

Verification of Mesoscale Numerical Weather Forecasts in Mountainous Terrain for Application to Avalanche Prediction

CLAUDIA ROEGER AND ROLAND STULL

Department of Earth and Ocean Sciences, The University of British Columbia, Vancouver, British Columbia, Canada

DAVID MCCLUNG

Department of Geography, The University of British Columbia, Vancouver, British Columbia, Canada

JOSHUA HACKER

National Center for Atmospheric Research, Boulder, Colorado

XINGXIU DENG AND HENRYK MODZELEWSKI

Department of Earth and Ocean Sciences, The University of British Columbia, Vancouver, British Columbia, Canada

(Manuscript received 18 June 2002, in final form 30 May 2003)

ABSTRACT

Two high-resolution, real-time, numerical weather prediction (NWP) models are verified against case study observations to quantify their accuracy and skill in the mountainous terrain of western Canada. These models, run daily at the University of British Columbia (UBC), are the Mesoscale Compressible Community (MC2) Model and the University of Wisconsin Nonhydrostatic Modeling System (NMS). The main motivations of this work are: 1) to extend the lead time of avalanche forecasts by using NWP-projected meteorological variables as input to statistical avalanche threat models; and 2) to create another tool to help avalanche forecasters in their daily decision-making process.

Observation data from the Whistler/Blackcomb ski area in the British Columbia (BC) Coast Mountains and from Kootenay Pass in the Columbia Mountains of southeast BC are used to verify the forecasts. The two models are run with grid spacings of 3.3 km (MC2) and 10 km (NMS) over Whistler/Blackcomb, and with 2, 10 (MC2), and 30 km (NMS) over Kootenay Pass. The quality of the forecasts is measured using standard statistical methods for those variables that are important for avalanche forecasting. It is found that the raw model output has biases that can be easily removed using Kalman filter predictor postprocessing. The resulting automatically corrected forecasts have quite small absolute errors in temperature (0.7°C).

It is also found that the coarser-resolution NMS model produces comparable results to the finer-resolution MC2 model for precipitation at Kootenay Pass. These objective forecast errors are of the same order of magnitude as the meteorological observation (sampling/representativeness) errors in the snowy, windy mountainous terrain, resulting in forecasts that have value in extending the range of avalanche forecasts for locations such as Kootenay Pass, as discussed in a recent study by Roeger et al.

1. Introduction

Forecast verification, as discussed in this paper and understood in meteorological literature, is concerned with measuring the quality of a forecast. In general, “the process and practice of determining the quality and value of forecasts” is called forecast evaluation (Murphy and Daan 1985). Two types of forecast evaluation with different goals can be distinguished: empirical evalua-

tion (verification) with the goal to determine the quality of a forecast; and decision-theoretic (or operational) evaluation, which is important to relate the value of a forecast to its users. Work in this latter area has been concerned with the development of measures of the monetary value of forecasts. For avalanche forecasting, the value of the forecast depends highly on the quality of the forecasts.

Along with accuracy, skill is also an important measure of the quality of a forecast. In this context, accuracy is the ability of a forecast to match the observation and the extent to which a forecast agrees with the measurement (Roeger et al. 2001). A forecast of good quality may also show skill, which is the degree of correctness

Corresponding author address: Roland Stull, Atmospheric Science Programme, Dept. of Earth and Ocean Sciences, 6339 Stores Rd., Vancouver, BC V6T 1Z4, Canada.
E-mail: rstull@eos.ubc.ca

above some reference baseline, such as a climatological average. Thus, by determining the accuracy and skill of a forecast, one can improve it and use it with confidence in the future.

Although these theoretical ideas about weather forecast verification are well known, not many verification results are actually published or are easily accessible for mesoscale models in complex terrain. While most weather forecast centers have their own routine model verification schemes using several statistical methods for their global- and regional-scale models, results of those verifications often appear only in technical reports or internal Web pages. Therefore, it was not possible to compare our results to other results. Similarly, we know through personal communications with colleagues at workshops and conferences that attempts are being made to apply mesoscale models to complex terrain, but these ideas are young and few verification results have been published except for a handful of "golden" case study days. We encourage the routine use and verification of mesoscale models in complex terrain.

The accuracy of the Mesoscale Compressible Community (MC2) Model and the University of Wisconsin Nonhydrostatic Modeling System (NMS) is determined with statistical verification against both manual and automatic surface weather observations for continuous as well as categorical variables at two avalanche sites in British Columbia (BC), Canada. Numerical weather forecasts depend highly on the initial conditions and the topography estimation in mountainous terrain, and hence, the resolution of the model grid. To estimate this dependence—which is especially important in mountainous terrain—numerical weather prediction (NWP) models are run at the University of British Columbia (UBC) with slightly different initial conditions and with different grid resolutions for the same forecast period. This is done to estimate the improvement using a higher-resolution grid and to reveal the effect of different topography approximations from each model.

Snow avalanche forecasting is a complex problem, based on the interaction of weather, terrain, and the snowpack. It is defined as the prediction of current and future snow instability in space and time relative to a given triggering level. The goal of avalanche forecasting is to minimize the uncertainty about instability introduced by the temporal and spatial variability of the snow cover (including terrain influences), any incremental changes in snow and weather conditions, and any variations in human perception (McClung 2000).

Statistical avalanche prediction refers to the organization of a database of previously measured parameters, including avalanche occurrences, for use with a computer to help compare current or expected future conditions with past ones. There are many different parameters that contribute toward snowpack instability, but primary emphasis is on meteorological data (McClung and Schaerer 1993), not only because they are usually measured by instruments at regular intervals and there-

fore are relatively easy to get, but also because snow avalanche forecasting is a multiscale problem (La-Chapelle 1980; McClung and Schaerer 1993; McClung 2000). Office-based forecasters often need to predict avalanches for an entire mountain range or parts of ranges, for which high-quality meteorological information is more relevant and can assume greater importance than local snow-stability information.

Precipitation and temperature are the key variables for dry or wet avalanche forecasting, respectively. Dry avalanches are most often slab avalanches that occur due to an initial failure underneath a wind-packed layer of snow. This slab may be of several centimeters to more than a meter in thickness, and its fracture line can reach over entire mountain slopes. Once in motion, the slab breaks into blocks and particles, which—if the original snow is very dry—may result into the separation of a dust cloud with very low density.

Wet avalanches are often loose avalanches, usually triggered by heavy melt due to warming. Loose snow avalanches—as opposed to the cohesive nature of slab avalanches—start from a point at or near the surface snow and spread out in a triangular pattern as they move down the slope. For this avalanche type the snow must have low cohesion. The cohesion of snow decreases with increasing water content; namely, wet snow has less cohesion than dry snow. Warm-up-related avalanching can abruptly occur when the air temperature warms to 0°C in the initiation areas (McClung and Schaerer 1993).

Due to the great variety of climate zones in Canada, the demand for avalanche prediction is at the mesoscale (horizontal scales of 2–1000 km), which requires more accurate finescale prediction than for synoptic-scale forecasts (1000–20 000 km). The avalanche hazard is concentrated in local areas where people and facilities are present in mountainous regions (McClung 1995). Any avalanche model is dominated by the interaction of weather with terrain and the physical processes in the snow cover, which leads to avalanche formation. Therefore, detailed networks of meteorological and snowpack measurements combined with avalanche observations are necessary for good avalanche forecasts (Foehn 1998).

The comparison of output variables from NWP models and input variables for statistical avalanche forecasting models (AFM) shows that a lot of the NWP variables can be directly applied into an AFM or can easily be derived. The remaining AFM variables are usually measured in the field and cannot be directly received from standard weather forecasts. But they can be estimated or approximated with empirical relationships. When weather forecasts are reasonably accurate on the local scale and they are included in avalanche forecasting models, the two fields may be combined successfully, allowing the prediction of future snowpack instabilities and avalanches (Roeger et al. 2001).

This paper contains comprehensive results of the

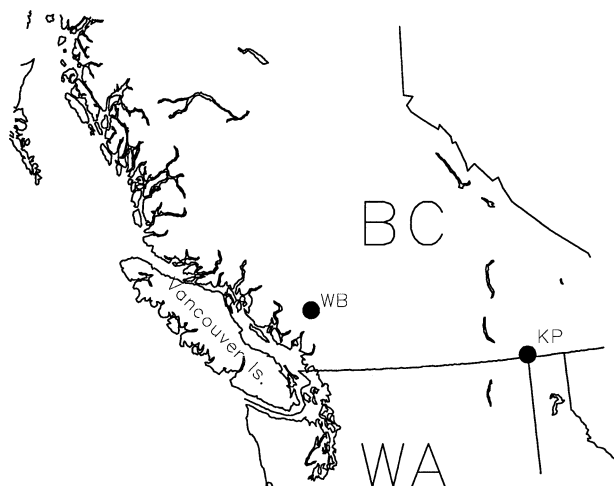


FIG. 1. Map of southwestern BC, and northern Washington (WA), indicating the site locations. WB: Whistler/Blackcomb, KP: Kootenay Pass.

weather forecast verification and the methods used. The avalanche sites and numerical model characteristics are identified in section 2, and statistical verification methods are discussed in section 3. Verification results for the key avalanche-prediction variables of wind, precipitation, and temperature are presented in section 4, with conclusions in section 5.

2. Data

Data from two different sites are used. The ski area Whistler/Blackcomb (50.05°N, 122.9°W) in the Coast Mountains in BC represents a maritime mountain climate, which is characterized by relatively heavy snowfall and relatively mild temperatures, resulting in deep snow covers and the possibility of rain at any time during the winter. Kootenay Pass (49.05°N, 117.0°W) in the southern Selkirk Mountains (Columbia Mountain Range) of southeastern BC represents a transitional climate zone, midway between a maritime and a continental climate (McClung and Schaerer 1993; Armstrong and Armstrong 1987). While a continental snow climate is characterized by relatively low snowfall (shallow snow covers), cold temperatures, and a location considerably inland from coastal areas, the transitional snow climate zone shows higher precipitation amounts resulting in middeep snow covers and temperatures cold enough for only snow events during midwinter, but also mostly located inland from coastal areas. Figure 1 shows a map with the two sites indicated.

In addition to two different climate zones, these sites represent two different types of operations (ski area versus highway operation) affected by avalanches. The ski area Whistler/Blackcomb is concerned about avalanches that may start on or above ski runs. While most ski runs experience relatively low avalanche danger due to constant grooming and skier traffic throughout the season,

TABLE 1. Parameters used from each station at Whistler/Blackcomb and Kootenay Pass and their type of observation (M: manual; R: remote); asl: above sea level.

| Weather station | | Parameters |
|--------------------|------------------------------|----------------------|
| Whistler/Blackcomb | Catskiller (1550 m asl) | Temperature (R) |
| | Horstman Hut (2240 m asl) | Precipitation (M) |
| | | Temperature (R) |
| | | Wind speed (R) |
| | | Wind direction (R) |
| Whistler/Blackcomb | Whistler Alpine (1825 m asl) | Temperature (R) |
| | | Wind speed (R) |
| | | Wind direction (R) |
| | Pig Alley (1650 m asl) | Precipitation (M) |
| | | |
| Kootenay Pass | Kootenay Pass (1780 m asl) | Temperature (R) |
| | | Precipitation (R, M) |
| | | Wind speed (M) |
| | | Wind direction (M) |
| | Stagleap (2140 m asl) | Temperature (R) |
| | | Wind speed (R) |
| | | Wind direction (R) |

steep less-trafficked slopes higher above may require systematic avalanche control programs. In order to avoid large hazardous avalanches, some ski runs must be closed regularly in order to intentionally trigger smaller avalanches. These closures should be short in time and locally limited. The time of concern is during ski hours, roughly between 0900 and 1600 local time.

The highway operation at Kootenay Pass is concerned with avalanches large enough to cover parts of the highway. Avalanche mitigation efforts primarily consist of either hand or artillery control. Because large avalanches need to be avoided due to the high costs associated with highway closures, avalanche control is a 24 h day⁻¹ concern.

With data from six meteorological observation stations, as described in detail later, a wide range of different locations is covered. Elevation of the six stations varies from 1550 m (Catskiller) to 2240 m (Horstman Hut); their surrounding topography varies from a partly sheltered location at midmountain (Pig Alley) to a location on top of a mountain ridge, well exposed to the wind (Stagleap). By using data from these six stations the behavior of the models in complex terrain is tested. Of most interest is model performance and output quality for different elevations and topographical characteristics. This information is critical for both model developers and end users.

a. Meteorological observations

Observation data from Whistler/Blackcomb were from automatic weather stations as well as manual observations taken by ski-patrol avalanche forecasters. Remote, automatic weather stations record weather conditions hourly or every 15 min, depending on the station. Manual observations are done twice daily. Table 1 lists the parameters used from each weather station. Except

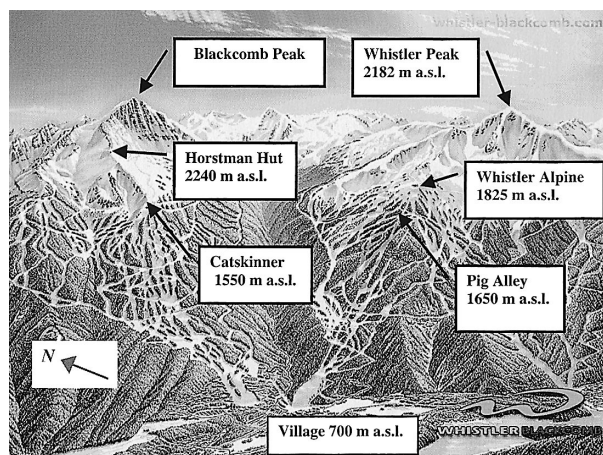


FIG. 2. Map of ski area Whistler/Blackcomb with weather stations indicated. Not to scale. Distance between the two peaks is ≈ 6500 m.

for precipitation, which is observed manually, data from remote stations have been available for two winters, 1998/99 and 1999/2000. Precipitation was verified with data from 1999/2000.

Precipitation rate (mm h^{-1}) was collected hourly from gauge measurements (remote observations) and twice daily with snow-measurement boards (manual observations: solid precipitation) at Kootenay Pass. At Whistler/Blackcomb, snow measurements were assessed with manual snowboards at the weather stations Pig Alley and Catskinner.

For better orientation Fig. 2 shows a drawing map (not to scale) of the ski area Whistler/Blackcomb with the weather stations and their altitude indicated. Figure 3 shows a topographic map of the area with the scale given. The distance between Blackcomb Peak and Whistler Peak is 6.6 km. From Whistler Village in the valley, the distances are 6.8 km to Blackcomb Peak and 5.9 km to Whistler Peak.

Catskinner (1550 m asl) and Horstman Hut (2240 m asl) are on Blackcomb Mountain (Fig. 2). Precipitation is measured at Catskinner, which is on the southwest side at midmountain elevation, neither particularly sheltered nor exposed. (Note that an ideal site would have had sheltered locations for precipitation, and well-exposed locations for wind; however, such designs are difficult to implement in some places due to the lay of the land and the forest cover.) Horstman Hut is located on the NW–SE-aligned ridge, northwest of Blackcomb Peak. The station is well exposed to winds from all directions.

At Whistler Mountain, temperature, wind speed, and wind direction were measured at Whistler Alpine at the ski-patrol building near the Roundhouse Lodge (1825 m asl). Since the site is in open terrain, neither sheltered nor particularly exposed, it is an ideal location for wind data verification. The avalanche forecasters also use wind data from other stations, representing more specific locations on the mountain.

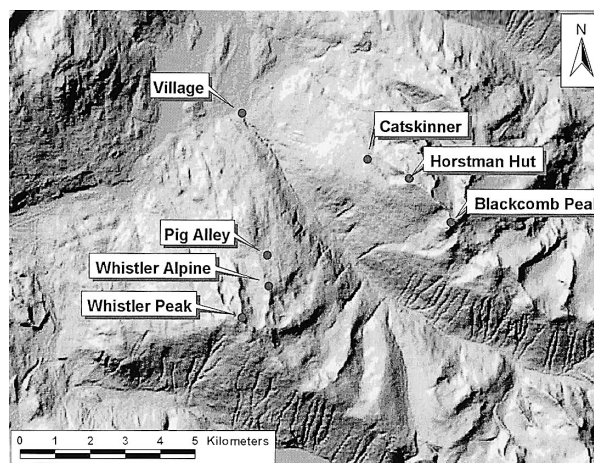


FIG. 3. Topographic map of Whistler/Blackcomb with locations of weather stations and other points of interest.

The field site for precipitation at Whistler (Pig Alley) is at 1650 m asl elevation in a central location of the ski area. The site is surrounded by trees, but fairly open so that the trees have little influence on the measurements. The location has proven to be representative of snow amount at midelevation (J. Tindle, avalanche forecaster, 1999, personal communication).

Observation data from the Kootenay Pass site are described in detail in Roeger et al. (2001). The operation consists of two weather stations collecting manual and remote data: Kootenay Pass and Stagleap. A topographic map of the area with the scale is given in Fig. 4. The manual observation site at the summit of Kootenay Pass is located at 1780 m asl elevation in an open area surrounded by trees. It is fairly sheltered and therefore wind observations here might be too slow, with direction that is less meaningful. Precipitation measurements are representative for the area, and temperatures are typical for this elevation. Temperature is measured at shelter height

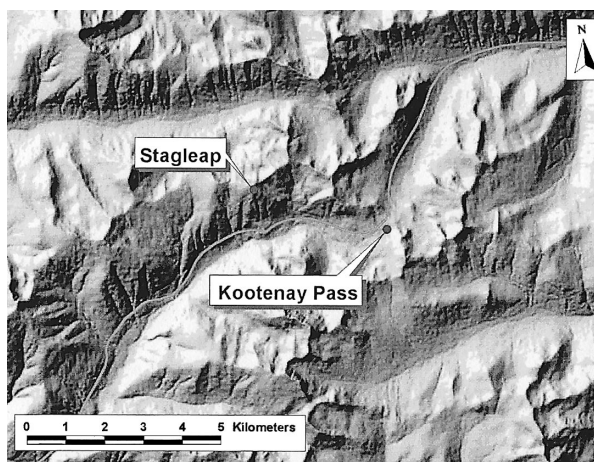


FIG. 4. Topographic map of Kootenay Pass area with the highway and locations of weather stations indicated.

above the ground or snow surface. Stag Leap is a remote weather station at the top of a ridge (2140 m asl) and well exposed to the wind. Wind speeds are therefore typical for this mountain ridge and elevations, but are not representative for some avalanche starting zones at midmountain elevation, especially on the lee side. At this station, winds are measured remotely (anemometer) atop a 10-m-high tower. Data from both areas are gathered according to the guidelines from the Canadian Avalanche Association (CAA 1995).

In general, dysfunctional measurement devices can be a problem especially related to winter weather, for example, frozen anemometers due to significant rime effects. At Kootenay Pass, information about the working condition of the instruments and a first check of the measured value (within a certain range depending on the variable) is done automatically with the measurements (true/false signal). At Whistler/Blackcomb, the avalanche forecasters regularly check their remote, automatic measurements with additional manual measurements as well as examine the significant values by eye. Therefore, the avalanche forecasters know which data are reliable, and a certain standard is maintained at both sites. For this project, all measured data were again examined in detail and only correctly measured data (within its range of uncertainty related to the measurement itself) have been chosen for verification. As a result of this process, 10% of the data (some temperatures at Whistler/Blackcomb, and some winds at Kootenay Pass) were rejected based on observer discretion and evaluation, in order to ensure reliable results.

b. Meteorological forecasts

The two research NWP models used here are the MC2 (version 4.8), refined by Environment Canada's Numerical Prediction Research group (RPN), and the University of Wisconsin NMS. Both were run in real-time for this verification study, making daily forecasts on multiple grids out to 48 h into the future with no manual *ex post facto* tuning, in order to simulate operational conditions.

The MC2 model (Benoit et al. 1997, 2002) utilizes nonhydrostatic, fully compressible, non-Boussinesq dynamics, and is discretized on an Arakawa C grid using semi-Lagrangian numerics and semi-implicit time differencing. The coordinate system is polar stereographic in the horizontal, and modified Gal-Chen in the vertical. The top boundary utilizes an absorbing layer, while lateral boundaries are nested with a "sponge" region. Bottom-boundary fluxes of heat, moisture, and momentum are parameterized using bulk-transfer and similarity algorithms between a force-restore soil layer and a 1.5-order closure turbulence scheme with turbulence kinetic energy prediction in a diffusive boundary layer. Cumulus convection is parameterized using the Zhang and Fritsch method; mixed-phase microphysics with the Sundqvist scheme; and radiation with Foucart-Bonnell

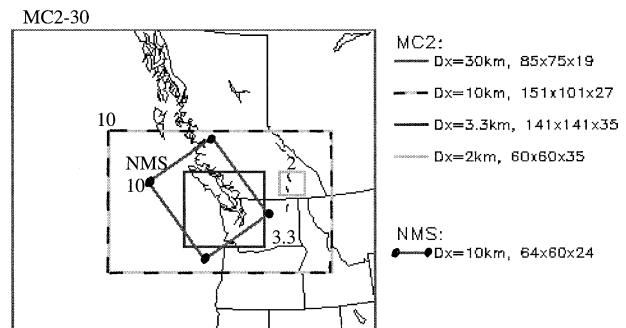


FIG. 5. Forecast grids of MC2 and NMS models run at UBC. The NMS 30-km grid covers the same domain as the MC2 30-km grid.

and Garand schemes (see Benoit et al. 1997 for details). Surface conditions (vegetation, snow cover, sea surface temperature, albedo, etc.) are from climatology fields from the Canadian Meteorological Centre (CMC) of Environment Canada. Benoit and colleagues (1997) have used the model both for real-time operational forecasts over all of North America, and for very fine resolution (3-km horizontal grid spacing) forecasts over the Alps for the Mesoscale Alpine Experiment (MAP; see Benoit 2002).

MC2 was run at UBC with horizontal gridpoint spacings of 90, 30, 10, 3.3, and 2 km, where the finer grids in small domains were one-way nested inside coarser, larger-domain grids (see Fig. 5). The two highest resolutions (smallest grid spacing) have been used for verification, and these grids have 35 layers in the vertical. These horizontal grid spacings are 3.3 and 10 km over Whistler/Blackcomb, and 2 and 10 km over Kootenay pass. Two resolutions were used in order to compare the specific improvements related to increasing resolution. The 10-km grid has $X \times Y \times Z = 85 \times 60 \times 19$ grid points, the 3.3-km grid has $141 \times 141 \times 35$ grid points, and the number of grid points of the 2-km grid is $60 \times 60 \times 35$ (all resolutions are true at 60°N).

The NMS model was developed primarily by G. Tripoli at the University of Wisconsin (Tripoli 1992). It uses a nonhydrostatic, quasi-compressible, non-Boussinesq formulation on local spherical horizontal coordinates and Gal-Chen vertical coordinates. Dynamics utilize an enstrophy-conserving second-order leapfrog scheme on an Arakawa C grid, while the thermodynamics use a flux-conservative sixth-order Crawley scheme. The upper boundary has an absorbing layer, while radiative lateral boundaries are used. A multilayer soil model is used with a Tremback and Kessler parameterization, with Louis surface layer similarity, 1.5-order turbulent kinetic energy (TKE) turbulence closure, a cumulus convection scheme by Kuo and Anthes, mixed-phase microphysics of Flatau et al., and radiation parameterization of Chen and Cotton (see Tripoli 1992 for details). Tripoli has used this model to simulate convection and banding in hurricanes, for daily real-time forecasts for the midwestern United States, and to forecast and sim-

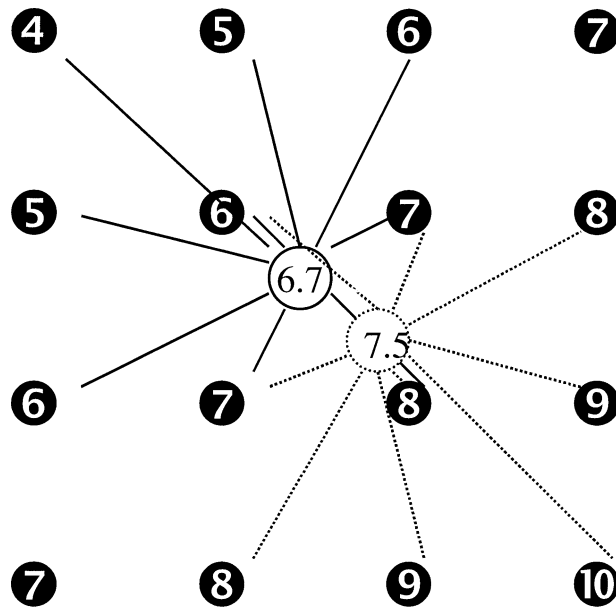


FIG. 6. Highly simplified scheme of interpolation (nonlinear). Two verification points within the same grid cell have different values. The value 6.7 derives from the interpolation between its nine surrounding grid cells, sketched with the solid line. The dashed lines show interpolation for the value 7.5.

ulate winter convection snowbands over Lake Michigan (Kristovich et al. 2000).

NMS was run at UBC for two-way interactive nests with 90-, 30-, and 10-km grid spacing. For verification, 10 km was used for Whistler/Blackcomb and 30 km for Kootenay Pass (because the latter location was outside the operational 3.3-km domain, which was limited by computer power). The number of grid points is $50 \times 68 \times 24$ for the 10-km gridpoint spacing and $68 \times 80 \times 28$ for the 30-km spacing. The vertical domain is also nested (viz., the top of the finest-mesh grid is below the top of the coarser grids). For each weather station, forecast values from the surrounding four or nine grid points have been interpolated to calculate the forecast for the exact location. Figure 6 shows a highly simplified scheme of interpolation (nonlinear), which explains why two verification points have different values although they may be located within the same grid cell.

Initial and boundary conditions for MC2 and NMS coarse grids (90-km grid spacing) are from Eta Model forecasts (U.S. National Centers for Environmental Prediction), valid every 3 h from 0 to 48 h. In turn, forecasts from MC2 and NMS coarse meshes provide the boundary conditions for the embedded finer meshes.

From the MC2 model, not only the raw NWP forecasts were verified with observations, but also forecasts that have been improved by the Kalman-predictor postprocessing correction method (Bozic 1979) have been compared to observation data. The Kalman-predictor correction is an automatic postprocessing method (a type of model-output statistics) that uses the observation

and the original forecast from the day before to calculate the model error. It then predicts the model error for the next day and uses it to correct the forecasts. This recursive, adaptive method “learns” on the fly (see description in appendix B) and does not need an extensive, static database to be trained. It can be used for every forecast where observation data are also available. For days of missing observations, it uses the unaltered corrections from the day before. The Kalman-predictor correction method applied to output from both the NWP models has been tested for all parameters to measure its overall improvement compared to the raw model output.

For the verification of temperature, wind speed, and wind direction, the forecasts were divided into two forecast time periods. The first includes forecasts from 0 to 24 h. The second covers forecasts that are valid 24–48 h into the future. For precipitation, only 0–24-h forecasts could be verified because of gaps in the MC2 forecasts during the 2- and 3.3-km grid test period at the beginning of this project.

3. Evaluation methods

a. Evaluation methods for continuous variables

For continuous variables, standard statistical methods as well as graphical techniques have been used. Emphasis was on robust and resistant mathematical measures. The mathematical measures include interquartile range (IQR) for information about the variation/spread of the dataset, and the median (0.5 quantile $q_{0.5}$) as a single representative number for the dataset. Descriptive statistical parameters (mean M , standard deviation s , variance v) have been calculated as well, but they may be neither robust nor resistant. Robustness and resistance are two aspects of insensitivity to assumptions about the nature of a set of data. Robust methods are generally not sensitive to particular assumptions about the overall nature of the data (e.g., it is not necessary to assume that the data have a Gaussian distribution). A resistant method is not strongly influenced by outliers. As an example, the data series $A = [12; 14; 13; 15; 12; 14; 13; 123]$ contains an outlier (123) about which we do not know if it is a correct measurement (physically possible) or even a typo. The data series B is the same but without the outlier: $B = [12; 14; 13; 15; 12; 14; 13]$. While the mean $M_A = 27$ is strongly affected by the outlier (the mean $M_B = 13.3$), the median $q_{0.5A} = 13.5$ is not (median $q_{0.5B} = 13$); hence, the median is resistant.

A summary of statistical verification equations is given in appendix A. For information about the linear relationship between two datasets, the correlation coefficient [Pearson product-moment; Eq. (A1)] was used. Basic absolute measures for ordinal predictands are the mean error [ME; Eq. (A2)], the mean absolute error [MAE; Eq. (A3)], the mean square error [MSE; Eq.

(A4)] and the root-mean-square error [RMSE; Eq. (A5)].

b. Evaluation methods for categorical variables

For nominal predictands, contingency tables were used for measurements of accuracy (see Table A1 in appendix A, illustrating a 2×2 contingency table). Our measurements include the hit rate (H), the probability of detection (POD), the false-alarm ratio (FAR), and the bias ratio (BIAS). These quantities are given as Eqs. (A6)–(A9).

The hit rate (or the percentage of forecast correct) is the ratio of correct forecast events to the total number of events. The worst possible hit rate is zero. A value of 1 would represent a “perfect forecast.”

The bias ratio is the comparison of the average forecast with the average observation. It is the ratio of the “yes” forecasts to the number of yes observations. The value $\text{BIAS} = 1$ indicates that the event was forecast correctly the same number of times that it was observed. Bias ratios greater than 1 indicate that the event was forecast more often than it was observed (overforecasting). Conversely, bias ratios less than 1 indicate underforecasting. The bias is not an accuracy measure because it says nothing about the correspondence between the forecasts and observations of the event on particular occasions (Wilks 1995).

Equations (A10) and (A11) show the Heidke skill score (HSS) and the true skill score (TSS). They are derived by contingency table analysis as well.

The Heidke skill score is based on the actual forecast hit rate relative to the hit rate expected for random forecasts, which is used as a baseline or reference accuracy measure. Forecasts equivalent to the reference forecasts receive 0 scores. Negative scores are given to forecasts that are worse than the reference forecasts. Perfect forecasts receive a Heidke score of 1 (Wilks 1995).

TSS is a measure of true forecast skill. In short, the true skill score is the POD, adjusted by the POFD (probability of false detection); namely, $\text{TSS} = \text{POD} - \text{POFD}$. It was originally proposed by Peirce (1884), then known as the Hanssen–Kuipers discriminant or Kuipers’ performance index (Murphy and Daan 1985), or referred to as the true skill score as discussed in Flueck (1987) (Wilks 1995). It is similar to the Heidke skill score but the random forecast that is taken into account is constrained to be unbiased. Similarly, a value of 1 represents a perfect forecast, 0 is random/neutral, and negative values indicate forecasts that are inferior to a random forecast.

c. Time series analysis

In order to assess phase errors, time series analyses were performed on two variables: precipitation and temperature. Here, the main interest was to look at specific storm cycles to see how the models perform in terms

TABLE 2. Indicators for phase and magnitude errors.

| |
|---|
| Precipitation |
| Storm duration (h) |
| Start time of storm cycle |
| Time of max precipitation rate |
| Accumulated precipitation (mm) |
| Max precipitation rate (mm [3 h]^{-1}) |
| Avg precipitation rate (mm h^{-1}) |
| Temperature |
| Time of max temperature |
| Time of min temperature |
| Max temperature ($^{\circ}\text{C}$) |
| Min temperature ($^{\circ}\text{C}$) |

of the timing and amount of precipitation. Precipitation-event timing is the key for dry avalanche forecasting. Temperature was also chosen for this analysis because it is the parameter with the most complete and continuous time series for the field sites studied here, and it is the key to wet avalanche forecasting (see section 1).

First, cross correlation as a function of phase lag was calculated using the statistical package Systat. The inputs are the two time series that one would like to compare. The output gives the correlation values for each phase lag and the standard error. Significantly correlated time series can be identified by comparing the correlation with the standard error of the time series. Two time series are significantly correlated when their correlation exceeds 2 times the standard error. Therefore, two time series have a phase lag when their correlation exceeds 2 times the standard error for any lag not equal to 0. A case where the correlation does not exceed 2 times the standard error for any lag would indicate that the two time series are not significantly correlated. A time lag refers to the time period between two forecasts, which is 3 h for precipitation and 1 h for temperature.

Second, a more descriptive analysis was done for each storm. The time difference between precipitation and temperature peaks as well as their difference in magnitude were compared subjectively. Table 2 contains a list of the different phase and magnitude indicators.

4. Results

a. Precipitation rate

Contingency table analysis was used as a verification method, as outlined in section 3b. First, two categories (precipitation yes/no) were chosen. Second, precipitation rate was divided into seven categories (Table 3), depending on the type of precipitation and the type of observations. The categories intense [$(12.5\text{--}80 \text{ mm h}^{-1}$ or $37.5\text{--}250 \text{ mm 3 h}^{-1})$] and extreme [$(>80 \text{ mm h}^{-1}$ or $>250 \text{ mm 3 h}^{-1})$] did not occur during these winters and are therefore not mentioned any further. Heavy precipitation [$(1.7\text{--}12.5 \text{ mm h}^{-1}$ or $5\text{--}37.5 \text{ mm 3 h}^{-1})$] was forecast and observed only once or twice at each station. For such a small number of events, the correct or in-

TABLE 3. Precipitation rate in categories (water equivalent).

| Category | mm h ⁻¹ | mm (3 h) ⁻¹ | mm (12 h) ⁻¹ |
|------------|--------------------|------------------------|-------------------------|
| None | 0 | 0 | 0 |
| Very light | 0–0.4 | 0–1.25 | 0–5 |
| Light | 0.4–0.8 | 1.25–2.5 | 5–10 |
| Moderate | 0.8–1.7 | 2.5–5 | 10–20 |
| Heavy | 1.7–12.5 | 5–37.5 | 20–150 |
| Intense | 12.5–80 | 37.5–250 | 150–1000 |
| Extreme | >80 | >250 | >1000 |

correct forecast may be coincidence and gives no meaningful information. Therefore, the bias of this precipitation category is not included. The remaining four categories used for the verification of precipitation rate are: none, very light, light, and moderate (Table 3). As mentioned before, only 0–24-h forecasts could be verified for precipitation.

Details of verification results for only the *remote observations* (total precipitation: liquid + solid) from *Kootenay Pass* were published in Roeger et al. (2001), and are briefly summarized here (see Table 4). Both models (24-h forecast) underforecast precipitation events, which means that precipitation was observed more often than it was forecast. The best value for the bias ratio is achieved from the MC2 2-km grid (0.90). The MC2 10-km grid and the NMS 30-km grid achieved 0.71 and 0.75.

The hit rate is close to 0.75 for all forecasts ($H = 0.74$ for the MC2 10 km; $H = 0.72$ for the MC2 2 km; and $H = 0.73$ for the NMS 30 km), which shows that in almost 75% of all cases, precipitation events were forecast as such and nonprecipitation events were forecast as such. Regarding this fairly high hit rate and the bias ratio lower than one for both models, we conclude that most precipitation events that were forecast did indeed occur, but on the other hand, precipitation also occurred that was not forecast.

Both models show some skill, with skill scores (HSS and TSS) of 0.4–0.5 (see Table 4 for details), but could be improved. The 2-km grid shows no improvement for the skill scores compared to the 10-km grid, but the bias

ratio is somewhat better (closer to one) than the MC2 10-km grid. For most cases, the NMS model with the significantly lower resolution produces comparable results to the MC2 model with the higher-resolution grids.

All statistical results from the original MC2 forecasts were improved when the original forecasts were automatically corrected with the Kalman-predictor correction method (see section 2b). Results are included in Table 4. Whereas only minor improvements are shown for the hit rate (from 0.75 to 0.76 for the 10-km grid and from 0.73 to 0.74 for the 2-km grid) and the skill scores (10%–15% improvement), the bias ratio was significantly better using this method. However, the trend for precipitation rate goes in the opposite direction: precipitation events are overforecast with the Kalman-predictor-corrected forecast, whereas they are underforecast by the original forecast at 24 h. For the MC2 10 km, bias ratios are 1.08 (Kalman corrected) versus 0.73 (raw forecast). The MC2 2 km achieved values of 1.10 (Kalman corrected) versus 0.86 (raw forecast) for the bias ratios.

New results from verification with *manual observations* (solid precipitation in mm water equivalent) for *Kootenay Pass* are shown in Fig. 7 and listed in Table 4, again for two categories (precipitation yes/no) and 24-h forecast period. The MC2 10-km grid gives the best results for both of the skill measurements, but not for the bias ratio. The hit rate is fairly high for all models (0.73–0.79), similar to the results with remote observations. Also similar to those results, the bias ratio is highest for the MC2 2-km grid, with a value of 0.95 very close to a perfect forecast. The NMS model has the lowest bias ratio.

When the precipitation rate was divided into more than two categories (see Table 3), the hit rates with manual observations are 0.55, 0.51, and 0.53 for the MC2 10-km, MC2 2-km, and NMS 30-km grids, respectively.

At *Whistler/Blackcomb*, the results of all three models are similar to each other for solid precipitation (mm water equivalent), as given in Figs. 8 and 9 for Pig Alley

TABLE 4. Results from contingency table analysis of precipitation at Kootenay Pass, Nov–Dec 1999, remote and manual observations. NMS, MC2 original, and MC2 Kalman-predictor-corrected forecast. Corr. = corrected.

| | | Precipitation vs nonprecipitation | | | | Precipitation rate in categories |
|--|--------------|-----------------------------------|------|------|------|----------------------------------|
| | | Hit rate | Bias | HSS | TSS | Hit rate |
| Remote observations (liquid and solid precipitation) | | | | | | |
| MC2 10-km grid | Original | 0.75 | 0.73 | 0.47 | 0.46 | 0.56 |
| | Kalman-corr. | 0.76 | 1.08 | 0.51 | 0.52 | 0.56 |
| MC2 2-km grid | Original | 0.73 | 0.86 | 0.43 | 0.42 | 0.56 |
| | Kalman-corr. | 0.74 | 1.10 | 0.48 | 0.49 | 0.53 |
| NMS 30-km grid | Original | 0.73 | 0.75 | 0.43 | 0.44 | 0.56 |
| Manual observations (solid precipitation) | | | | | | |
| MC2 10-km grid | Original | 0.79 | 0.83 | 0.59 | 0.60 | 0.55 |
| MC2 2-km grid | | 0.73 | 0.95 | 0.46 | 0.46 | 0.51 |
| NMS 30-km grid | | 0.73 | 0.71 | 0.45 | 0.54 | 0.53 |

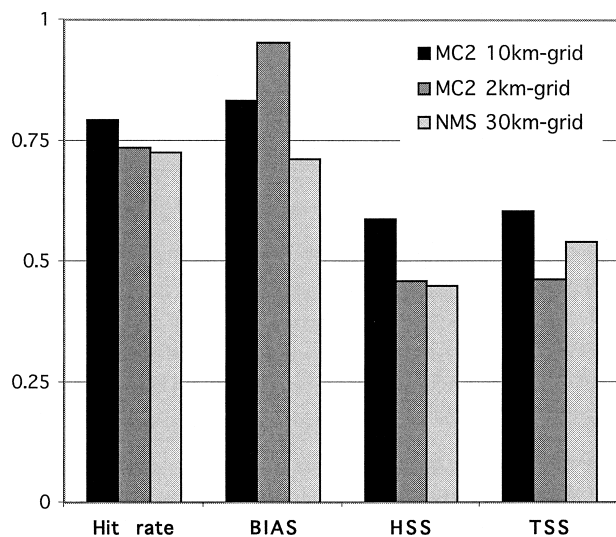


FIG. 7. Verification results for solid precipitation rate at Kootenay Pass, manual observations, Nov 1999–Apr 2000. Results from contingency table analysis. Perfect forecasts have a value of 1.

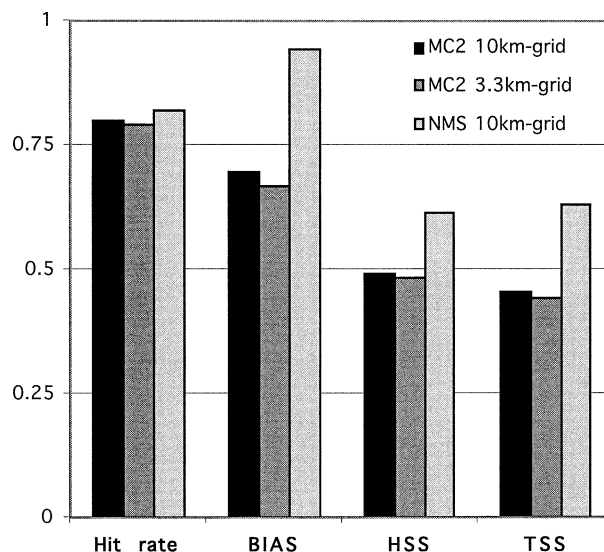


FIG. 9. Same as in Fig. 8 but for Catskiner.

(Whistler Mountain) and Catskiner (Blackcomb Mountain); only precipitation versus nonprecipitation is compared. The NMS model shows better results for the skill score measurements at both stations, most distinct at Catskiner. At Pig Alley, the bias ratio is better with the MC2 models; both MC2 grids have a perfect value of 1. The bias ratio of the NMS model is also very close to 1 (0.95). At Catskiner, the NMS model achieves a similar high value (0.94), whereas the bias ratio of both MC2 grids is much lower.

Between the two MC2 models, no improvement from the 10-km grid to the 3.3-km grid can be seen. The 10-

km grid has better values in all statistics except for the bias ratio at Pig Alley, which has a perfect value of 1 for both MC2 grids (see Figs. 8 and 9).

The false-alarm ratio (viz., when precipitation was forecast but not observed; Table 5) is lowest (best) for the NMS 10-km grid at both mountains. The MC2 grids have a FAR about twice as high as the NMS model. This implies that nonprecipitation events were forecast as precipitation events from the MC2 model, which agrees with the hit rate of these two grids. Because the NMS model only slightly underforecasts precipitation events (bias ratio = 0.95) and the FAR is only 0.12, this model shows no trend toward one event, but underforecasts both categories.

At Catskiner, the values for the MC2 models, together with the hit rate, suggest that most of the nonprecipitation events were forecast as such, but precipitation events were not always predicted. The bias ratio shows the same result. Since the NMS model has a low false-alarm ratio and a bias ratio below one, but close to one, this model predicts precipitation events better than the MC2 model. It does not capture all nonprecipitation events as such, because the hit rate is not very close to 1 (0.82).

Figures 10 and 11 show the bias ratio when the solid precipitation rate is divided into four categories. At Pig Alley, the MC2 model achieves a perfect forecast of 1

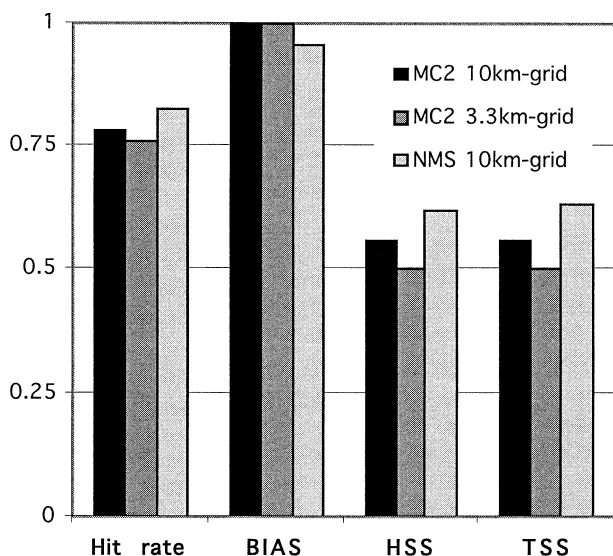


FIG. 8. Verification results for solid precipitation rate at Pig Alley, Nov 1999–Apr 2000. Results from contingency table analysis. Perfect forecasts have a value of 1.

TABLE 5. FAR and H for solid precipitation at Pig Alley and Catskiner.

| | FAR | | H | |
|-----------------|-----------|-----------|-----------|-----------|
| | Pig Alley | Catskiner | Pig Alley | Catskiner |
| MC2 10-km grid | 0.19 | 0.24 | 0.65 | 0.72 |
| MC2 3.3-km grid | 0.21 | 0.23 | 0.62 | 0.71 |
| NMS 10-km grid | 0.12 | 0.12 | 0.62 | 0.66 |

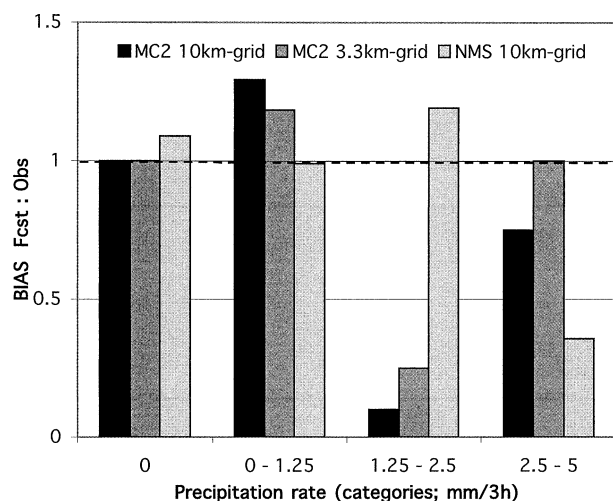


FIG. 10. Bias ratio: Solid precipitation rate in four categories. Pig Alley, Nov 1999–Apr 2000, 24-h forecasts. Dashed line shows a perfect forecast.

for no precipitation, but it performs poorly for light precipitation $[(1.25\text{--}2.5 \text{ mm } 3\text{h})^{-1}]$, which is highly underforecast. Comparing the MC2 10-km grid with the 3.3-km grid shows improvement from the lower to the higher resolution in all categories at Pig Alley. At Catskinner, the MC2 3.3-km grid does a slightly better job than the 10-km grid in certain categories [very light: $0\text{--}1.25 \text{ mm } (3\text{h})^{-1}$ and moderate: $2.5\text{--}5 \text{ mm } (3\text{h})^{-1}$]. The NMS model shows values very close to one for the bias ratio in the categories very light $[(0\text{--}1.25 \text{ mm } (3\text{h})^{-1})]$ and light $[(1.25\text{--}2.5 \text{ mm } (3\text{h})^{-1})]$ at Catskinner, but performance drops off considerably for heavier precipitation rates. At Pig Alley, the NMS model has acceptable results for the first three categories, but performs poorly for moderate precipitation $[(2.5\text{--}5 \text{ mm } (3\text{h})^{-1})]$.

The hit rate is highest for the MC2 10-km grid at

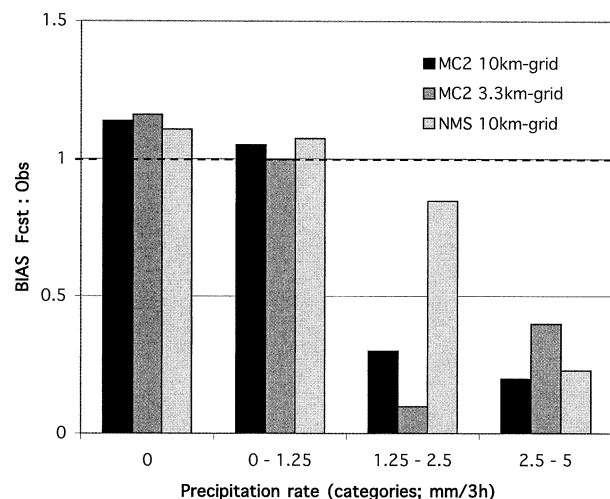


FIG. 11. Same as in Fig. 10 but for Catskinner.

TABLE 6. Wind speed categories (km h^{-1}) according to CAA (1995).

| Category | Wind speed (km h^{-1}) |
|----------|-----------------------------------|
| Calm | 0–1 |
| Light | 1–25 |
| Moderate | 25–40 |
| Strong | 40–60 |
| Extreme | >60 |

both stations. Values are given in Table 5. Pig Alley shows lower values than Catskinner. Generally, $H = 0.72$ and 0.65 is good, considering default (equi-likelihood; no skill) of 0.20 for this statistical measurement.

b. Wind speed

Wind speed was verified with categories according to the Canadian Avalanche Association (CAA 1995), as given in Table 6. Wind speed is generally underpredicted at both study areas. For this variable, results from the NMS model are not as good as from the MC2 model. At *Whistler/Blackcomb*, hit rates are very high with 0.80 (NMS) to 0.88 (MC2) at Whistler Alpine, but much lower at Horstman Hut ($0.33\text{--}0.38$ in 1999/2000 and $0.51\text{--}0.52$ in 1998/99). For both stations, the higher grid resolution (3.3 km) shows no significant improvement compared to the next lower resolution (10 km). Figure 12 shows the wind speed distribution of the 24-h forecasts from the MC2 10-km, MC2 3.3-km, and NMS 10-km grid at Whistler Alpine. Figure 13 shows the wind speed distribution at Horstman Hut from the MC2 model, with both resolutions (3.3-- and 10-km grid) segregated into $0\text{--}24\text{-h}$ and $24\text{--}48\text{-h}$ forecast periods. All models lack realistic variability. Only light and calm winds are predicted. Light winds are highly overforecast, whereas higher wind speeds are not captured at all. Significant differences cannot be seen, either be-

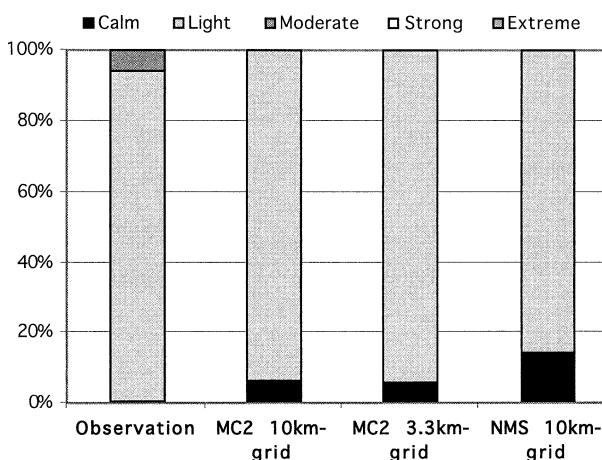


FIG. 12. Wind speed distribution at Whistler Alpine, 24-h forecast, Nov 1999–Jan 2000.

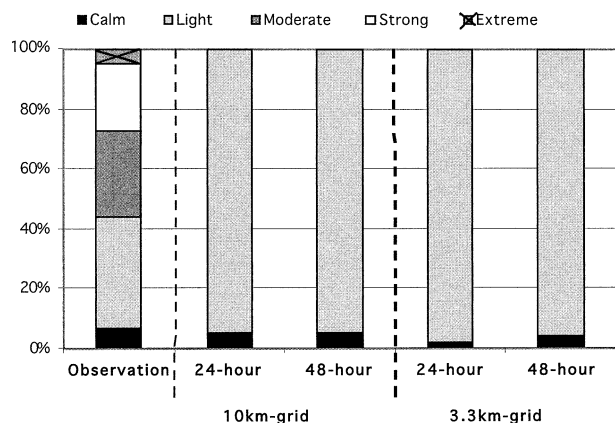


FIG. 13. Wind speed distribution at Horstman Hut, MC2 forecast, Oct 1999–Jan 2000.

tween the two grid resolutions or between the two forecast periods.

Good improvements can be seen with the Kalman-predictor correction method. Figure 14 shows an example of 24-h forecasts at Horstman Hut. Moderate, strong, and extreme wind events are captured as well. The overall improvement of hit rate with this automatic postcorrection method is shown in Table 7. The H values for verification of the Kalman-corrected forecasts range between 0.59 and 0.62, compared to the results from the original forecasts of 0.51–0.52. Table 8 shows error reduction and higher correlation coefficients compared to the original MC2 forecasts (wind speed analyzed as continuous variable). The Pearson correlation coefficient is increased from 0.50 to 0.70 (original MC2 forecasts) to 0.78–0.80 (Kalman-predictor corrected forecasts). Mean absolute errors are reduced from 18.7 to 19.4 km h⁻¹ from the original forecast to 8.0–8.5 km h⁻¹ from the corrected forecast. Even more significant are the improvements of the mean error. Values between 18.1 and 18.7 km h⁻¹ from the original MC2 forecasts are reduced to 2.9 and 3.0 km h⁻¹ with the Kalman-

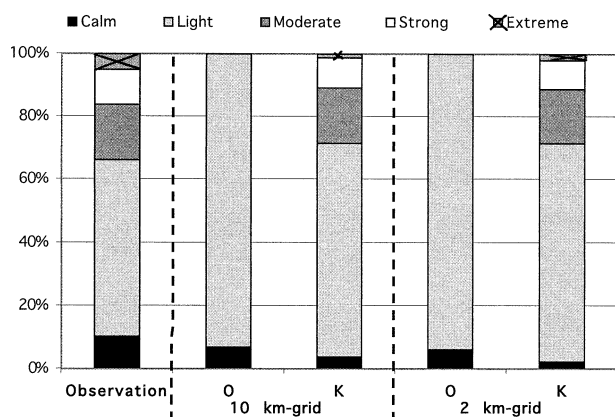


FIG. 14. Wind speed distribution at Horstman Hut, MC2 original (O) vs Kalman-predictor corrected (K) 24-h forecast, Feb–May 1999.

TABLE 7. Wind speed: H for Horstman Hut: Results from 1998/99 and 1999/2000. (Kalman correction was tested only for 1998/99 forecasts here.)

| H | Horstman Hut Feb–May 1999 | | Horstman Hut Oct 1999–Jan 2000 | |
|-----------------|------------------------------|------|-----------------------------------|------|
| | 24 h | 48 h | 24 h | 48 h |
| MC2 10-km grid | | | | |
| Original | 0.52 | 0.52 | 0.36 | 0.38 |
| Kalman-corr. | 0.59 | 0.60 | — | — |
| MC2 3.3-km grid | | | | |
| Original | 0.52 | 0.51 | 0.33 | 0.35 |
| Kalman-corr. | 0.62 | 0.61 | — | — |

predictor correction method. This shows that the Kalman-predictor correction method is of high value for wind speeds.

Comparing the results of the 24-h forecast period with results of the 48-h forecast period show slightly higher correlation coefficients for the 48-h period. No real difference in MAE and ME can be seen.

The results from *Kootenay Pass* are discussed in detail in Roeger et al. (2001). In summary, wind speed is also underforecast at this study site. Figure 15 (Stagleap) shows that the wind speed distribution is similar to Whistler/Blackcomb; namely, there is also a lack of variability, with prediction of only light and calm winds, and overforecasts of light winds. However, the MC2 2-km grid does significantly better than the MC2 10-km grid for this location. The original MC2 forecasts are also highly improved with the Kalman-predictor correction method, as shown in Fig. 16.

The results in Fig. 17 suggest that the topography approximation plays an important role in model performance. The plot gives median values of absolute error (AE; differences between observation and forecast) and their spread (lower and upper quartile). The NMS model performs worst at Stagleap, where the grid spacing is 30 km. At Whistler Alpine, where the NMS model has the 10-km grid, it has comparable results to the MC2 model (lower median absolute error but larger spread). Both MC2 grids have evidently higher median errors at Stagleap than at Whistler Alpine, where the anemometer measurements are not as strongly influenced by local

TABLE 8. Results for wind speed as continuous variable, Pearson correlation coefficient r , MAE and ME in km h⁻¹. MC2 original (O) vs Kalman-predictor-corrected (K) forecast, Horstman Hut, Feb–May 1999.

| | r | | MAE | | ME | |
|-----------------|------|------|------|-----|------|-----|
| | O | K | O | K | O | K |
| 24 h | | | | | | |
| MC2 10-km grid | 0.50 | 0.78 | 19.4 | 8.5 | 18.7 | 3.0 |
| MC2 3.3-km grid | 0.60 | 0.78 | 19.0 | 8.5 | 18.5 | 3.0 |
| 48 h | | | | | | |
| MC2 10-km grid | 0.65 | 0.80 | 19.0 | 8.1 | 18.4 | 2.9 |
| MC2 3.3-km grid | 0.70 | 0.80 | 18.7 | 8.0 | 18.1 | 2.9 |

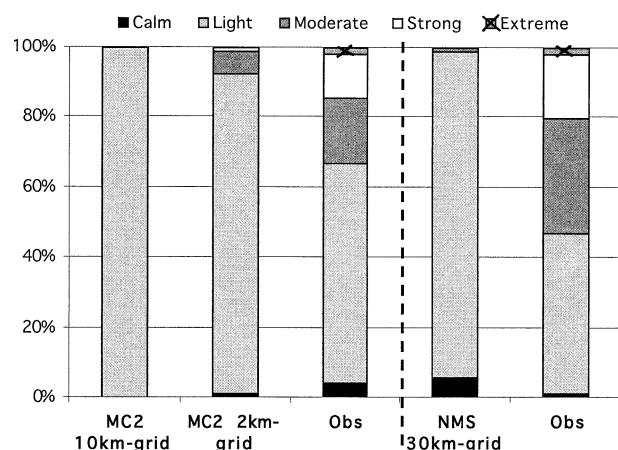


FIG. 15. Wind speed distribution at Stagleap, remote observations, 24-h forecast, Nov 1999–Jan 2000.

terrain, that is, where the topography approximation of the models is not as significant.

c. Wind direction

Wind direction has been verified with contingency table analysis in eight categories (45° angle section: N, NE, E, SE, S, SW, W, NW) or in four categories (90° angle section: N, E, S, W). The different models and grids were compared with wind roses, which represent the prevailing wind as a percentage of time/observations that the wind blows from different directions, as well as the bias ratio for each wind direction and the hit rate.

Figure 18 shows the wind rose for Whistler Alpine. Prevailing winds are from the south. The bias ratio for wind direction divided into the four main aspects as well as the percentage of occurrence is given in Fig. 19 for Whistler Alpine. It can be seen that the 3.3-km grid

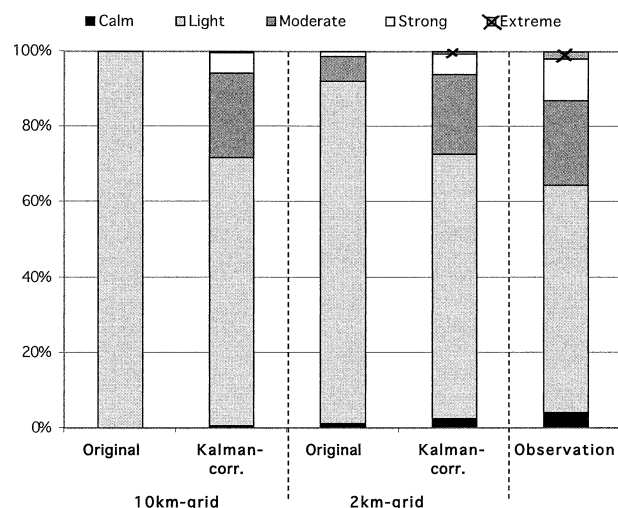


FIG. 16. Wind speed distribution at Stagleap, MC2 original vs Kalman-predictor corrected 24-h forecast, Jan–Apr 2000.

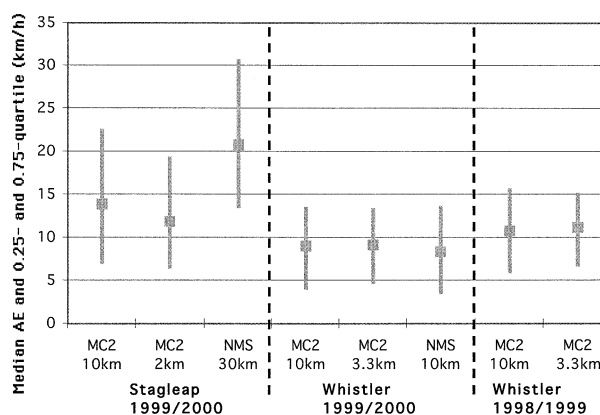


FIG. 17. Median values of absolute differences between observation and forecast and their spread for Stagleap, Jan–Apr 2000, and Whistler Alpine, Nov 1999–Jan 2000 and Feb–Apr 1999.

has better results than the 10-km grid for southerly and westerly winds, which together make 69% of all observations. North winds are badly captured by both model resolutions, but with 2% occurrence this result is not meaningful. However, the 10-km grid is more *accurate* because the overall H values are higher (0.57) than from the 3.3-km grid (0.44; Table 9).

At Horstman Hut, the bias ratios do not show significantly better performance from the 3.3-km grid (Figs. 20 and 21). The H values suggest that the 10-km grid performs better than the 3.3-km grid, because the 10-km grid has a higher hit rate overall (see Table 9). Comparing the 24-h forecast period with the 48-h forecast period (Fig. 20 vs Fig. 21, Table 9) shows subtle differences for all aspects, with the 24-h forecast being better than the 48-h forecast at Horstman Hut (48-h forecasts for Whistler Alpine have not been verified here).

These two figures also show an improvement for this

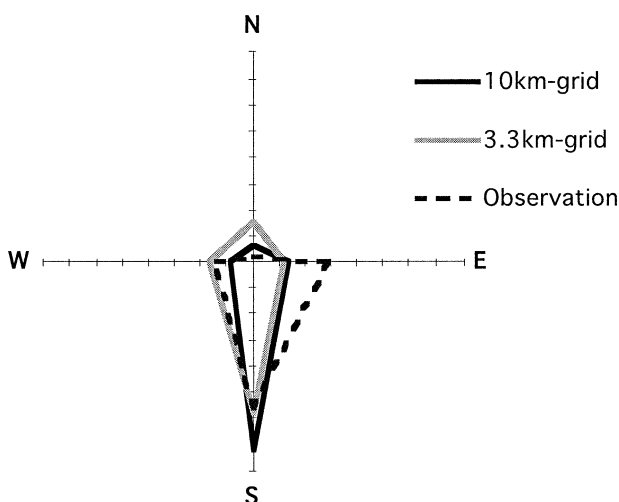


FIG. 18. Wind rose for Whistler Alpine, MC2 24-h forecast, Feb–Apr 1999.

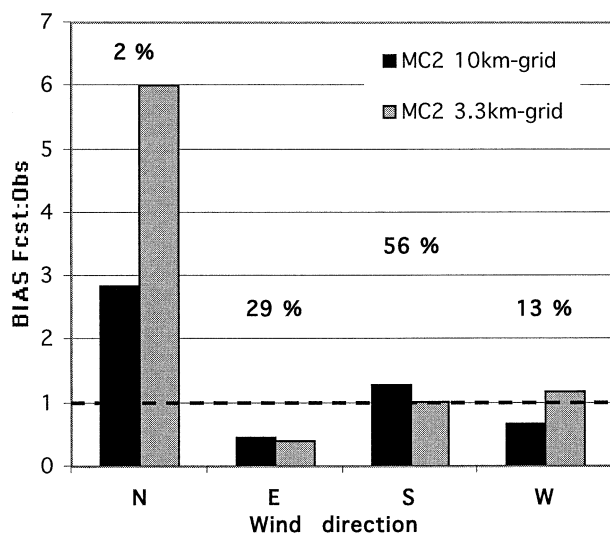


FIG. 19. Bias for wind direction in four categories. Whistler Alpine, MC2 24-h forecast, Feb–Apr 1999. Perfect forecasts have a bias of 1.

variable at Horstman Hut using the Kalman-predictor correction method. Northerly and easterly winds are not improved, but the bias ratio for southerly winds with the highest percentage of occurrence (76%) is better. Similarly, westerly winds are highly overpredicted by the original forecasts, but refined to a large extent with the Kalman prediction. However, westerly winds occur only 4% or 5% of all times at this location. The wind rose for Horstman Hut is given in Fig. 22. Improvement for both grids can be seen for southerly and westerly aspects.

At Stagleap (*Kootenay Pass* study site), prevailing winds are generally from the west (SW: 25%, W: 27%, and NW: 21%), which is mainly due to the general flow pattern (midlatitudes in Northern Hemisphere) but may also be partly influenced by the east–west alignment of the ridge. The wind rose is given in Fig. 23. Figure 24 gives the bias ratio for the four aspects with percentage of occurrence for the MC2 10-km and 2-km grid, and with their equivalent Kalman-predictor correction. Improvement for both grids can be seen for all aspects with the Kalman correction. The H values have also increased: the MC2 10-km grid originally has $H = 0.55$

TABLE 9. Wind direction: Results from contingency table analysis: H for Whistler Alpine and Horstman Hut, Feb–Apr 1999.

| H | Whistler Alpine | Horstman Hut original | Horstman Hut Kalman-corr. |
|-----------------|-----------------|-----------------------|---------------------------|
| 24 h | | | |
| MC2 10-km grid | 0.570 | 0.61 | 0.71 |
| MC2 3.3-km grid | 0.44 | 0.47 | 0.50 |
| 48 h | | | |
| MC2 10-km grid | — | 0.52 | 0.68 |
| MC2 3.3-km grid | — | 0.48 | 0.68 |

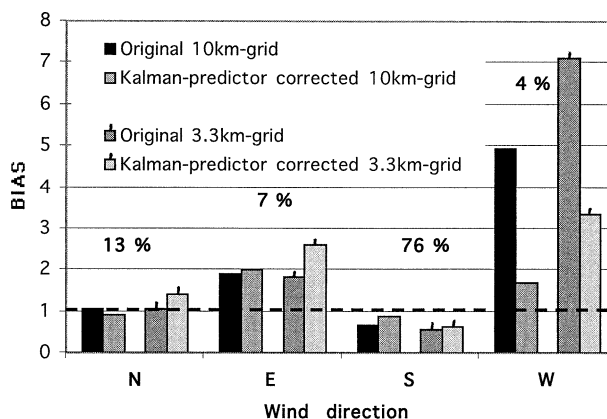


FIG. 20. Bias for wind direction in four categories. Horstman Hut, MC2 original vs Kalman-predictor-corrected 24-h forecast, Feb–Apr 1999. Perfect forecasts have a value of 1.

versus 0.61 with the correction method; the 2-km grid has 0.53 (original) versus 0.57 (corrected). More details are published in Roeger et al. (2001).

d. Temperature

Temperature forecasts (for the specific hour) are generally very good. All models and grids achieve high correlation between forecast and observation values. Predicted temperature is generally too high with mean absolute errors between 1° and 3°C at Kootenay Pass and 2° – 6°C at Whistler/Blackcomb. Figures 25, 26, and 27 show MAE results from Whistler Alpine, Horstman Hut, and Catskiner. The higher-resolution MC2 grid performs better than its lower-resolution grid in all cases. The NMS model has lower errors except for Catskiner (with its remote observations). The temperature MAE results of the 24-h forecast are better than those of the 48-h forecast for all cases.

Correlation coefficients are graphically shown in Figs. 28, 29, and 30. The highest correlation coefficient

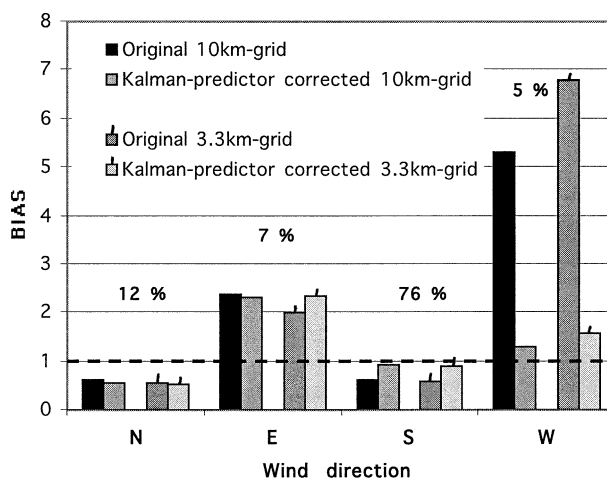


FIG. 21. Same as in Fig. 20 but for 24–48-h forecast period.

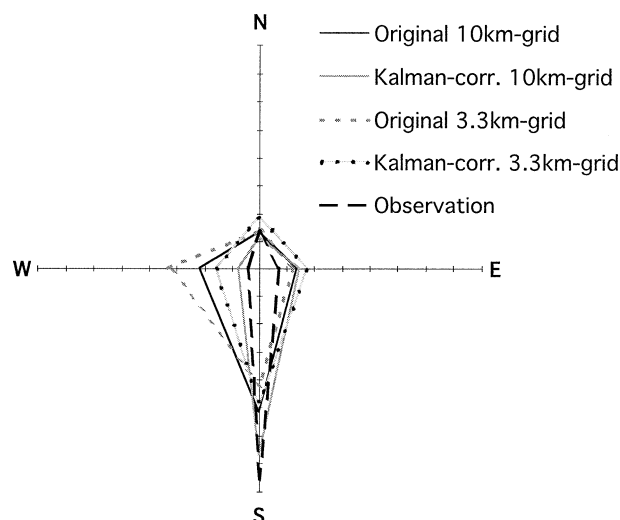


FIG. 22. Wind rose for Horstman Hut, MC2 original vs Kalman-predictor-corrected 24-h forecast. Feb–Apr 1999.

is achieved by the NMS model. It is above 0.8 in almost all cases. The results of the MC2 model are similarly high in some cases, but can be lower than 0.6 in other cases. The 3.3-km grid has better results than the 10-km grid except with observations from 1998/99.

For both years, the MC2 original forecasts at Cat-skiner were automatically corrected with the Kalman prediction. The correlation coefficient is significantly improved, as illustrated in Fig. 31, which shows results from 1999/2000. The mean absolute error is also significantly reduced—in several cases by more than 40%. Results from 1999/2000 are shown in Fig. 32.

Summarized results from *Kootenay Pass* (Roeger et al. 2001) give correlation coefficients and mean absolute errors that are in the same range as those for Whistler/Blackcomb. Figure 33 shows the correlation coefficient for temperature at Kootenay Pass. The NMS model performs not as well as the MC2 model for this study site,

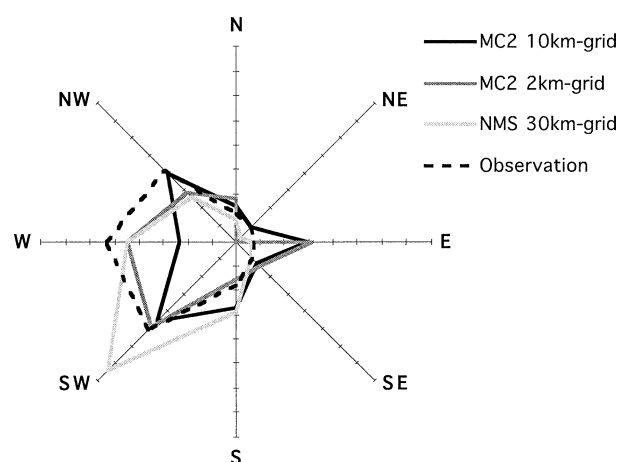


FIG. 23. Wind rose for Stagleap, MC2: Jan 2000; NMS: Nov 1999–Jan 2000.

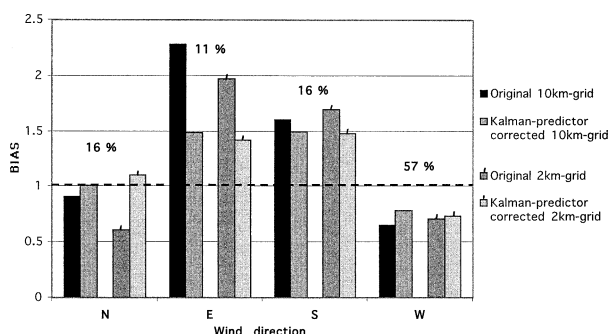


FIG. 24. Bias for wind direction in four categories. Stagleap, MC2 original vs Kalman-predictor corrected 24-h forecast, Jan–Apr 2000. Perfect forecasts have a value of one.

but results are still good. Again, Kalman-predictor corrected MC2 forecasts are significantly better than their original forecasts. Correlation coefficients are increased and achieve values up to 0.97 (see Fig. 34). Mean absolute errors show up to 50% error reduction by using the Kalman corrector.

e. Results from time series analysis

Time series analysis could be done only for *Kootenay Pass* since it is the only station with sufficient precipitation records during any one storm cycle. At Whistler/Blackcomb, manual observations give values only twice a day, which is not enough data for storms that last only 1–2 days. Temperature was chosen as a second variable for time series analyses. These two are the most significant variables for dry and wet avalanche forecasting (as explained in section 1).

Eight storms have been chosen, according to their precipitation patterns. Figure 35 shows the time series

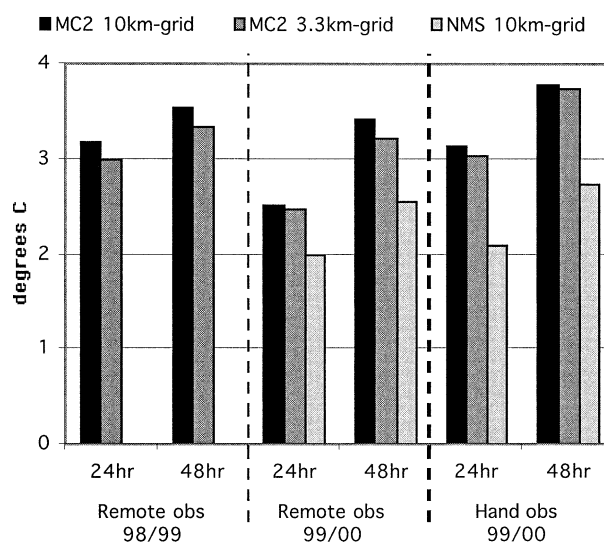


FIG. 25. Mean absolute error for temperature at Whistler Alpine, MC2: Feb–Apr 1999 and Nov 1999–Jan 2000, respectively; NMS: Nov 1999–Mar 2000.

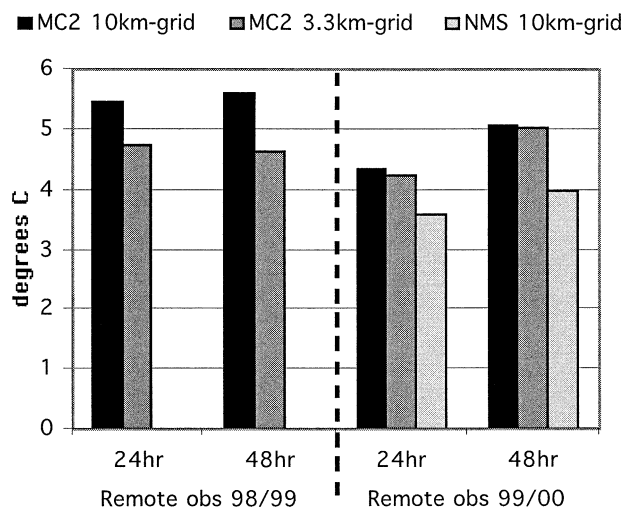


FIG. 26. Mean absolute errors for temperature at Horstman Hut, MC2: Feb–Jun 1999 and Oct 1999–Jan 2000, respectively; NMS: Oct 1999–Mar 2000.

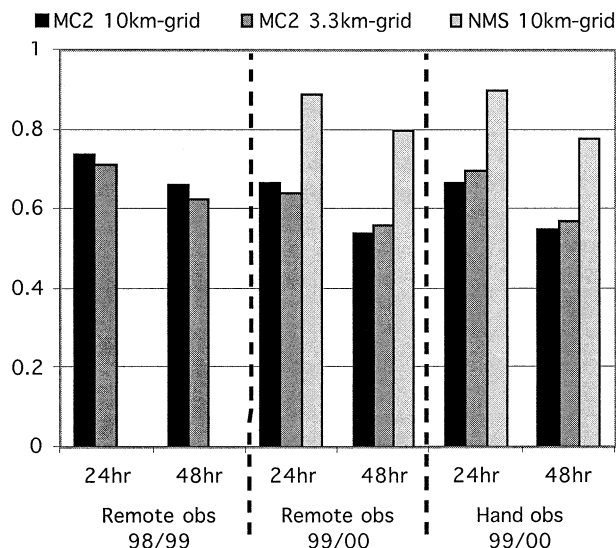


FIG. 28. Correlation coefficient for temperature at Whistler Alpine, MC2: Feb–Apr 1999 and Nov 1999–Jan 2000, respectively; NMS: Nov 1999–Mar 2000.

for these eight storms, which are named accordingly to their start date.

Cross-correlation analysis showed no obvious time lag between any forecast model and any forecast period with the observations for Kootenay Pass for both variables. Almost every correlation that was significant was for 0 time lag. For precipitation, only 6 out of 21 cases were found with significant correlation at non-0 time lags; 5 of them at time lag units of -1 or $+1$, only 1 of them at time lags $+2$ and $+3$ (where each lag unit equals 3 h). For temperature, only 1 case (out of 28) shows significant correlation at nonzero time lag. This

suggests that the timing between forecast and observation is correct for the analyzed eight storms.

The more descriptive time series analysis (not shown here) confirmed that all forecasts underpredict precipitation amount, and most of them underpredict precipitation intensity. The NMS 24-h forecast and the original MC2 10-km grid 24- and 48-h forecasts show extreme values of 41%–50% for the difference in accumulated precipitation. The Kalman-predictor correction method improves precipitation amount for both MC2 forecast grids. The corrected forecasts underpredict accumulated precipitation by 13% and 21%, respectively. For the NMS model, no time trend is obvious, unlike the MC2

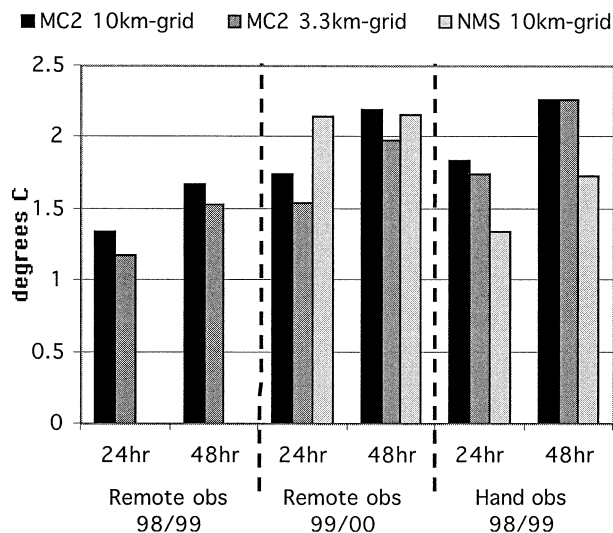


FIG. 27. Mean absolute errors for temperature at Catskinner, MC2: Feb–Apr 1999, Nov–Dec 1999 (remote) and Nov 1999–Jan 2000 (manual), respectively; NMS: Nov 1999–Mar 2000 (remote and manual).

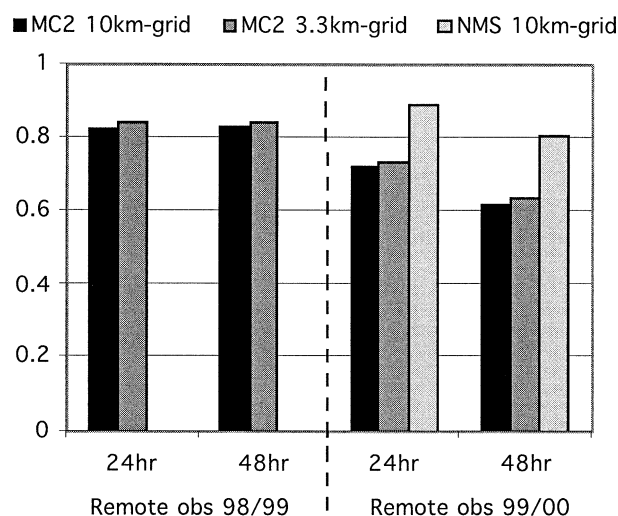


FIG. 29. Correlation coefficients for temperature at Horstman Hut, MC2: Feb–Jun 1999, Oct 1999–Jan 2000, respectively; NMS: Oct 1999–Mar 2000.

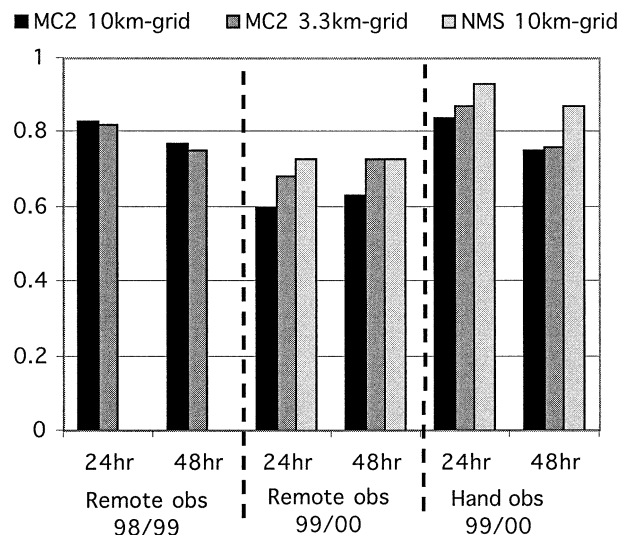


FIG. 30. Correlation coefficients for temperature at Catskinner, MC2: Feb–Apr 1999, Nov–Dec 1999 (remote) and Nov 1999–Jan 2000 (manual), respectively; NMS: Nov 1999–Mar 2000 (remote and manual).

model that starts storms too late, and continues them too long.

The NMS model as well as the original forecast from the MC2 model overforecast temperature magnitude. The Kalman-corrected forecasts reduce this difference but also indicate a reverse trend toward underforecasting. These conclusions are valid for the analyzed eight storms only. To justify those conclusions as a general behavior of the models, more data are needed.

An obvious phase shift in forecasting maximum and minimum temperature cannot be identified. For maximum temperature, all forecasts seem to predict it too early. However, a similar statement cannot be made for

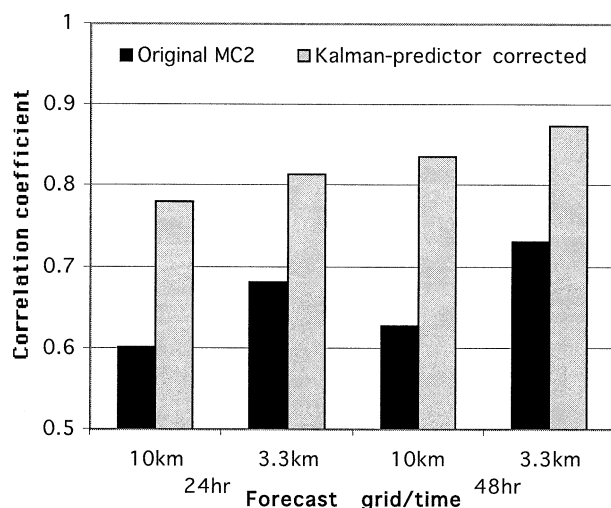


FIG. 31. Correlation coefficient for Kalman-predictor-corrected temperature (°C). Catskinner, Nov–Dec 1999.

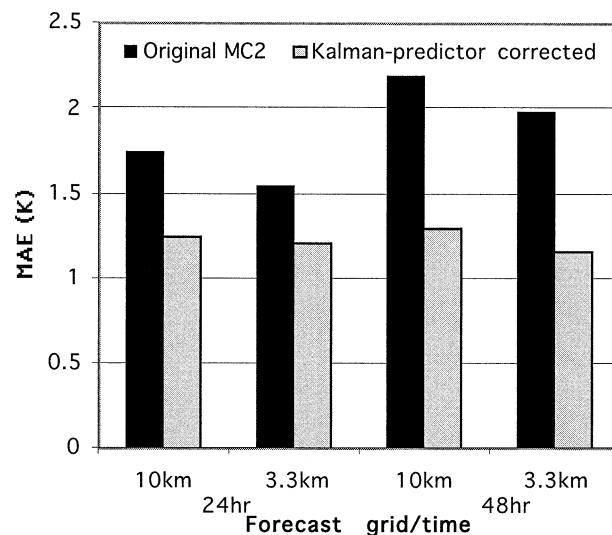


FIG. 32. MAE for Kalman-predictor-corrected temperature (°C). Catskinner, Nov–Dec 1999.

minimum temperature, and therefore no conclusions regarding the timing are possible.

5. Conclusions and outlook

Detailed verification of two high-resolution, real-time, numerical weather prediction (NWP) models was performed with case study observations from two winters: 1998/99 and 1999/2000. The main goal of this research project was to assess the accuracy and the bias of the weather predicted by two models with regard to potential applications such as avalanche forecasting. Verification was against standard meteorological variables from surface observations. Two winters are a rel-

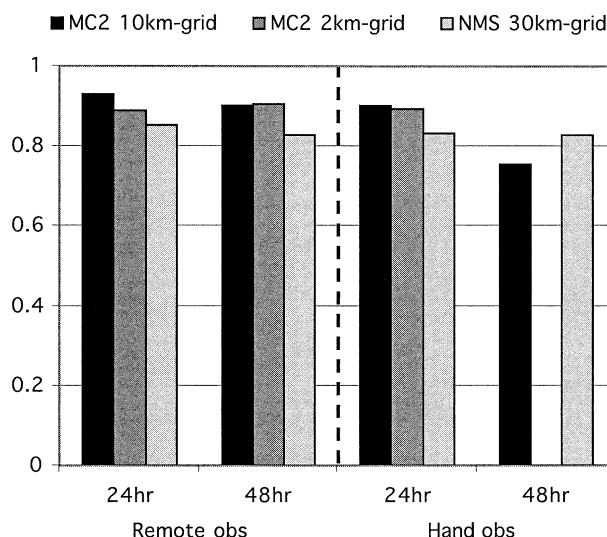


FIG. 33. Correlation coefficient for temperature, Kootenay Pass, Nov 1999–Jan 2000.

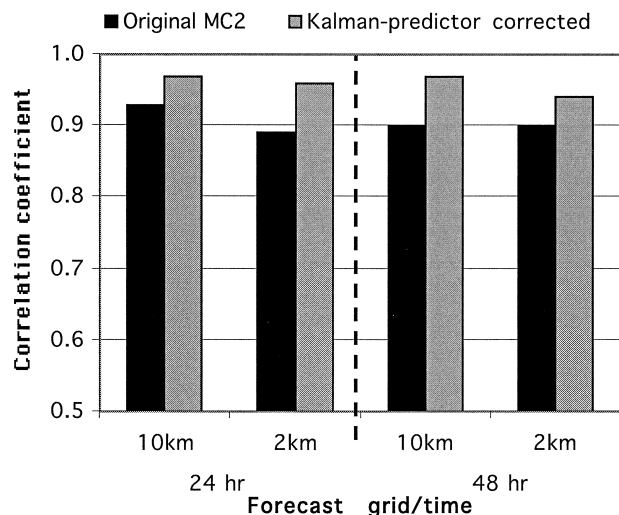


FIG. 34. Correlation coefficient for Kalman-predictor-corrected temperature ($^{\circ}\text{C}$), Kootenay Pass, Nov 1999–Jan 2000.

atively short time period, and it should be kept in mind that the interpretation of the results outlined here can only represent the weather of these two winters.

While this project focused on detailed quantitative verification, some possible explanations for the performance of the two models are suggested here. Comparing MC2 versus NMS, similarly good results were obtained from both. Differences based on grid resolution can be seen between the two study locations. At the Kootenay Pass area, where the NMS has a low grid resolution of 30 km, it does not perform as well as the MC2 model in wind speed (Stagleap) and temperature (Kootenay Pass), whereas the differences are fairly subtle at the latter station. At Whistler/Blackcomb, where the NMS grid resolution is 10 km, its results are at least as good as the results from the MC2 model. For temperature and

partly for precipitation rate, the NMS model performs better than the MC2 model at this study area.

Comparing the two MC2 model resolutions shows somewhat better results for precipitation rate from the finer grid spacing at Kootenay Pass. At Whistler/Blackcomb, the 3.3-km grid has a higher hit rate but the 10-km grid has better bias ratios. Overall, no significant improvement can be seen from the lower to the higher resolution for this parameter. For wind speed at Stagleap and for temperature at Whistler/Blackcomb, the 2-km or 3.3-km grid perform significantly better than the 10-km grid. The results for wind direction show better bias ratios from the 3.3-km grid at Whistler Alpine, but the 10-km grid is more accurate at this location.

The 24-h forecasts are overall more accurate than 48-h forecasts for the events and locations studied here. Results are slightly better for wind speed (correlation coefficients) and wind direction, and significantly better for temperature with the shorter forecast period, as mentioned earlier. No comparison could be done for precipitation (see section 2b).

Time series analysis showed that the timing between forecast and observation is correct for the analyzed eight storms (within the 3-h time resolution of the forecasts). For more general conclusions about the models behaviors regarding correct timing, more storms from several years as well as summer storms should be investigated.

The results also show that the Kalman-predictor correction method is highly suitable for all tested variables. The verification results were improved at all study locations with this automated correction method. This method is a very successful tool in improving the original forecast and should be further developed to use in real time.

In general, precipitation events are underforecast. The results from Kootenay Pass and Catskiner show that

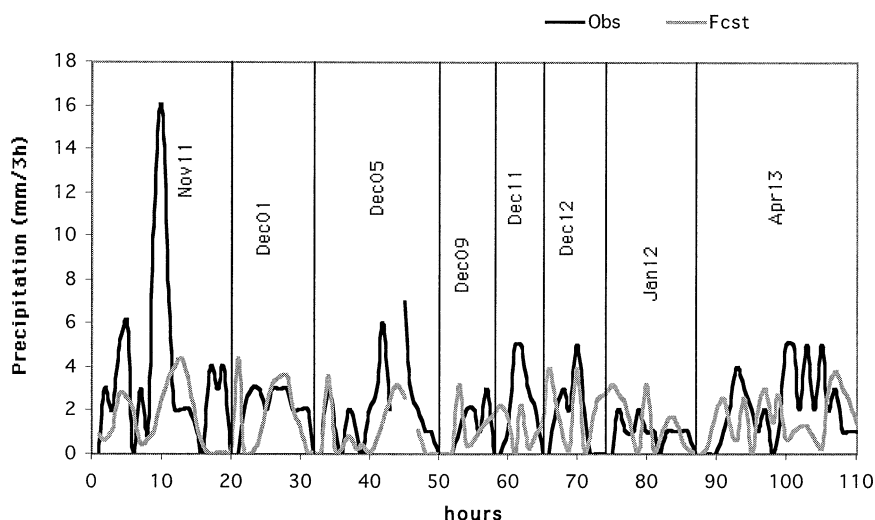


FIG. 35. Three-hour precipitation for the eight chosen storms. Example shows observations and NMS 24-h forecast.

most precipitation events that were forecast also occurred, but on the other hand, additional precipitation events also occurred that were not forecast. This, as well as underpredicted precipitation intensity (results from time series analysis) may be dangerous for the application in avalanche forecasting because this may result in an unexpected increase of avalanche risk. At Pig Alley, the FAR values together with the hit rate suggest that nonprecipitation events were forecast as precipitation events from the MC2 model, which would at least mean that avalanche forecasters are “on the safe side.” The NMS model, with a low false-alarm ratio and a bias ratio below one, predicts precipitation events better than the MC2 model at this location.

The difference in temperature mean absolute errors between Kootenay Pass (1° – 3° C) and Whistler/Blackcomb (2° – 6° C) may be due to an incorrect elevation approximation at Whistler/Blackcomb. The reason is that although the forecast is made for the correct longitude and latitude, the elevation of the model could be off because the model smooths the topography within its grid resolution. This can have a large effect on the temperature field in locations of steep topography. Biases of temperature at Kootenay Pass may also be influenced by poor integration of the model with continental air masses, but this idea has not been further investigated. Temperatures are generally predicted as too warm, but the small MAE values around 0.7° C, achieved with the Kalman-corrected MC2 forecast, suggest that this forecast can be used in further applications, such as avalanche forecasting.

The difference in hit rate of wind speed between Whistler Alpine and Horstman Hut is due to the different distribution of observed wind speeds. At Whistler Alpine, which shows significantly higher hit rates than Horstman Hut, light winds were observed in 94% of all cases, while predictions of the models vary between 86% and 93%. At Horstman Hut, the distribution of the observed wind speeds is quite different: 3% calm, 37% light, 30% moderate, 22% strong, and even 8% extreme. The models, however, predict 94%–97% light winds. Therefore, the models either have a systematic error of not capturing wind speeds greater than light, or they have similar topography approximations for the two stations that both differ from reality.

Underpredicted wind speeds at Stagleap may occur because of the local topography. The weather station is located on top of an east–west aligned ridge and therefore fairly wind-exposed. The topography approximation of the models might not capture this. In addition, a systematic error is possible because the wind speed is also underpredicted at Whistler Alpine.

Another reason why we think that the topography approximation of the models plays an important role are the results shown in Fig. 17. At Whistler Alpine, where the anemometer measurements are not as strongly influenced by local terrain (i.e., where the topography approximation of the models is not as significant), both MC2 grids have lower median errors than at Stagleap.

This implies that the different topography approximations of the two models significantly affect the results. This effect is somewhat larger than the effect of increased grid resolution with the MC2 model. A higher resolution should improve the results because the topography is captured more accurately. However, for British Columbia, improved forecasts for all parameters might not be realized for finer grids (i.e., for better representations of topographic effects), because a limiting factor is the dearth of weather observations upstream (west of) BC. This “data void” over the NE Pacific must be remedied before more accurate forecasts are possible. Boundary effects (boundary value problems due to a closed domain in the numerical models) might be another reason for the bias. A third factor might be the numerical approximations made by the MC2 developers to improve execution speed.

It was shown that each model has different strengths and weaknesses. Neither one of the models is best for all variables. This indicates that, in general, a single model should not be used for all variables. An ensemble forecast that combines several models may do a better job than only one, when all parameters are considered.

While this verification project focused on basic steps in model verification, many more meteorological features are yet to be verified with measurements. For example, for snow avalanching it is of high importance to have information about the extent, timing, and magnitude of temperature inversions and cold frontal passages, both of which have a significant effect on snowpack stability. Therefore not only temperature but also temperature change is very important and it is one of the significant variables of numerical avalanche models.

Although not shown in this paper, Roeger et al. started to take the next step by using this numerical forecast output as input to a statistical avalanche threat model. So far, the resulting 24-h forecasts of avalanche threat seem to be as skillful as traditional 6–12-h avalanche forecasts based only on weather observations.

Thus, we recommend that NWP forecasts be used to increase the lead time for avalanche forecasts. By increasing the advanced warning, avalanche and resource managers can take mitigation action to better protect lives and property, and reduce avalanche closures of key transportation corridors.

Acknowledgments. This research was sponsored by Canadian Mountain Holidays, the Natural Sciences and Engineering Research Council of Canada (NSERC), Forest Renewal BC, Environment Canada, the Canadian Foundation for Climate and Atmospheric Sciences, and BC Hydro. Claudia Roeger was supported by the German Academic Exchange Service (DAAD). The data for this study are provided by the Ministry of Transportation and Highways (MoTH) of British Columbia, and the ski resort Intrust Whistler/Blackcomb. Research coordination, infrastructure, and computer support was provided by the Geophysical Disaster Com-

putational Fluid Dynamics Centre at UBC. We would like to thank John Tweedy and Ted Weick from MoTH, as well as the avalanche forecasters from Whistler/Blackcomb for their great help. We are extremely grateful for the support from all these organizations.

APPENDIX A

Equations for Statistical Analysis

Pearson correlation coefficient r :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (\text{A1})$$

where x_i are forecast data values, y_i are observed data values, \bar{x} : mean forecast value, \bar{y} : mean observed value, and n : number of data pairs.

Mean error (ME):

$$\text{ME} = \bar{x} - \bar{y}, \quad (\text{A2})$$

mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{k=1}^n |x_k - y_k|, \quad (\text{A3})$$

mean-square error (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{k=1}^n (x_k - y_k)^2, \quad (\text{A4})$$

root-mean-square error (RMSE):

$$\text{RMSE} = \sqrt{\text{MSE}}. \quad (\text{A5})$$

Contingency table analysis equations for the 2×2 table of Table A1.

| | Range | Perfect forecast | |
|---|----------------|------------------|-------|
| Hit rate H : | | | |
| $H = \frac{A + D}{N}$ | 0 to 1 | 1 | (A6) |
| Probability of detection (POD): | | | |
| $\text{POD} = \frac{A}{A + C}$ | 0 to 1 | 1 | (A7) |
| False-alarm rate (FAR): | | | |
| $\text{FAR} = \frac{B}{A + B}$ | 0 to 1 | 0 | (A8) |
| BIAS: | | | |
| $\text{BIAS} = \frac{A + B}{A + C}$ | 0 to $+\infty$ | 1 | (A9) |
| Heidke skill score (HSS): | | | |
| $\text{HSS} = \frac{2(AD - BC)}{(A + C)(C + D) + (A + B)(B + D)}$ | -1 to +1 | 1 | (A10) |
| True skill score (TSS): | | | |
| $\text{TSS} = \frac{AD - BC}{(A + C)(B + D)} = \frac{A}{A + C} - \frac{B}{B + D}$ | -1 to +1 | 1 | (A11) |

TABLE A1. Contingency table definition, where A to D are the counts of events in each category, out of N total events.

| | | Observation | |
|----------|-----|-------------|-----|
| | | Yes | No |
| Forecast | Yes | A | B |
| | No | C | D |

APPENDIX B

Kalman Filter Basics

Kalman filtering is used as an adaptive, recursive method (Bozic 1979) to optimally estimate the bias and reduce rms error between raw, noisy NWP forecasts and noisy verification observations. It is recursive because it carries only a filtered summary of the past input signals, into which it can incorporate new inputs to create a modified filter. It is adaptive in the sense that any changes in the stationarity of the input signals is quickly incorporated into the modified filter, causing information about the old filter to be gradually lost with succeeding time steps. These attributes are desirable because the filter adapts to changing climate, changing seasons, or even changing NWP model versions without requiring one to first accumulate a large database of historical data.

Kalman (1960) and Kalman and Bucy (1961) showed how this approach can also be used as a statistical *predictor* to estimate future forecast bias. This is particularly useful for real-time weather forecasting, where the raw NWP projection can be corrected with the Kalman-estimated bias projection to create a corrected forecast. We used this objective, linear approach during postprocessing of each model forecast, in place of traditional model-output statistics (MOS). This operation is fully automated with no manual tweaking or bogusing.

Let e_k be the bias between the forecast and the verifying observation valid today (for time step k), such as for temperature [$e_k = T_k(\text{fcst}) - T_k(\text{obs})$] at one weather station location. This e_k is the signal that we would like to predict (i.e., estimate) for tomorrow (at $k + 1$). Kalman designed his filter/predictor for a first-order autoregressive system of the form $e_{k+1} = ae_k + w_k$, where w_k is a Gaussian-distributed random term of variance σ_w^2 . The meteorological interpretation of this “signal model” or “system model” is that a portion (a) of the future bias of the weather forecast is successfully described by persistence of the current bias, but with the addition of a random term that is related to the fundamental deterioration of weather predictability with increasing lead time. This system model applies not only to the actual system, but to our estimate of the system.

Similarly, the input observations are assumed to be noisy (with random error v_k) that can be described by: $e_k = c\hat{e}_k + v_k$, where the error variance is σ_v^2 , factor c indicates the relationship between the filtered expected value and the actual observation, and the hat indicates

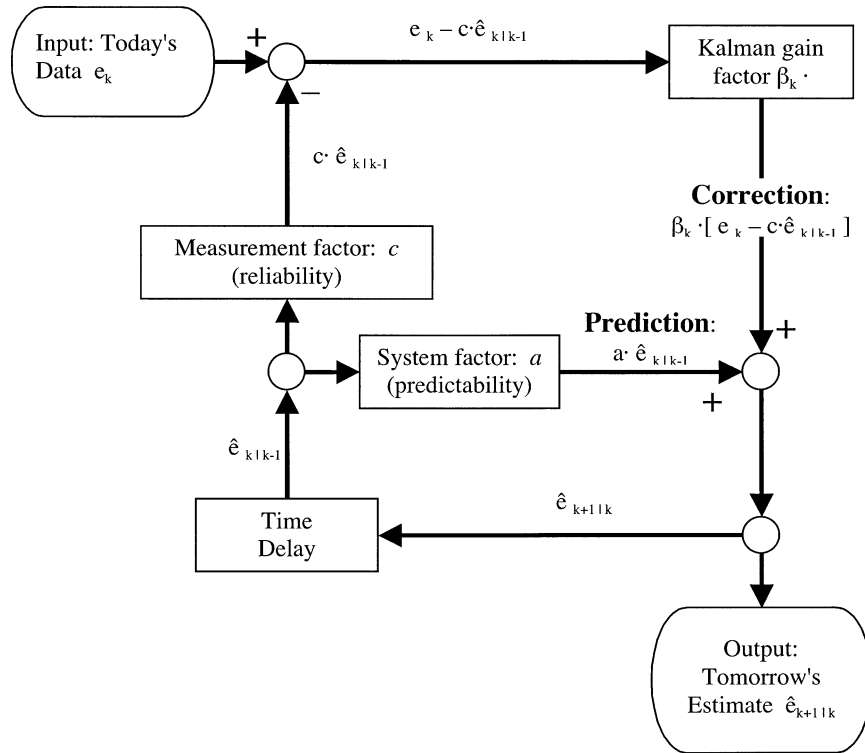


FIG. B1. Flow diagram for the Kalman predictor.

expected value. Meteorologically, the random error can be due to subgrid terrain influences, spurious numerical artifacts, inadequacies of the physical parameterizations, and errors in the observations themselves.

Flow diagram Fig. B1 illustrates that the Kalman approach is basically an optimal predictor-corrector method. The *prediction* of tomorrow's bias uses the bias from today, which is assumed to persist with the loss of skill associated with predictability. The difference between today's observed bias and the reliable (nonrandom) portion of today's bias that was estimated yesterday, when weighted by a factor β called the Kalman gain so as to project to tomorrow, gives the *correction* that was "learned" from previous errors. This correction is added to the prediction, to give the final estimate of the bias for tomorrow. We use this bias estimate to adjust our raw numerical forecasts. Also, this bias estimate is saved for one day (i.e., the time delay operator), to be recycled into the Kalman algorithm to estimate the bias for the subsequent day. This cycle repeats every day, as counted by index k .

Combining the previous equations yields the resulting predictor equation (Bozic 1979):

$$\hat{e}_{k+1|k} = a\hat{e}_{k|k-1} + \beta_k[e_k - c \cdot \hat{e}_{k|k-1}], \quad (\text{B1})$$

where the Kalman gain β is found from:

$$\beta_k = acp_{k|k-1}[c^2p_{k|k-1} + \sigma_w^2]^{-1}, \quad (\text{B2})$$

and where p is the prediction mean-square-error from:

$$p_{k+1|k} = a^2p_{k|k-1} - ac\beta_k p_{k|k-1} + \sigma_w^2. \quad (\text{B3})$$

Subscripts such as $k|k-1$ indicate the value for today (index k) as estimated from yesterday's ($k-1$) value. The parameters a and c are found from the covariance matrices of bias, and the whole system is started on the first day using a 0 initial-bias estimate. Within the first several days of operation of this postprocessing system, (B1) approaches the best estimate of forecast bias. This approach is used separately for each weather station in this study; namely, different parameters a , c , β , and p can evolve for the different stations.

REFERENCES

- Armstrong, R. L., and B. R. Armstrong, 1987: Snow and avalanche climates of the western United States: A comparison of maritime, intermountain and continental conditions. *Avalanche Formation, Movement and Effects: Proceedings of a Symposium Held at Davos*, B. Salm and H. Gubler, Eds., International Association of Hydrological Sciences (IAHS) Publ. 162, 686 pp. [Available from IAHS Press, Centre for Ecology and Hydrology, Wallingford, Oxfordshire OX10 8BB, United Kingdom.]
- Benoit, R., M. Desgagne, P. Pellerin, S. Pellerin, and Y. Chartier, 1997: The Canadian MC2: A semi-Lagrangian, semi-implicit wideband atmospheric model suited for finescale process studies and simulation. *Mon. Wea. Rev.*, **125**, 2383–2415.
- , and Coauthors, 2002: The real-time ultrafinescale forecast support during the special observing period of the MAP. *Bull. Amer. Meteor. Soc.*, **83**, 85–109.
- Bozic, S. M., 1979: *Digital and Kalman Filtering*. John Wiley & Sons, 153 pp.
- CAA, 1995: *Observation Guidelines and Recording Standards for*

- Weather, Snowpack and Avalanches*. Canadian Avalanche Association (CAA), 99 pp. [Available from The Canadian Avalanche Centre, P.O. Box 2759, Revelstoke, BC, Canada, V0E 2S0.]
- Flueck, J. A., 1987: A study of some measure of forecast verification. *10th Conf. on Probability and Statistics in Atmospheric Sciences*. Edmonton, AB, Canada, Amer. Meteor. Soc., 64–68. [Available from American Meteorological Society, 45 Beacon Street, Boston, MA 02108-3693.]
- Foehn, P. M. B., 1998: An overview of avalanche forecasting models and methods. Norwegian Geotechnical Institute Publ. 203, 256 pp. [Available from the Norwegian Geotechnical Institute, Sognsveien 72, 0806 Oslo, Norway.]
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.*, **82**, 35–45.
- , and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *Trans. ASME, J. Basic Eng.*, **83**, 95–108.
- Kristovich, D. A. R., and Coauthors, 2000: The Lake—Induced Convection Experiment and the Snowband Dynamics Project. *Bull. Amer. Meteor. Soc.*, **81**, 519–542.
- LaChapelle, E. R., 1980: The fundamental processes in conventional avalanche forecasting. *J. Glaciol.*, **26**, 75–84.
- McClung, D. M., 1995: Computer assistance in avalanche forecasting. *Proc. Int. Snow Science Workshop*, Snowbird, Salt Lake City, UT, American Avalanche Institute, 310–313. [Available from Dave McClung, Dept. of Geography, UBC, 1984 West Mall, Vancouver, BC V6T 1Z2, Canada.]
- , 2000: Predictions in avalanche forecasting. *Ann. Glaciol.*, **31**, 377–381.
- , and P. Schaerer, 1993: Avalanche prediction II: Avalanche forecasting. *The Avalanche Handbook*. The Mountaineers, 272 pp.
- Murphy, A. H., and H. Daan, 1985: Forecast evaluation. *Probability, Statistics, and Decision Making in the Atmospheric Sciences*, A. H. Murphy, and R. W. Katz, Eds., Westview Press, Inc., 379–437.
- Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science*, **10**, 453–454.
- Roeger, C., D. McClung, R. Stull, J. Hacker, and H. Modzolewski, 2001: A verification of numerical weather forecasts for avalanche prediction. *Cold Reg. Sci. Technol.*, **33**, 189–205.
- Tripoli, G. J., 1992: A nonhydrostatic mesoscale model designed to stimulate scale interaction. *Mon. Wea. Rev.*, **120**, 1324–1359.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. International Geophysical Series, Vol. 59, Academic Press, 468 pp.