A Hidden Markov Model for avalanche forecasting on Chowkibal–Tangdhar road axis in Indian Himalayas

Jagdish Chandra Joshi[1,*] and Sunita Srivastava[2]

[1]*Snow and Avalanche Study Establishment, Him-Parisar, Sector 37-A, Chandigarh, India.*
[2]*Department of Physics, Panjab University, Chandigarh 160104, India.*
[*]*Corresponding author. e-mail: joshjagdish@gmail.com*

A numerical avalanche prediction scheme using Hidden Markov Model (HMM) has been developed for Chowkibal–Tangdhar road axis in J&K, India. The model forecast is in the form of different levels of avalanche danger (no, low, medium, and high) with a lead time of two days. Snow and meteorological data (maximum temperature, minimum temperature, fresh snow, fresh snow duration, standing snow) of past 12 winters (1992–2008) have been used to derive the model input variables (average temperature, fresh snow in 24 hrs, snow fall intensity, standing snow, Snow Temperature Index (STI) of the top layer, and STI of buried layer). As in HMMs, there are two sequences: a state sequence and a state dependent observation sequence; in the present model, different levels of avalanche danger are considered as different states of the model and Avalanche Activity Index (AAI) of a day, derived from the model input variables, as an observation. Validation of the model with independent data of two winters (2008–2009, 2009–2010) gives 80% accuracy for both day-1 and day-2. Comparison of various forecasting quality measures and Heidke Skill Score of the HMM and the NN model indicate better forecasting skill of the HMM.

1. Introduction

Avalanche forecasting has been done worldwide using deterministic as well as statistical methods. In the deterministic methods, relationship between variables is described by equations which lead to the formulation of physical laws. These methods are capable of predicting the behaviour of a physical system with the help of given initial conditions. Statistical methods, on the other hand, can be employed when an event to be forecasted is considered to be dependent on one or more parameters. Statistical avalanche forecasting models have been developed over the globe using various techniques such as nearest neighbour analysis, discriminant analysis, cluster analysis, classification, and regression trees, etc. Obled and Good (1980) tested different statistical methods to address the challenging problem of avalanche forecasting and compared their respective potential in operational forecast. They concluded that nearest neighbour analysis and discriminant analysis appear more promising than cluster analysis but require further developments. Buser (1983) worked on the development of multivariate 'nearest neighbours' technique which provides an indication of the actual situation. This method compares the actual situation with the specific situations in the past and yields a set of previous situations from which a probabilistic value can be determined. He pointed out that the nearest neighbour technique has its advantages but also its limitations. It does not relieve the forecaster from making his own decision but helps to make the forecast sounder one. Buser (1989) after

Keywords. Snow temperature index; avalanche activity index; Hidden Markov Model.

developing the nearest neighbours method used it for operational avalanche forecasting in the ski area of the Parsenn region and found quite useful for operational purpose. Following the suggestions of Obled and Good (1980) and Buser *et al.* (1987), McClung and Tweedy (1994) derived a numerical avalanche prediction scheme for avalanche forecasting on Kootenay Pass, British Columbia, Canada. In their model, they used parametric discriminant analysis techniques (incorporating Bayesian statistics), cluster techniques and 'nearest neighbours' calculated in discriminant space using the Mahalanobis distance as the distance metric. Besides these attempts many studies have been done in recent years on 'nearest neighbours' method for short term avalanche forecasting, e.g., by Gassner *et al.* (2000), Barbec and Meister (2001), Purves *et al.* (2003), Mc Collister *et al.* (2003) and Singh and Ganju (2008).

Avalanche prediction over the Indian Himalayas began with the work of Agrawal and Ganju (1994) and Bhatnagar *et al.* (1994). Agrawal and Ganju (1994) used a semi-quantitative approach for avalanche forecasting. This approach is based on the basic understanding of the physical processes affecting snow pack stability in conjunction with the use of avalanche forecaster's experience. Bhatnagar *et al.* (1994) used nearest neighbourhood criterion using snow and meteorological data. Naresh and Pant (1999) developed a knowledge-based system for forecasting snow avalanches of Chowkibal–Tangdhar Axis (J&K). Singh *et al.* (2005) used nearest neighbour technique and meso-scale model (MM5) for prediction of avalanches three days in advance. Singh and Ganju (2008) developed a rule based expert system for operational purpose over the Himalaya. Joshi *et al.* (2010) used Avalanche Activity Index to categorize levels of avalanche danger on different road axes over the Himalayas. All these studies tend to enhance the confidence and skill of operational avalanche forecasting over the Himalayas.

None of the studies has been reported in the literature on avalanche forecasting using Hidden Markov Models (HMMs). However, a lot of work has been done on HMMs for weather. Zucchini and Guttorp (1991) applied HMM to describe pattern of precipitation in space and time. They introduced unobserved climate states, which had different rainfall distributions associated with them. The transitions between these climate states were assumed to follow a Markov chain, with stationary transition probabilities. Hay *et al.* (1991) introduced 'weather state' models, which link synoptic-scale information such as sea level pressure with local precipitation assuming precipitation to be conditionally (temporally) independent given the weather states. Hughes and Guttorp (1994) described a Nonhomogeneous Hidden Markov model (NHMM) to relate broad scale atmospheric circulation pattern to local rainfall.

The HMMs are well adopted in the forecasting because of their capability of recognition (forward algorithm), optimization (Viterbi algorithm), and training (Baum–Welch algorithm) the model. The advantage of using HMM for prediction is its capability to capture inherent nonlinearity of the data. The HMM based avalanche forecasting model differs from the other models developed so far in two ways: (a) It uses initial dataset only for all the advance predictions (day-1, day-2, etc.) whereas the models developed so far need future dataset for advance prediction. (b) Once the model, has been developed, there is no further requirement of historical database for future predictions.

The present work aims at the development of an HMM for avalanche forecasting on Chowkibal–Tangdhar road axis (C–T axis) in Indian Himalayas for two days in advance. In this model, a new derived snowpack factor – Snow Temperature Index (STI) has been used as one of the input parameters. As the snowpack factors are more relevant for avalanche forecasting compared to snow and meteorological factors, introduction of STI into the model enhances confidence level of the model forecast. The model forecast is in the form of different levels of avalanche danger (no, low, medium and high) with a lead time of two days. There are six input parameters (fresh snow during 24 hrs, standing snow, average temperature, snow fall intensity, and STI of top and buried layers), four observation symbols derived from the input parameters and four states (no, low, medium, and high danger levels) in the model. For an initial observation, the model first computes most probable observation sequence with the help of forward algorithm. The state sequences, i.e., avalanche danger levels have been computed by using the most probable observation sequence in the Viterbi algorithm.

2. Study area and data

Chowkibal–Tangdhar road is the only route that connects district Tithwal with Kupwara of Jammu & Kashmir (J&K). This road axis falls in the lower Himalayan zone, crosses Pir–Panjal range at Nasta–Chun Pass (N–C Pass) and affected by 25 registered avalanche sites (figure 1) within a stretch of 36 km. This region receives moderate to heavy snow fall during winter due to extratropical cyclonic circulations called western disturbances. Weather of C–T axis and surrounding region is represented by snow and meteorological data collected at Stage-II, an observatory of Snow and Avalanche

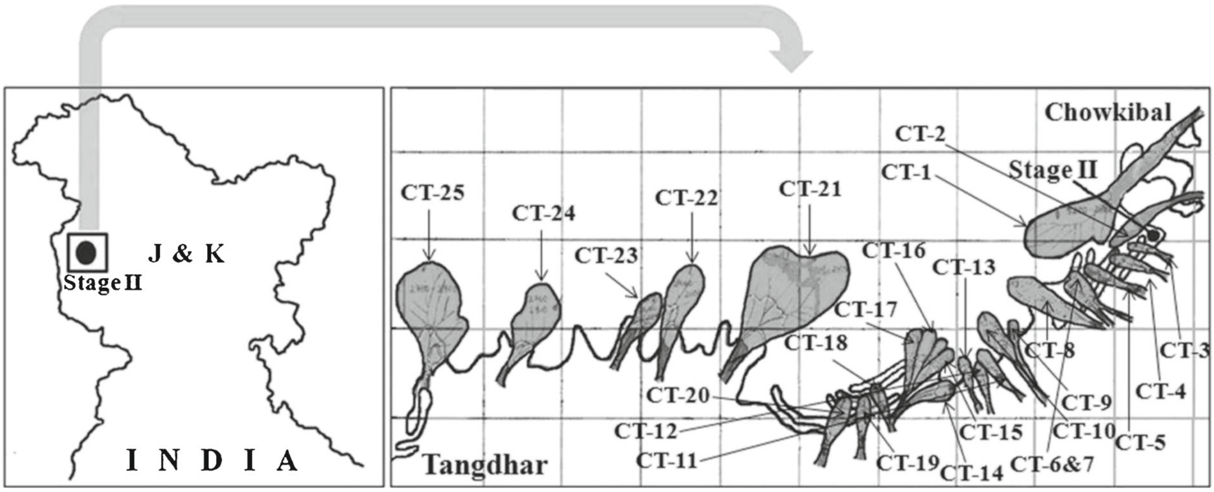

Figure 1. Registered avalanche sites on Chowkibal–Tangdhar road axis (CT-axis) in J&K, India.

Table 1. *Summary statistics of the data used in the model.*

Summary statistics	Average temperature (°C)	Snowfall in 24 hrs (cm)	Snowfall intensity (cm/hrs)	Snowpack height (cm)	Temperature index of top layer	Temperature index of buried layer
Minimum value	–10.5	0	0	0	0	0
Maximum value	12	134	6.67	335	470.8	470.8
Mean value	0.08	7.5	0.5	93	58.3	26.5
Standard deviation	3.3	16.8	0.97	70.2	76.3	49.3

Study Establishment (SASE), functional at an altitude of 2650 m on this road axis. This road axis has been chosen for study because of availability of continuous record of snow, meteorological and avalanche data from its representative observatory, Stage-II.

The data consist of records of snow and meteorological observations of 12 winters (December–March 1992–1993 through 2007–2008), recorded daily at 0830 hrs forming a total of 1812 records. In addition to snow and meteorological data, avalanche occurrence, and avalanche warning bulletins of the same duration have also been taken in the analysis. The avalanche warning bulletin (AWB) of SASE gives avalanche warning in the form of different levels of avalanche danger (no, low, medium, and high). These danger levels have been taken as four state variables of the HMM. The data used in the analysis consist of routine snow-meteorological observations taken at 0830 hrs and the data derived from these routine observations. The snow and meteorological data include a record of maximum temperature, minimum temperature, fresh snow, fresh snow duration, and standing snow collected during the past 12 winters (1992–2008). The derived model input parameters consist of average temperature, fresh snow in 24 hrs, snow fall intensity, standing snow, Snow Temperature Index (STI) of top and buried layer. Summary statistics of the data has been given in the table 1.

Snow pack factors such as snow layer hardness, shear strength of layers, etc., are more relevant to avalanche forecasting as compared to snow and meteorological factors (McClung 1994). The snow layer hardness is affected by the temperature of the surroundings. To explain the effect of temperature on snow layer hardness, the STI is introduced. The STI attempts to summarize a physical effect that occurs over a period of time into a single index (Kozak *et al.* 2002). A degree-day method is used as the conceptual foundation for the temperature index. For a day when maximum temperature exceeds a base temperature (−10°C), it is summed for each day within the temperature index period. The base temperature of −10°C is chosen for the temperature index based on the finding that sintering increases rapidly above −10°C (Gubler 1982). The STI of a layer exposed for 'D' days is thus given by:

$$\mathrm{STI} = \sum^{D} (T_{\max} + 10°\mathrm{C}),$$

when $T_{\max} > -10°\mathrm{C}$.

By this method, a warmer day would have a higher individual day index than a colder day. Adding the daily indices describe the cumulative effect that temperature has on changes in hardness over a multiple-day period. The introduction of STI in avalanche forecasting model is an attempt to introduce snow pack information into the model.

3. Methodology

The HMM consists of two parts:

- an unobserved state sequence $S_1, S_2, \ldots, S_n$ of 'n' states satisfying the following property of Markov chain (Rabiner 1989):

$$P(S_t|S_{t-1}; S_{t-2}, S_{t-3}, \ldots) = P(S_t|S_{t-1}), \quad t = 2, 3, \ldots$$

where P is the probability and t is the state index. This equation signifies that the probability of transition from previous state to the current state depends on the previous state only.

- The state dependent observation sequence $O_1, O_2, \ldots, O_m$ of 'm' observations, such that when S_t is known, the distribution of O_t depends only on the current state S_t and not on previous states or observations.

In the present study, different levels of avalanche danger as stated in the AWB of SASE have been taken as the state variables of the HMM. The model observations have been derived from the snow-meteorological data collected during 1992–2008 and the AWBs issued during the same duration. The derivation of the model observations have been explained in the following steps:

Step 1: All the model input parameters are categorized into different ranges and for each range of these parameters an index of avalanche (IA) which is a ratio of avalanche days to total days is calculated.
Step 2: A linear sum of the IAs of all the parameters of a day gives Avalanche Activity Index (AAI) for that day (Joshi *et al.* 2010). An average value of the AAI has been computed for each level of avalanche danger and used to categorize the AAI into four ranges, thus forming four observations of the model.

The model has been defined by four states and four observations along with three type of probabilities (state transition probabilities, observation probabilities in different states, and initial state probability). All these probabilities have been computed separately for December, January, February, and March. The state transition probabilities have been computed by taking the ratio of the number of transitions from one state to all the other states and the total number of transitions from that state to the other states. The probabilities of the observations in a state have been computed by taking the ratio of the number of observations and total observations in that state. Similarly initial state probabilities have been computed by taking the ratio of the number of counts of a state and the total number of counts. For prediction of state sequence, i.e., avalanche danger, the preliminary step is to predict observation sequence. In this study, a sequence of two states and two observations is taken for delivering avalanche warning two days in advance. With four observations, taking a sequence of two at a time, there are totally 16 combinations. The probability of these observation sequences is computed by using the forward algorithm.

3.1 *Forward algorithm*

The AAI of a day is one of the observations for that day. A sequence of 'j' observations is required for predicting 'j' days in advance. To predict the best observation sequence $(\mathrm{O}_1, \mathrm{O}_2, \ldots, \mathrm{O}_j)$ based on the information of initial state and initial observation, a forward variable $\alpha_t(i)$ is defined as:

$$\alpha_t(i) = \mathrm{P}(\mathrm{O}_1, \mathrm{O}_2, \ldots, \mathrm{O}_j), \quad t = 1, 2, \ldots, j$$

This forward variable represents the probability of observation sequence $(\mathrm{O}_1, \mathrm{O}_2, \ldots, \mathrm{O}_j)$. For a given sequence of two observations $(\mathrm{O}_1, \mathrm{O}_2)$, initial state probabilities (π), state transition probabilities (T) and observation probabilities (O), the probability of the sequence is calculated in the following steps:

Initialization: Probability of initial observation in different states is given by:

$$\alpha_1(\mathrm{N}) = \pi(\mathrm{N}) * \mathrm{O}_1(\mathrm{N})$$
$$\alpha_1(\mathrm{L}) = \pi(\mathrm{L}) * \mathrm{O}_1(\mathrm{L})$$
$$\alpha_1(\mathrm{M}) = \pi(\mathrm{M}) * \mathrm{O}_1(\mathrm{M})$$
$$\alpha_1(\mathrm{H}) = \pi(\mathrm{H}) * \mathrm{O}_1(\mathrm{H})$$

where N, L, M and H stand for no, low, medium, and high dangers, respectively.
Induction: Probability of the observation sequence '$\mathrm{O}_1\mathrm{O}_2$' in different states is given by:

$$\alpha_2(\mathrm{N}) = [\alpha_1(\mathrm{N}) * \mathrm{T}_{\mathrm{NN}} + \alpha_1(\mathrm{L}) * \mathrm{T}_{\mathrm{LN}} + \alpha_1(\mathrm{M}) * \mathrm{T}_{\mathrm{MN}} + \alpha_1(\mathrm{H}) * \mathrm{T}_{\mathrm{HN}}] * \mathrm{O}_2(\mathrm{N})$$
$$\alpha_2(\mathrm{L}) = [\alpha_1(\mathrm{N}) * \mathrm{T}_{\mathrm{NL}} + \alpha_1(\mathrm{L}) * \mathrm{T}_{\mathrm{LL}} + \alpha_1(\mathrm{M}) * \mathrm{T}_{\mathrm{ML}} + \alpha_1(\mathrm{H}) * \mathrm{T}_{\mathrm{HL}}] * \mathrm{O}_2(\mathrm{L})$$
$$\alpha_2(\mathrm{M}) = [\alpha_1(\mathrm{N}) * \mathrm{T}_{\mathrm{NM}} + \alpha_1(\mathrm{L}) * \mathrm{T}_{\mathrm{LM}} + \alpha_1(\mathrm{M}) * \mathrm{T}_{\mathrm{MM}} + \alpha_1(\mathrm{H}) * \mathrm{T}_{\mathrm{HM}}] * \mathrm{O}_2(\mathrm{M})$$
$$\alpha_2(\mathrm{H}) = [\alpha_1(\mathrm{N}) * \mathrm{T}_{\mathrm{NH}} + \alpha_1(\mathrm{L}) * \mathrm{T}_{\mathrm{LH}} + \alpha_1(\mathrm{M}) * \mathrm{T}_{\mathrm{MH}} + \alpha_1(\mathrm{H}) * \mathrm{T}_{\mathrm{HH}}] * \mathrm{O}_2(\mathrm{H})$$

where T_{NN} represents probability of transition from no to no danger state and so on.

Termination: Probability of the observation sequence 'O_1O_2' is given by:

$$P(O_1, O_2) = \alpha_2(N) + \alpha_2(L) + \alpha_2(M) + \alpha_2(H).$$

Using this algorithm, probability of all the possible observation sequences is computed. The observation sequence having highest probability among all the sequences is used to predict the most probable state sequence. To know the probability of state sequence, Viterbi algorithm is used.

3.2 *Viterbi algorithm*

The Viterbi algorithm is a dynamic programming algorithm that computes the most likely state transition path given an observation sequence. It is actually very similar to the forward algorithm, except that it uses a 'max' rather than summing over all the possible ways to arrive at the current state under consideration and keeps track of states in every step to predict most probable state sequence, the Viterbi path. The best state sequence, $S = (S_1, S_2)$ corresponding to the most probable observation sequence (O_1, O_2) as computed by forward algorithm, is calculated in the following steps:

Initialization: Probability of initial observation in different states is given by:

$$\delta_1(N) = \pi(N) * O_1(N)$$
$$\delta_1(L) = \pi(L) * O_1(L)$$
$$\delta_1(M) = \pi(M) * O_1(M)$$
$$\delta_1(H) = \pi(H) * O_1(H)$$

Induction: Probability of the state sequence 'S_1S_2' in different states is given by:

$$\delta_2(N) = \mathrm{Max}\,[\delta_1(N)*T_{NN}, \delta_1(L)*T_{LN}, \delta_1(M)*T_{MN}, \delta_1(H)*T_{HN}]*O_2(N)$$
$$\delta_2(L) = \mathrm{Max}\,[\delta_1(N)*T_{NL}, \delta_1(L)*T_{LL}, \delta_1(M)*T_{ML}, \delta_1(H)*T_{HL}]*O_2(L)$$
$$\delta_2(M) = \mathrm{Max}\,[\delta_1(N)*T_{NM}, \delta_1(L)*T_{LM}, \delta_1(M)*T_{MM}, \delta_1(H)*T_{HM}]*O_2(M)$$
$$\delta_2(H) = \mathrm{Max}\,[\delta_1(N)*T_{NH}, \delta_1(L)*T_{LH}, \delta_1(M)*T_{MH}, \delta_1(H)*T_{HH}]*O_2(H)$$

Termination: Probability of the state sequence 'S_1S_2' is given by:

$$P(S_1, S_2) = \mathrm{Max}\,[\delta_2(N), \delta_2(L), \delta_2(M), \delta_2(H)].$$

4. Result and discussion

In the present model, a sequence of two observations and a corresponding sequence of states give avalanche forecast in the form of no, low, medium, and high dangers for two days in advance. The model observations and corresponding index of avalanche for different ranges of observations are listed in table 2. The index of avalanche is higher for average temperature range −8 to 0°C. This is because most of the avalanche activities take place during or immediately after snowfall when the temperature remains within the range −8 to 0°C. Both below −8°C and above 0°C, the IA gradually decreases. During mid-winter (January and February) when there is clear sky after snow storm, the average temperature remains below −8°C. After 2–3 days of the snow fall, snow pack becomes stable due to compaction and internal metamorphic processes leading to less avalanche activities. In the month of March, average temperature rises above 0°C and because of less average snow fall and shallow snow pack due to melting, less avalanche activities have been reported. For fresh snow, snowfall intensity, standing snow, and STI of buried layer, the IA increases when traversing from lower to higher range of the parameters. This is because the parameters such as fresh snow, snowfall intensity, and standing snow are directly related to the amount of snow available for avalanching. Higher the values of these parameters the higher will be the probability of avalanching. The STI tells about the melt-freeze cycles faced by the snow pack layers. Higher value of the STI means the snow pack layer has undergone more melt-freeze cycles and has become unstable, triggering an avalanche. For the STI of top layer, the IA decreases while moving through lower to higher range of parameters. This is because the top thick layer of fresh snow having low STI triggers direct action avalanches.

Present model is dependent on three types of probability matrices – the initial state probability matrix (π), the state transition probability matrix (T) and the observation probability matrix (O) as given in table 3. The probability of transition from one state of danger to the other has been computed by taking the ratio of the number of transitions from one state to all the other states and total number of transitions from that state to the other states. The probability of an observation in a state has been computed by taking the ratio of the number of the observations and total observations in that state. Similarly initial state probabilities have been computed by taking the ratio of the counts of a state and total counts of the states. These probability matrices have been computed separately for different months of winter. The months from December to March have been analyzed separately because of variability of avalanche occurrence data in these months. In the month of December, the snowpack starts building up and a critical roughness height of snow has to be reached for initiation of avalanches.

Table 2. *Model input variables and their index of avalanche (IA) in different categories.*

Model input variables		Different categories of variables and corresponding index of avalanches					
Average temperature (°C)	Categories	<–8	–8 to –4	–4 to 0	0–4	4–8	>8
	IA	0.14	0.24	0.20	0.09	0.09	0.08
Fresh snow in 24 hrs (cm)	Categories	<20	20–40	40–60	60–80	80–100	>100
	IA	0.11	0.29	0.41	0.38	0.67	0.57
Snowfall intensity (cm/hr)	Categories	<1	1–2	2–3	3–4	4–5	>5
	IA	0.12	0.23	0.32	0.33	0.29	0.55
Standing snow (cm)	Categories	<50	50–100	100–150	150–200	200–250	>250
	IA	0.03	0.11	0.20	0.31	0.32	0.35
Snow temperature index of top layer	Categories	<20	20–40	40–60	60–80	80–100	> 100
	IA	0.20	0.25	0.06	0.09	0.06	0.06
Snow temperature index of buried layer	Categories	<20	20–40	40–60	60–80	80–100	> 100
	IA	0.11	0.25	0.31	0.25	0.45	0.26

Table 3. *Probability matrices in different months.*

State transition probabilities					Probability of observations in different states					Initial state probabilities	
	No	Low	Medium	High		No	Low	Medium	High		
December											
No	0.721	0.216	0.045	0.009	O_1	0.792	0.192	0.015	00.00	No	0.590
Low	0.448	0.466	0.086	00.00	O_2	0.235	0.735	00.00	0.029	Low	0.310
Medium	0.143	0.429	0.214	0.214	O_3	0.00	0.421	0.579	0.00	Medium	0.070
High	0.400	0.200	0.200	0.200	O_4	0.00	0.00	0.200	0.800	High	0.030
January											
No	0.676	0.261	0.051	0.011	O_1	0.733	0.204	0.063	0.00	No	0.470
Low	0.395	0.353	0.160	0.084	O_2	0.259	0.722	0.019	0.00	Low	0.320
Medium	0.179	0.446	0.286	0.089	O_3	0.00	0.460	0.492	0.048	Medium	0.150
High	0.00	0.250	0.600	0.150	O_4	0.00	0.182	0.303	0.515	High	0.050
February											
No	0.720	0.121	0.116	0.043	O_1	0.780	0.168	0.037	0.014	No	0.480
Low	0.436	0.351	0.149	0.064	O_2	0.297	0.527	0.176	0.00	Low	0.220
Medium	0.157	0.313	0.313	0.205	O_3	0.200	0.240	0.533	0.027	Medium	0.190
High	0.065	0.217	0.413	0.304	O_4	0.045	0.015	0.328	0.612	High	0.110
March											
No	0.779	0.149	0.050	0.018	O_1	0.795	0.183	0.019	0.003	No	0.620
Low	0.450	0.422	0.119	0.009	O_2	0.364	0.600	0.036	00.00	Low	0.240
Medium	0.295	0.318	0.250	0.136	O_3	0.102	0.306	0.571	0.020	Medium	0.100
High	0.00	0.333	0.333	0.333	O_4	0.00	0.077	0.308	0.615	High	0.040

Very few avalanche activities have been reported in the month of December. January and February have almost similar snow and meteorological conditions and statistics of avalanche occurrence data. In the month of March because of slightly lower frequency of snowfall, rising temperature, and snowpack settlement, delayed action avalanches have been reported. In this way except in peak winter (January and February) all the months behave differently towards avalanche activity.

Prediction of the state sequence in the HMM depends on initial observation. For different initial observations, predicted state sequences are given in table 4. This table shows that initial observation O_1 is associated with 'no' danger of avalanche for both day-1 and day-2 irrespective of the winter months. In December, initial observation O_2, is associated with 'low' danger for day-1 and 'no' danger for day-2, observation O_3 with 'low' danger for both day-1 and day-2 and observation O_4 with 'medium' danger for day-1 and 'low' danger for day-2. In January, with initial observation O_2, there is 'low' danger for both day-1 and day-2, with O_3, 'medium' danger for day-1 and 'low' danger for day-2 and with O_4, 'high' danger for day-1 and 'medium' danger for day-2. In February,

Table 4. *Forecast of states (avalanche danger levels) for different initial observations.*

Months	Forecasting days	Sequence of states for different initial observations			
		Observation-1	Observation-2	Observation-3	Observation-4
December	Day-1	No	Low	Low	Medium
	Day-2	No	No	Low	Low
January	Day-1	No	Low	Medium	High
	Day-2	No	Low	Low	Medium
February	Day-1	No	Low	Medium	High
	Day-2	No	No	Low	Medium
March	Day-1	No	Low	Medium	Medium
	Day-2	No	No	Low	Low

observation O_2 is associated with 'low' danger for day-1 and 'no' danger for day-2, observation O_3 with 'medium' danger for day-1 and 'low' danger for day-2 and observation O_4 with 'high' danger for day-1 and 'medium' danger for day-2. In March, for observation O_2 there is 'low' danger for day-1 and 'no' danger for day-2, for O_3, 'medium' danger for day-1 and 'low' danger for day-2 and for O_4, 'medium' danger for day-1 and 'low' danger for day-2.

The accuracy or percent correct is the ratio of the number of days when there is correct forecast and total number of days. There were a total of 242 days during winters 2008–2009 and 2009–2010, out of which 194 days of model forecast (26 days of danger warnings and 168 days of no danger warnings) match with actual observed states, thus giving a reasonably good accuracy (0.80) for day-1. Similarly there are 193 days of the model forecast (15 days of danger warnings and 178 days of no danger warning) that match with the observed states giving forecast accuracy of 0.80 for day-2 also.

The probability of detection is the ratio of the number of 'hits' and number of 'yes observed' events. During validation period, there are 11 days when observed avalanche occurrences do not match with the model warnings giving 0.7 probability of detection for day-1. For day-2, there are 22 such mismatched days giving 0.41 probability of detection.

The false alarm is the ratio of the number of days when the model predicted 'danger', but actually it was 'no danger' and the total number of days of the model prediction when actually no danger was observed. The bias score is the ratio of the number of days when the model predicted 'danger' and the total number of avalanche occurrence events. Out of 242 days of the winter, there are 37 nonoccurrence days for which the model predicted occurrence and 168 nonoccurrence days for which the model also predicted nonoccurrence for day-1, giving a false alarm of 0.18 and model biasing of 1.7. For day-2, there are 27 nonoccurrence days for which the model predicted occurrence and 178 nonoccurrence days for which the model also predicted nonoccurrence resulting in false alarm of 0.14 and model biasing of 1.14.

The HSS measures the fractional improvement of the forecast over random forecast. The range of the HSS is $-\infty$ to 0. Negative values indicate that the chance forecast is better than model forecast, 0 means no skill, and 1 indicates a perfect forecast. The HSS of the model is 0.41 for day-1 and 0.26 day-2.

4.1 *Comparison of HMM with nearest-neighbour (NN) model*

The NN model is based on calculating Euclidian distance between current day observations and all the past observations. A smaller value of Euclidian distance indicates better nearest neighbours. While searching neighbours from historical database, 10 nearest neighbours are selected to predict the avalanche danger. A threshold value of three avalanche days out of 10 nearest days has been taken for warning in favour of an avalanche day.

The HMM as well as the NN model has been run using independent data of winters 2008–2009 and 2009–2010 and assessed their performance by computing various forecasting quality measures such as percent correct or accuracy, probability of detection, false alarm rate, bias, and Heidke Skill Score (table 5) using 2 × 2 contingency matrix (tables 6 and 7). The performance assessment of NN model is done only for day-1 because of its ability to predict only one day in advance with a given dataset.

As far as the data of winters 2008–2009 and 2009–2010 is concerned, forecast accuracy of the NN model is 0.83 whereas that of the HMM is 0.80. However, probability of detection for the HMM (0.70) is higher than that for the NN (0.41). Bias value of the HMM (1.7 for day-1 and 1.14 for day-2) indicates slightly overforecasting, whereas that of the NN (0.86) indicates slightly underforecasting.

Table 5. *Forecast quality measures and Heidke Skill Score of the HMM and the NN models.*

Forecasting days	Percent correct/ accuracy	Probability of detection	False alarm rate	Bias	Heidke Skill Score
HMM based avalanche forecasting model					
Day-1	0.80	0.70	0.18	1.7	0.41
Day-2	0.80	0.41	0.14	1.14	0.26
Nearest-neighbour technique based avalanche forecasting model					
Day-1	0.83	0.41	0.08	0.86	0.32

Table 6. *Contingency matrices for comparison of the forecasts of HMM and NN models.*

		Contingency matrix for HMM				Contingency matrix for NN	
		Day-1		Day-2		Day-1	
		Observed					
		Yes	No	Yes	No	Yes	No
Forecasted	Yes	26	37	15	27	15	19
	No	11	168	22	178	22	186

Table 7. *Forecast verification statistics.*

Event forecasted	Event observed: Yes	No	Marginal total
Yes	a	b	a + b
No	c	d	c + d
Marginal total	a + c	b + d	a + b + c + d

Bias score (B) = (a + b)/(a + c); B =1 unbiased, B<1 under forecast, B>1 over forecast

Probability of detection (POD) = a/(a + c); $0 \leq \text{POD} \leq 1$

False alarm rate (FAR) = b/(b + d); $0 \leq \text{FAR} \leq 1$

Accuracy (A) = (a + d)/(a + b + c + d); $0 \leq A \leq 1$

Heidke Skill Score (HSS) = 2(ad − bc)/(a + b)(b + d) + (a + c)(c + d); $-\infty \leq \text{HSS} \leq 1$

The HSS for the NN (0.32) is relatively less than that for the HMM (0.41) indicating better forecasting skill of the HMM.

In operational avalanche forecasting, cost of type-II error (avalanche danger warning not issued when an avalanche actually occurred) is extremely high (life of human beings involved) as compared to type-I error (avalanche danger warning issued when actually avalanche did not occur) and therefore it is preferred to have type-II error as small as possible. It is obvious (table 6) that smaller value of type-II error corresponds slightly to overforecasting, i.e., bias greater than one. Thus a bias greater than one with smaller type-II error is better than a bias smaller than one with larger type-II error. As far as type-I and type-II errors are concerned, the performance of the HMM is better than the NN model.

Verification of the models with independent data shows that the approach of avalanche forecasting using HMM gives reasonably good results for both the days and can be used as one of the tools for advance prediction of avalanche danger.

5. Conclusion

In this study snow and meteorological variables are used for prediction of avalanche danger on a scale of no, low, medium, and high danger for two days in advance. The model is capable of forecasting with an accuracy of 80% for both day-1 and day-2 with a false alarm – 18% for day-1 and 14% for day-2 and probability of detection–70% for day-1 and 41% for day-2. The lead time of the forecast can be increased by increasing the length of observation as well as state sequence. However, by increasing length of the sequences, the accuracy will deteriorate because probability of the sequences is sensitive to initial observation. As the snow pack factors are more relevant to avalanche forecasting, the accuracy of forecast may further be improved by incorporating additional snow pack factors into the model. This model can be coupled with an appropriate weather prediction model to increase lead time of prediction with reasonable accuracy and can be done in future.

Acknowledgements

The authors acknowledge all data collection teams of SASE for their precious contribution in the collection of valuable snow and meteorological data in the harsh climatic conditions of the Himalayas. Authors would like to give due credit to Sh. Dan Singh who helped in the comparative study of the HMM with the NN model. They also acknowledge their colleagues for valuable technical suggestions and critical comments.

References

Agrawal K C and Ganju A 1994 A semi-quantitative approach to avalanche forecasting; *Proc. Snow Symposium*, pp. 430–439.

Barbec B and Meister R 2001 A nearest neighbor model for regional avalanche forecasting; *Ann. Glaciol.* **32** 130–134.

Bhatnagar V, Goyal R K, Chaudhury S, Gupta H M and Mondal D 1994 On avalanche prediction using neighbourhood criterion; *Proc. Snow Symposium*, pp. 414–416.

Buser O 1983 Avalanche forecast with the method of nearest neighbors: An interactive approach; *Cold Region Sci. Technol.* **8** 135–163.

Buser O 1989 Two years experience of operational avalanche forecasting using the nearest neighbor method; *Ann. Glaciol.* **13** 31–34.

Buser O, Butler M and Good W 1987 Avalanche forecast by the nearest method; *IAHS Publ. No.* **162** 557–569.

Gassner M, Etter H and Birkelend K 2000 An improved avalanche forecasting program based on the nearest-neighbor method; *Proc. International Snow Science Workshop, Big Sky, Montana, USA*, pp. 52–59.

Gubler H 1982 Strength of bonds between ice grains after short contact times; *J. Glaciol.* **28(100)** 457–473.

Hay L, McCabe J, Wolock D M and Ayers M A 1991 Simulation of precipitation by weather type analysis; *Water Resour. Res.* **27** 493–501.

Hughes J P and Guttorp P A 1994 Class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena; *Water Resour. Res.* **30** 1535–1546.

Joshi J C, Ganju A and Sharma V 2010 A new approach to avalanche warning over Indian western Himalaya; *Curr. Sci.* **91(11)** 69–72.

Kozak M C, Elder K, Birklend K and Chapman P 2002 Predicting snow layer hardness with meteorological factors; *Proc. International Snow Science Workshop, Penticton, British Columbia*.

McClung D M 1994 Use of expert knowledge in avalanche forecasting; Proc. Snow Symposium, pp. 417–424.

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

Mc Collister C, Birklend K, Hansen K, Aspinall R and Comy R 2003 Exploring multi-scale spatial pattern in historical avalanche data, Jackson Hole ski area, Wyoming; *Cold Region Sci. Technol.* **37(3)** 299–313.

Naresh P and Pant L M 1999 Knowledge based system for forecasting snow avalanches of Chowkibal–Tangdhar Axis (J&K); *Defense Sci. J.* **49(5)** 381–391.

Obled and Good 1980 Recent developments of avalanche forecasting by discriminant techniques: A methodological review and some applications to the Parsenn area (Davos, Switzerland); *J. Glaciol.* **25(92)** 315–345.

Purves R S, Morrison K W, Moss G and Wright D S B 2003 Nearest neighbors for avalanche forecasting in Scotland–development, verification and optimization of a model; *Cold Region Sci. Technol.* **37(3)** 343–355.

Rabiner L R 1989 A tutorial on Hidden Markov Model and selected applications in speech recognition; *Proceedings of IEEE* **73**, No. 2.

Singh D and Ganju A 2008 Expert system for prediction of avalanche; *Curr. Sci.* **94** 1076–1081.

Singh A, Srinivasan K and Ganju A 2005 Avalanche forecast using numerical weather prediction in Indian Himalaya; *Cold Reg. Sci. Technol.* **43** 83–92.

Zucchini W and Guttorp P 1991 A hidden Markov model for space-time precipitation; *Water Resour. Res.* **27** 1917–1923.

MS received 21 November 2012; revised 23 June 2014; accepted 24 June 2014