ELSEVIER

Cold Regions Science and Technology 33 (2001) 189–205

cold regions science and technology

www.elsevier.com/locate/coldregions

A verification of numerical weather forecasts for avalanche prediction

Claudia Roeger [a,b,*], David McClung [a,1], Roland Stull [b,2], Joshua Hacker [b,2], Henryk Modzelewski [b,2]

[a] Avalanche Research Group, Department of Geography, University of British Columbia, 1984 West Mall, Vancouver, BC, Canada, V6T 1Z2

[b] Weather Forecast Research Team, Department of Earth and Ocean Sciences University of British Columbia, Vancouver, BC, Canada

Received 30 September 2000; accepted 23 July 2001

Abstract

Past, current, and future meteorological conditions are key parameters for snowpack instability and hence, the risk of avalanches. For regional, computer-assisted avalanche forecasts, highly accurate weather predictions are needed. The objective of this research is to help weather and avalanche forecasters in their decision-making process based on meteorological predictions by (1) quantifying some of the many uncertainties of meteorological forecasts and (2) combining numerical weather and avalanche prediction. Case studies were used to verify and quantify output from high-resolution numerical weather prediction (NWP) models as input for avalanche forecasting models. At the University of British Columbia, two high-resolution, real-time, numerical weather forecast models are currently run every day. Their output of the fine grid spacing of 3.3 km for the Whistler/Blackcomb ski area in the British Columbia Coast Mountains, and 2 km for Kootenay Pass in the Columbia Mountains, are verified here. The forecasts are compared with surface observations of manual and automatic weather stations using standard statistical methods. The results are very good, especially regarding the mountainous terrain: all forecasted surface parameters are within range to their observed value. Precipitation rate has results in the same order of magnitude, which is very good for this variable that is very difficult to forecast. Here, the MC2 2-km grid has a better bias ratio than the MC2 10-km grid. In general, the NMS model produces comparable results even though the resolution is lower. For temperature, an error reduction as much as 50% was achieved using the post-processing Kalman-predictor correction method. With such small errors (around 0.7 K), it looks quite promising that the forecast can be used for avalanche forecast models such as at Kootenay Pass where air temperature is a primary variable for wet avalanche prediction.

Keywords: Numerical weather prediction; High-resolution weather models; Weather forecast verification; Avalanche forecasting; Applied meteorology; Mountain weather in British Columbia, Canada

* Corresponding author. Graf-Eberstein Str 19, 76199 Karlsruhe, Germany. Fax: +1-604-822-6088.

E-mail addresses: croeger@eos.ubc.ca (C. Roeger), mcclung@geog.ubc.ca (D. McClung), rstull@eos.ubc.ca (R. Stull), jhacker@eos.ubc.ca (J. Hacker), hmodzelewski@eos.ubc.ca (H. Modzelewski).

[1] Fax: +1-604-822-6150.

[2] Fax: +1-604-822-6088.

PII: S0165-232X(01)00059-3

1. Introduction

Snow avalanche prediction is a complex problem, which is almost exclusively solved by experience. Avalanche forecasters developed good skills in evaluating current conditions and making short-time forecasts given the expected weather conditions. This type of prediction is called conventional avalanche forecasting, which is the most widely used method and it is regarded as the most successful. Numerical avalanche prediction refers to organization of a database of previously measured parameters, including avalanche occurrences, for use with a computer to help compare current conditions with past ones. Primary emphasis is on meteorological data (McClung and Schaerer, 1993). Computer-aided avalanche forecasting has the advantage to handle large organized data sets and to help people with limited field experience in their decision making process (even though the computer models developed up to this point should not be used by people without field experience, McClung, 1995).

The character and quality of data used in forecasting avalanches is determined by the scale of the forecasting problem. Due to the great variety of climate zones in Canada, the demand for avalanche prediction is at the meso scale, which implies more accurate weather prediction than for synoptic scale forecasts, which relies strongly on snow-stability information with less reliance on meteorological data.

The combination of weather forecast and avalanche prediction requires the combination of two different scales. Avalanches initiate in a snow cover that can change its layering and structure within hours or days across distances in the order of hundreds of meters. The solution of the mathematical equations expressing the physical laws for the atmosphere is too costly and time consuming for the local scale, and therefore extrapolations and simplifications have to be made. This results in a serious limitation on the accuracy of avalanche forecasting (Foehn, 1998). Any model is dominated by the interaction of weather with the physical processes in the snow cover, which lead to avalanche formation. Therefore, detailed networks of meteorological and snow pack measurements combined with avalanche observations are necessary.

The comparison of output variables from numerical weather prediction (NWP) models and input variables for avalanche forecasting models shows that a lot of the NWP variables can be directly applied into an avalanche forecasting model 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 instabilities and hence, avalanches.

McClung and Tweedy (1994) developed a numerical avalanche forecasting model, which is used operationally at Kootenay Pass for predictions 12 h into the future with current observations. It might be possible to predict avalanches up to 24 hours into the future, if meteorological forecast data are available and sufficiently accurate. Therefore, forecasts from two numerical weather research models were analyzed with respect to measurements at Kootenay Pass to assess the forecasts for eastern BC, and a similar study for Whistler/Blackcomb in western BC is in progress.

Snow avalanche forecasters use numerical weather and avalanche models to make daily decisions with important economic consequences. Avalanche forecasting is only one of many applications of weather predictions that require high accuracy. Uncertainties are found by evaluating forecast output against observed data.

Forecast evaluation can be described as “the process and practice of determining the quality and value of forecasts” (Murphy and Daan, 1985). Two types of forecast evaluation with different goals can be distinguished. Inferential (or empirical) evaluation, and decision-theoretic (or operational) evaluation. In meteorological literature, the former is called forecast verification (as discussed in this paper) and it is concerned with the quality of forecasts, whereas the latter is concerned with the value of the forecasts. Even though they are considered as two different topics, the two fields obviously overlap.

Decision-theoretic evaluation is important to relate the value of forecasts to users. Work in this area has been concerned with the development of measures of the monetary value of forecasts. For avalanche forecasting, the value of the forecast highly depends on the quality of the forecasts.

Within the field of forecast verification, it is possible—and necessary—to identify more specific purposes. Here, the purposes include (a) the determination of the state-of-the-art of weather forecasting at The University of British Columbia (UBC), (b) the comparison of different forecast models with variations in grid spacing and (c) the combination of weather forecasts and avalanche predictions to produce longer range (>1 day) avalanche hazard forecasts.

Weather forecast verification provides information about the quality of the forecast. There are many different ways to measure the quality of a forecast. In this context, not only accuracy but also skill is important. Accuracy is defined as "the ability of a measurement to match the actual value of the quantity being measured" (AHD, 2000) or as "the extent to which a given measurement agrees with the standard value for that measurement" (RHW, 1997). For prediction, one can say that accuracy is *the ability of a forecast to match the observation* and *the degree to which a forecast agrees with the measurement*.

But even a highly accurate forecast is not necessarily a "good" forecast. For example, using average weather conditions of a particular region (climatology) to make a forecast may give accurate results, but requires no skill. Therefore, a forecast is of good quality when it shows skill by being more accurate than climatology (Stull, 2000). The definition of human skill is "the proficiency, facility, or dexterity that is acquired or developed through training or experience" (AHD, 2000), and mathematically, "the degree of correctness of a quantity, expression, etc." (RHW, 1997). Thus, by determining the accuracy and skill of a forecast, one can improve it and use it with more confidence in the future. For this project, the accuracy of both models at UBC has been determined with statistical methods for continuous and categorical variables (see Section 3). For the MC2 model, forecasts from two different grid-spacings for the same forecast time have been analyzed. Section 2 describes both models and their grids.

Numerical weather forecasts depend significantly on the initial conditions and the topography estimation and hence, the resolution of the model grid. To estimate this dependence, weather models are run with slightly different initial conditions and with different grid resolutions for the same forecast period. This yields ensemble forecasts (Stull, 2000). For this project, output from different models with different grid resolutions has been used to estimate the improvement using a higher-resolution grid and to illustrate the effect of different topography approximations from each model. This is very important for mountainous regions such as British Columbia.

This paper contains the weather forecast verification and the methods used. First results from Kootenay Pass are presented and ideas for future work are mentioned.

2. Data

Data from two different sites are used. The ski area Whistler/Blackcomb in the Coast Mountains in British Columbia (50.05°N, 122.9°W) represents a maritime mountain climate. Kootenay Pass (49.05°N, 117.0°W) in the southern Selkirk Mountains of BC represents a transitional climate zone (Armstrong and Armstrong, 1987), mid-way between a maritime and a continental climate (McClung and Schaerer, 1993). In addition to two different climate zones, these sites represent two different types of operations (ski area and highway operation) affected by avalanches. Fig. 1 shows a map with the two sites indicated.

2.1. Meteorological observations

Hourly data from Kootenay Pass were collected automatically at two weather stations operated by the BC Ministry of Transportation and Highways for the winter 1999/2000. In addition, morning and afternoon manual observations from the station at the summit of the pass were used. The meteorological variables used are listed in Table 1.

The observation site at the summit of Kootenay Pass is located at 1780 m a.s.l. elevation in an open area surrounded by trees. It is fairly sheltered and therefore wind observations here might be biased. Precipitation measurements are representative for the

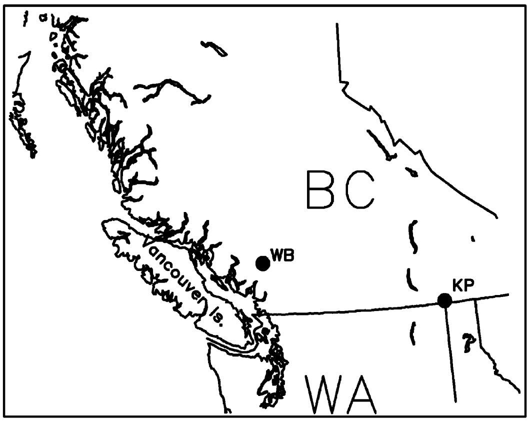

Fig. 1. Map of southwestern British Columbia (BC, Canada) and northern Washington (WA, USA) indicating the site locations. WB: Whistler/Blackcomb, KP: Kootenay Pass.

area and temperatures are typical for this elevation. Temperature is measured at shelter height above the ground or snow surface, respectively.

Stagleap is a remote weather station at the top of a ridge (2140 m a.s.l.) and well exposed to the wind. Wind speeds are therefore typical for this mountain ridge and elevation, but are not representative for some avalanche starting zones at mid mountain elevation, especially on the lee side. At this station, winds are measured remotely (anemometer) atop a 10-m high tower.

Observation data from Whistler/Blackcomb were from automatic weather stations as well as manual observations taken by ski patrol avalanche forecasters. The few results gathered so far are not shown in this paper and therefore we have not included a detailed description of this data set.

The data from both areas include standard weather and snowpack measurements as well as avalanche observations, gathered according to the guidelines from the Canadian Avalanche Association (CAA, 1995).

2.2. Meteorological forecasts

At UBC, two numerical weather prediction models are run real-time, making daily forecasts on multiple grids out to 48 h into the future. The two models are: (1) the Meso-scale Compressible Community (MC2), refined by Recherche en Prevision Numerique (RPN) in Canada, and (2) the University of Wisconsin Non-hydrostatic Modeling System (UW-NMS).

The MC2 model (Benoit et al., 1997) is run with grid-point spacing of 90, 30, 10, 3.3 and 2 km, where the finer grids in small domains are nested inside coarser, larger-domain grids. The highest resolutions (smallest grid-spacing) have been used for verification. These are 3.3 and 10 km over Whistler/Blackcomb, and 2 and 10 km over Kootenay Pass. 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 Greg Tripoli at the University of Wisconsin (Tripoli, 1992), and is run at UBC for nests with 90, 30, and 10 km grid spacing. For verification, 10 km is used for Whistler/Blackcomb and 30 km for Kootenay Pass. The number of grid points is $50 \times 68 \times 24$ for the 10 km grid-point spacing and $68 \times 80 \times 28$ for the 30 km spacing. Vertical domain is also nested, where the coarsest horizontal mesh has 32 vertical layers. For each weather station, forecast values from the surrounding four or nine grid points have been interpolated to calculate the forecast for the exact station location.

Initial and boundary conditions for MC2 and UW-NMS coarse grids are from Eta model forecasts (US National Centers for Environmental Prediction). In turn, forecasts from MC2 and NMS coarse meshes provide the boundary conditions for the imbedded finer meshes.

In NWP models, shelter height (2 m) temperatures, and anemometer height (10 m) winds are

Table 1
Parameters used from each station at Kootenay Pass and their type of observation

Weather station	Parameters
Kootenay Pass (1780 m a.s.l.)	Temperature (R)
	Precipitation (R, M)
	Wind speed (M)
	Wind direction (M)
Stagleap (2140 m a.s.l.)	Temperature (R)
	Wind speed (R)
	Wind direction (R)

M: manual; R: remote.

diagnosed by nonlinear interpolation between the lowest model level and the surface. The lowest model layer has a variable height above ground. A land-surface parameterization scheme determines true surface (skin) temperatures, while the surface wind speed at 0 m is necessarily 0. The variation of temperature and wind between the surface and the lowest model layer is based on momentum and heat transfer properties of the land-atmosphere interface, and values at any level can be determined.

Precipitation in a NWP model can be at the grid scale (resolved), or below the grid scale. Resolved scale precipitation is handled by explicit calculation of cloud microphysics. Sub-grid scale precipitation is calculated by a "convective" parameterization scheme that takes into account moisture and instability in a particular grid cell. Precipitation values from each process are added together to get the forecasted precipitation for a particular grid cell. In this case, it is assumed that all precipitation will be resolved by grid spacing less than 10 km, and the convective parameterization is not accounted for. One reason why precipitation forecasting with NWP models is challenging is because precipitation values are sensitive to adjustable parameters in both cloud microphysics and convective parameterization schemes. Another reason is that temporal and spatial scales are often small relative to model initialization data, even with forecasts of grid spacing of 2 km.

The MC2 original forecasts were also compared with forecasts that have been improved by the Kalman-predictor correction method (Bozic, 1979). The Kalman-predictor correction is a post-processing method 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. It can be used for every forecast when observation data are available simultaneously. A list of variables whose MC2 forecasts were corrected with the Kalman-predictor correction method and checked against observation data is given in Table 2.

Table 2
List of Kalman-predictor corrected MC2 forecasts used for verification at Kootenay Pass 1999/2000

Weather station	Parameters	Grid spacing	
		2 km	10 km
Kootenay Pass (1780 m a.s.l.)	Precipitation	24-h	24-h
	Temperature	24-h	24-h
		48-h	48-h
Stagleap (2140 m a.s.l.)	Wind speed	24-h	24-h
	Wind direction	24-h	24-h

For verification, the forecasts were divided into two forecast time periods. The first was from 0 to 24 h forecasts. The second time period covers forecasts for 24–48 h into the future.

3. Evaluation methods

Standard statistical methods as well as graphical techniques were used first. Emphasis was on robust and resistant mathematical techniques. Robustness and resistance are insensitive 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 follow a Gaussian distribution). A resistant method is not strongly influenced by a small number of outliers.

Mathematical techniques give information about the variation/spread of the data set (interquartile range IQR), smallest and largest values, and a single representative number for the data set (median; 0.5-quantile, $q_{0.5}$). Descriptive statistical parameters (mean, M; standard deviation, s; variance, υ) were also calculated, but they may be neither robust nor resistant.

The correlation coefficient (Pearson product-moment; Eq. (A.1) in Appendix A) gives information about a relationship between two data sets. It is a good measure of the strength of a linear relationship. Basic absolute measures for ordinal predictands are the mean error (ME; Eq. (A.2)), the mean absolute error (MAE; Eq. (A.3)), the mean square error (MSE; Eq. (A.4)) and the root mean square error (RMSE; Eq. (A.5)).

For nominal predictands, measurements of accuracy are best represented by contingency tables. Contingency tables display categorical verification data as an $I \times J$ table. They contain absolute frequencies of the $I \times J$ possible combinations of forecast and event pairs. An example for $I = J = 2$ is given in Fig. 2. The total number of events, N, is equal to the

		Observation	
		Yes	No
Forecast	Yes	A	B
	No	C	D

Fig. 2. Contingency table for $I = J = 2$.

sum $A + B + C + D$, which are the components a 2×2 contingency table. Common measurements of accuracy include the hit rate (H), the percentage of forecasts correct (PFC), the threat score (TS), the probability of detection (POD), the false-alarm rate (FAR), and the bias. These quantities are given as (Eqs. (A.6)–(A.6.10) in Appendix A.

The hit rate (or the percentage of forecast correct) is the most direct and intuitive measure of the accuracy of categorical forecasts. Simply, it is the ratio of correct forecast events $(A + D)$ to the total number of events N. The worst possible hit rate is zero. A value of one would represent a perfect forecast. The bias 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 $B = 1$ indicates that the event was forecast correctly the same number of times that it was observed. Bias greater than one indicates that the event was forecast more often than it was observed (over-forecasting). Conversely, bias less than one indicates under-forecasting. 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).

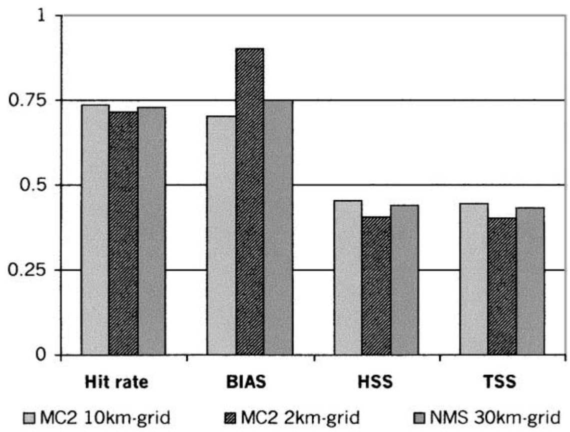

Fig. 3. Precipitation rate at Kootenay Pass: results from contingency table analysis. Perfect forecasts have a value of one.

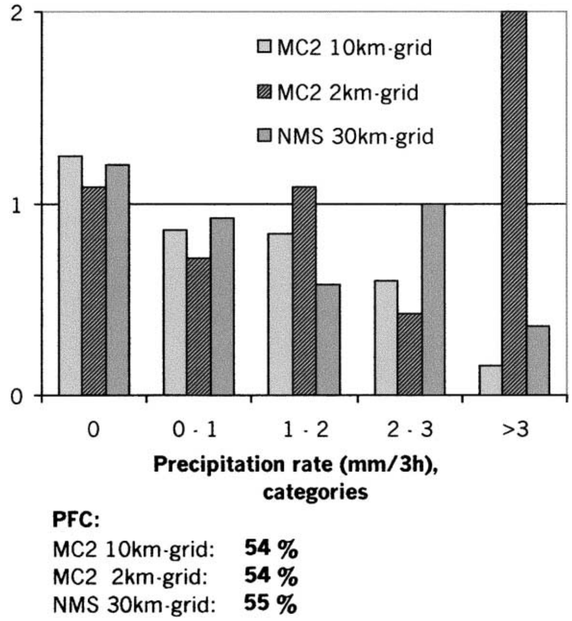

Fig. 4. BIAS: precipitation rate; MC2 24-h forecast, remote observations. Kootenay Pass, Nov.–Dec. 1999. A value of one represents a perfect forecast.

Skill scores are relative measures of data attributes. Eqs. (A.11) and (A.12) 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 hit rate as the basic accuracy measure, but it also takes the random nature of forecasts into account. The hit rate expected for random forecasts is taken as the reference accuracy measure. Forecasts equivalent to the reference forecasts receive zero scores. Negative

Table 3
Wind speed categories (km/h) according to CAA (1995)

Category	Wind speed (km/h)
Calm	0–1
Light	1–25
Moderate	25–40
Strong	40–60
Extreme	> 60

scores are given to forecasts that are worse than the reference forecasts. Perfect forecasts receive a Heidke score of one (Wilks, 1995).

The parameter TSS is a measure of true forecast skill. In short, the true skill score is the POD (probability of detection), adjusted by the POFD (probability of false detection). It is derived from Flueck, 1987. 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 one represents a perfect forecast, zero is random/neutral, and TSS less than one are inferior to a random forecast.

4. Results

4.1. Precipitation rate

Figs. 3 and 4 show results from contingency table analysis from Kootenay Pass data. First, the distinction between precipitation and non-precipitation events was made. Fig. 3 shows the results from the MC2 model 10 and 2 km grid-point spacing as well as the 30 km grid from the NMS model. A perfect forecast has a value of one in all categories. The Hit rate is close to 0.75 for all forecasts, which shows that in almost 75% of all cases, precipitation events were forecast as such and non-precipitation events were forecast as such.

Table 3 gives verification results from contingency table analysis. The bias in this case is the ratio of the number of events when precipitation was forecast to the number of events it was observed. It shows that all models under-forecast precipitation events, which means that precipitation was observed more often than it was forecasted. The best value here is achieved from the MC2 2 km-grid (0.90).

Both skill scores, HSS and TSS, are about 0.4–0.5. This means that both models show skill (greater than zero) but could be improved. The 2 km grid does not improve the results for the skill scores, but the bias is somewhat better (closer to one) than the MC2 10-km grid. It can be seen that the NMS model with the significantly lower resolution produces comparable results to the MC2 model with the higher resolution grids.

Table 4
Wind speed: percentage of forecast correct for kootenay pass and stagleap, 24-h forecast

PFC (%)	Kootenay Pass, Nov. 1999–Jan. 2000	Stagleap, Jan.–Apr. 2000
MC2 10-km grid	46	63
MC2 2-km grid	45	59
NMS 30-km grid	50	41

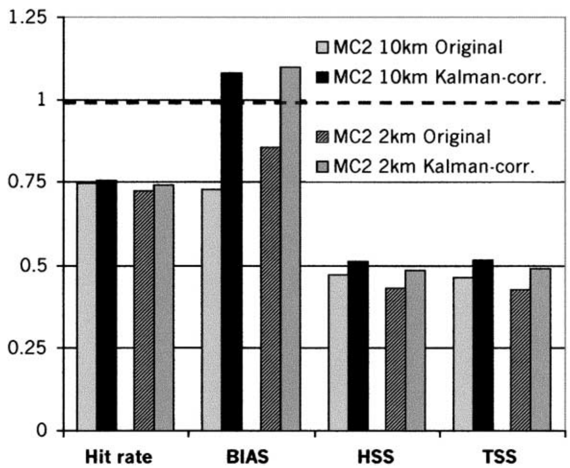

Fig. 5. Verification statistics for precipitation rate at Kootenay Pass. Results from contingency table analysis. MC2 Original vs. Kalman-predictor corrected 24-h forecast. Kootenay Pass, Nov.–Dec. 1999. Perfect forecasts have a value of one.

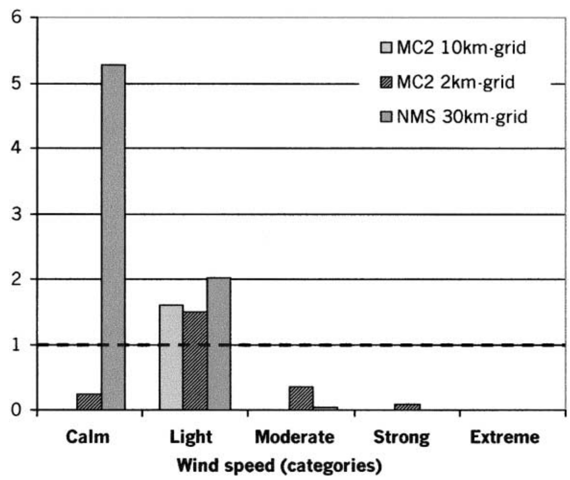

Fig. 6. Bias: wind speed at Stagleap, 24-h forecast, Jan.–Apr. 2000. A value of one represents a perfect forecast.

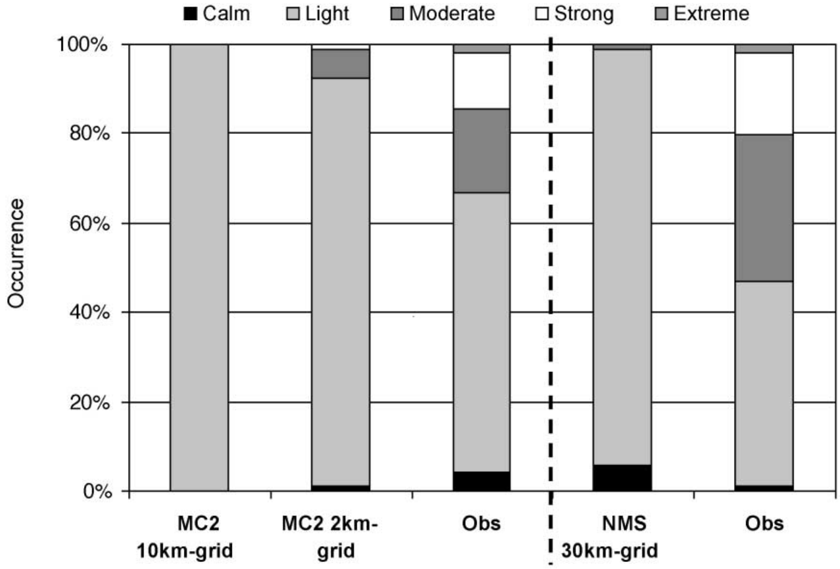

Fig. 7. Wind speed distribution at Stagleap, 24-h forecast, Nov. 1999–Jan. 2000.

Fig. 4 shows the bias ratio when the precipitation rate was distinguished into five categories. Again, results from the MC2 10-km and 2-km grid and the NMS 30-km grid are presented. It shows that the MC2 10-km grid over-forecasts non-precipitation events, and increasingly under-forecasts events with increasing precipitation rate. The 2-km grid significantly over-forecasts the category with the highest precipitation rate. However, this category represents only a small portion of all precipitation events.

Again, the NMS model demonstrates comparable accuracy to the MC2 model even though its resolution is much lower. In two categories (0–1 and 2–3 mm/3 h), its results are better than both MC2 grids.

Table 5
Results for wind speed as continuous variable, Original vs. Kalman-corrected MC2 forecasts

	Original			Kalman-corrected		
	r	MAE	ME	*r*	MAE	ME
MC2 10-km grid	0.56	16.6	15.5	0.65	8.2	2.3
MC2 2-km grid	0.46	14.5	11.8	0.64	8.5	2.0
NMS 30-km grid	0.44	22.6	22.4	–	–	–

Pearson correlation coefficient (r), MAE and ME in km/h. Stagleap, Jan.–Apr. 2000.

The scale of the categories must be considered as well. The categories are fine: The range of all five categories together covers only up to 3 mm/h, which

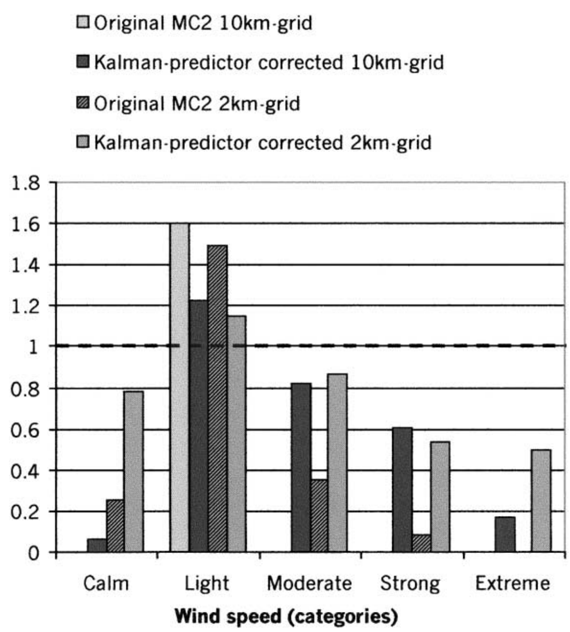

Fig. 8. Wind speed: bias. MC2 Original vs. Kalman-predictor corrected forecast. Stagleap, Jan.–Apr. 2000. A value of one is best.

is considered as *light* rainfall. The reason for choosing these small categories is the distribution of precipitation occurrence at both study sites and to ensure more accurate verification for precipitation rate. The Percentage of Forecast Correct (PFC; hit rate expressed in %) is 54% for the MC2 model and 55% for the NMS model. This is fairly good considering that a random forecast would have 20%. However, this should be improved because it is unsatisfying for daily applications if the precipitation amount can be trusted only a bit more than 50%. Also, the distribution of the total number of events in each category must be considered. More than 90% of all events belong to the first three categories, where the bias is fairly close to one. So, the models do quite well, even though the category with the highest precipitation rate is substantially under-forecast (MC2 10-km grid and NMS 30-km grid) or over-forecast (2-km grid).

At Kootenay Pass, the MC2 original precipitation forecasts were post-corrected with the Kalman-predictor correction method and also compared with observation data. Results of both, the 10-km and 2-km grid are shown in Fig. 5. For both grid spacings, the hit rate is slightly improved. The bias is significantly improved, but the trend for precipitation amount goes in the opposite direction: Precipitation events are over-forecast with the Kalman-predictor corrected forecast, whereas they are under-forecast by the original forecast. Both, the Heidke skill score and the true skill statistic are slightly improved with the Kalman-correction method.

4.2. Wind speed

Wind speed has been analyzed with methods for continuous as well as categorical variables. The categories are given in Table 3. Wind speed is observed manually at the weather plot at Kootenay Pass, and remotely at Stagleap. The verification results with remote data are better than with hand observations as shown in Table 4, which gives values for Percentage of Forecast Correct (PFC). These differences are probably due to biased wind speeds at the sheltered study plot at Kootenay Pass, and due to different topography approximations of the two different models.

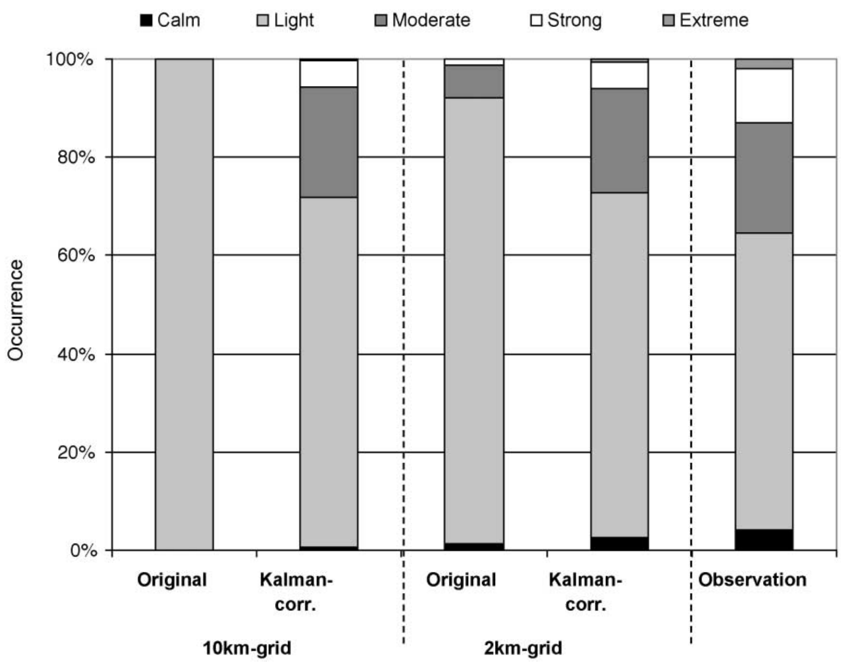

Fig. 9. Wind speed distribution at Stagleap. MC2 Original vs. Kalman-predictor corrected 24-h forecast. Jan.–Apr. 2000.

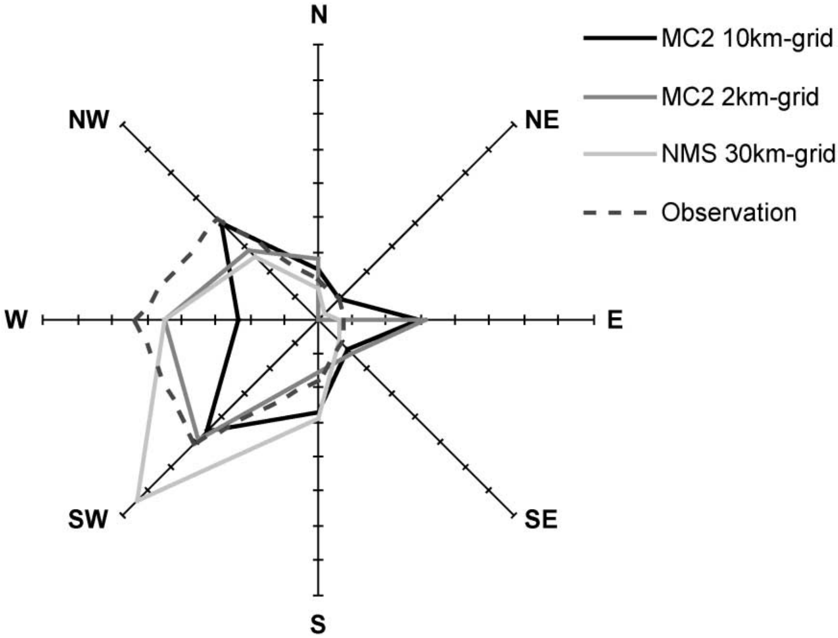

Fig. 10. Wind rose for Stagleap, MC2: Jan. 2000; NMS: Nov. 1999–Jan. 2000. Winds from SW, W, NW are prevailing.

Fig. 6 shows the bias ratio for Stagleap. Categories that are not shown have neither been observed nor forecasted. The MC2 2-km grid does better than the 10-km grid-point spacing. Indeed, the MC2 10 km-grid forecasts *light* winds only, whereas the 2-km grid forecasts 61% *light*, 19% *moderate*, and 13% *strong*. It is obvious that these categories are not as much under-forecast as with the 10-km grid. This lack of variability from both resolutions can be seen in Fig. 7. The NMS model with the 30-km grid-point spacing lacks variability as well. It over-forecasts *calm* winds by more than 400%. It hardly forecasts *moderate* events, which are observed about 33% of all times in that range. The NMS model does not forecast *strong* and *extreme* events at all within the analyzed time period.

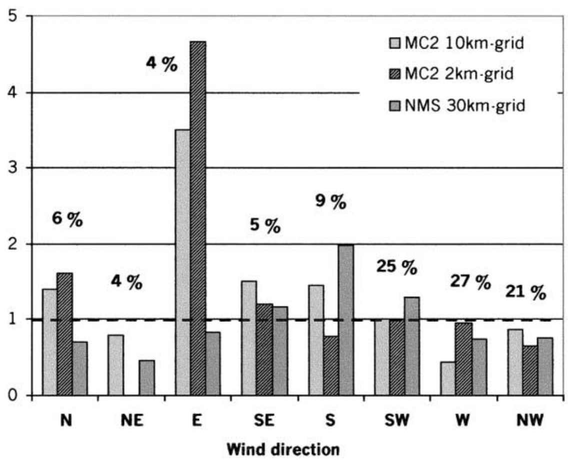

Fig. 11. Bias for wind direction in eight categories. Stagleap, MC2: Jan. 2000, NMS: Nov. 1999–Jan. 2000. A value of one represents a perfect forecast.

Mean absolute errors (MAE), mean errors (ME) and correlation coefficients r are given in Table 5. The positive mean error (observation–forecast) shows that all models under-forecast wind speed. Mean absolute errors are somewhat high. No general trend

Table 6
Wind direction: percentage of forecast correct (aspect observed was same than forecast), Kootenay Pass, Nov. 1999–Apr. 2000, and Stagleap, Jan. 2000 (MC2) and Nov. 1999–Jan. 2000 (NMS)

PFC (%)	Kootenay Pass	Stagleap
MC2 10-km grid	35	44
MC2 2-km grid	30	47
NMS 30-km grid	33	50

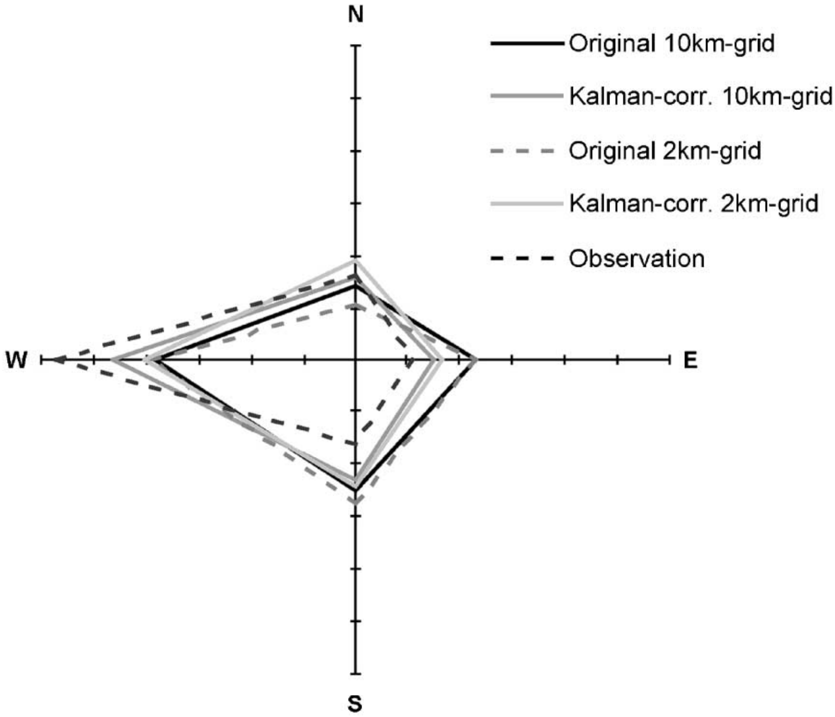

Fig. 12. Wind rose for Stagleap, MC2 Original vs. Kalman-predictor corrected 24-h forecast. Jan.–Apr. 2000.

for one model can be seen. The correlation coefficient is highest with the MC2 10-km grid, but still fairly low with a value of 0.56.

The Kalman-predictor correction method shows significant improvement for wind speed at Stagleap. The bias of the original MC2 vs. Kalman-predictor

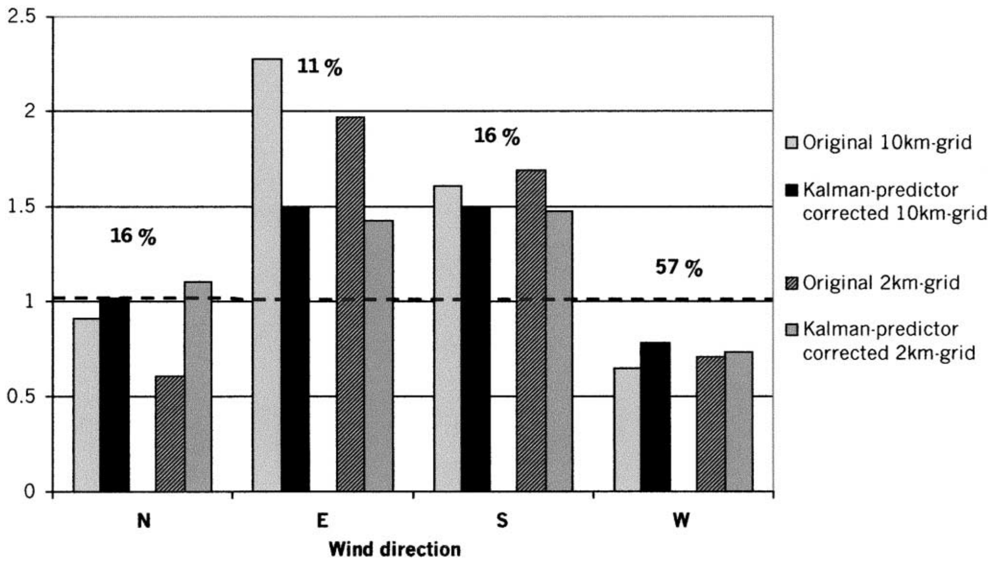

Fig. 13. Bias for wind direction at Stagleap. MC2 Original vs. Kalman-predictor corrected 24-h forecast, Jan. 2000. Perfect forecasts have a value of one.

Table 7
Wind direction: percentage of forecast correct for stagleap, Jan.–Apr. 2000, MC2 Original vs. Kalman-predictor corrected 24-h forecast

PFC (%)	Original	Kalman-predictor corrected
MC2 10-km grid	55	61
MC2 2-km grid	53	57

corrected forecast is shown in Fig. 8. Results from both grids are highly improved: the bias is reduced in all categories. The Kalman-predictor corrected 10-km grid forecast covers all wind speed categories, whereas the original predicted only *light* winds. Fig. 9 shows the improved distribution with the Kalman-predictor correction method compared to the original forecasts and the observed distribution on the very right. Values are given in Table 5.

4.3. Wind direction

Wind direction has been analyzed with contingency table analysis in eight and four categories, 45° or 90° angle section, respectively. The different models and grids were compared with wind roses, which represent the prevailing wind as percentage of time/observations the wind blows from different directions, as well as the bias for each wind direction and the Percentage of Forecast Correct (PFC).

Wind direction was observed remotely at Stagleap. The wind rose is given in Fig. 10. Prevailing winds are from western directions (SW: 25%, W: 27% and NW: 21%). This pattern is partly influenced by the east–west alignment of the ridge but is mainly due to the general flow pattern (mid-latitudes in Northern Hemisphere). The MC2 2-km grid predicts southwestern winds well, but under-forecasts western and northwestern winds. The NMS model over-forecasts winds from the southwest and also under-forecasts western and northwestern winds.

The bias ratio for wind direction at Stagleap of all three different resolutions is shown in Fig. 11. The NMS model with the 30-km grid has values close to one in almost all categories. Westerly and northerly aspects are under-forecast by all three models. Again, the NMS model has quite good results compared to the MC2 model with higher-resolution grids. Both

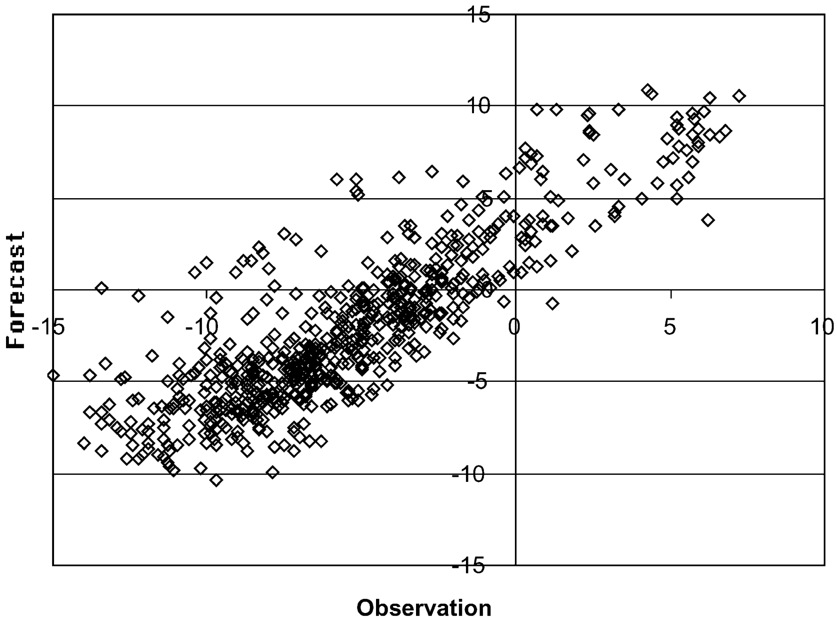

Fig. 14. Scatterplot for temperature (°C) at Kootenay Pass. NMS 30-km grid 24-h forecast, remote observations, $N = 635$.

MC2 grids over-forecast easterly winds considerably, but wind was observed only 4% from that direction in the analyzed time period. The results of the MC2 2- and 10-km grid are similar: neither is superior. The Percentage of Forecast Correct (Table 6) higher than 25% indicates that all models have skill. However, the NMS with the lowest resolution has the overall highest value here.

Wind direction forecasts have also been corrected with the Kalman-predictor correction method. The wind rose is shown in Fig. 12. Fig. 13 gives the bias ratio for each aspect with percentage of occurrence. Improvement for both grids can be seen for all aspects. Westerly winds are slightly under-forecast, whereas winds from the east and south are over-forecast. The Percentage of Forecast Correct has also increased (Table 7).

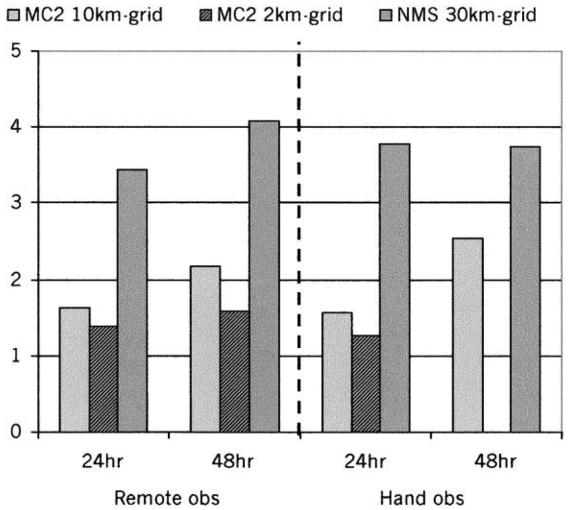

Fig. 16. Mean absolute error (MAE) for temperature (°C). Kootenay Pass, Nov. 1999–Jan. 2000.

4.4. Temperature

Generally, the temperature forecasts are very good. All models and grids achieve high correlation between forecast and observation values. Fig. 14 shows an example of a scatterplot. All points are close to the 1:1-line. Also, more points are above the line than below, which is valid for all forecasts. Together with the negative mean error ME (observation value–forecast value) in more than 90% of all forecasts, it indicates that the predicted temperature is generally too high.

The correlation coefficient of all different forecast–observation pairs for Kootenay Pass is shown in Fig. 15. The 24-h forecast shows generally better results. The MC2 2-km grid does not achieve higher correlations than the 10-km grid. Except for the 48-h

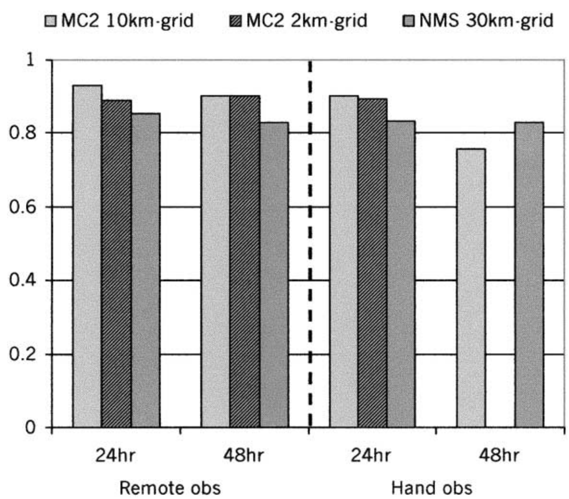

Fig. 15. Pearson correlation coefficient for temperature. Kootenay Pass, Nov. 1999–Jan. 2000.

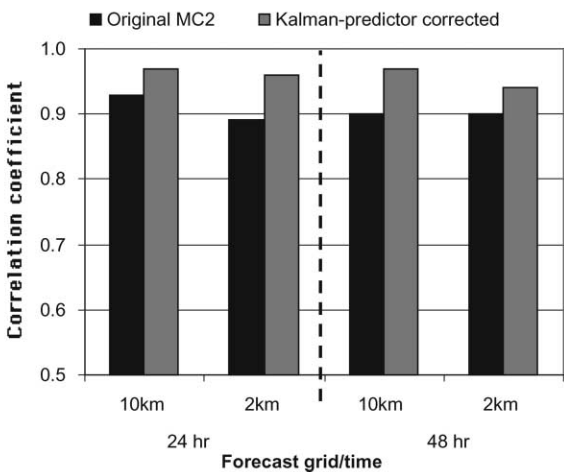

Fig. 17. Pearson correlation coefficient for Kalman-predictor corrected temperature (°C). Kootenay Pass, Nov. 1999–Jan. 2000.

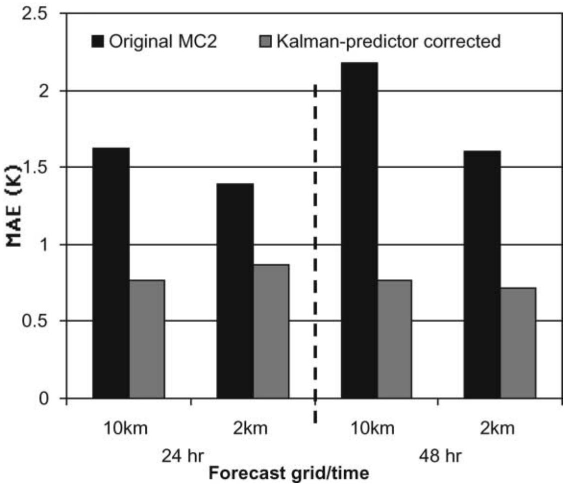

Fig. 18. Mean absolute error (MAE) for Kalman-predictor corrected temperature (°C). Kootenay Pass, Nov. 1999–Jan. 2000.

forecast verified with hand observations, the correlations with the NMS model are not as good as with the MC2 model, but values higher than 0.8 are still prevalent.

The results of MAE are more consistent. The MC2 results are significantly better than the NMS results. The histograms of MAE (Fig. 16) also show better results with the MC2 2-km grid compared to the 10-km grid. The 24-h forecasts have lower (better) values than the equivalent 48-h forecasts, but the differences are not very distinct.

The Kalman-predictor post-correction method was also tested with temperature data at Kootenay Pass. The correlation coefficients are given in Fig. 17. Fig. 18 shows mean absolute errors. An obvious improvement can be seen: correlation coefficients increase and achieve values up to 0.97 using the Kalman technique. The Kalman-predictor corrected forecasts have a significantly lower error than the original forecasts, in some cases as much as 50% error reduction compared to the original forecast error is achieved.

5. Conclusions and outlook

The verification results look very promising. It was shown that each model has different strengths and weaknesses. Neither one of the models is best for all variables. For example, the highest value for PFC for wind direction from the NMS model 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.

Precipitation rate: in almost 75% of all cases, precipitation events and non-precipitation events were forecast as such. All models slightly under-forecast precipitation events. Each model shows skill because the skill score is higher than zero, but improvement is needed. The NMS model has comparable results to the MC2 model even though its resolution is lower. With the Kalman-predictor correction method, an improvement can be seen in every category for precipitation divided into two and five categories.

Wind speed is generally under-forecast. For Stagleap, this might be 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. At Kootenay Pass, wind speed is observed manually in categories, which might account for most of the error. The Kalman-predictor correction method shows high improvement since *moderate*, *strong* and *extreme* winds are captured significantly better. PFC-values as well as statistic results for wind speed as a continuous variable (Pearson correlation coefficient and errors ME and MAE) are improved.

The results for *wind direction* confirm that the NMS model has comparable results to the MC2 model even though the resolution of the NMS model is much lower over the Kootenay Pass area. The NMS model with the 30-km grid has values for the bias ratio close to one in almost all categories at Stagleap. The 2-km grid of the MC2 model performs slightly better than the 10 km-grid. PFC-values higher than 25% indicate that all models have skill, but they are not satisfying. The NMS model with the lowest resolution has the overall highest value here. With the Kalman-predictor correction method, improvement for both MC2 grids can be seen for all aspects.

For *temperature*, results from the MC2 model are better than from the NMS model in most cases, which may be explained by the higher grid resolution. Whereas the MC2 model is run with 10- and 2-km grids, the NMS model has a grid spacing of 30 km over the Kootenay Pass area. Temperatures are generally predicted as too warm. The Kalman-predictor correction method significantly improves correlation coefficients and mean absolute errors of the MC2 forecasts at Kootenay Pass and Stagleap. The small MAE-values around 0.7 K suggest that this forecast can be tested as input for avalanche forecast models.

Twenty-four-hour forecasts are generally more accurate than 48-h forecasts. Comparing the two grids from the MC2 model showed that the 2-km grid has overall slightly better results than the 10-km grid, but not significantly. The NMS model showed very good results for precipitation. For temperature, the MC2 model is clearly better. This is also true for wind speed whereas for wind direction, the NMS model performs very well. This concludes that the different topography approximation of the two models have a large effect on 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 numeric models) might be another reason for the bias.

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

Due to the great variety of climate zones in Canada, the demand for avalanche prediction is at the meso scale, which implies more accurate weather prediction than for synoptic scale forecasts, which relies strongly on snow-stability information with less reliance on meteorological data.

These first results of the project are very successful and encourage to test the verified parameters as input for avalanche forecasting models. Due to the scale of these models and therefore the character of their input data, only numerically available weather forecast data have been verified. For conventional avalanche forecasting (synoptic scale), further important variables like cloud cover or type of precipitation are considered. For this project, our goal is to combine weather and avalanche forecasting, which is possible only at the meso scale where numerical data of the same range is available. The four verified parameters (precipitation rate, wind speed, wind direction, and temperature) are standard weather forecast output and are measured regularly at surface stations so that enough data were available.

For future work, the project includes time series analysis for the identification of phase and amplitude errors. Also, the output of the numerical weather models will then be directly applied for numerical avalanche forecasting, using the model developed by McClung and Tweedy (1994).

For this combination, as well as for applications in other fields, we suggest to use a combination of the two models (an ensemble forecast), depending on the significance of the meteorological variables. The MC2 2-km grid 24-h forecast has overall best results and is our suggestion when only one forecast can be chosen.

Acknowledgements

This research was sponsored by Canadian Mountain Holidays, the Natural Sciences and Engineering Research Council of Canada (NSERC), Forest Renewal BC, Environment Canada 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 Intrawest Whistler/Blackcomb. 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)

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{A.1}$$

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

Mean error (ME)

$$\mathrm{ME} = \bar{x} - \bar{y} \tag{A.2}$$

Mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n} \sum_{k=1}^{n} |x_k - y_k| \tag{A.3}$$

Mean square error (MSE)

$$\mathrm{MSE} = \frac{1}{n} \sum_{k=1}^{n} (x_k - y_k)^2 \tag{A.4}$$

Root mean square error (RMS)

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{A.5}$$

Contingency table analysis: equations for $I = J = 2$ table.

			Range	Perfect Forecast
Hit rate (H)	$H = \frac{A+D}{N}$	(A.6.1)	0–1	1
Percentage of forecast correct (PFC)	$\mathrm{PFC} = H \times 100\%$	(A.6.2)	0–100%	100%
Threat score (TS)	$\mathrm{TS} = \frac{A}{A+B+C}$	(A.7)	0–1	1
Probability of detection (POD)	$\mathrm{POD} = \frac{A}{A+C}$	(A.8)	0–1	1
False-alarm rate (FAR)	$\mathrm{FAR} = \frac{B}{A+B}$	(A.9)	0–1	0
BIAS	$\mathrm{BIAS} = \frac{A+B}{A+C}$	(A.10)	$-\infty$–$+\infty$	0
Heidke skill score (HSS)	$\mathrm{HSS} = \frac{2(AD-BC)}{(A+C)(C+D)+(A+B)(B+D)}$	(A.11)	-1–$+1$	1
True skill score (TSS)	$\mathrm{TSS} = \frac{AD-BC}{(A+C)(B+D)} = \frac{A}{A+C} - \frac{B}{B+D}$	(A.12)	-1–$+1$	1

References

AHD, 2000. The American Heritage® Dictionary of the English Language, 4th edn. Houghton Mifflin, Boston, www.bartleby.com/61/.

Armstrong, R.L., Armstrong, B.R., 1987. Snow and avalanche climates of the western United States: a comparison of maritime, intermountain and continental conditions. IAHS Publ. 162, 281–294.

Benoit, R., Desgagné, M., Pellerin, P., Pellerin, S., Chartier, Y., Desjardins, S., 1997. The Canadian MC2: a semi-Lagrangian, semi-implicit wideband atmospheric model suited for finescale process studies and simulations. Mon. Weather Rev. 125, 2382–2415.

Bozic, S.M., 1979. Digital and Kalman Filtering. Wiley, New York, 164 pp.

CAA, 1995. Canadian Avalanche Association: Observation Guidelines and Recording Standards for Weather, Snowpack and Avalanches. CAA, Revelstoke, BC, Canada, 99 pp.

Flueck, J.A., 1987. A study of some measures of forecast verification. 10th Conference on Probability and Statistics in Atmospheric Sciences. October 6–8, 1987, Edmonton, Alberta, Canada.

Foehn, P.M.B., 1998. An overview of avalanche forecasting models and methods. Norwegian Geotechnical Inst. Publ. No. 203, Oslo, pp. 19–27.

McClung, D.M., Schaerer, P., 1993. The Avalanche Handbook. The Mountaineers, Seattle, WA, 272 pp.

McClung, D.M., Tweedy, J., 1994. Numerical avalanche prediction: kootenay Pass, British Columbia, Canada. J. Glaciol. 40 (135), 350–358.

Murphy, A.H., Daan, H., 1985. Forecast evaluation. In: Murphy, A.H., Katz, R.W. (Eds.), Probability, Statistics, and Decision Making in the Atmospheric Sciences. Westview Press, Boulder, Colorado, USA, 547 pp.

RHW, 1997. The Random House Webster's Unabridged Dictionary, 2nd edn., 2256 pp.

Stull, R.B., 2000. Meteorology Today For Scientists and Engineers, 2nd edn. Brooks/Cole, Pacific Grove, CA, 502 pp.

Tripoli, G.J., 1992. A nonhydrostatic mesoscale model designed to simulate scale interaction. Mon. Weather Rev. 120, 1324–1359.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. International Geophysics Series, vol. 59. Academic Press, San Diego, California, USA, 468 pp.