7. von Recum, A. F., *Handbook of Biomaterials Evaluation, Scientific Techniques and Clinical Testing of Implant Materials*, Macmillan, New York, 1986.
8. International Standard ISO 10993, Biological evaluation of medical devices, International Organization for Standardization, Switzerland, 1997.
9. Class II special controls guidance document: Resorbable calcium salt bone void filler device: Guidance for Industry and FDA. US Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Document issued on 2 June 2003.
10. Aoki, H., *Medical Applications of Hydroxyapatite*, Takayama Press, Tokyo, 1994.

ACKNOWLEDGEMENTS. We thank the staff of Toxicology Division and Implant Biology Division of our Institute for help during evaluation.

Received 3 April 2006; revised accepted 23 August 2006

Improvement in nearest neighbour weather forecast model performance while considering the previous day's forecast for drawing forecast for the following day

Dan Singh* and Ashwagosha Ganju

Research and Development Centre, Snow and Avalanche Study Establishment, Chandigarh 160 036, India

Nearest neighbour model for prediction of weather in terms of snow day/no snow day for consecutive three days in advance (lead time up to 72 h) was tested in two different modes of prediction for two different stations; Dhundi in Himachal Pradesh and Stage-II in Jammu and Kashmir (J&K), in the Pir Panjal range of NW Himalaya, with two different types of data. The data of station Stage-II are incomplete with less data of 12 winters (winter 1991–92 to winter 2003–04, missing data of 1994–95) and those of station Dhundi are complete with more data of 15 winters (winter 1989–90 to winter 2003–04). The model performance was tested with incomplete and complete data respectively, in two different modes. First, in mode I prediction of weather is made based on the probability of snowfall calculated from nearest days/nearest situations. Secondly, in mode II the prediction was made considering the previous day's probability of snowfall also, along with the probability of snowfall calculated from nearest days/situations, i.e. while forecasting for day-2 (lead time 48 h), probability of snowfall for day-1 (lead time 24 h) is also taken into account.

The model performance is found to be better for mode II compared to mode I for all three days except for day-1 forecast with incomplete data. The model performance is better for Stage-II compared to Dhundi in both the modes. Significant difference in the model performance for day-1 and day-2 forecasts is found between those with incomplete data compared to those with complete data. The model results are briefly discussed here.

Keywords: Nearest neighbour technique, snow and no snow day, weather prediction.

MANY case-based techniques have been proposed for prediction of weather and weather parameters[1–4]. These methods use different pattern-recognition techniques. Riordan and Bjarne[3] proposed a fuzzy case-based system for weather prediction. Following the approach of case-based reasoning, Singh *et al.*[5] proposed the nearest neighbour model for prediction of weather at a station in terms of snow day/no snow day and the expected snowfall amount under different well-established snowfall categories. The proposed model predicts weather for consecutive three days in advance based on the probability of snowfall calculated from nearest days/situations, where the previous day's probability of snowfall is not considered for drawing forecast for following day, i.e. while forecasting for day-2 (lead time 48 h), probability of snowfall for day-1 (lead time 24 h) is not taken into consideration.

In this communication, nearest neighbour model for weather prediction, where the previous day's probability of snowfall also has been taken into consideration along with the probability of snowfall calculated from nearest days/situations, for drawing forecast for the following day is proposed. The developed model has been tested for two different stations with two different types of data in NW Himalaya. The model performance improves for both the stations while considering the previous day's probability of snowfall along with the probability of snowfall calculated from nearest days/situations for drawing forecast for the following day, i.e. while forecasting for day-2, day-1 forecast has been taken into account. The model performance with incomplete data is found better for mode II compared to mode I for all three days, except for day-1 forecast. The model performance significantly differs for both the stations. A significant difference in the model performance is found for day-1 and day-2 forecasts for incomplete data compared to those with complete data.

The western Himalayan range comprises diverse climatic zones. The climatic conditions in these zones are briefly presented in Sharma and Ganju[6]. The present study area lies in Pir Panjal range of NW Himalaya. The model has been tested for two different stations, i.e. Dhundi in Himachal Pradesh and Stage-II in Jammu and Kashmir (J&K). The climatic conditions at Stage-II have been presented in Singh *et al.*[5].

Dhundi is one of the main research and field study stations of Snow and Avalanche Study Establishment (SASE),

*For correspondence. (e-mail: dan_@rediffmail.com)

Table 1. Climatic conditions at Dhundi (winter 1989–90 to winter 2003–04)

Month	Temperature (°C) Mean (max)	Lowest (max)	Highest (max)	Mean (min)	Lowest (min)	Highest (min)	Mean wind speed (km/h)	Mean cumulative snowfall (cm)	Mean standing snow (cm)
November	10.9	–1.0	20	0.9	–13.0	13.0	4.0	59.7	7.8
December	7.0	–10.0	16.0	–1.8	–14.0	4.5	4.6	155.6	23.8
January	4.0	–5.0	15.0	–4.5	–14.0	3.0	5.6	283.6	86.5
February	4.7	–3.0	14.0	–4.7	–20.0	6.0	5.7	363.3	164.7
March	7.8	–0.5	18.5	–2.3	–22.0	5.5	5.9	298.1	168.8
April	11.9	0.0	25.0	2.0	–18.0	10.0	6.5	56.4	98.8

Manali, at an altitude of 3050 m amsl. The winter precipitation (November to April) in Dhundi area is mainly recorded in the form of snow, and heavy snow storms are recorded during mid-winter (February/March). Snowfall of the order of 200–210 cm has been recorded in 24 h during extreme snow-storm conditions during the past 15 winters (winter 1989–90 to winter 2003–04). The mean winter ambient temperature at Dhundi remains around –10°C and minimum temperature dips to –22°C during peak winter. The climatic conditions at Dhundi are presented in Table 1.

The primary aim of this study was to check the nearest neighbour model performance, while considering the previous day's probability of snowfall along with the probability of snowfall calculated from the nearest days/situations, for drawing forecast for the following day and to assess the model performance in two different areas in NW Himalaya, with respect to actual weather recorded in those areas. In addition, the study is aimed at comparing the model performance in two different modes, for both the areas.

The present study refers to the winter weather prediction in two different areas of NW Himalaya. Therefore, data for winter months (November through April) of the two representative stations are used for the study. Data at Dhundi station are available for 15 winters (winter 1989–90 to 2003–04) measured daily at 0830 h (time 'F') and 1730 h (time 'A') as twice daily measurement by the winter study team. The data were more or less complete, i.e. all surface weather parameters were mostly available for all the winters. However, for a few days one or two surface weather parameters were missing. Data for such days were removed from the original database, to create the model database. Thus, the model database consists of 3472 records (winter 1989–90 to winter 1998–99) for Dhundi station. The developed model was tested independently for the next five consecutive winters (winter 1999–2000 to 2003–04) for Dhundi station.

The data for station Stage-II are more incomplete, i.e. all surface weather parameters were not available for all the winters. There were few days with all the surface weather parameters available during the past 12 winters (winters 1991–92 to 2003–04), while the entire data of winter 1994–95 were missing. For most places in J&K, the problem was more or less the same due to sudden failure of the instruments during winter. Repairing/replacement of the instruments could not be undertaken due to remoteness, inaccessibility, harsh and hazardous climatic conditions prevalent during winter. The model database for Stage-II station consists of 2453 records measured daily at 0830 h (time 'F') and 1730 h (time 'A') for the winters of 1991–99. The developed model was also tested independently for the next five consecutive winters (winter 1999–2000 to winter 2003–04) for Stage-II station.

Many similarity metrics have been defined for searching nearest days from the database[7]. For the present model development, we have taken Euclidean distance metric as proposed by Buser[8] for prediction of avalanches[9–11].

Let x_i be a vector of length p which consists of measurements of p different parameters for day i, and x_i^k be the kth element of vector x_i. The distance between x_i and x_j is defined as

$$d_{ij} = \sqrt{\sum_{k=1}^{p} w_k (x_i^k - x_j^k)^2},$$

where w_k is the weight assigned to the kth parameter. For the present model, equal weights have been assigned to all parameters, although attempts were made to optimize the weights for prediction of avalanches[12].

Ten nearest days, prior situations to the nearest situations, next situations to the nearest situations and next to next situations to the nearest situations are searched from the database and are shown in Figure 1.

Based on ten nearest days/situations, the probability of snowfall is calculated for day-1, day-2 and day-3 forecasts (Figure 1). For calculation of probability of snowfall, ten nearest days/situations are rated differently according to their nearness to the current day, i.e. the first nearest day (nearest day 1) was given maximum weight and the tenth nearest day was given minimum weight. The probability of snowfall (in per cent) is calculated using the following formula[13]

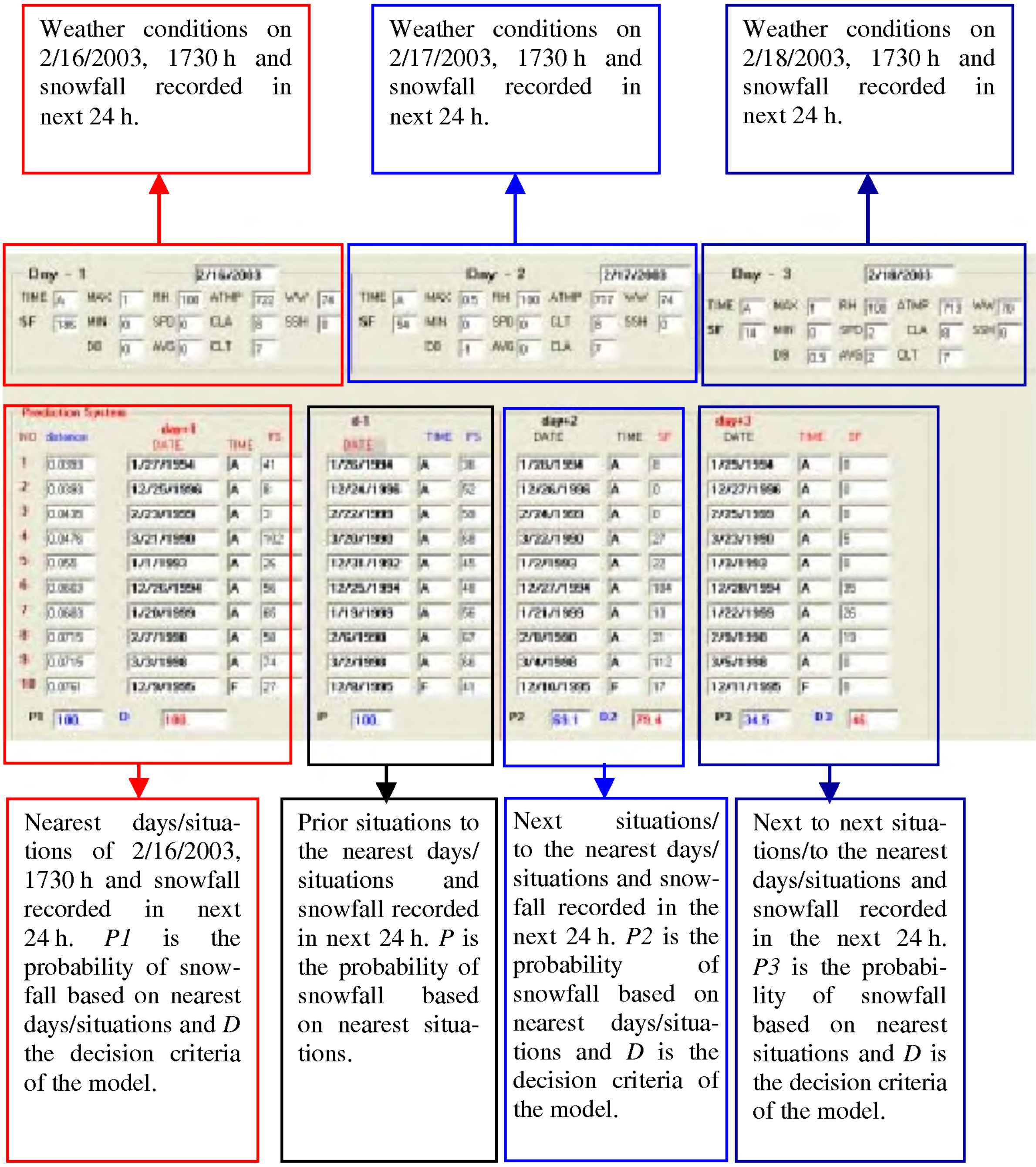

Figure 1. Nearest days/situations, probability of snowfall and decision criteria of model for 2/16/2003, 1730 h.

$$P = \frac{\sum_{r=1}^{N}(N-r+1)\times n_r}{\sum_{r=1}^{N}(N-r+1)} \times 100\%,$$

where N is the number of nearest neighbours selected (10 in the present case), r is the rank of nearest day/situation (1 to 10 in the present case), n_r is 1 if the rth nearest day/situation is snow day and 0 if rth nearest day/situation is no snow day.

In this communication, an attempt has been made to include the previous day's probability of snowfall also along with the probability of snowfall calculated from nearest days/situations, for drawing forecast for following day, i.e. while forecasting for day-2, probability of snowfall on day-1 is also taken into consideration (Figure 1). This has been achieved by taking the weighted average of

Figure 2. Final weather forecast generated by the model (mode II) for Dhundi and recorded weather for 2/16/2003, 1730 h.

probability of snowfall for that day and that of the previous day, i.e. while forecasting for day-2, weighted averages of probability of snowfall for day-1 and day-2 are taken. This may account for small-scale changes in atmospheric processes. The atmosphere remains in a continual state of change. Thus an assumption is made here that day-2 weather (while predicting) will depend on day-1 weather and gradually change to the atmospheric state achieved on day-2. We exactly do not know the atmospheric state (in our case described in terms of surface weather parameters) on day-2. Therefore, weighted average of probability was taken and one-third weight (about 33.33% importance in the final prediction) was given to probability of snowfall on the previous day and two-third weight (about 66.67 % importance in the final prediction) was given to the probability of snowfall on the day under consideration. Thus, the decision criteria of the model was defined as

$$D = \tfrac{1}{3} \times P_{n-1} + \tfrac{2}{3} \times P_n,$$

where P_n is the probability of snowfall for the day under consideration calculated from the nearest days/situations and P_{n-1} is the probability of snowfall for the previous day/situation.

The present model predicts a day as snow/no snow day based on the value of D. Thus, the value of D becomes the decision criterion of the model.

The model test run on 2/16/2003, 1730 h (time 'A') for Dhundi station is shown in Figure 1. The snow and weather conditions for three consecutive days, nearest days/situations, probability of snowfall based on nearest days/situations and decision criteria of the model, i.e. the value of D, are shown in Figure 1. The final weather forecast generated by the model three days in advance for Dhundi station and the recorded weather are shown in Figure 2.

For present study the nearest neighbour model has been tested in two different modes described below:

Mode I: Nearest neighbour model which predicts weather based on the probability of snowfall calculated from the nearest days/situations.

Mode II: Nearest neighbour model which predicts weather considering previous day's probability of snowfall along with the probability of snowfall calculated from nearest days/situations.

The nearest neighbour model in both the modes was validated independently for the next five winters (winter 1999–2000 to 2003–2004) for both the stations. The model has been developed and validated with the actual weather recorded in both the areas and is compared independently in both the modes for both the stations.

The decision label of the model for both the modes was kept at 40% for all three days, based on Singh *et al.*[5] and after training for winter 1998–99, i.e. if the value of D is greater than 40%, then the model predicts the day as snow day, otherwise it is no snow day for both the modes. This has been done to compare the model performance in both the modes.

The model in both the modes has been validated for 896 days for Dhundi station (Tables 2 and 3) and for 907 days for Stage-II station (Tables 4 and 5) for generating day-1 forecast. The model performance has been presented in terms of per cent correct[14,15]. Significant difference of about 9% is found in the model performance between

Table 2. Model performance with mode I for Dhundi and its validation with actual weather recorded in the area

	Day-1 forecast		Day-2 forecast		Day-3 forecast	
Recorded	Snow day	No snow day	Snow day	No snow day	Snow day	No snow day
Snow day	83	113	48	148	37	158
No snow day	111	589	134	561	96	595
Total days	896		891		886	
Overall model performance (%)	75.0		68.35		71.33	

Table 3. Model performance with mode II for Dhundi and its validation with actual weather recorded in the area

	Day-1 forecast		Day-2 forecast		Day-3 forecast	
Recorded	Snow day	No snow day	Snow day	No snow day	Snow day	No snow day
Snow day	88	108	52	144	41	154
No snow day	99	601	132	563	96	595
Total days	896		891		886	
Overall model performance (%)	76.90		69.02		71.78	

Stage-II and Dhundi stations for day-1 forecast for mode I. The model performance for Stage-II is about 84% and for Dhundi it is about 75% for day-1 forecast for mode I. The difference in the model performance between the two stations may be due to the difference in perturbation on surface meteorological observations of the approaching westerly disturbance over the stations[5] and may be due to significant role of topography of the area. Dhundi area is widely covered with forest cover compared to the Stage-II area. Further, the Stage-II station directly lies in the path of the westerly disturbance, i.e. westerly disturbances hit the Stage-II station more frequently compared to Dhundi. Generally, when an intense and widespread westerly disturbance approaches in the NW Himalayan region, heavy precipitation is recorded in Dhundi area.

The model performance degrades for both the stations as the lead time in forecast increases i.e. from day-1 to day-2 forecast for mode I (Tables 2 and 4) and mode II (Tables 3 and 5). The maximum difference in the model performance for both the modes between day-1 and day-2 forecasts for the incomplete data, i.e. Stage-II station, is about 18% (Table 4). With complete data, for Dhundi station, the maximum difference in model performance in both the modes between day-1 and day-2 forecasts is about 8 % (Table 3). The model performance increases for both the stations from day-2 to day-3 forecast for both modes, while it is found to be more for mode I.

The model performance for both the stations for both the modes for the past five winters is given in Tables 2 (mode I), and 3 (mode II) for Dhundi station (complete data) and in Tables 4 (mode I), and 5 (mode II) for Stage-II station (incomplete data).

Total 196 snow days and 700 no snow days were recorded for Dhundi area during the past five winters (Table 2 and Table 3). Out of 196 snowfall days, 83 snow days and out of 700 no snow days, 589 no snow days were correctly predicted by the model for day-1 forecast for mode I (Table 2). Accurately predicted days by the model are 672 (83 snow days and 589 no snow days) for day-1 forecast and model was test run for a total 896 days for generating day-1 forecast. Thus the overall model performance (percent correct[14,15] works out to 75% [$100 \times (672/896) = 75\%$) for day-1 forecast (Table 2).

In mode II, the model predicts 88 snow days and 601 no snow days correctly for Dhundi station for day-1 forecast (Table 3). In total the model predicts 689 days (88 snow days and 601 no snow days) correctly out of 896 days. Model performance for mode II is 76.90% (Table 3) for day-1 forecast.

The model in mode II predicts 17 more days accurately (5 snow days and 12 no snow days) compared to model I for Dhundi station for day-1 forecast.

Similar analysis has been done for Stage-II station and the results are presented in Tables 4 and 5.

The model performance is found better for mode II compared to mode I for day-2 and day-3 forecasts with incomplete data (Tables 4 and 5). With incomplete data the model performance decreases for day-1 forecast (Table 5) for mode II compared to mode I (Table 4). This may be due to the incomplete character of data.

The model performance for mode II with complete data is found consistent compared to the incomplete data, i.e. the model performance for Dhundi area increases for all three days for mode II and for Stage-II station the model performance increase for day-2 and day-3 forecasts and decreases for day-1 forecast.

The performance of nearest neighbour model for weather prediction improves on considering previous day's prediction along with the probability of snowfall based on the nearest days/situations. The difference in the model

Table 4. Model performance for Stage-II with mode I and its validation with actual weather recorded in the area

	Day-1 forecast		Day-2 forecast		Day-3 forecast	
Recorded	Snow day	No snow day	Snow day	No snow day	Snow day	No snow day
Snow day	108	64	67	105	71	100
NSF	79	656	202	528	176	550
Total days	907		902		897	
Overall model performance (%)	84.23		65.96		69.23	

Table 5. Model performance for Stage-II with mode II and its validation with actual weather recorded in the area

	Day 1-forecast		Day-2 forecast		Day-3 forecast	
Recorded	Snow day	No snow day	Snow day	No snow day	Snow day	No snow day
Snow day	102	70	67	105	76	95
No snow day	77	658	181	549	166	560
Total days	907		902		897	
Overall model performance (%)	83.79		68.29		70.90	

performance between day-1 and day-2 forecasts is found to be less with complete data compared to incomplete data.

Model performance is found better for the stage-II station compared to Dhundi station. This may be due to topographical differences and difference in approach of westerly disturbance in both the areas. More detailed investigation of the precipitation characteristics of Dhundi area is needed. Further improvement and training of the model for Dhundi area are needed.

Complete data are needed for consistent and more accurate prediction of weather in terms of snow day/no snow day. Attempts are being made to collect complete data at different stations in J&K with the help of state-of-the-art equipment and automatic weather stations.

The different parameters have been assigned equal weights for searching the nearest days/situations. Different weights can be assigned to the different parameters for better prediction of weather in terms of snow day/no snow day.

1. Chevallier, F., Cheruy, F., Scott, N. A. and Chedin, A., A neural network approach for a fast and accurate computation of a long wave radiative budget. *Mon. Weather Rev.*, 1997, **37**, 1385–1397.
2. Kaufmann, P. and Whiteman, C., Cluster-analysis classification of winter wind patterns in the Grand Canyon region. *J. Appl. Meteorol.*, 1999, 1131–1147.
3. Riordan, D. and Bjarne, H. K., A fuzzy case based system for weather prediction. *Eng. Int. Syst.*, 2002, **10**, 139–145.
4. Carter, M. M. and Elsner, J. B., A statistical method forecasting rain over Puerto Rico. *Weather Forecast*, 1997, 515–525.
5. Singh, D., Ganju, A. and Singh, A., Weather prediction using nearest neighbour model. *Curr. Sci.*, 2005, **88**, 1283–1289.
6. Sharma and Ganju, A., Complexities of avalanche forecasting in Western Himalaya: An overview. *Cold Region Sci. Technol.*, 2001, **31**, 95–102.
7. Fukunaga, K., *Statistical Pattern Recognition*, Academic Press, New York, 1991.
8. Buser, O., Avalanche forecast with the method of nearest neighbours: An interactive approach. *Cold Region Sci. Technol.*, 1983, **8**, 115–163.
9. Buser, O., Fohn, P., Gubler, H. and Salm, B., Different methods for assessment of avalanche danger. *Cold Region Sci. Technol.*, 1985, **10**, 199–218.
10. Buser, O., Monika, B. and Walter, G., Avalanche forecast by nearest neighbour method. IAHS Publ. No. 162, 1987, pp. 557–567.
11. Buser, O., Two years experience of operational avalanche forecasting using nearest neighbours method. *Ann. Glaciol.*, 1989, **13**, 31–34.
12. Gassner, Martin, Hans, J. E., Karl, B. and Tom, L., NXD 2000 an improved avalanche forecasting programme based on nearest neighbour method. *ISSW*, 2000, 2–59.
13. Singh, A., Srinivasan, K. and Ganju, A., Avalanche forecast using numerical weather prediction in Indian Himalaya. *Cold Region Sci. Technol.*, 2005, **43**, 83–92.
14. Wilks, D. S., *Statistical Methods in Atmospheric Sciences*, Academic Press, New York, 1995.
15. Mohanty, U. C. and Dimri, A. P., Location-specific prediction of the probability of occurrence and quantity of precipitation over western Himalaya. *Weather Forecast*, 2004, **19**, 420–429.

ACKNOWLEDGEMENTS. We thank R. N. Sarwade, Director, SASE, Chandigarh for encouragement during this study. We also acknowledge critical comments by the anonymous reviewers. The study would not have been possible without the efforts of scientists and technical assistants of the Avalanche Forecasting Group, SASE who collected data under harsh climatic conditions in remote snow-covered areas.

Received 8 September 2005; revised accepted 18 August 2006