



Hidden Markov Model for quantitative prediction of snowfall and analysis of hazardous snowfall events over Indian Himalaya

J C JOSHI^{1,*} , K TANKESHWAR² and SUNITA SRIVASTAVA²

¹*Snow and Avalanche Study Establishment, Chandigarh 160 036, India.*

²*Department of Physics, Panjab University, Chandigarh 160 014, India.*

*Corresponding author. e-mail: joshjagdish@gmail.com

A Hidden Markov Model (HMM) has been developed for prediction of quantitative snowfall in Pir-Panjol and Great Himalayan mountain ranges of Indian Himalaya. The model predicts snowfall for two days in advance using daily recorded nine meteorological variables of past 20 winters from 1992–2012. There are six observations and six states of the model. The most probable observation and state sequence has been computed using Forward and Viterbi algorithms, respectively. Baum–Welch algorithm has been used for optimizing the model parameters. The model has been validated for two winters (2012–2013 and 2013–2014) by computing root mean square error (RMSE), accuracy measures such as percent correct (PC), critical success index (CSI) and Heidke skill score (HSS). The RMSE of the model has also been calculated using leave-one-out cross-validation method. Snowfall predicted by the model during hazardous snowfall events in different parts of the Himalaya matches well with the observed one. The HSS of the model for all the stations implies that the optimized model has better forecasting skill than random forecast for both the days. The RMSE of the optimized model has also been found smaller than the persistence forecast and standard deviation for both the days.

1. Introduction

The Himalayan region, during winter is prone to severe weather due to large amount of snowfall. The snowfall occurs during southward excursions of the subtropical westerly jet (SWJ) (Schiemann *et al.* 2009) associated with terrain-locked, low pressure systems (WDs) at the notch formed by the Himalayas and the Hindu Kush mountains (Lang and Barros 2004). Snowfall due to WD is the principal source for snow cover evolution and glaciers of western Himalaya (Dimri *et al.* 2015). Himalayan snow cover is feeding source of many major rivers in the central Asia with drainage basins reaching a total area of 6.7×10^9 km², and supplying water to more than 1.2 billion people (Revenga *et al.* 2003).

While snowfall is vital for survival of glaciers and glacier-fed rivers, extreme snowfall give birth to severe snow avalanches. Snow avalanches toll lives as well as damage property worth millions every year in western Himalaya (Gusain *et al.* 2009). Therefore, snowfall prediction over Himalaya is of prime importance for avalanche-hazard mitigation. Snowfall prediction is also important in hydrology, climate change, glaciology and crop models in agriculture.

Different modelling approaches have been used over the globe for the prediction of meteorological variables. General circulation models (GCMs) are used to simulate large-scale atmospheric circulation patterns and for determining the effect of changes in the atmosphere on those patterns

Keywords. Precipitation forecast; forward algorithm; Viterbi algorithm; Baum–Welch algorithm.

(Zorita *et al.* 1995). However, the GCMs are not adequate for reproducing local and regional phenomena, such as precipitation (Rind *et al.* 1989). In areas where local topography strongly influences precipitation pattern, stochastic models are used to improve the quality of precipitation sequence (Hay *et al.* 1991). Stochastic models are statistical models that aim at quickly simulating realistic random sequences of atmospheric variables such as temperature, precipitation and wind (Wilks and Wilby 1999). There are mainly three features that distinguish stochastic models from numerical weather prediction models. The stochastic models focus on small spatial scales, they have to be computationally fast to provide numerous random realizations and their outputs have the same distributional properties as observed time series. Because of these reasons, stochastic models have been adopted in impact studies as computationally inexpensive tools to generate synthetic daily time series of atmospheric variables at local sites. Such simulated outputs can be used in any process-based models, typically, electricity demand models or crop models (Launay *et al.* 2009; Kolokotroni *et al.* 2012).

In stochastic models, weather states can be considered as observed or latent depending on the availability of good descriptors of weather patterns. The weather states are said to be observed when they are extracted from external variables such as descriptors of large scale synoptic climatological patterns (Bardossy and Plate 1991, 1992; Wilson *et al.* 1992). Weather types are considered as latent variables when they are estimated on local variables by means of an *a priori* clustering algorithm (Flecher *et al.* 2010), or when they are estimated as a hidden variable in the statistical model.

The weather generators were first developed for hydrological application (Gabriel and Neumann 1962; Todorovic and Woolhiser 1975). Rainfall occurrences at a single site were described by a two-state Markov chain and their intensities by independent exponential or Gamma random variables, leading to the so-called ‘chain dependent model’ (Katz 1977). In this simple model, weather states correspond to the states of the Markov chain, i.e., to dry and wet states. Richardson (1981) added the modelling of daily minimum and maximum temperature and solar radiation to the weather generator in Katz (1977).

Wilks and Wilby (1999) gave a detailed presentation of Richardson’s model and its extensions, with a discussion of the advantages and drawbacks of these models and some application issues. Srikanthan and McMahon (2001) provided a quite comprehensive list of models for annual, monthly and daily climate variables at a single site together with some remarks on multisite models.

Maraun *et al.* (2010) and Wilks (2010, 2012) discussed in detail the strong links between downscaling approaches and stochastic models, mainly focussing on how to make the connection between circulation patterns and local atmospheric variables at the daily scale. The linking between weather types and large scale synoptic atmospheric patterns has been explored by researchers working on statistical downscaling (Hughes and Guttorp 1994; Wilks 2012; Haberlandt *et al.* 2014) and sea state condition generators (Guanche *et al.* 2013). Hughes and Guttorp (1994) described a non-homogeneous hidden Markov model (NHMM) to relate broad scale atmospheric circulation pattern to local rainfall. They classified atmospheric patterns into classes that are associated with particular precipitation pattern based on unobserved weather states. In a study by Mares *et al.* (2003), HMM and NHMM have been used successfully for prediction of quantitative precipitation during spring season over Danube Basin.

A few attempts in the past have been made for numerical as well as statistical prediction of snowfall over Himalaya during winter. Mohanty and Dimri (2004) used perfect prognostic method for probability of precipitation forecasting and quantitative precipitation forecast over western Himalayan region. They used forecast variables of different NWP models as predictor. Dimri *et al.* (2008) predicted probability of precipitation and quantitative precipitation for two days using k-Nearest Neighbour (k-NN) sampling method. Singh *et al.* (2008) used analog method for site-specific snowfall forecast for northwest Himalaya with a lead time of three days. Menegoz *et al.* (2013) applied a Regional Climate Model (RCM) to simulate precipitation and snow cover over the Himalaya. Though this model underestimated precipitation at the foothills of the Himalaya and in its eastern part, it provided a first estimation of liquid and solid precipitation in high altitude areas, where satellite and rain gauge networks are not very reliable. Gerlitz *et al.* (2015) presented a neural-network based approach to generate spatially high resolution precipitation fields in central Asia, covering the Tibetan Plateau and the adjacent mountain ranges and lowlands. They estimated local-scale precipitation rates as a function of large-scale atmospheric conditions.

In the present work, the HMM has been developed for quantitative prediction of snowfall in two different mountain ranges of the Himalaya. In Pir-Panjol range, snowfall has been predicted for eight locations (Station-1, Station-3 to Station-9) and in Great Himalayan range for two locations (Station-2 and Station-10). The geographic location of these stations is shown in figure 1. Historical data of past 20 winters (1992–2012)

collected at these stations has been used to predict quantitative snowfall for two days in advance with a temporal resolution of 12 hrs. The HMM has been developed using nine meteorological variables: (i) maximum temperature, (ii) minimum temperature, (iii) atmospheric pressure, (iv) relative humidity, (v) wind speed, (vi) wind direction, (vii) cloud amount, (viii) cloud type and (ix) sun shine hours. The model output is quantitative snowfall in one of the six different categories;

(i) 0, (ii) 1–15, (iii) 16–30, (iv) 31–45, (v) 46–60, and (vi) >60 cm. The HMM parameters has been optimized using Baum–Welch algorithm. The model has been validated using data of two winters (2012–2013 and 2013–2014) by computing root mean square error (RMSE), percent correct (PC), critical success index (CSI) and Heidke skill score (HSS). The RMSE of the model has been compared with the persistence forecast and standard deviation of snowfall. Leave-one-out cross-validation method has also been applied to carry out RMSE analysis of the model.

2. Data and methodology

A homogeneous HMM has been developed for forecasting of quantitative snowfall at 10 meteorological stations in Pir-Panjal and Great Himalayan mountain ranges of Indian Himalaya. At these stations of Snow and Avalanche Study Establishment (SASE), snow and meteorological data are recorded twice daily at 08:30 and 17:30 hrs since more than last four decades. The dataset used in the model consists of maximum temperature, minimum temperature, atmospheric pressure, relative humidity, average wind speed, wind direction, cloud amount, cloud type and sun shine hours for the winters (November–April) 1992/1993 to 2013/2014. In addition to daily recorded data, derived data such as deviation of maximum temperature from the average maximum temperature, deviation of minimum temperature from the average minimum temperature, deviation of pressure from the standard atmospheric pressure at the station and deviation of relative humidity from saturation have also been used in the model as input. The plots of average maximum temperature, average minimum temperature and average snowfall and average snow

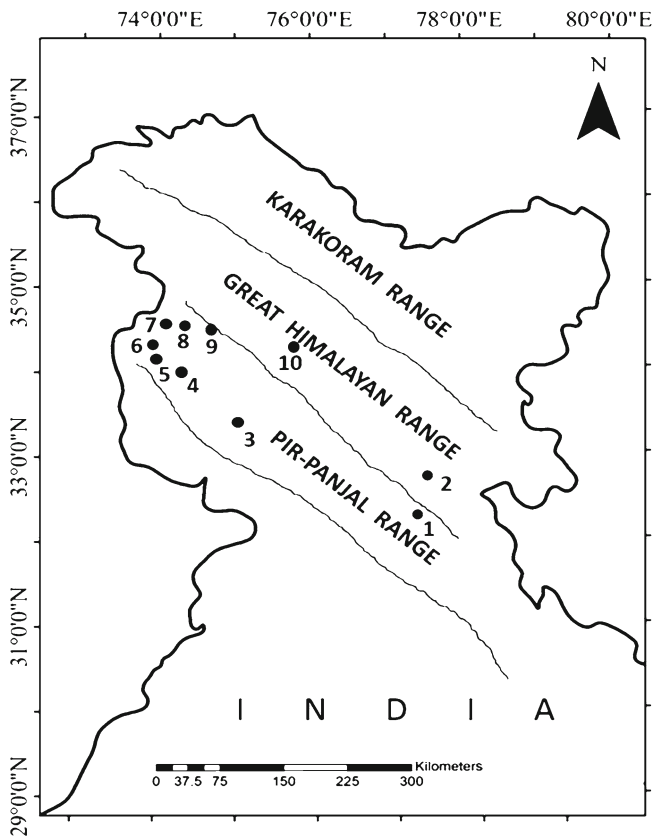


Figure 1. Meteorological stations of SASE in Pir-Panjal and Great Himalayan mountain ranges.

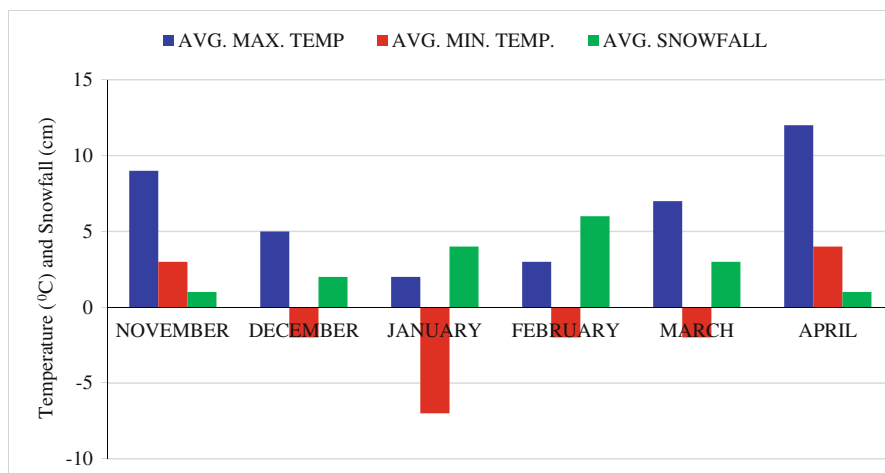


Figure 2. Average weather condition of the Station-6 in Pir-Panjal range of the Himalaya.

depth of representative Station-6 in Pir-Panjal range and that of Station-10 in Great Himalayan ranges for the winters 1992/1993 to 2013/14 are shown in figures 2 and 3, respectively.

Snow and meteorological data collected at 10 meteorological stations of SASE has been used to derive the model input data. The model input variables have been categorized into different ranges having equal intervals and calculated precipitation index (PI) for each range of all the variables. The PI of a range of the variable is the ratio of precipitation days and total days in that range. The PI values of different variables have been normalized between 0 and 1 for making them scale invariant. The variables have been attributed to weights based on correlation of the variables with their PIs. The square of the correlation has been considered as the weight of the variable. The weighted sum of the PIs of all the variables has been grouped into six categories to define six observations of the model. The snowfall amount has also been

categorized into six ranges (0, 1–15, 16–30, 31–45, 46–60 and >60 cm) defining six states of the HMM. The observations and states of the model for different stations are summarized in table 1. The number of snowfall events in each category of the snowfall for all the stations are given in table 2. The average meteorological characteristics of each state for Station-6 in Pir-Panjal range and Station-10 in Great Himalayan range of the Himalaya are given in table 3. The model parameters (initial state probability, state transition probability and probability of observations in different states) have been computed after defining states and observations of the model. The initial state probability denoted as π_i , represents ratio of the number of times state i visited and total number of states. The state transition probability is denoted as A_{ij} that represents ratio of number of transitions from state i to state j and total number of transitions taking place from state i . The probability of observations in different states is the ratio of number of

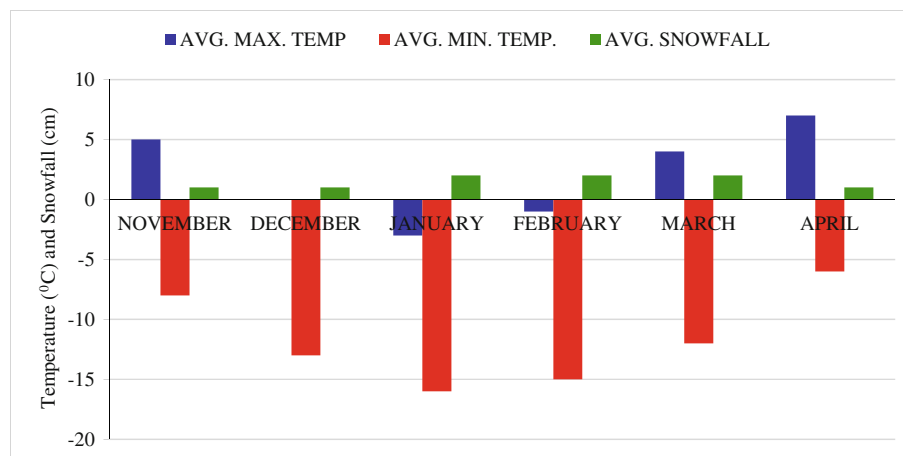


Figure 3. Average weather condition of the Station-10 in Greater-Himalayan range.

Table 1. Ranges of observations and states of HMM for different stations in Indian Himalaya.

Station	Ranges of states of HMM for snowfall (cm)					
	S1 (0)	S2 (1–15)	S3 (16–30)	S4 (31–45)	S5 (46–60)	S6 (>60)
	Ranges of observations for different stations in Indian Himalaya					
	O1	O2	O3	O4	O5	O6
Station-1	<0.80	0.80–1.00	1.00–1.30	1.30–1.50	1.50–1.70	>1.70
Station-2	<0.80	0.80–1.00	1.00–1.20	1.20–1.30	1.30–1.40	>1.40
Station-3	<0.40	0.40–0.45	0.45–0.50	0.50–0.55	0.55–0.60	>0.60
Station-4	<0.90	0.90–1.00	1.00–1.10	1.10–1.15	1.15–1.20	>1.20
Station-5	<0.60	0.60–0.65	0.65–0.70	0.70–0.75	0.75–0.80	>0.80
Station-6	<0.65	0.65–0.70	0.70–0.74	0.74–0.78	0.78–0.82	>0.82
Station-7	<0.50	0.50–0.55	0.55–0.60	0.60–0.65	0.65–0.70	>0.70
Station-8	<1.30	1.30–1.40	1.40–1.50	1.50–1.60	1.60–1.70	>1.70
Station-9	<0.95	0.95–1.05	1.05–1.10	1.10–1.15	1.15–1.20	>1.20
Station-10	<0.60	0.60–0.65	0.65–0.70	0.70–0.75	0.75–0.80	>0.80

observations in state i and total number of observations. These three sets of probabilities have been used to compute the most probable observation as well as state sequence. Sequence of two observations has been used to predict the sequence of two states. The most probable observation sequence has been computed by using Forward algorithm and corresponding most probable state sequence by Viterbi algorithm. The flow chart of the model development is shown in figure 4.

While developing HMM, it is supposed that the model parameters are good to explain the model observations, but this is not true every time. There is a need to adjust the model parameters to maximize the probability of observation sequence. This is done through an iterative method called

Baum–Welch algorithm. The forward, Viterbi and Baum–Welch algorithms have been discussed below.

2.1 Forward algorithm

In the forward algorithm, a forward variable $\alpha_t(i)$ is initialized, which represents probability of initial observation for the state ‘ i ’. Initially, the forward variable is calculated by using initial state probability ‘ π_i ’ and observation probability $b_i(O_t)$ in that state with ‘ t ’ representing time. The probability of observation sequence, $P(O_t O_{t+1})$ can be calculated inductively in the following steps:

Initialization:

$$\alpha_t(i) = \pi_i \times b_i(O_t), \quad 1 \leq i \leq 6 \quad (1)$$

Table 2. Number of snowfall events in different categories at different stations over Himalaya.

Station	Number of snowfall events in different categories of snowfall					
	Category-1 (0 cm)	Category-2 (1–15 cm)	Category-3 (16–30 cm)	Category-4 (31–45 cm)	Category-5 (46–60 cm)	Category-6 (>60 cm)
Station-1	5442	1279	464	248	113	68
Station-2	5849	1500	211	41	11	2
Station-3	5238	1477	503	207	99	90
Station-4	5721	991	434	230	130	108
Station-5	6421	720	225	113	58	77
Station-6	5778	943	339	217	131	206
Station-7	5347	1397	491	219	94	66
Station-8	5886	1085	360	144	81	58
Station-9	5676	1551	278	82	18	9
Station-10	5443	1255	496	221	102	97

Table 3. Average meteorological characteristics in different states at representative stations in Pir-Panjal and Great Himalaya.

State	Average meteorological characteristics						
	Maximum temperature deviation	Minimum temperature deviation	Atmospheric pressure deviation	Relative humidity deviation from saturation	Average wind speed	Cloud amount (okta)	Sun shine hours
Representative station in Pir-Panjal (Station-6)							
State-1	5.0	1.8	2.0	36	2.5	4	5.0
State-2	−0.6	2.0	−0.5	13	1.2	6	2.0
State-3	−1.7	2.5	−2.0	8	0.7	7	0.5
State-4	−2.5	3.5	−2.5	5	0.5	8	0.3
State-5	−3	4.0	−4	4	0.6	8	0.0
State-6	−4	4.0	−5	2	0.3	8	0.0
Representative station in Great Himalaya (Station-10)							
State-1	5.0	3.0	4.0	40	2.8	4	6.0
State-2	1.0	4.0	0.0	22	1.8	6	2.0
State-3	2.0	7.0	−1.0	20	2.6	7	1.0
State-4	3.0	10.0	−6.0	10	2.5	8	0.0
State-5	3.0	10.0	−6.0	5	3.0	8	0.0
State-6	4.0	12.0	−8.0	2	3.2	8	0.0

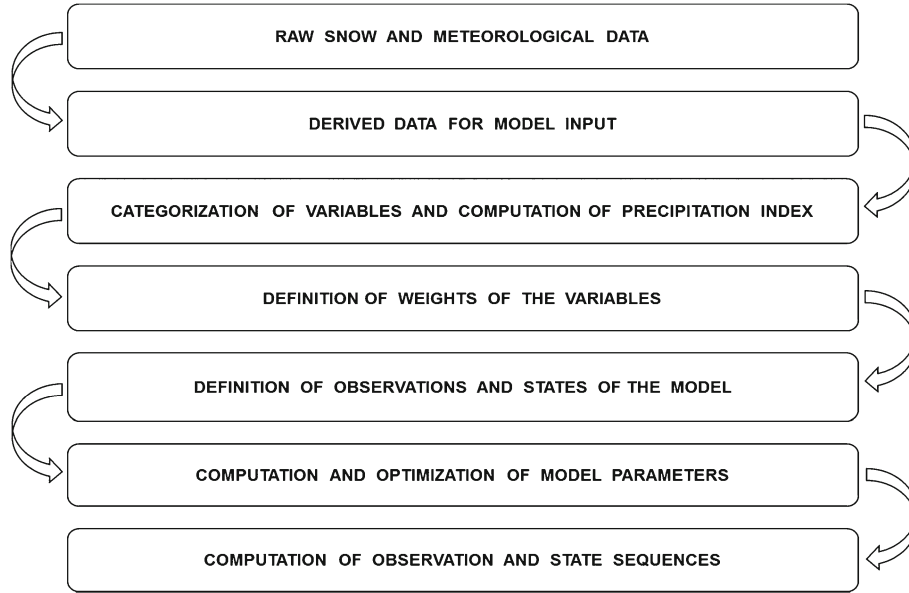


Figure 4. Flow chart of the HMM development for snowfall forecast.

Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^6 \alpha_t(i) \times a_{ij} \right] \times b_j(O_{t+1}), \quad 1 \leq j \leq 6 \quad (2)$$

Termination:

$$P(O_t O_{t+1}) = \sum_{i=1}^6 \alpha_{t+1}(i). \quad (3)$$

After the most probable observation sequence has been calculated, most probable state sequence can be calculated by using Viterbi algorithm as stated below.

2.2 Viterbi algorithm

This algorithm is similar to that of Forward algorithm except that it uses maximization rather than summing over all the possible states and it keeps track of states in every step to predict most probable state sequence. A new variable $\delta_t(i)$ has been introduced that represents highest probability along a single path at time t . It accounts for the first t observations and ends in state i . The most probable state sequence $P(S_t S_{t+1})$ corresponding to the most probable observation sequence $O_t O_{t+1}$ can be calculated iteratively in the following steps:

Initialization:

$$\delta_t(i) = \pi_i \times b_i(O_t), \quad 1 \leq i \leq 6. \quad (4)$$

Induction:

$$\delta_{t+1}(j) = \text{Max}[\delta_t(i) \times a_{ij}] \times b_j(O_{t+1}), \quad 1 \leq j \leq 6. \quad (5)$$

Termination:

$$P(S_t S_{t+1}) = \text{Max}[\delta_{t+1}(j)], \quad 1 \leq j \leq 6. \quad (6)$$

2.3 Baum–Welch algorithm

It is based on both forward and backward variables. The backward variable $\beta_t(i)$ represents probability of partial observation sequence from $t+1$ to the end, given state i at time t . $\beta_t(i)$ can be calculated inductively in the following steps:

Initialization:

$$B_t(i) = 1, \quad 1 \leq i \leq 6 \quad (7)$$

Induction:

$$\beta_t(i) = \sum_{j=1}^6 a_{ij} \times b_j(O_{t+1}) \times \beta_{t+1}(j),$$

$$t = T - 1, T - 2, \dots, 1, \quad 1 \leq i \leq 6. \quad (8)$$

From the forward and backward variables, two new variables can be calculated. The first one is $\xi_t(i, j)$, the probability of being in state S_i at time t , and state S_j at time $t+1$, given the model and the observation sequence, i.e.,

$$\xi_t(i, j) = \alpha_t(i) \times a_{ij} \times b_j(O_{t+1}) \times \beta_{t+1}(j) / P(O/\lambda). \quad (9)$$

The second one is $\gamma_t(i)$, the probability of being in state S_i at time t , given the observation sequence O , and the model, i.e.,

$$\gamma_t(i) = \alpha_t(i) \times \beta_t(i) / P(O/\lambda). \quad (10)$$

The variables ξ and γ satisfy the relationship

$$\gamma_t(i) = \sum \xi_t(i, j), \quad 1 \leq j \leq 6. \quad (11)$$

Now, if $\gamma_t(i)$ is summed over all instants (excluding instant T), expected number of times that the state S_i has been visited is obtained. On the other hand, if $\xi_t(i, j)$ is summed over all instants (excluding instant T), expected number of transitions that have been made from state i to j is obtained. From this behaviour of $\gamma_t(i)$ and $\xi_t(i, j)$, the following re-estimations of the model parameters could be deduced:

$$\pi'_i = \gamma_t(i), \quad (12)$$

$$a'_{ij} = \frac{\left(\sum_{t=1}^{T-1} \xi_t(i, j)\right)}{\left(\sum_{t=1}^{T-1} \gamma_t(i)\right)} \quad (13)$$

$$b'_j(O_k) = \frac{\left(\sum_{\substack{t=1 \\ O_t=W_k}}^T \gamma_t(i)\right)}{\left(\sum_{t=1}^T \gamma_t(i)\right)}. \quad (14)$$

After re-estimations of the model parameters, a new model λ' is obtained, which is more likely than model λ , producing observation sequence O . This process of re-estimation is continued till no improvement in the probability of observation sequence reached.

3. Results and discussion

Hidden Markov Model has been developed for quantitative snowfall at different stations in Pir-Panjaj and Great Himalayan mountain ranges in Indian Himalaya. The HMM uses nine input variables that have been categorized into different ranges. For each range of all the variables, PI has been calculated. The PI of a range of a parameter represents likelihood of precipitation in that range. The model input variables have been assigned weights according to their correlation with corresponding precipitation index. The square of the correlation has been considered as the weight of the variable. The weighted sum of the PI of the model input variables have been categorized into six different ranges to define six observations of the model. The snowfall amount has been categorized into six categories to represent six states of the model (table 1). The probability matrices (initial state probability, state transition probability, probability of observations in different states) have been computed for all the stations to obtain most

probable observation and state sequences. The initial state probability is the ratio of count of a state during winter period and total number of counts of all the states during that period. The state transition probability is the ratio of number of state transitions from one state to the other and total number of transitions from that state. The probability of an observation in a state is the ratio of number of that observation and total number of observations in that state.

The probability matrices as computed from the database can be modified to achieve best possible observation sequence. This can be achieved iteratively by adjusting the model parameters using Baum-Welch algorithm. The probability matrices have been optimized for all the stations to achieve optimal state sequence. The most probable observation sequence and corresponding state sequence have been computed for all the stations with and without optimization.

The model has been validated using independent data (not used in the model development) of two winters (2012–2013 and 2013–2014) by computing skill scores and other verification measures using 6×6 contingency table (Appendix). These scores are summarized in table 4. The RMSE of the model with and without optimization has been compared with the persistence forecast and standard deviation of snowfall for both the days as shown in figures 5 and 6.

As far as the data used in model validation is concerned, the critical success index (CSI) is higher in category-1 as compared to that in other categories for all the stations implying that the model can predict snowfall in lower categories better than in higher categories for both the days. The CSI for all the stations in category-1 vary from 0.63 to 0.87 for day-1 and 0.10 to 0.41 for day-2. The HSS for these stations vary from 0.19 to 0.37 for day-1 and 0.06 to 0.16 for day-2 implying that the model is superior to the random forecast. The overall performance of the model in terms of PC with the validation data vary from 60.3 to 79.6% for day-1 and 15.9 to 38.2% for day-2. All the accuracy measures and skill scores show that the model performance for these stations is reasonably good for day-1 and can be used reliably in other applications such as avalanche forecasting.

The RMSE of the model computed using data of two winters reveal that for all stations, RMSE of the model with optimization has been found smaller than that without optimization and persistence forecast for both the days. The RMSE of the model with optimization has been found <5 cm for Stations 9 and 10, between 5 and 10 cm for Stations 2, 4 and 9 and 10–15 cm for Stations 1, 3, 5, 6 and 7 for day-1 forecast. For day-2 forecast, it has been found <5 cm for Station 10,

Table 4. *RMSE, accuracy measures and Heidke skill score of optimized model for different stations in Indian Himalaya.*

		Skill scores of the optimized model for snowfall								
Station		CSI (Cat-1)	CSI (Cat-2)	CSI (Cat-3)	CSI (Cat-4)	CSI (Cat-5)	CSI (Cat-6)	PC	HSS	RMSE (cm) (leave-one-out)
Station-1	Day-1	0.77	0.09	0.06	0.06	0.08	0.20	66.7	0.31	16.0
	Day-2	0.15	0.04	0.06	0.06	0.07	0.10	30.1	0.10	16.5
Station-2	Day-1	0.77	0.26	0.12	0.09	0.05	0.10	72.2	0.37	8.9
	Day-2	0.10	0.18	0.08	0.10	0.09	0.08	15.9	0.14	9.4
Station-3	Day-1	0.87	0.08	0.09	0.11	0.09	0.10	79.6	0.35	10.0
	Day-2	0.35	0.05	0.05	0.10	0.09	0.08	29.9	0.12	12.1
Station-4	Day-1	0.83	0.13	0.12	0.05	0.09	0.06	78.5	0.32	9.9
	Day-2	0.23	0.12	0.02	0.05	0.07	0.06	16.3	0.15	10.3
Station-5	Day-1	0.63	0.18	0.11	0.07	0.05	0.09	60.5	0.19	11.8
	Day-2	0.19	0.09	0.11	0.08	0.02	0.09	26.2	0.06	12.0
Station-6	Day-1	0.83	0.15	0.11	0.13	0.12	0.18	72.8	0.34	11.3
	Day-2	0.41	0.11	0.09	0.08	0.10	0.12	38.2	0.16	13.0
Station-7	Day-1	0.72	0.15	0.09	0.09	0.10	0.10	60.3	0.23	13.5
	Day-2	0.15	0.15	0.11	0.06	0.12	0.16	19.5	0.13	16.1
Station-8	Day-1	0.71	0.10	0.09	0.09	0.10	0.14	62.9	0.22	12.6
	Day-2	0.26	0.05	0.11	0.06	0.12	0.12	22.2	0.10	12.8
Station-9	Day-1	0.82	0.09	0.10	0.16	0.17	0.10	72.7	0.33	12.3
	Day-2	0.30	0.09	0.07	0.08	0.22	0.12	30.6	0.08	12.5
Station-10	Day-1	0.72	0.20	0.10	0.04	0.06	0.08	68.3	0.27	11.3
	Day-2	0.40	0.08	0.08	0.05	0.10	0.09	16.6	0.09	12.9

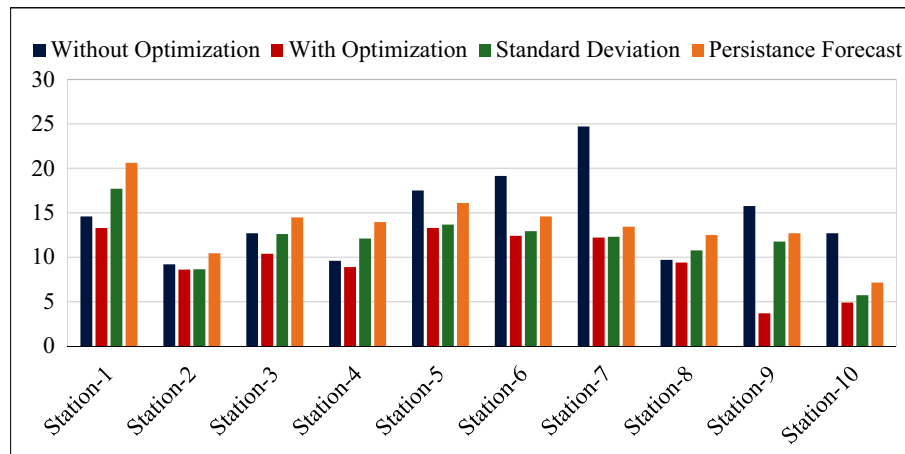


Figure 5. Forecast validation at different stations in Indian Himalaya for day-1.

5–10 cm for Stations 2, 4 and 8 and 10–15 cm for Stations 1, 3, 5, 6, 7 and 9 with model optimization. Thus at 50% stations, the RMSE has been found <10 cm and between 10 and 15 cm for rest of the stations for day-1. For day-2, the RMSE has been found <10 cm for 40% stations and 10–15 cm for rest of the stations. The RMSE of the model with optimization has also been found less than the standard deviation of snow fall.

Leave-one-out cross-validation method has also been applied to carry out RMSE analysis of the model. The RMSE computed through this method vary from station to station from 8.9 to 16 cm for day-1 and 9.4 to 16.5 for day-2. For 60% stations, the RMSE lies between 10 and 15 cm for both the days. 30% of the stations have RMSE lying between 5 and 10 cm and 10% stations have RMSE >15 cm. The RMSE calculated through both the

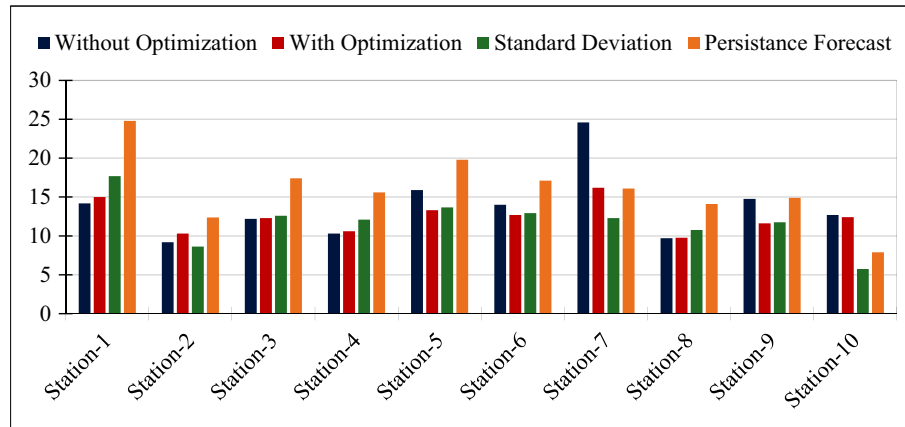


Figure 6. Forecast validation at different stations in Indian Himalaya for day-2.

methods reveal that there is no significant variation in the results, thus implying towards stability and robustness of the model.

3.1 Performance analysis of the model during hazardous snowfall events over Himalaya

Some of the hazardous avalanche activities due to excessive snowfall in different parts of Indian Himalaya have been discussed in this section. The cumulative snowfall during complete snow storm period that resulted into an avalanche has been predicted using the HMM. The observed and predicted cumulative snowfall during the hazardous snow storms are summarized in table 5. The analysis of observed and HMM predicted snowfall during some of the recent hazardous avalanche activities has been discussed below.

3.1.1 Avalanche hazard in Gulmarg, J&K, India (8 February 2010)

A western disturbance hit the Indian Himalaya on 5th February, 2010 that caused a total 92 cm snow fall during 5–8 February, 2010 at Gulmarg. Before the snowfall started in the region, there was 34 cm standing snow existing on the ground. The existing snow faced a number of melt–freeze cycles before it got buried under new snowfall. At the time of snowfall, the existing snowpack was loose due to daytime melting as the snow fall started in the afternoon. This loose snowpack acted as a weak layer when got buried under the new snow and triggered avalanche. The avalanche triggered during snow storm at Khilanmarg (Gulmarg) in J&K, India at 10:30 hrs on 8th February 2010. There were a total of 226 victims of this avalanche accident and 17 lost their lives. Cumulative snowfall during the snow storm was 92 cm and standing snow was 215 cm at Gulmarg at the time

of avalanche accident. The HMM predicted 72 cm cumulative snowfall during the storm from 5 to 8 February 2010.

3.1.2 Avalanche hazard in Drass Sector, J&K, India (17 February 2012)

Snowfall in the month of February in 2012 started on 12th February in the morning at Drass and continued till 17th February. Before the snow storm started, a shallow snow pack of 35 cm was existing on the ground. Within this snowpack at the bottom, there were weak depth hoar crystals formed due to temperature gradient. There was 35 cm cumulative snowfall and 100 cm standing snow at the Drass observatory when an avalanche triggered at 08:30 hrs on 17th February 2012 at Kill Nala in Mashkoh Valley, Drass. Four army persons were victims of this avalanche and two lost their lives. The HMM predicted 52 cm cumulative snowfall over a period of 24 hrs for Drass region.

3.1.3 Avalanche hazard at Sonamarg and Dawar, J&K, India (22 February 2012)

Moderate to heavy snowfall during the snowstorm during 17–22 February 2012, triggered major avalanches at Sonamarg and Dawar in Indian Himalaya. The avalanche triggered at Sonamarg (Srinagar–Leh road axis, J&K) at 16:45 hrs and that at Dawar (Bandipur–Gurej road axis, J&K) at 22:15 hrs on 22nd February 2012. At Sonamarg, three army persons lost their lives out of the nine affected. At Dawar, there were 29 army persons affected by avalanche out of which 16 lost their lives. During this snow storm, there was 132 cm cumulative snowfall at Kanzalwan and 39 cm at Sonamarg. The HMM predicted 180 cm cumulative snowfall at Kanzalwan (Dawar) and 66 cm at Sonamarg during the storm.

Table 5. Observed and model predicted cumulative snowfall during Hazardous snowfall events in different regions over Himalaya.

Region/station	Observed cumulative snowfall	Model predicted cumulative snowfall
Gulmarg (Station-4)	92 cm	72 cm
Drass (Station-10)	35 cm	52 cm
Sonamarg (Near Station-9)	132 cm	180 cm
Dawar (Near Station-9)	39 cm	66 cm

4. Conclusion

In the present study, a Hidden Markov Model has been developed for quantitative prediction of snowfall at different stations in Pir-Panjal and Great Himalayan mountain ranges of Indian Himalaya. The HMM predicts quantitative snowfall for two days in advance. The root mean square error of snow fall with optimization has been found smaller than that without optimization for both the days. It has also been found smaller than the persistence forecast and standard deviation of the snow fall for both the days. The RMSE of snowfall computed through leave-one-out cross validation method for different stations vary from 8.9 to 16 cm for day-1 and 9.4 to 16.5 for day-2. It lies between 10 to 15 cm for six stations, 5–10 cm for three stations and >15 cm for one station. The comparison of RMSE of the model and persistence forecast reveal that the model performance with optimization has been found better than the persistence forecast. The HSS for all the stations vary from 0.19 to 0.37 for day-1 and 0.06 to 0.16 for day-2 implying that the model forecast is better than the random forecast. The overall performance of the model in terms of percent correct (PC) vary from 60.3% to 79.6% for day-1 and 15.9% to 38.2% for day-2. The model predicted cumulative quantitative snowfall during the hazardous snow storms in different parts of Indian Himalaya with fairly good accuracy. The model forecast validation show that the model performance for all the stations over Himalaya is reasonably good for day-1 and can be used reliably in other applications such as avalanche forecasting.

The forecast using HMM in this study is based on fixed model parameters (probability matrices) derived from the database. The forecast can be made more realistic by making transition probabilities vary with time and forming a Non-homogeneous Hidden Markov Model (NHMM). The NHMM can be developed by incorporating large scale synoptic weather conditions into the HMM for more accurate long-term forecast and can be done in future.

Acknowledgements

The authors acknowledge Director, SASE for facilitating and encouraging this work. Dr Hemendra Gosain is duly acknowledged for his valuable suggestions to improve the paper. Field data collection teams of Snow and Avalanche Study Establishment are acknowledged for collecting data in adverse climatic conditions of the Himalaya during winter. The authors are thankful to the reviewers for their valuable suggestions and positive criticism, which helped to improve the quality of this paper.

Appendix (Categorical forecast verification of six category event)

In table A1, the total number of observed events in category-1 are given by:

$$A1 = C11 + C12 + C13 + C14 + C15 + C16.$$

The total number of forecasted events in category-1 are given by:

$$B1 = C11 + C21 + C31 + C41 + C51 + C61.$$

In a similar fashion, total number of observed and forecasted events in all the six categories are calculated. The total number of events are given by:

$$\begin{aligned} T &= C11 + C12 + C13 + C14 + C15 + C16 \\ &= C11 + C21 + C31 + C41 + C51 + C61. \end{aligned}$$

The forecast verification measures, derived from table A1 are:

Percentage correct (PC)

$$PC = ((C11 + C22 + C33 + C44 + C55 + C66) \times 100/T)\%.$$

Critical success index (CSI) in different categories

$$CSI-1 = \frac{C11}{(A1 + B1 - C11)}$$

$$CSI-2 = C22/(A2 + B2 - C22)$$

Table A1. 6 × 6 contingency table of observed vs. forecasted events.

Observed	Forecasted						Total
	Category-1	Category-2	Category-3	Category-4	Category-5	Category-6	
Category-1	C11	C12	C13	C14	C15	C16	A1
Category-2	C21	C22	C23	C24	C25	C26	A2
Category-3	C31	C32	C33	C34	C35	C36	A3
Category-4	C41	C42	C43	C44	C45	C46	A4
Category-5	C51	C52	C53	C54	C55	C56	A5
Category-6	C61	C62	C63	C64	C65	C66	A6
Total	B1	B2	B3	B4	B5	B6	T

$$CSI-3 = C33 / (A3 + B3 - C33)$$

$$CSI-4 = C44 / (A4 + B4 - C44)$$

$$CSI-5 = C55 / (A5 + B5 - C55)$$

$$CSI-6 = C66 / (A6 + B6 - C66)$$

Heidke skill score (HSS)

$$HSS = ((C11 + C22 + C33 + C44 + C55 + C66) - (A \times B1 + A2 \times B2 + A3 \times B3 + A4 \times B4 + A5 \times B5 + A6 \times B6) / T) / (T - (A1 \times B1 + A2 \times B2 + A3 \times B3 + A4 \times B4 + A5 \times B5 + A6 \times B6) / T).$$

References

Bardossy A and Plate E 1991 Modeling daily rainfall using semi-Markov representation of circulation pattern occurrence; *J. Hydrol.* **122** 33–47.

Bardossy A and Plate E 1992 Space–time model for daily rainfall using atmospheric circulation patterns; *Water Resour. Res.* **28(4)** 1247–1260.

Dimri A P, Niyogi D, Barros A P, Ridley J, Mohanty U C, Yasunari T and Sikka D R 2015 Western disturbances: A review; *Rev. Geophys.* **53**, doi: [10.1002/2014RG000460](https://doi.org/10.1002/2014RG000460).

Dimri A P, Joshi P and Ganju A 2008 Precipitation forecast over western Himalaya using k-Nearest Neighbor Method; *Int. J. Climatol.* **28** 1921–1931.

Flecher C, Naveau P, Allard D and Brisson N 2010 A stochastic daily weather generator for skewed data; *Water Resour. Res.* **46** W07519.

Gabriel K R and Neumann J 1962 A Markov chain model for daily rainfall occurrences at Tel-Aviv; *Quart. J. Roy. Meteorol. Soc.* **88** 85–90.

Gerlitz L, Conrad O and Bohner J 2015 Large-scale atmospheric forcing and topographic modification of precipitation rates over high Asia – a neural-network-based approach; *Earth Syst. Dyn.* **6** 61–81.

Guanche Y, Minguez R and Mendez F 2013 Climate-based Monte Carlo simulation of trivariate sea states; *Coast. Eng.* **80** 107–121.

Gusain H S, Chand D, Thakur N, Singh A and Ganju A 2009 Snow avalanche climatology of Indian western Himalaya; *International symposium on snow and avalanches (ISSA)*, 6–10 April, Manali, India.

Haberlandt U, Belli A and Bardossy A 2014 Statistical downscaling of precipitation using a stochastic rainfall model conditioned on circulation patterns – an evaluation of assumptions; *Int. J. Climatol.* **35(3)** 417–432.

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 1994 Incorporating spatial dependence and atmospheric data in a model of precipitation; *J. Appl. Meteorol.* **33** 1503–1515.

Katz R 1977 Precipitation as a chain-dependent process; *J. Appl. Meteorol.* **16** 671–676.

Kolokotroni M, Ren X, Davies M and Mavrogianni A 2012 London’s urban heat island: Impact on current and future energy consumption in office buildings; *Energy and Buildings* **47** 302–311.

Launay M, Brisson N, Beaudoin N and Mary B 2009 Conceptual basis, formalizations and parameterization of the STICS crop model; *Editions Quae.* **3(7)** 949–961.

Lang T J and Barros A P 2004 Winter storms in central Himalayas; *J. Meteorol. Soc. Japan* **82(2)** 829–844.

Maraun D, Wetterhall F, Ireson A, Chandler R, Kendon E, Widmann M, Brienen S, Rust H, Sauter T and Themel M 2010 Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user; *Rev. Geophys.* **48(3)**.

Mares C, Mares I, Huebener H, Mihailescu M, Cubasch U and Stanciu P 2003 A Hidden Markov Model applied to the daily spring precipitation over the Danube Basin; Technical report UCI-ICS 03-27, Information and Computer Science, University of California.

Menegoz M, Gallee H and Jacobi H W 2013 Precipitation and snow cover in the Himalaya: From reanalysis to regional climate simulations; *Hydrol. Earth Syst. Sci.* **17** 3921–3936.

Mohanty U C and Dimri A P 2004 Location specific prediction of probability of occurrence and quantity of precipitation over western Himalayas; *Wea. Forecast.* **19(3)** 520–533.

Richardson C 1981 Stochastic simulation of daily precipitation, temperature, and solar radiation; *Water Resour. Res.* **17(1)** 182–190.

Rind D, Goldberg R and Ruedy R 1989 Change in climate variability in the 21st century; *Climate Change* **14** 5–37.

Revenga C, Nackoney J, Hoshino E, Kura Y and Maidens J 2003 Watersheds of the world CD; IUCN, IWMI, Ramsar, World Resources Institute, Washington.

Schiemann R, Luthi D and Schar C 2009 Seasonality and interannual variability of the westerly in the Tibetan Plateau region; *J. Climate* **22** 2940–2957.

Singh D, Dimri A P and Ganju A 2008 An analogue method for simultaneous prediction of surface weather parameters at a specific location in the western Himalaya in India; *Meteorol. Appl.* **15(4)** 491–496.

- Srikanthan R and McMahon T 2001 Stochastic generation of annual, monthly and daily climate data: A review; *Hydrol. Earth Syst. Sci. Discuss.* **5(4)** 653–670.
- Todorovic P and Woolhiser D A 1975 A stochastic model of n-day precipitation; *J. Appl. Meteorol.* **14** 17–24.
- Wilks D and Wilby R 1999 The weather generation game: A review of stochastic weather models; *Progr. Phys. Geogr.* **23(3)** 329–357.
- Wilks D 2010 Use of stochastic weather generators for precipitation downscaling; *Wiley Interdisciplinary Reviews: Climate Change* **1(6)** 898–907.
- Wilks D 2012 Stochastic weather generators for climate-change downscaling, part II: Multivariable and spatially coherent multisite downscaling; *Wiley Interdisciplinary Reviews: Climate Change* **3(3)** 267–278.
- Wilson L, Lettenmaier D and Skillingstad E 1992 A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation; *J. Geophys. Res.* **97(D3)** 2791–2809.
- Zorita E, Hughes J P, Lettenmaier D P and Von Storch H 1995 Stochastic characterization of regional circulation pattern for climate model diagnosis and estimation of local precipitation; *J. Climate* **8** 1023–1042.

MS received 4 May 2016; revised 11 August 2016; accepted 30 December 2016

Corresponding editor: ASHOK KARUMURI