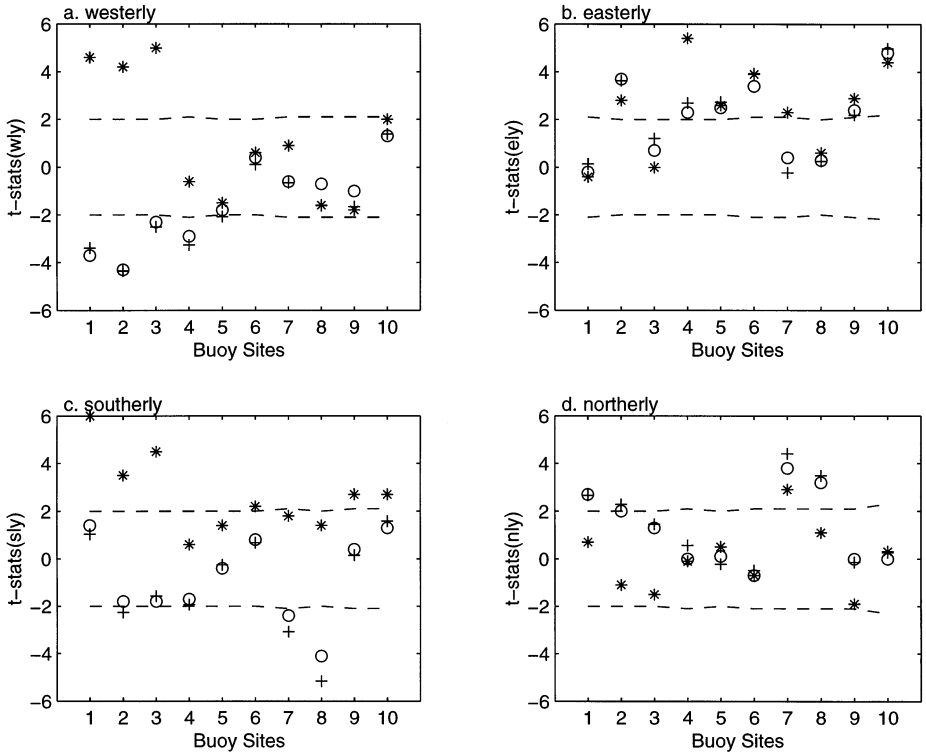


Fig. 13. Values of *t*-statistics (*t*-stats) for each direction: (a) westerly (wly), (b) easterly (ely), (c) southerly (sly) and (d) northerly (nly). (○) Observations versus CANFIS; (+) Observations versus MLR; (✳) observations versus REAN. Dashed lines correspond to a 5% significance level of the tests for 28 to 202 equivalent degrees of freedom. Buoy site numbers as in Fig. 8

rate than REAN data, but differences between CANFIS and MLR results are small.

6. CONCLUSION

The purpose of this project was to produce a reliable series of continuous 6-hourly wind data for the Canadian west coast waters for the 40 yr period from 1958 to 1997. Since available wind data set do not resolve coastal winds with great accuracy (Faucher & Pandolfo unpubl.), a statistical downscaling technique called CANFIS was used to generate wind data from large-scale reanalyzed atmospheric variables.

The CANFIS-generated winds were compared with independent buoy observations to verify the accuracy of the wind models. Winds computed using a simpler technique (MLR) were also included to show the greater efficacy of CANFIS in producing reliable winds when small-scale effects are present. Reanalyzed surface winds were included in the comparison for reference. In general, CANFIS and MLR produce winds that seem to include the effects of mesoscale phenomena often present in coastal areas and the effects of the dominant synoptic systems further offshore. On the other hand, reanalyzed winds misrepresent observed winds in areas close to the coast and between islands. Further offshore, however, they are almost as accurate as CANFIS or MLR data. When comparing both sets of model-generated winds, small improvements of CANFIS over MLR winds are found only for sites close to the coastline.

Finally, preliminary analysis of the auto-correlation functions of the residuals of CANFIS winds indicates serial correlations between 18 and 30 h. This suggests there are missing predictor(s) in our CANFIS modeling and/or that the statistical models are not able to capture all of the non-linear dynamics of surface marine winds. Similar serial correlations are also found in the reanalyzed winds residuals. It would be possible to improve the CANFIS winds by modeling the residuals with a first-order auto-regression algorithm. However, this is beyond the scope of the present work. The next step is to use these CANFIS winds to force models of ocean currents and ecosystems in order to investigate the dynamics of fish populations in relation to climate change and variability.

Acknowledgements. This study was sponsored mainly by the Environmental Adaptation Research Group of Environment Canada, and partially by West Coast - GLOBal Ocean Ecosystems Dynamics and an NSERC research grant to L.P. The

REAN data were obtained from the Canadian Centre for Climate Modelling and Analysis, Environment Canada (Victoria, BC). The buoy data were obtained from the Institute of Ocean Sciences (Sidney, BC). CART software is available from Salford Systems, 8880 Rio San Diego Drive, Suite 1045, San Diego, CA 92108 USA. Software for NFIS and FKM is included in the MATLAB Fuzzy Logic Toolbox, available from The MathWorks, Inc., 24 Prine Park Way, Natick, MA 01760-1500 USA. We thank Dr Francis Zwiers from the Canadian Centre for Climate Modelling and Analysis (CCCMA) and Ted Lord from the Pacific Weather Centre (PWC) for helpful comments.

LITERATURE CITED

Bluestein HB (1992) Synoptic-dynamic meteorology in midlatitudes, Vol 1. Principles of kinematics and dynamics. Oxford University Press, New York

Brieman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. CRC Press, Boca Raton, FL

Burrows WR (1997) CART regression models for predicting UV radiation at the ground in the presence of cloud and other environmental factors. J Appl Meteorol 36:351–544

Burrows WR (1998) CART Neuro-fuzzy statistical data modeling, Part 1. Preprints, 14th Conference on Probability and Statistics in the Atmospheric Sciences, Phoenix, AZ. American Meteorological Society, Boston, MA, p J105–J112

Burrows WR, Benjamin M, Beauchamp S, Lord ER, McCollor D, Thomson B (1995) CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada. J Appl Meteorol 34:1848–1862

Burrows WR, Walmsley J, Montpetit J, Faucher M (1998) CART-NEURO-FUZZY statistical data modeling, Part 2: Results. Preprints, 14th Conference on Probability and Statistics in the Atmospheric Sciences, Phoenix, AZ. American Meteorological Society, Boston, MA, p 160–167

Cherniawsky JY, Crawford WR (1996) Comparison between weather buoy and Comprehensive Ocean-Atmosphere Data set wind data for the west coast of Canada. J Geophys Res 101(C8):18377–18389

Chiu S (1994) Fuzzy model identification based on cluster estimation. J Intelligent Fuzzy Syst 2:269–278

Dillon WR, Goldstein M (1984) Multivariate analysis, methods and applications. John Wiley & Sons Inc, New York

Enke W, Spekat A (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression. Clim Res 8:195–207

Environment Canada (1995) Annual report. In: 1995 ODAS Buoy Service Reports, Buoy Operations, Atmospheric Data Acquisition (Ocean) Division, Environment Canada, P & Y Region. Environment Canada, Vancouver, BC

Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybernetics 23: 665–685

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin A, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetma A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–471

Kidson JW, Thompson CS (1998) A comparison of statistical and model-based downscaling techniques for estimating local climate variations. J Clim 11:735–753

Overland JE (1984) Scale analysis of marine winds in straits and along mountainous coasts. Mon Weather Rev 112: 2530–2534

Overland JE, Bond NA (1995) Observations and scale analysis of coastal wind jets. Mon Weather Rev 123:2934–2941

Saucier WJ (1955) Principles of meteorological analysis. Dover edition, 1989, Dover Publications, Inc, New York

Steinberg D, Colla P (1995) CART: tree-structured non-parametric data analysis. Salford Systems, San Diego

Stull RB (1988) An introduction to boundary layer meteorology. Kluwer Academic Publishers, Dordrecht

Thomson RE (1983) A comparison between computed and measured oceanic winds near the British Columbia Coast. J Geophys Res 88(C4):2675–2683

Editorial responsibility: Hans von Storch, Geesthacht, Germany

Submitted: August 24, 1998; Accepted: December 6, 1998
Proofs received from author(s): March 5, 1999