https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

GDHPM: A Geostatistical Disaggregation approach for generating hourly Precipitation in Mountainous regions preserving complex temporal patterns

Nibedita Samal [1], Meenakshi KV [1], Akshay Singhal [1], Sanjeev K. Jha [1], and Fabio Oriani [2]

[1]Indian Institute of Science Education and Research Bhopal, Madhya Pradesh, India
[2]Agroscope, Division Agroecology and Environment, 8046 Zürich, Switzerland

Correspondence: Sanjeev K. Jha (sanjeevj@iiserb.ac.in)

Abstract. Accurate precipitation estimation with high temporal resolution is crucial to monitor and predict natural hazards in mountainous regions. While rain gauges are the reliable source of precipitation data, they lack continuous fine resolution at desired locations such as avalanche and landslide sites. In this context, temporal disaggregation approaches can be used to obtain continuous hourly precipitation time series that account for the issues observed in mountainous regions, such as, (i) filling gaps in the data, (ii) capturing fine resolution statistical properties using available nearest station record, and (iii) availing longer historical records for better hindcasting. Multiple Point Geostatistics (MPS) approaches are known to mimic spatial patterns from the observed physical reality. This study introduces GDHPM, a temporal disaggregation approach, that investigates the possibility of MPS to search and generate complex temporal patterns. The objective is to simulate hourly precipitation time series from observed daily precipitation at multiple avalanche sites. Moreover, combinations of auxiliary time series from different locations and in varying numbers are tested as covariates. The results reveal that GDHPM produces hourly precipitation ensembles of realistic time patterns over a complex and extensive mountain terrain to improve avalanche and landslide forecasting.

1 Introduction

Precipitation estimates are arguably the most essential element in hydrological studies. The diurnal temporal variability, diverse spatial patterns, and varying intensities of precipitation play a pivotal role in shaping the hydrological dynamics of a region. Hence, understanding precipitation intricacies with fine temporal resolution can advance the understanding and preparedness for natural disasters. The repetitive occurrence of natural disasters such as floods and avalanches causes extensive damage to infrastructures, affects local transport, and causes substantial loss of life (Gardner et al., 2002; Gardner and Saczuk, 2004). The increasing temperature, specific humidity, and water vapour availability has increased frequency and intensity of extreme precipitation events (Kumar et al., 2018; Mayowa et al., 2015; jun Wang et al., 2016). For example, in past few decades, the Indian Himalayan regions have experienced increased extreme weather events and variability in precipitation (Dash et al., 2007). According to Karuna Sagar et al. (2017), during 1951-2015 the number of storm days have increased from 4 to 6 to 12

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

to 27 rainstorm days per year in India. Increase in rainstorm events can create added risk of large-scale flood. Hence, ensuring the availability of fine-resolution precipitation data can be helpful to disaster management authorities in mitigating risks.

The historical way of precipitation measurement is mainly by rain gauges or weather stations. Rain gauge or weather station measurements have limitations, such as availability only at specific points and inadequate spatial coverage (Margulis and Entekhabi, 2001). In regions with extreme weather conditions, the rain gauges are sparse and frequent sampling is difficult. For instance, for better monitoring and predictions in mountainous regions, building and maintaining weather stations at adequate number of places and sampling them frequently is difficult due to rugged terrain and inhabitable weather conditions (Jing et al., 2016). The gridded datasets available from different numerical weather prediction models are available for larger spatial and temporal extent. However, the model generated datasets are biased and need pre-processing before using for monitoring and prediction of hazards (Depina et al., 2020; Samal and Jha, 2022). For monitoring hazards such as avalanche and landslides, high temporal resolution data is required at specific locations for which spatial averaging techniques are used on both gauge and gridded data. However, spatially averaged datasets can contain large biases due to under sampling of spatial fields (Prein and Gobiet, 2017; Schneider et al., 2014). Hence, statistical temporal disaggregation approaches can be helpful in producing fine temporal resolution datasets from available coarse temporal resolution datasets.

There were significant developments in the field of temporal disaggregation from the 1970s where (Valencia and Schakke Jr, 1973) has introduced linear disaggregation model. Linear disaggregation has been widely used in the field of temporal down-scaling to produce fine scale weather data (Firor et al., 1996; Maheepala and Perera, 1996; Obeysekera and Salas, 1986). After linear disaggregation many researchers have started improving the models for simulating finer scale data using two stage method (Guenni and Bárdossy, 2002), stepwise or cascade disaggregation (Santos and Salas, 1992), disaggregation with near-est neighbour resampling techniques (Wójcik and Buishand, 2003), disaggregation using k-nearest neighbour technique (Sharif et al., 2007), non-parametric disaggregation (Lee and Jeong, 2014; Srikanthan et al., 2006) etc. Following these developments in disaggregation approaches, researchers have started to produce fine temporal resolution precipitation data using annual or monthly precipitation (Connolly et al., 1998; Koutsoyiannis, 1992; Papalexiou et al., 2018; Pui et al., 2012; Thober et al., 2014). Recent studies were aimed at inferring fine resolution hourly meteorological estimates using statistical approaches (Acharya et al., 2022). Such statistical solutions are helpful in producing fine resolution data from coarse resolution data with less in-vestment of time and computational power (Dash et al., 2007; Koutsoyiannis, 1992; Lee and Jeong, 2014). Further, multiple hydro-meteorological variables are simulated together in multivariate disaggregation approaches to produce fine temporal scale weather data (Debele et al., 2007; Firor et al., 1996; Sharif et al., 2013). A majority of the statistical rainfall disaggregation models are implemented on a single site, which limits the rainfall property and makes it spatially noncoherent. This means that the spatial correlation and rainfall patterns observed in multi-site rainfall data is not preserved in the temporal disaggregation. To overcome this limitation, multi-site disaggregation approaches have started simulating finer temporal resolution precipita-tion data for multiple sites together (Kumar et al., 2000; Lu and Qin, 2014). Recently Acharya et al. (2022) have used a simple temporal disaggregation approach to produce hourly rainfall using daily rainfall at multiple rain gauge sites of Australia, which could be helpful in hydrological modelling.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

Multiple Point Statistics (MPS) is widely known for its spatial data simulation and it focuses on data pattern reproduction. Mariethoz et al. (2010) has used direct sampling (DS) algorithm that is suitable for multiple sites and grid data simulation together. Meerschman et al. (2013) have shown different usage of MPS, including 3D simulation, post-processing, multivariate simulation, and data conditioning. MPS has also been used to simulate spatial patterns of multiple meteorological variables for downscaling and extrapolation of the datasets (Jha et al., 2015; Singhal et al., 2022). Singhal and Jha (2021) applied the DS technique to spatially downscale and demonstrated its effectiveness in generating reliable fine-resolution weather information. Recently, the DS technique has been used in the simulation of rainfall time series from different climatic regions (Oriani et al., 2014) and for stream-flow gap filling (Dembélé et al., 2019) With respect to traditional nearest-neighbor resampling techniques, DS applies a variable-size pattern dependence. This methodological strong point allows preserving the statistical structure of rainfall at multiple scales and in the long term, which is critical for seasonal-variability impact studies and multi-scale modelling.

The aim of this study is to generate stochastic ensembles of hourly precipitation data at single as well as multiple avalanche sites of the northwest Himalayan region using the DS approach. The study is novel since, this is the first time, MPS has been used to temporally disaggregate daily precipitation to simulate hourly precipitation time series. Also, to the best of our knowledge, no study has undertaken statistical temporal disaggregation of precipitation from the complex weather of the Himalayan region. From the presented tests, we identify a standard setup of covariates and parameter values. In particular the goals of the study are: (a) to set up a statistical temporal disaggregation model based on the DS algorithm for producing hourly precipitation using hindcast daily and hourly precipitation at multiple avalanche sites; (b) to finalize a fixed parameter set for DS for temporal disaggregation of daily precipitation; (c) to select one or more neighbouring site precipitation as appropriate covariates of the model using statistical approaches – nearest neighbour, correlation coefficient and complex networks; and (d) to explore appropriate temporal moving average window for improving model performance.

In this paper, we address the aforementioned four objectives by introducing GDHPM, a geo-statistical temporal disaggregation approach for generating hourly precipitation in Mountainous region preserving complex temporal patterns. The GDHPM formulated here produces hourly precipitation, but can be adapted to other time resolution easily. The remainder of the paper is structured as follows. Section 2 describes the details related to the study area and datasets. Section 3 describes the methodology implemented for temporal disaggregation of precipitation using MPS. Section 4 presents the results of GDHPM along with their evaluation. Section 5 discusses the results, and Section 6 deals with conclusions and the future scope of this study.

2 Study Area and Datasets

2.1 Study Area

The Indian Himalayan Region (IHR) witnesses multiple avalanche events in most parts of the year. Many districts of Himachal Pradesh see such events very often. Based on the past 20 years of historical records of avalanche events in the area, 20 avalanche sites are identified in the districts of Kullu and Lahul Spiti and taken as the study locations. All these sites are located between 31˚N to33˚N latitude and 75˚E to 81˚E longitude. Figure 1(a) shows the study area with details of the locations in district

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

boundaries of Himachal Pradesh. The reasons for selecting the study area are: (1) in past few decades the IHR has experienced heavy to extreme precipitations frequently during monsoon season i.e., June, July, August and September (JJAS); (2) due to the complex topography several areas are vulnerable to landslides, avalanches and flash floods which cause large destruction of life and property; (3) the hourly meteorological datasets currently available in IHR are insufficient for monitoring or modelling extreme weather phenomena. The availability of a reliable hourly precipitation hindcast could be crucial in mitigating the losses.

2.2 Datasets

In this study, we use the High Asia Refined analysis version 2 (HAR-v2) data at daily and hourly time scales. The HAR-v2 data is extracted at the centre of the grid nearest to 20 selected avalanche sites of two districts of Himachal Pradesh (Kullu and Lahaul & Spiti). The HAR-v2 is a publicly available dataset obtained by running a Weather Research and Forecasting model. The model takes ERA5 data from European centre for medium-range weather forecasts (ECMWF) as forcing data to produce multiple meteorological variables for a time span of 1980 to 2020 (Wang et al., 2021). The gridded dataset is available in monthly, daily, and hourly temporal scales for northern India and the Tibetan region at a spatial resolution of 10×10 km^2. Figure 1(b) demonstrates the characteristics of a hourly and daily precipitation time series obtained from HAR-v2 for a specific avalanche site of Himalayas. The difference in variability and the unpredictable fluctuation of hourly time series can be observed in the Figure 1(b).

3 Methodology

3.1 Workflow of the Direct Sampling (DS) Algorithm

DS is one of the many algorithms developed under the family of MPS. DS follows the same rationale of the MPS which is to search for the required fine-scale information from a large training database. This training database is called Training Data (TD) and consists of both the coarse and fine-scale data. With reference to a data pattern called Conditioning Data (CD), a similar pattern is searched in the TD. Upon successful search, the fine-scale found pattern is transferred to the target location in a separate database called the simulation grid (SG). In this application, prior to the run, SG is populated with coarse-scale precipitation data. SG also includes auxiliary data, which supports the accurate searching of information from the TD. More details about TD, CD, and SG are provided in Section 3.2. The working principle of DS is explained following the previous studies (Mariethoz et al., 2010; Meerschman et al., 2013; Oriani et al., 2016).

Here, we present the working principle of DS in the context of GDHPM of a continuous variable, precipitation. Figure 2 shows the workflow of the DS algorithm for this study. Figure 2 includes graphical presentation of the inputs required for DS i.e., (i) a Simulation Grid (SG), a time vector hosting the target variable to be simulated, (ii) training data (TD), target variable and auxiliary variable at informed time step, from which the unknown information is sampled. In both TD and SG, the time steps are identical and uniformly spaced. Target variable simulation is carried out in non-consecutive time steps along a random

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

path. Gradually, SG is filled producing the actual output of the simulation. We denote the SG as X = X_1,……,X_n and TD as Y = Y_1,……,Y_m.For an easier comparison between patterns, the variables in X and Y are linearly normalized to a range of [0, 1] using the transformation $Z \rightarrow Z.(max(Z) - min(Z))^{-1}$.In a random order, a uninformed X_t is visited and the algorithm executes the following steps:

1. The data event d (X_t) = (Z(X_{t+h1}),……, Z(X_{t+hn})) is retrieved from the SG by selecting a fixed neighbourhood of radius R centered on X_t. The size of X_t is provided by the user defined parameter N. It consists of at most N informed time steps, closest to X_t. This defines a set of lags L = l_1,l_2,…l_n with $|l_i| \leq$ R and n $\leq$ N. For example, if R = 50 and N = 10, the pattern is composed by the 10 informed time steps closest to t inside the time span $\pm$t[50]. The size of d(X_t) is therefore limited by N and the available informed time steps inside the search neighbourhood window of R2.

2. A random time-step Y_i in TD is visited and the corresponding data event d(Y_i), defined according to the same L, is retrieved to be compared with d(X_t).

3. A distance D(d(X_t), d(Y_i)), i.e. a measure of dissimilarity between the two data events, is calculated. For continuous variables, it is given by the formula

$$D(d(X_t), d(Y_i)) = \frac{1}{n}\sum_{j=1}^{n} |z(x_j) - z(y_j)| \quad (1)$$

where n is the number of elements of the data event. Independently from their position, the elements of d(X_t) play an equivalent role in conditioning the simulation of Z(X_t). The normalization at step 1 ensures the distances to be in the range [0, 1].

4. If D(d(X_t), d(Y_i)) is below a defined threshold T [0, 1], meaning that the two data events are sufficiently similar, the iteration stops and the datum Z(Y_i) is assigned to Z(X_t) for all uninformed variables. Otherwise, the procedure is repeated from step 1 to 4 until a suitable d(Y_i) is found or a prescribed TI fraction F is scanned.

5. In case no time step corresponding to D(d(X_t), d(Y_i)) is found, the datum Z(Y_i^*) minimizing this distance is assigned to Z(Xt).

6. This procedure is iterated for the simulation at each X_t until the entire SG is completely informed. The variables are then linearly back transformed to their original range.

Figure 2 illustrates the above procedure including two time series at both coarse and fine scale. Here, a X_{t1} data event is selected randomly from the SG. The distance D is calculated between the random data events of Y_t. When (D < T), conveying that the dissimilarity between the two events are lowest between X_{t1} and Y_{t2} (matching data event), Y_{t2} is assigned to the SG in place of X_{t1}. If the patterns do not match, and the search window moves to another random location and repeats the process. All other unknown values (X_{t2} and X_{t3}) in the SG are simulated using the same process. This procedure is repeated for all the data events of SG and the fine scale product is simulated. In the context of this study, the unknown values denote the hourly precipitation data of the target year sampled from the TD consisting of both daily and hourly data of previous years. The details of all the fixed and varied parameters used to set up the model for this study are presented in Table 1 and Table 2. The fixed

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

parameters are used following Meerschman et al. (2013) and the varied parameters are finalised after conducting sensitivity study.

3.2 Arrangement of training and conditioning data

The DS algorithm requires the input datasets in the form of training data (TD) and conditioning data (CD). The TD is used to train the model using DS algorithm and CD is used as a reference to generate the simulations. The splitting of both the coarse and fine temporal resolution data into TD and CD is done in ratio of 95:5. In this study, as we have taken 41 years of data (1980-2020), the first 39 years of data (1980-2018) is used for training the model with both coarse and fine resolution data, and the next 2 years (2019-2020) of coarse resolution data is used as CD. The fine resolution data for the 2 years is used as target data to validate the model simulations. The DS algorithm gives the option to put any year of data as the target year by choosing the year in the CD, which means we can generate a simulation for any year from the available time period. We have chosen the most recent two years as the simulation period out of the whole period of 1980 to 2020, as with the DS algorithm, it is possible to rely on the historical datasets to predict the recent past events. Further, we believe that 2 years validation period is enough to represent the hourly precipitation simulations. Also, we have explored statistics like annual maxima, minima, extremes, and made sure that the two years are neither exceptionally wet nor dry and can represent other years. Furthermore, the fine and coarse resolution data including P_hourly and P_daily are arranged together in TD and CD by replicating P_daily 24 times. As P_daily contains a repetition of same values 24 times, a smoothening of P_daily time series was needed. We applied the moving mean technique to every 3-hour- and 6-hour P_daily values using a sliding window across neighbouring daily values.

3.3 Statistical identification of covariates

Covariates are an integral part of the DS simulations as they assist in matching of patterns. Hence, identification and selection of suitable covariates become important. In this study, we identify the most relevant neighbouring sites and use the corresponding precipitation data as covariates. The covariates are added in the training and conditioning datasets to support the simulations. In this study, we identify the neighbouring sites using three statistical techniques (Nearest neighbour, Pearson correlation coefficient, and the theory of complex networks). A brief description of the techniques is given below-

3.3.1 The nearest neighbour technique

A nearest neighbour is identified by calculating the smallest Euclidean distance between the sites. The Euclidean distance is calculated in two-dimensional space by using the latitude and longitude of all the 20 sites taken in this study. Given sites i = 1, 2, 3,. . . , n, the Euclidean distance is calculated as follows,

$$d_E = \sqrt{\sum_{i=1}^{n} (x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \tag{2}$$

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

Where x_i = latitude and y_i = longitude of each site. The site with smallest dE to (x_i, y_i) is selected as the nearest neighbour and taken as covariate in the training and conditioning dataset. The nearest neighbour technique is chosen for covariate selection to test whether the closest site precipitation information is helpful for the model in the mountainous region.

3.3.2 The Pearson Correlation Coefficient:

The Pearson correlation coefficient (PCC) between two variables, X and Y, is calculated as the ratio of their covariance to the product of their standard deviations (Asadollah et al., 2021; Zhou et al., 2016). PCC is a measure of the strength of the relationship between two variables. The following simulation shows the mathematical expression of PCC:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\sum_{i=1}^{n}(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (3)$$

PCC ranges from -1 to 1, where a value of -1 indicates a perfect negative correlation, a value of 0 indicates no correlation, and a value of 1 indicates a perfect positive correlation. In this study, PCC is calculated using the daily precipitation data of 20 sites and the sites with highest PCC values are taken as covariates in the training and conditioning datasets.

3.3.3 The theory of complex networks

The theory of complex networks is a widely used concept to study the spatio-temporal connections in precipitation data. The concept of complex network considers a system as a graph, which is a set of points (nodes) connected together by a set of lines (links). In our case, the selected avalanche sites are the nodes and the connection between them act as the links as shown in Figure 1(c). There are various measures of complex networks such as the degree centrality, clustering coefficient, betweenness centrality etc. which are used to evaluate the connection in precipitation. Previous studies provide the detail description about the applications of complex networks in hydrology (Jha and Sivakumar, 2017; Singhal et al., 2024; Sivakumar, 2014). In this study, we use the network measure of clustering coefficient as it serves the purpose of identifying connections.

The clustering coefficient (CC) is a measure of the local density of a network (assigns a score for each node) and quantifies the network's tendency to cluster. Each site's data i is considered a node of a complex network, so the 20 sites are considered as 20 nodes, and the connection between them acts as links. The connection between each node (avalanche site) is determined based on a certain correlation threshold (CT). The links to the nodes that have correlations exceeding CT are considered to be connected. Consider the node i in the network connects to k_i other nodes, where these nodes are neighbours to node i. If these neighbours i are considered as part of a cluster, finding the number of possible links existing E_i,

$$E_i = \frac{k_i(k_i - 1)}{2} \quad (4)$$

and the clustering coefficient

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

$$CC_i = \frac{2E_i}{k_i(k_i-1)} \tag{5}$$

Here, E_i is the actual number of links which exist between the k_i neighbours of node i and k_i $(k_i-1)/2$ are the total number of possible links which may exist among the k_i neighbours of node i. The neighbouring nodes of a given node i are considered to be well connected if the CC values are higher (close to 1) and less connected if the CC values are low (close to 0).

3.4 Experimental design

The basic aim of the experiments performed in GDHPM is to obtain reasonably reliable and accurate hourly precipitation data at the selected sites over the Himalayan region. To this end, we explore the different combinations of the number of covariates, method of their selection, and the two temporal windows of the moving mean. We performed three sets of experiments in GDHPM to generate future hourly precipitation data using the hindcast daily and hourly precipitation data over the IHR. The experiments are based on the number of sites used as covariates. The arrangement of TD and CD data and the workflow of this study is presented through the flowchart in Figure 3. The experiments performed in this study are based on selection of covariates using different techniques.

Under each experiment (E1, E2, ...), multiple experimental runs (R1, R2, ...) are undertaken as presented in the Table 3. The different runs are based on the moving-mean method applied on P_daily data. For example, experiment_1 consists of two runs which implies that the first run is denoted as E1_R1 and the second run is denoted as E1_R2. In each run, 39 years of training (P_daily and P_hourly) and conditioning data (P_daily) of the target years, i.e., 2019 and 2020 is given as input data to the statistical model. The DS disaggregates the input data to generate P_hourly for the target year in each run. The generated P_hourly data from the model is verified with the available P_hourly data of the target year which is not used as the input data. As shown in Table 3, the different combinations of covariates are tested systematically, considering different conditions: number of covariates (none, single, or multiple), covariate selection criteria (station proximity, data correlation, or data clustering), and size of moving-average window applied (3-day or 6-day).

A set of error metrics are employed to measure the accuracy and the bias of the estimated data with respect to the reference: Bias percentage (BIAS%) , Root-mean-square error (RMSE) and mean-absolute error (MAE).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y(i)-X(i))^2} \tag{6}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|Y(i)-X(i)| \tag{7}$$

$$\text{BIAS\%} = \frac{\sum_{i=1}^{n}(X(i)-Y(i))}{\sum_{i=1}^{n}Y(i)} \times 100 \tag{8}$$

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

4 Results

In this section, we evaluate the hourly precipitation data generated in GDHPM using the hindcast P_hourly and P_daily data through time series plots and error statistics. For each simulation, we generated 50 realizations taking into account the time pattern variability. The generated hourly time series ensemble is visually compared with the reference HAR P_hourly, and evaluated through error statistics.

4.1 GDHPM using no covariate (Experiment 1)

4.1.1 Visual comparison of the timeseries plots

Here, we evaluate the time series plots of the statistically simulated P_hourly with 50 ensembles; obtained using P_hourly and P_daily in the TD and CD for the monsoon months of 2019 and 2020. The results are presented at hourly scale for the periods where ample rain is observed in the year 2019. All the time series plots are presented for the time period of 15th to 30th June, 15th to 31st July, and 1st to 15th August and for random selected sites 1, 6, 11 and 16. The effect of using different moving average technique on daily precipitation is insignificant on the P_hourly produced. Hence, we are presenting time series from 6-hourly moving average in the supplementary file.

The time series plots (Figure 4 and S1) presents the P_hourly from experiment 1 generated using no covariates. The experiment includes two runs, the first run (E1_R1) uses a 3-hr moving average technique and the second run (E1_R2) uses a 6-hr moving average technique to smoothen the P_daily in the TD and CD for better simulation from the model. Figure 4 and Figure S1 presents the results in which each subplot includes P_hourly reference data from HAR, model simulated ensembles of P_hourly, and the mean of ensembles P_hourly.

Using 3-hr moving average

Figure 4 shows the results from E1_R1 where the temporal fluctuations of mean ensembles of model simulated P_hourly are matching with available reference P_hourly at all the four avalanche sites. The mean of the ensembles plotted above the reference P_hourly clearly shows the matching simulation timings of the precipitation peaks. The mean ensemble precipitation smoothens out the high and low peaks in the simulations and hence shows a lower local maximum. Conversely, the ensemble envelope covers the maximum P_hourly at most of the precipitation peaks in all four avalanche sites. Both the timing and the local variability of the events are overall preserved inside the simulations. The occurrence of the simulated peaks varies approximately of ± 3 hours. This is part of the uncertainty which is reduced by the moving mean covariates. At site_1 and site_16 around 200 hrs in august, at site_6 around 150 hrs in July, at site_11 after 200 hrs in July, DS has simulated precipitation without corresponding reference values as daily data. This shows model uncertainty in identifying the events while simulating P_hourly.

Using 6-hr moving average

Figure S1 presents results from run E1_R2 where mean of the ensembles of model simulated P_hourly is shown with the reference P_hourly. The temporal fluctuations of the simulations are matching with the reference P_hourly in the month of June in all four avalanche sites. At all four sites, the model simulated ensembles are covering the precipitation peaks of reference

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

P_hourly. However, an overestimation of the P_hourly values are seen in most places. The mean of the ensemble P_hourly has missed the reference P_hourly peaks at many places e.g., around 200 hrs in the month July (column 2 of Figure S1). After 200 hrs, false simulation of precipitation peaks is also seen in the P_hourly ensembles.

A comparison between Figure 4 and Figure S1 suggests that the temporal fluctuations are more distinct and similar to reference P_hourly in Figure 4 than in Figure S1. The results of site_1 show that the ensemble mean matches the maximum P_hourly values better in Figure 4. Also, overestimation of the maximum P_hourly is observed at all four sites in Figure S1. Hence, visual inspection of single site simulation depicts that, results from E1_R1 is producing more accurate results at most of the sites.

4.1.2 Evaluation of results using the error statistics

Here, we present a quantitative assessment of the accuracy of the model simulated results using deterministic error statistics. The assessment is done using ensemble P_hourly from MPS and P_hourly from HAR for the simulated period at the 20 avalanche sites. The results are presented using boxplots for each error statistics (RMSE, MAE, and BIAS %). Figure 5 presents a comparison of error statistics obtained from E1_R1 and E1_R2. The mean error statistics of the simulated period are calculated between each of the 50 realizations obtained for model simulated P_hourly and HAR P_hourly. The RMSE values shown in Figure 5 are fairly low ranging for E_R1 between 0.16 (site_20) and 0.41 (site_2) and for E1_R2 between 0.16 (site_20) and 0.42 (site_1). Similarly, the MAE values shown in Figure 5 are fairly low at all avalanche sites in both the runs, in the 0.03 to 0.12 range for both runs. The percent bias values obtained from model simulated P_hourly values seem to be low at all the avalanche sites. The maximum and minimum bias % values obtained from E1_R1 is 86 at site_1 and 0.4 at site_5 respectively. The maximum and minimum bias % values obtained from E1_R2 is 88 at site_10 and -0.16 at site_2 respectively. The median bias values from E1_R2 are lesser at 65% of the avalanche sites than bias From E1_R1. The quantitative assessment of E1_R1 and E1_R2 suggests that the P_hourly obtained from 6-hourly moving mean has higher range of error than the P_hourly values obtained from the 3-hourly moving mean simulations. By taking MAE as a measure of accuracy, it is observed that 3-hourly moving average is producing more accurate values than 6-hourly moving average. By taking bias % as a measure of precision, it is observed that E1_R1 has produced more precise simulations with lesser range of bias in all the ensembles compared to E1_R2. Hence, Overall E1_R1 has produced more reliable simulations then E1_R2.

4.2 GDHPM with one-site covariate (Experiment 2)

4.2.1 Visual comparison of the timeseries plots

The experiment 2 with one-site covariate is divided into two sub-experiments producing run E2_R3, E2_R4, E2_R5, and E2_R6. E2_R3 and E2_R4 choose covariate using the nearest neighbour technique and E2_R5 and E2_R6 choose covariate using the correlation coefficient technique. Figure 6 and S2 presents time series plots of P_hourly simulated with MPS from E2_R3 and E2_R4. Figure 7 and S3 presents time series plots of P_hourly simulated with MPS from E2_R5 and E2_R6. These runs simulate hourly precipitation simultaneously for two avalanche sites. Each subplot in Figures include hourly reference

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

data from HAR, model simulated ensembles of the forecast, and the average mean of ensembles of model simulated hourly data.

Using 3-hr moving average with the nearest neighbour technique

Figure 6 shows model simulated P_hourly with ensembles and mean of ensembles in comparison to the reference P_hourly at the four avalanche sites using 3-hour moving average technique on HAR P_daily. The model simulated P_hourly is showing temporal fluctuations which are completely different from the reference P_hourly at all the sites. The magnitude of mean of ensemble has similar magnitude to the P_hourly peaks, but the temporal variability does not match with the reference P_hourly. The ensembles are also not being able to cover the maximum P_hourly values. Figure 6 shows that taking nearest site as covariate has not helped the MPS simulations of P_hourly values by finding out accurate space of events.

Using 6-hr moving average with the nearest neighbour technique

Figure S2 presents the model simulated P_hourly with ensembles from E2_R4 where 6-hour moving average is applied to the daily training and conditioning data. Figure S2 shows similar results as Figure 6 where the mean of the ensembles of model simulated P_hourly do not match with the temporal fluctuations of available reference P_hourly at all four avalanche sites. The ensembles are not fitting the places of P_hourly peaks from the reference P_hourly. The reason of this abrupt difference between the simulations and the reference could be due to inappropriate selection of the covariate in the model TD. As the Himalayas have undulating topography, the nearest site may not exhibit similar geographical properties and precipitation patterns. Hence, both the sites are not complementing model simulation to produce accurate results according to the reference P_hourly.

As taking covariates using nearest neighbour selection is not a successful attempt, we have taken the next covariate set using correlation coefficient. The correlation coefficient is calculated between the 20 avalanche sites using the precipitation data and the sites who have highest correlation coefficient (>0.95) are taken as covariates in simulation. Figure 7 and S3 presents time series plots from E2_R5 and E2_R6.

Using 3-hr moving average with the PCC technique

Figure 7 shows that at all four sites, the temporal variations of the mean of ensembles of the MPS simulated P_hourly is matching with available reference P_hourly. The temporal variability of the ensembles is precise with the variability of reference P_hourly and the simulations produced the peaks accurately. In Figure 7, all the ensembles produced the temporal variability and precipitation magnitude quite accurately with no false simulation of peaks and no overestimation (e.g., site_16). This suggests the correlated covariate provides a high predictive power to reduce the uncertainty in the simulations. Around 325 hrs in the month of August, maximum P_hourly is observed. This P_hourly could be simulated at site_11 and site_16, however at site_1 and site_6 the ensembles are not being able to simulate the values.

Using 6-hr moving average with the PCC technique

Figure S3 presents the model simulated P_hourly with ensembles from E2_R6 where mean of the ensembles of model simulated P_hourly is matching with the temporal variability of available reference P_hourly at all four avalanche sites. The magnitude of the mean of the P_hourly is lowered by averaging the lows and highs of P_hourly simulated in ensembles. The reference P_hourly peaks are covered by model simulated ensembles and are helpful in predicting the uncertainties in the

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

hourly precipitation. At site_1 and site_6, some P_hourly peaks are seen to be missing in the simulations and could not be covered by ensembles (around 325 hrs in the month of August).

A comparison between Figure 7 and Figure S3 suggests that the mean ensemble precipitation generated through both the runs (E2_R5 and E2_R6) are very similar to each other irrespective of the use of different moving average technique applied to the P_daily. The temporal fluctuations generated for P_hourly through E2_R5 produces mean ensemble precipitation more accurately than E2_R6. For example, in Figure 7, at site_1, around 160 hrs in July the mean ensemble precipitation shows a peak which matches the reference P_hourly more accurately than in Figure S3. If compared the results from E1_R1, the magnitude of mean of the ensembles have been simulated better in E2_R5. Compared to E1_R1, E2_R5 has produced more accurate mean of ensemble and produced ensembles with lesser spread. The model is producing hourly precipitation quite accurately using both 3-hourly and 6-hourly simulations after taking a site as covariate using correlation coefficient technique.

4.3 4.2.2 Evaluation of results using the error statistics

This section presents a quantitative assessment of results obtained from runs E2_R3 to E2_R6 using error statistics (RMSE, MAE, and BIAS %). Figure 8 and 9 presents boxplots showing mean error statistics of the simulation period calculated using MPS P_hourly obtained from 50 realizations and HAR P_hourly. Results from E2_R3 and E2_R4 with nearest site as covariate is presented in Figure 8, and E2_R5 and E2_R6 with most correlated site as covariate is presented in Figure 9. Figure 8 presents RMSE values at all avalanche sites that are significantly in both runs E2_R3 and E2_R4. The RMSE values obtained range in 0.17 to 0.39 for E2_R3 and 0.16 to 0.39 E2_R4 with a tendency to higher RMSE when using a 6-hourly moving average. Figure 8 presents MAE values that are fairly low at all avalanche sites and both the runs E2_R3 and E2_R4. The MAE values obtained range in 0.038 - 0.112 for E2_R3. The MAE values range in 0.029 - 0.110 for E2_R4. The fluctuations in MAE values at different avalanche sites is low, and the MAE values from both the runs are similar when compared at each site. Figure 8 presents bias percentage significantly low by both the runs at every avalanche sites. The bias% values obtained from E2_R3 range from -2.77 to 65.27. The bias% values obtained from E2_R4 range in 0.22 - 69.99. The quantitative assessment of E2_R3 and E2_R4 suggests that the P_hourly obtained from 3-hourly moving mean has lesser error values than the P_hourly values obtained from the 6-hourly moving mean simulations.

Figure 9 presents the error evaluation with most correlated site as covariate. From the Figure, we observe that the RMSE values at all avalanche sites are significantly in both runs E2_R5 and E2_R6. The RMSE values obtained for E2_R5 range 0.40 - 0.16. The RMSE obtained for E2_R6 range 0.16 – 0.39. A comparison of RMSE values from E2_R5 and E2_R6 shows that, at each site the values are of similar magnitude, however, the range of RMSE is higher in 6-hourly (E2_R6) at some sites such as site_1 and site_2. Figure 9 presents MAE values that are fairly low at all avalanche sites from both the runs (E2_R5 and E2_R6). The MAE values obtained from E2_R5 range 0.033 – 0.116. The MAE values obtained for E2_R6 range 0.117 - 0.030. The fluctuations in MAE values at different avalanche sites is low, and the MAE values from both the runs are similar when compared at each site. Figure 9 presents bias % values significantly low by both the runs at every avalanche sites. The bias values obtained from E2_R5 and E2_R6 is (-0.314 %, 1.537 %) and (78.63, 81.86) at site_14 and site_10 respectively. From the quantitative assessment of E2_R5 and E2_R6 the P_hourly obtained from 3-hourly moving mean presents a lower

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

error then the P_hourly values obtained from the 6-hourly moving mean simulations. The lower MAE and bias % values in E2_R5 compared to E2_R6 suggests that E2_R5 is producing more accurate and precise simulations. Compared to runs presented earlier, the E2_R5 and E2_R6 have shown the lowest values in all error statistics, suggesting that MPS-based model simulations with most correlated site as covariate has produced the highest quality simulations.

4.4 GDHPM with multi-site covariates (Experiment 3)

4.4.1 Identifying covariate-sites using clustering coefficient

Experiment 3 involves using multi-site covariates to generate the P_hourly precipitation data using the DS algorithm. To this end, we introduce the concept of network measure of clustering coefficient in identifying optimum covariates for the DS simulations. The idea here is to identify the avalanche sites which may be spatially connected to each other. The sites which are more spatially connected are likely to record similar precipitation data and hence, can be considered as covariates. For the calculation of CC, we generally consider seven correlation thresholds, viz. 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9, however, in the present case, results show that the CC values are prominent between the CTs 0.6 to 0.9. Figure 10 shows the results obtained from the application of clustering measure where different clusters obtained at CTs 0.6 to 0.9. The first step in calculating the CC is to identify an optimum CT which provides reasonable information about the characteristics of the network.

Results show that at CT=0.6, the network is reasonably strong as all the 20 avalanche sites belong to the highest CC (green circles). The network begins to show some variation at CT = 0.7 and more prominently at 0.8. At CT = 0.9, the network begins to disintegrate as all the CC values show different CC range, thus providing no distinct information. Hence, based on the inference, CT = 0.8 can be the optimum CT to deduce reasonable information about the characteristics of the network. At CT = 0.8, we identify three clusters of avalanche sites represented as green (CC = 0.9 to 1), blue (CC = 0.8 to 0.9), and pink circles (CC = 0.7 to 0.8). With CC= 0.9 to 1, three sites are forming a cluster, and a three-site MPS simulation is performed in E3_R7 and E3_R8. With CC= 0.6 to 0.8, 5 sites are forming a cluster, and a five-site MPS simulation is performed in E3_R9 and E3_R10. With CC= 0.8 to 0.9, 7 sites are forming a cluster, and a seven-site MPS simulation is performed in E3_R11 and E3_R12. The results obtained from the above experiments are discussed in the sections below.

4.5 Evaluation of results using the timeseries plots

Figures 11 and S4 presents time series plots from E3_R7 to E3_R12 obtained from experiment 3 using multiple site precipitation data as covariate in the model simulation. The experiment is run using precipitation data using 7 sites, 5 sites, and 3 sites precipitation data as covariates. The three groups are made by taking clustering threshold of 0.8 in clustering coefficient method and creating three groups based on correlation coefficient of 0.6 to 0.8, 0.8 to 0.9, and 0.9 to 1. In Figure 11 and S4, we are presenting P_hourly time series for the months of JJAS of 2019.

Using 3-hr moving average with clustering coefficient technique

Figure 11a presents a time series plot for site_6 using 7 sites as covariates where the model simulated P_hourly is matching with the reference P_hourly quite well. Figure 11b presents a time series of model simulated ensembles and mean of ensem-

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

bles for site 1 using 5 sites as covariate. The model simulated P_hourly in this run has produced fluctuations which are not synchronized with the reference P_hourly. Figure 11c presents time series of model simulations with ensembles using 3 sites precipitation as covariate for site 16. The temporal fluctuations produced through MPS is matching the reference hourly and the ensembles are capturing the P_hourly peaks accurately.

Using 6-hr moving average with clustering coefficient technique

Figure S4 shows results from runs E3_R8, E3_R10, and E3_R12 using 6-hour moving average technique on P_daily data. Figure S4a presents a time series plot for site_6 using 7 sites as covariates where the model simulated P_hourly is matching with the reference P_hourly quite well. Figure S4b presents a time series of model simulated ensembles and mean of ensembles for site_1 using 5 sites precipitation as covariates. The model simulated P_hourly in this run has produced temporal fluctuations matching the reference P_hourly. Figure S4c presents time series of model simulations with ensembles using 3 sites precipitation as covariates for site_16. The temporal fluctuations produced through model is matching the reference hourly and the ensembles are capturing the P_hourly peaks accurately. The E3_R7 has produced P_hourly which are simulating the lower and higher peaks efficiently.

4.6 Evaluation of results using the error statistics

Table 4 presents the error statistics from E3_R7 to E3_R12 where 7, 5, and 3 site precipitation data is used for P_hourly simulations. At site_6, where 7 sites are used as covariates, the 3-hour and 6-hour moving average technique is producing similar error statistics. The RMSE, MAE, and bias % values for E3_R7 and E3_R8 is (0.1651, 0.1822), (0.0321, 0.0347), and (-5.12, -5.09) respectively. Considering the closer value of RMSE, based on MAE and bias value, it is said that E3_R7 is performing better than E3_R8. At site_1, where 5 sites are used as covariates, the RMSE and MAE values for both E3_R9 and E3_R10 are 0.217 and 0.040 respectively. The bias % in E3_R10 (-2.84 %) is producing better results than E3_R9 (-13.08%). At site_16, where 3 sites are used as covariates, the RMSE, MAE, and bias % values for E3_R11 and E3_R12 are (0.1367, 0.1282), (0.0205, 0.0209), and (-11.12, 72.43) respectively. Hence, considering overall error statistics in E3_R7 and E3_R11, we can say that 3-hour moving mean is performing better than 6-hour moving mean.

4.7 Reproducing extreme rainfall events

Figure 12 shows the performance of GDHPM in producing the extreme rainfall events while simulating hourly rainfall events. To calculate the rainfall extremes, 95 percentile rainfall value is calculated using 39 years (1980-2018) of rainfall. It is to be noted that, for accounting the $\pm$3-hour uncertainty in the simulations, the hourly precipitation is converted into a 6-hourly sum. Figure 12 compares extreme events from experiments E1_R1, E2_R3, and E3_R11, i.e., with no covariate, correlated site as covariate, and seven clustered sites as covariate respectively. The sites presented in figure 12 are used in E3_R11 experiment. The result shows that the E2_R3 and E3_R11 experiments are able to predict the extremes more accurately in comparison to the no covariate run. If compared between the E2_R3 and E3_R11, at most of the sites E2_R3 captured the extremes more accurately. The selection of most correlated site as covariate in the TI provided precise patterns to reproduce the extremes accurately. However, the maximum 1% rainfall events are not well reproduced using DS. As DS works on pattern reproduction,

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

producing rare events is difficult for DS. Largely, the results demonstrate that DS can handle high peak events even at these high altitudes; however, the training data must contain enough high peak information as a prerequisite.

5 Discussions

The key motivation of this study is to demonstrate the efficacy of MPS approach-based DS algorithm in effectively generating the hourly precipitation data through GDHPM. The MPS approach is tested at the complicated terrains of the IHR. Specifically, the approach generates future P_hourly data using hindcast P_hourly and P_daily data. The MPS approach-based DS algorithm identifies the temporal patterns from hindcast data, extracts them, and generates similar future patterns for the simulated P_hourly. To help the model generate accurate patterns, we have introduced covariates into the model with similar topographical and locational properties. For this, we have conducted 12 experimental runs using none, one, and multiple site precipitation data as covariate for simulations using three statistical techniques - nearest neighbour, correlation coefficient and complex networks. The overall results show that DS is capable of simulating realistic P_hourly time series using hindcast P_hourly and P_daily. The 50 realizations produced for each experimental run cover the uncertainty in the prediction with different accuracy depending on the simulation setup adopted. The 3-hourly and 6-hourly moving average applied on the P_daily data smoothens the daily time series and helps the model produce finer hourly variability.

Specifically, in the simulations using no covariates, results suggests that the run E1_R1 provides mean ensemble P_hourly which matches better to the reference P_hourly than the run E1_R2. The time series plots and the error statistics of both the runs suggest that 3-hourly moving average is producing more accurate simulations in case of no covariate simulations. The MAE values indicate more accurate simulations and the bias % values indicate more precise simulation in E1_R1. Visual analysis of time series can indicate overestimation of P_hourly in run E1_R2. Further, in case of two-site simulations, both the time series plots and the error statistics show that taking the most correlated site as covariate helps the model in producing better simulations than taking the nearest site as covariate. Close inspection of the time series simulations generated from E2_R3 and E2_R4 suggest that the nearest site covariate is not helpful to the model for generating temporal patterns. Moreover, the hourly time series simulated for E2_R5 and E2_R6 show substantial match with the reference. Comparison between E2_R5 and E2_R6 using both the time series plots and the error statistics suggests that E2_R5 has produced better results than E2_R6. The E2_R5 has simulated the small and high peaks accurately with less overestimation through ensembles. Moreover, the introduction of complex networks in identifying the covariates shows encouraging results as the P_hourly precipitation data is precisely simulated using 7 sites, 5 sites, and 3 sites as covariates. Results suggest that out of the six runs from experiment 3, E3_R7 and E3_R11 (using 3-hour moving average) have outperformed the other runs. Also, the RMSE, MAE and bias values of these runs are significantly lower than other runs. Hence, we observe that with a higher number of covariates, the MPS approach is being able to capture the precipitation peaks better with mean ensemble simulations. This implies that the clustering coefficient method is able to identify optimum sites for the DS algorithm to generate reliable hourly precipitation data. Furthermore, the efficiency of these experimental runs is tested in reproducing the extreme events (>95 %). The results suggested that, correlated site covariate and multi-site covariate experiments are doing well in producing the extreme events.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

DS is capable of exploring the variability of intense events within the ensemble envelope. However, as DS is based on pattern reproduction, identification and simulation of rare extreme events is difficult.

The complexity of generating hourly precipitation for the Himalayas region is due to highly variable topography and complex local climate which influences the characteristics of precipitation. GDHPM is significant in producing realistic and reliable hourly precipitation ensembles at avalanche sites. GDHPM can be highly beneficial to develop future systems to predict and monitor avalanche and landslide events in the complex terrains of Himalayas for early warning, mitigation and adaptation. Moreover, the study creates greater avenues to study the behaviour of critical glaciers especially in the context of climate change.

6 Conclusions

The primary aim of this study is to set up a robust MPS-based geo-statistical model in mountainous region to produce future hourly precipitation with statistical ensembles, resulting in GDHPM. The approach is based on the Direct Sampling algorithm, applied to temporal disaggregation problem for the first time to our knowledge. The technique resamples the given training datasets using a pattern-similarity rule. DS searches for relevant patterns throughout the available data and assigns them to the centre of the datum at the simulated time step. The process is repeated for the whole simulation grid. In this study, one or more precipitation patterns obtained from neighbouring sites are conditioned in the simulation grid for better simulation of the precipitation patterns. The MPS-based model is set up on 20 avalanche sites of Himachal Pradesh where frequent avalanche/landslide events are reported during the monsoons. The training dataset for model setup includes P_hourly and P_daily data for a period of June to September of 1980 to 2018. The conditioning dataset involves the P_daily data for a period of June to September of 2019 to 2020. The P_daily is considered to be used with a 3-hour or 6-hour moving average technique for smoothening the time series. Instead of simulating precipitation with same site data, for making the model more robust, single, double, and multiple site data is used as covariates and 12 experimental runs are created. The results suggests that the DS algorithm is capable of producing the temporal patterns of the P_hourly accurately using P_daily. The predictive power of the model is shown in terms of low bias, RMSE, and MAE values obtained from the simulations. In all the experimental runs, we have observed that the statistical ensemble generated explored the uncertainty of rainfall with realism compared to observations. The performance of the model has improved with addition of covariates using various statistical techniques. As the Himalayas present a very complex topography, taking the nearest site in space as covariate is not useful for the model. Analysing the individual time series obtained from all the experimental runs, at individual sites suggest that higher peaks and smaller variabilities in the time series are better simulated while used with multiple sites precipitation as covariates. It is also noticed that, 3-hour and 6-hour moving average applied on P_daily has helped the model in generating efficient hourly precipitation, out of all the experimental runs, the following setups have helped the model producing best results:

-Single site simulation with 3-hourly moving average on P_daily

-Two site simulation using the most correlated site precipitation and both 3-hourly and 6-hourly moving average on P_daily

-Multi-site simulation using minimum 3 sites as covariate and 3-hourly moving average on P_daily

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024


Hydrology and Earth System Sciences
Discussions

-Multi-site simulation using 7 sites as covariate and 3-hourly moving average on P_daily

The study also demonstrates that extreme events can also be captured using DS technique with selection of appropriate covariates in the training data. The GDHPM is an easy, cost-effective, and robust way to produce reliable ensembles of hourly precipitation which is crucial for monitoring vulnerable regions, better forecasting of avalanches and landslides in any mountainous region. The application of the generated hourly precipitation data can be highly beneficial in the field of hydro-meteorological modelling and predictions. Future research will focus on producing sub-hourly hydro-meteorological data at fine spatial scales for the complex terrains of Himalayas.

Data availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contributions. **NS:** Conceptualization, Formal analysis, Writing – original draft, Investigation, Visualization. **MKV:** Visualization, Formal analysis, Writing and editing manuscript. **AS:** Methodology, Formal analysis, Visualization, Writing and editing manuscript. **SJ:** Methodology, Conceptualization, Visualization, Validation, Writing and editing manuscript; Supervision.**FO:** Methodology, Conceptualization, Visualization, Validation, Writing and editing manuscript.

Competing interests. The authors declare they have no conflict of interest.

Acknowledgements. We thank the scheme for Transformational and Advanced Research in Sciences, Ministry of Education (MoE STARS) [grant number: STARS/APR2019/DS/391/FS] for the research support it awarded to Dr. Sanjeev Kumar Jha.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

References

Acharya, S. C., Nathan, R., Wang, Q. J., and Su, C.-H.: Temporal disaggregation of daily rainfall measurements using regional reanalysis for hydrological applications, Journal of Hydrology, 610, 127 867, 2022.

Asadollah, S. B. H. S., Sharafati, A., and Shahid, S.: Application of ensemble machine learning model in downscaling and projecting climate variables over different climate regions in Iran, Environmental Science and Pollution Research 2021, 1, 1–20, https://doi.org/10.1007/S11356-021-16964-Y, 2021.

Connolly, R., Schirmer, J., and Dunn, P.: A daily rainfall disaggregation model, Agricultural and Forest Meteorology, 92, 105–117, https://doi.org/10.1016/S0168-1923(98)00088-4, 1998.

Dash, S. K., Jenamani, R. K., Kalsi, S. R., and Panda, S. K.: Some evidence of climate change in twentieth-century India, Climatic Change, 85, 299–321, https://doi.org/10.1007/s10584-007-9305-9, 2007.

Debele, B., Srinivasan, R., and Parlange, J. Y.: Accuracy evaluation of weather data generation and disaggregation methods at finer timescales, Advances in Water Resources, 30, 1286–1300, https://doi.org/10.1016/j.advwatres.2006.11.009, 2007.

Dembélé, M., Oriani, F., Tumbulto, J., Mariéthoz, G., and Schaefli, B.: Gap-filling of daily streamflow time series using Direct Sampling in various hydroclimatic settings, Journal of Hydrology, 569, 573–586, 2019.

Depina, I., Oguz, E. A., and Thakur, V.: Novel Bayesian framework for calibration of spatially distributed physical-based landslide prediction models., Computers and Geotechnics, 125, 103 660, https://doi.org/10.1016/j.compgeo.2020.103660, 2020.

Firor, S. E., Finney, B. A., Willis, R., and Dracup, J. A.: Disaggregation Modeling Process for Climatic Time Series, Journal of Water Resources Planning and Management, 122, 205–212, https://doi.org/10.1061/(ASCE)0733-9496(1996)122:3(205), 1996.

Gardner, J., Sinclair, J., Berkes, F., and Singh, R. B.: Accelerated tourism development and its impacts in Kullu-Manali, H.P., India, Tourism Recreation Research, 27, 9–20, https://doi.org/10.1080/02508281.2002.11081370, 2002.

Gardner, J. S. and Saczuk, E.: Systems for hazards identification in high mountain areas: An example from the Kullu District, western Himalaya, Journal of Mountain Science, 1, 115–127, https://doi.org/10.1007/bf02919334, 2004.

Guenni, L. and Bárdossy, A.: A two steps disaggregation method for highly seasonal monthly rainfall, Stochastic Environmental Research and Risk Assessment, 16, 188–206, https://doi.org/10.1007/S00477-002-0094-4, 2002.

Jha, S. K. and Sivakumar, B.: Complex networks for rainfall modeling: Spatial connections, temporal scale, and network size, Journal of Hydrology, 554, 482–489, https://doi.org/10.1016/j.jhydrol.2017.09.030, 2017.

Jha, S. K., Mariethoz, G., Evans, J., McCabe, M. F., and Sharma, A.: A space and time scale-dependent nonlinear geostatistical approach for downscaling daily precipitation and temperature, Water Resources Research, 51, 6244–6261, https://doi.org/10.1002/2014WR016729, 2015.

Jing, W., Yang, Y., Yue, X., and Zhao, X.: A spatial downscaling algorithm for satellite-based precipitation over the Tibetan plateau based on NDVI, DEM, and land surface temperature, Remote Sensing, 8, https://doi.org/10.3390/rs8080655, 2016.

jun Wang, X., yun Zhang, J., Shahid, S., hong Guan, E., xiang Wu, Y., Gao, J., and min He, R.: Adaptation to climate change impacts on water demand, Mitigation and Adaptation Strategies for Global Change, 21, 81–99, https://doi.org/10.1007/S11027-014-9571-6/FIGURES/3, 2016.

Karuna Sagar, S., Rajeevan, M., and Vijaya Bhaskara Rao, S.: On increasing monsoon rainstorms over India, Natural Hazards, 85, 1743–1757, 2017.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Koutsoyiannis, D.: A nonlinear disaggregation method with a reduced parameter set for simulation of hydrologic series, Water Resources Research, 28, 3175–3191, https://doi.org/10.1029/92WR01299, 1992.

Kumar, A., Gupta, A. K., Bhambri, R., Verma, A., Tiwari, S. K., and Asthana, A. K.: Assessment and review of hydrometeorological aspects for cloudburst and flash flood events in the third pole region (Indian Himalaya), https://doi.org/10.1016/j.polar.2018.08.004, 2018.

Kumar, D. N., Lall, U., and Petersen, M. R.: Multisite disaggregation of monthly to daily streamflow, Water Resources Research, 36, 1823–1833, https://doi.org/10.1029/2000WR900049, 2000.

Lee, T. and Jeong, C.: Nonparametric statistical temporal downscaling of daily precipitation to hourly precipitation and implications for climate change scenarios, Journal of Hydrology, 510, 182–196, https://doi.org/10.1016/j.jhydrol.2013.12.027, 2014.

Lu, Y. and Qin, X. S.: Multisite rainfall downscaling and disaggregation in a tropical urban area, Journal of Hydrology, 509, 55–65, https://doi.org/10.1016/j.jhydrol.2013.11.027, 2014.

Maheepala, S. and Perera, B. J.: Monthly hydrologic data generation by disaggregation, Journal of Hydrology, 178, 277–291, https://doi.org/10.1016/0022-1694(95)02834-X, 1996.

Margulis, S. A. and Entekhabi, D.: Temporal disaggregation of satellite-derived monthly precipitation estimates and the resulting propagation of error in partitioning of water at the land surface, Hydrology and Earth System Sciences, 5, 27–38, https://doi.org/10.5194/hess-5-27-2001, 2001.

Mariethoz, G., Renard, P., and Straubhaar, J.: The direct sampling method to perform multiple-point geostatistical simulations, Water Resources Research, 46, 1–14, https://doi.org/10.1029/2008WR007621, 2010.

Mayowa, O. O., Pour, S. H., Shahid, S., Mohsenipour, M., Harun, S. B., Heryansyah, A., and Ismail, T.: Trends in rainfall and rainfall-related extremes in the east coast of peninsular Malaysia, Journal of Earth System Science, 124, 1609–1622, https://doi.org/10.1007/S12040-015-0639-9/FIGURES/11, 2015.

Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Meirvenne, M. V., and Renard, P.: A practical guide to performing multiple-point statistical simulations with the Direct Sampling algorithm, Computers and Geosciences, 52, 307–324, https://doi.org/10.1016/j.cageo.2012.09.019, 2013.

Obeysekera, J. T. and Salas, J. D.: Modeling of aggregated hydrologic time series, Journal of Hydrology, 86, 197–219, https://doi.org/10.1016/0022-1694(86)90165-4, 1986.

Oriani, F., Straubhaar, J., Renard, P., and Mariethoz, G.: Simulation of rainfall time series from different climatic regions using the direct sampling technique, Hydrology and Earth System Sciences, 18, 3015–3031, https://doi.org/10.5194/hess-18-3015-2014, 2014.

Oriani, F., Borghi, A., Straubhaar, J., Mariethoz, G., and Renard, P.: Missing data simulation inside flow rate time-series using multiple-point statistics, Environmental Modelling and Software, 86, 264–276, https://doi.org/10.1016/j.envsoft.2016.10.002, 2016.

Papalexiou, S. M., Markonis, Y., Lombardo, F., AghaKouchak, A., and Foufoula-Georgiou, E.: Precise Temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for Stationary and Nonstationary Processes, Water Resources Research, 54, 7435–7458, https://doi.org/10.1029/2018WR022726, 2018.

Prein, A. F. and Gobiet, A.: Impacts of uncertainties in European gridded precipitation observations on regional climate analysis, International Journal of Climatology, 37, 305–327, https://doi.org/10.1002/joc.4706, 2017.

Pui, A., Sharma, A., Mehrotra, R., Sivakumar, B., and Jeremiah, E.: A comparison of alternatives for daily to sub-daily rainfall disaggregation, Journal of Hydrology, 470-471, 138–157, https://doi.org/10.1016/j.jhydrol.2012.08.041, 2012.

Samal, N. and Jha, S. K.: Bias correction of WRF output for operational avalanche forecasting in the Indian Himalayan region, Journal of Earth System Science, 131, 1–19, https://doi.org/10.1007/S12040-022-01899-W/FIGURES/20, 2022.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Santos, E. G. and Salas, J. D.: Stepwise Disaggregation Scheme for Synthetic Hydrology, Journal of Hydraulic Engineering, 118, 765–784, https://doi.org/10.1061/(ASCE)0733-9429(1992)118:5(765), 1992.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theoretical and Applied Climatology, 115, 15–40, https://doi.org/10.1007/s00704-013-0860-x, 2014.

Sharif, M., Burn, D., and Wey, K.: Daily and Hourly Weather Data Generation using a K-Nearest Neighbour Approach, 18th Canadian Hydrotechnical Conference, pp. 22–24, https://www.researchgate.net/publication/253996137_Daily_and_Hourly_Weather_Data_Generation_using_a_K-Nearest_Neighbour_Approach/citations, 2007.

Sharif, M., Burn, D. H., and Hofbauer, K. M.: Generation of Daily and Hourly Weather Variables for use in Climate Change Vulnerability Assessment, Water Resources Management, 27, 1533–1550, https://doi.org/10.1007/S11269-012-0253-4, 2013.

Singhal, A. and Jha, S. K.: An application of multiple-point statistics downscaling approach over North-West Himalayas in avalanche-prone areas, International Journal of Climatology, pp. 1–20, https://doi.org/10.1002/joc.7342, 2021.

Singhal, A., Cheriyamparambil, A., and Jha, S. K.: Spatial extrapolation of statistically downscaled weather data over the Northwest Himalayas at major glacier sites, Environmental Modelling and Software, 149, 105 317, https://doi.org/10.1016/j.envsoft.2022.105317, 2022.

Singhal, A., Jaseem, M., Divya, Sarker, S., Prajapati, P., Singh, A., and Jha, S. K.: Identifying potential locations of hydrologic monitoring stations based on topographical and hydrological information, Water Resources Management, 38, 369–384, 2024.

Sivakumar, B.: Networks: a generic theory for hydrology?, Stochastic Environmental Research and Risk Assessment, 29, 761–771, https://doi.org/10.1007/s00477-014-0902-7, 2014.

Srikanthan, R., Sharma, A., and McMahon, T.: Comparison of Two Nonparametric Alternatives for Stochastic Generation of Monthly Rainfall, Journal of Hydrologic Engineering, 11, 222–229, https://doi.org/10.1061/(ASCE)1084-0699(2006)11:3(222), 2006.

Thober, S., Mai, J., Zink, M., and Samaniego, L.: Stochastic temporal disaggregation of monthly precipitation for regional gridded data sets, Water Resources Research, 50, 8714–8735, https://doi.org/10.1002/2014WR015930, 2014.

Valencia, R. D. and Schakke Jr, J. C.: Disaggregation processes in stochastic hydrology, Water Resources Research, 9, 580–585, 1973.

Wójcik, R. and Buishand, T. A.: Simulation of 6-hourly rainfall and temperature by two resampling schemes, Journal of Hydrology, 273, 69–80, 2003.

Zhou, H., Deng, Z., Xia, Y., and Fu, M.: A new sampling method in particle filter based on Pearson correlation coefficient, Neurocomputing, 216, 208–215, https://doi.org/10.1016/J.NEUCOM.2016.07.036, 2016.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

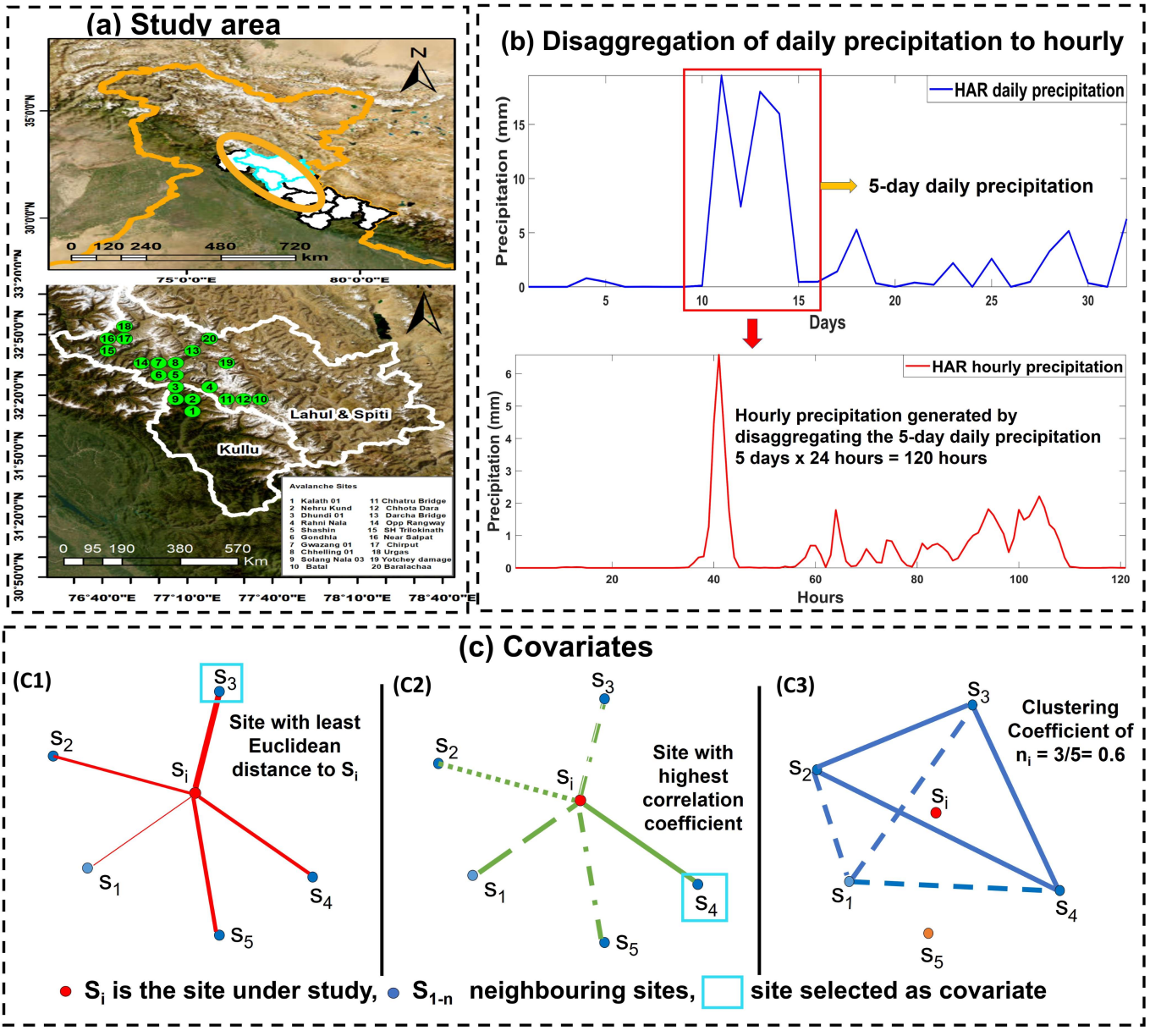

Figure 1. (a) Domain representing the study area with the location of the 20 avalanche sites in Kullu and Lahul & Spiti districts of Himachal Pradesh (topographic basemap is obtained from ESRI); (b) a sample representation of temporal disaggregation approach to generate hourly precipitation time series from daily; (c) graphical illustration of the three approaches used to select covariates in the mountainous region namely nearest neighbour (c1), correlation coefficient (c2), and clustering coefficient measure of complex networks (c3)

.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

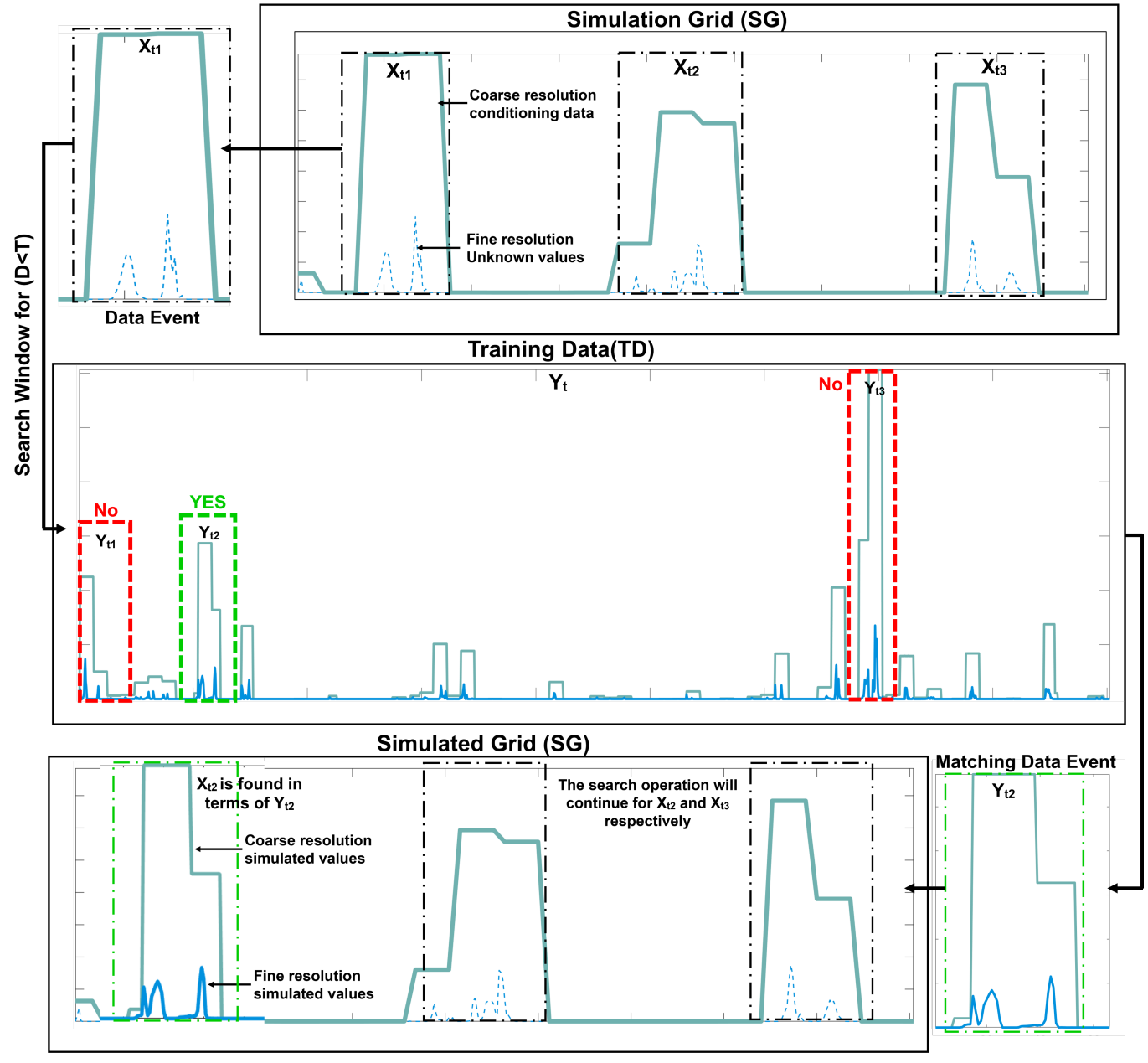

Figure 2. Working of the DS algorithm for temporal disaggregation where the black dashed rectangle represents the search window, D is the measure of dissimilarity between data events and T is a defined distance threshold between 0 and 1. The cyan line represents the coarse scale data and the dashed cyan line represents the fine resolution data to be simulated.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

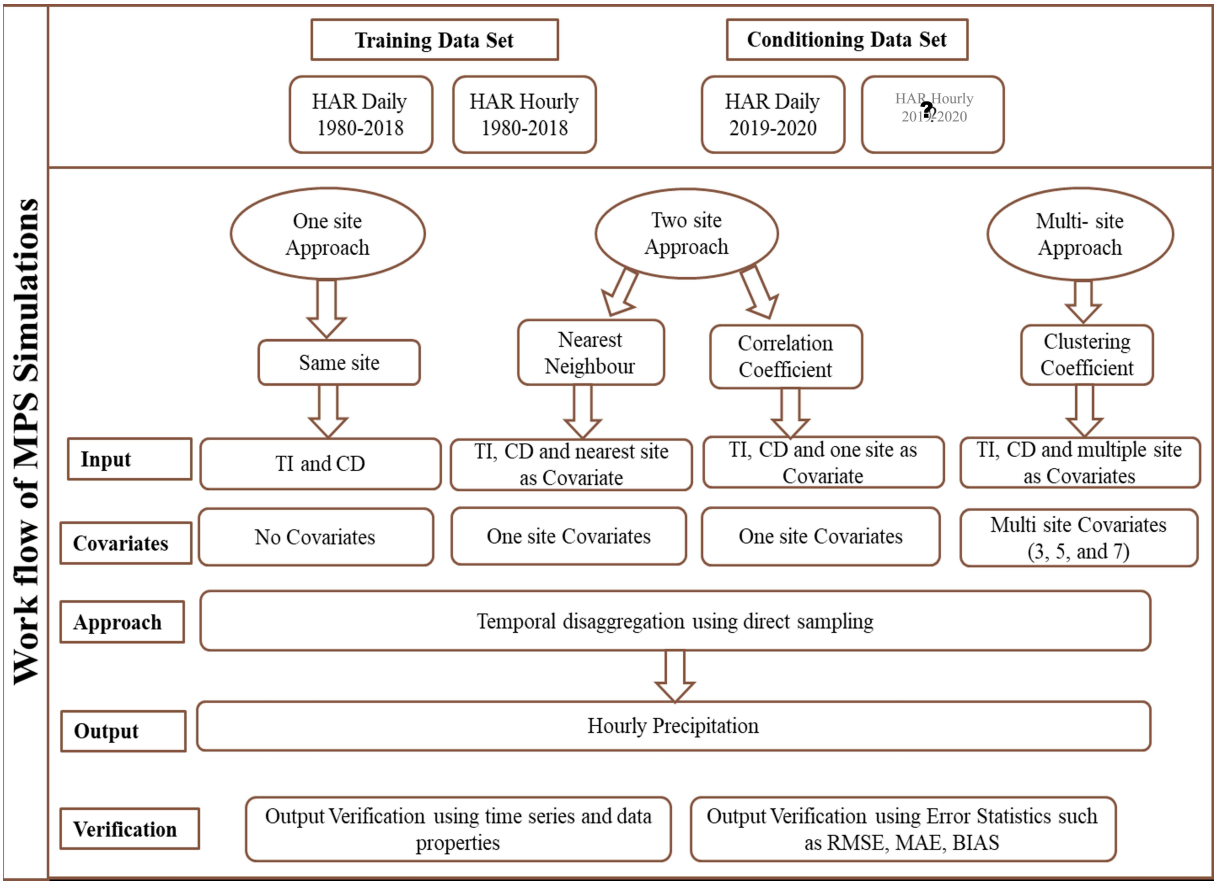

Figure 3. Flowchart showing the workflow of GDHPM model for hourly precipitation simulation.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

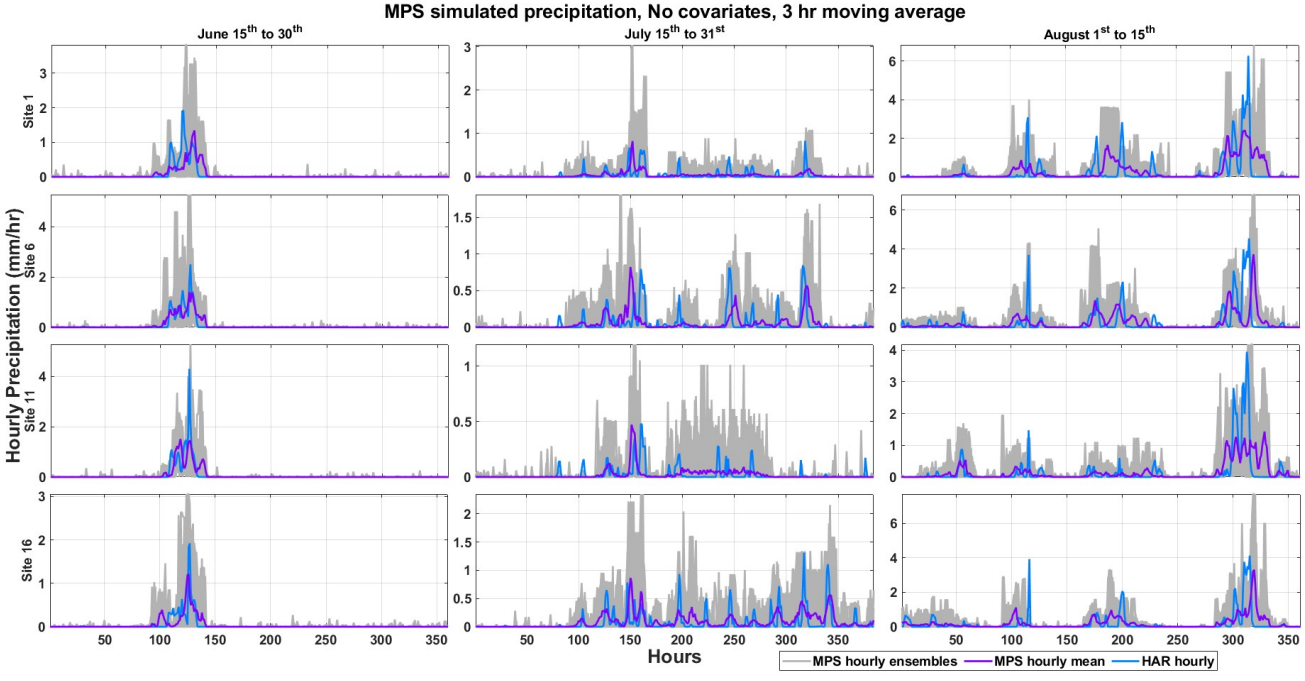

Figure 4. Hourly simulated precipitation for the monsoon season under experimental run E1_R1 for four avalanche sites. Each subplot represents time series of the HAR hourly precipitation, MPS hourly precipitation with 50 ensembles and the mean of the ensembles for a time period of June 15th to 30th, July 15th to 31st, and August 1st to 15th.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

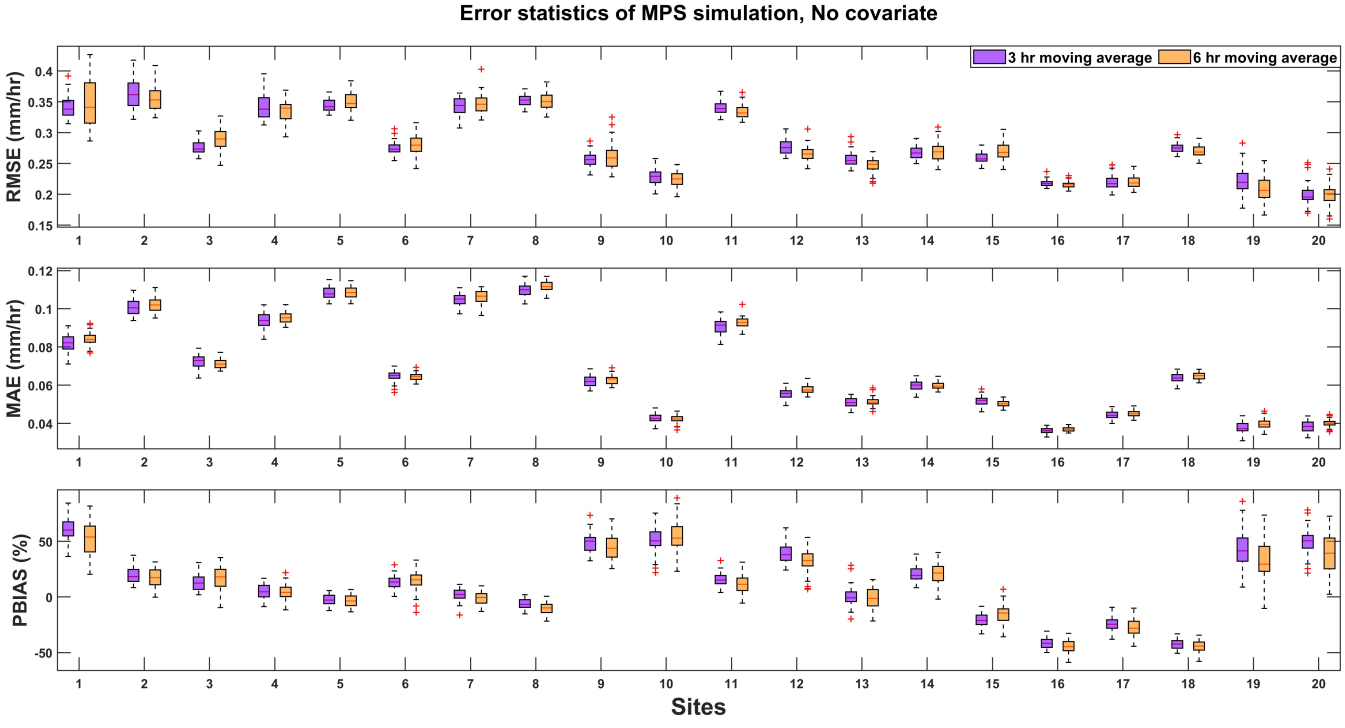

Figure 5. Boxplots presenting RMSE (mm/hr), MAE (mm/hr), and bias% values of each ensemble of MPS simulated hourly precipitation obtained from experimental runs E1_R1 and E1_R2 in comparison to the reference hourly precipitation at twenty avalanche sites for the whole simulation period

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

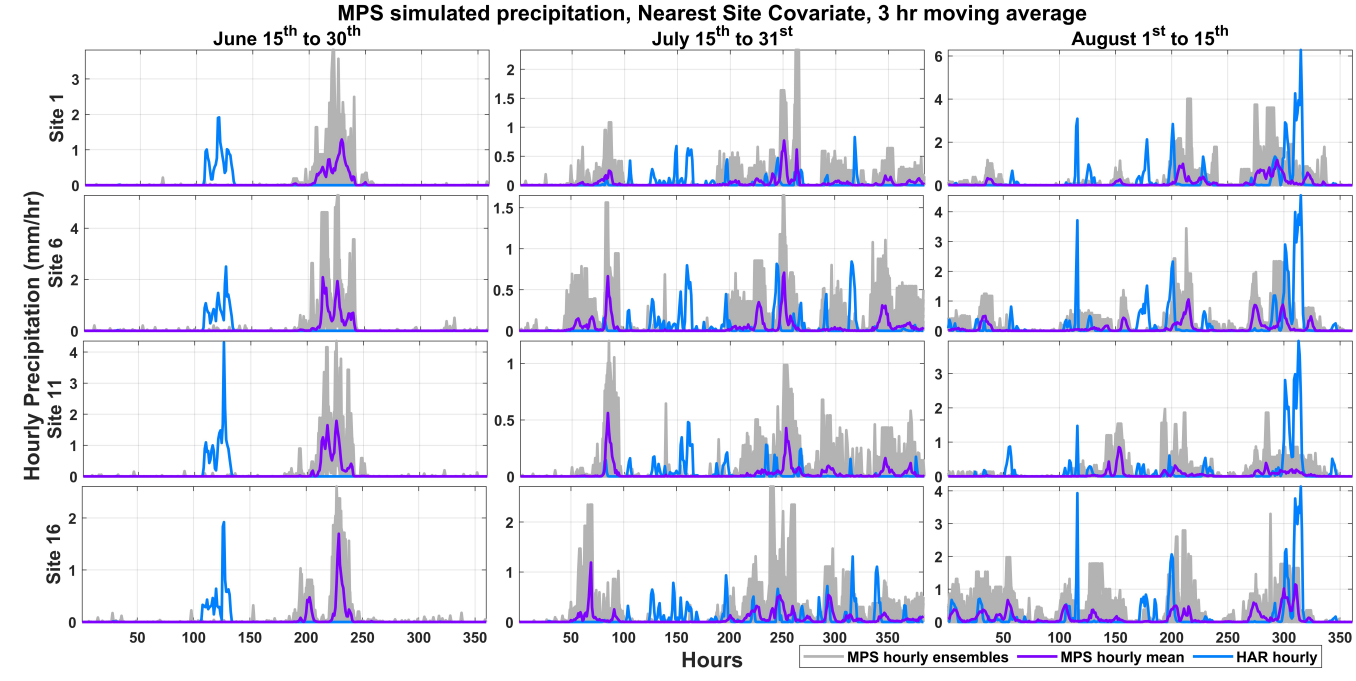

Figure 6. Hourly simulated precipitation for the monsoon season under experimental run E2_R3 for four avalanche sites. Each subplot represents time series of the HAR hourly precipitation, MPS hourly precipitation with 50 ensembles and mean of the ensembles for a time period of June 15th to 30th, July 15th to 31st, and August 1st to 15th.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

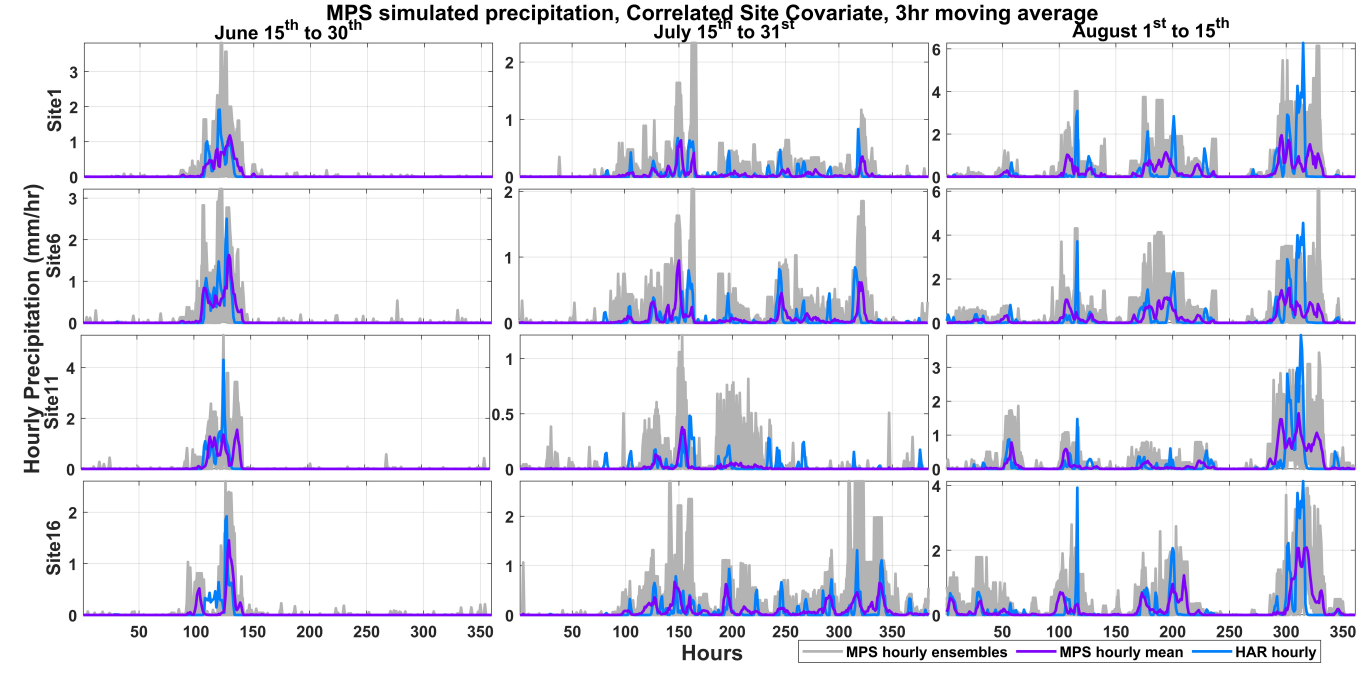

Figure 7. Hourly simulated precipitation for the monsoon season under experimental run E2_R5 for four avalanche sites. Each subplot represents time series of the HAR hourly precipitation, MPS hourly precipitation with 50 ensembles and mean of the ensembles for a time period of of June 15th to 30th, July 15th to 31st, and August 1st to 15th.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

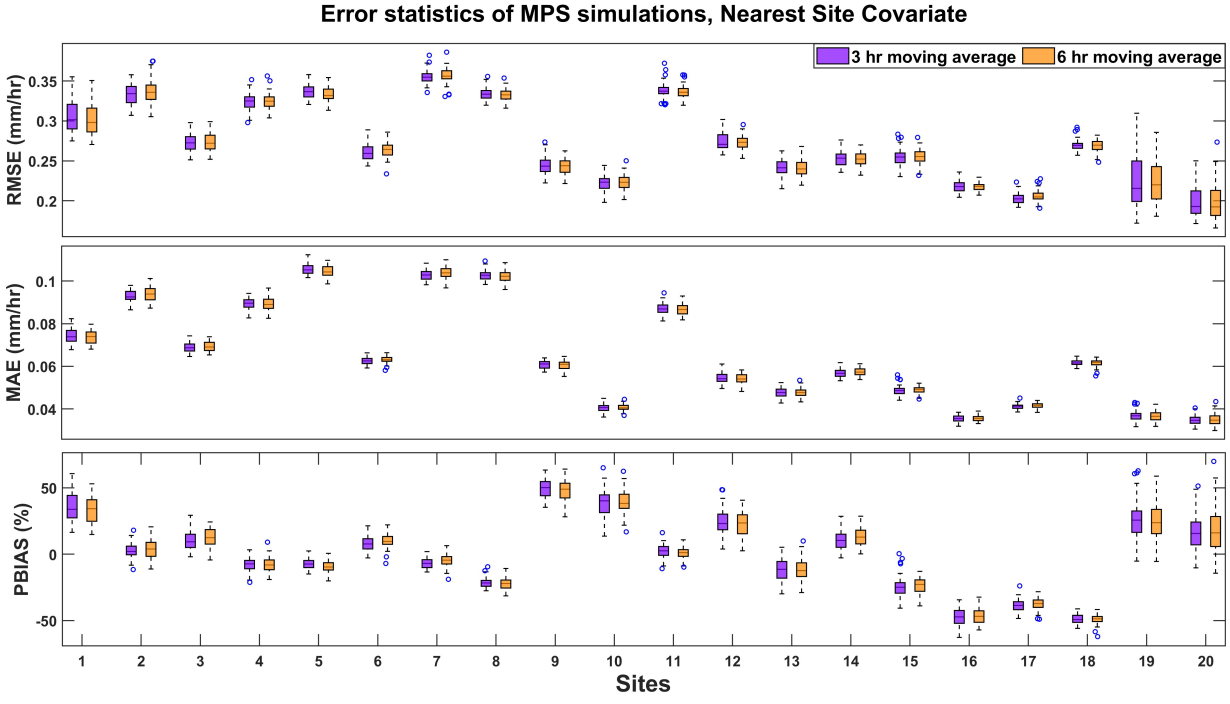

Figure 8. Boxplots presenting RMSE (mm/hr), MAE (mm/hr), and bias% values of each ensemble of MPS simulated hourly precipitation obtained from experimental runs E2_R3 and E2_R4 in comparison to the reference hourly precipitation at twenty avalanche sites for the whole simulation period.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

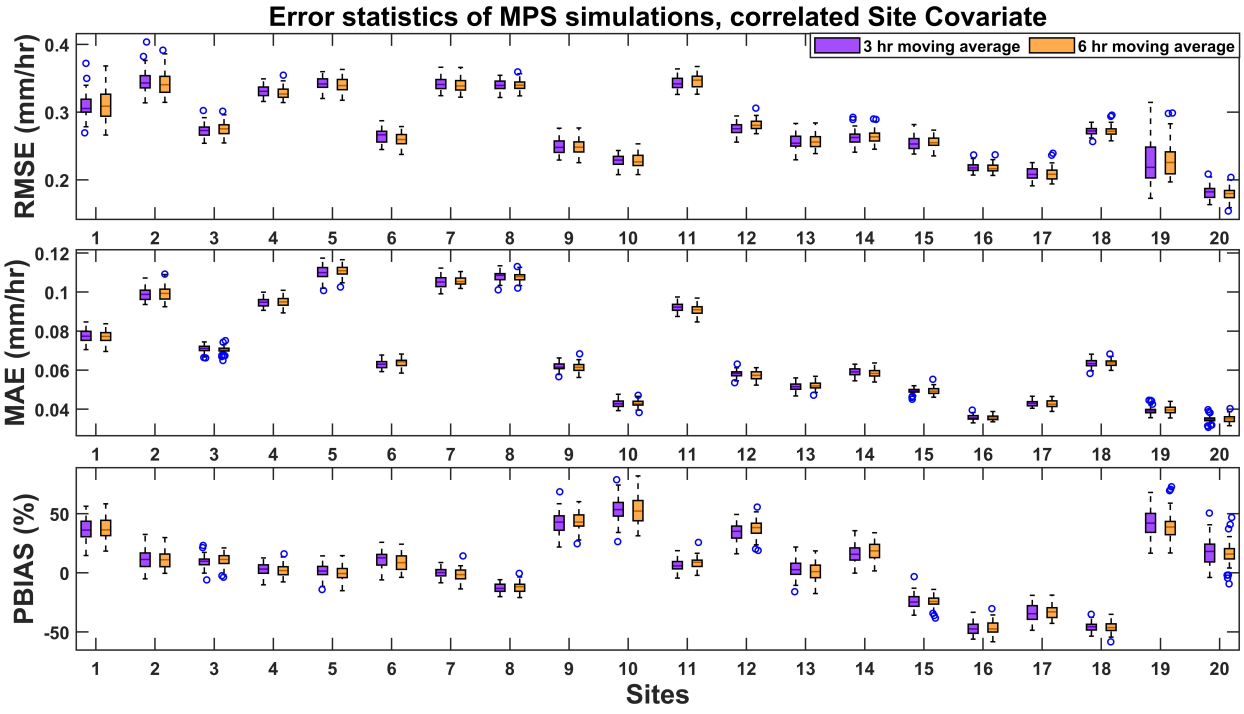

Figure 9. Boxplots presenting RMSE (mm/hr), MAE (mm/hr), and bias% values of each ensemble of MPS simulated hourly precipitation obtained from experimental runs E2_R5 and E2_R6 in comparison to the reference hourly precipitation at twenty avalanche sites for the whole simulation period.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

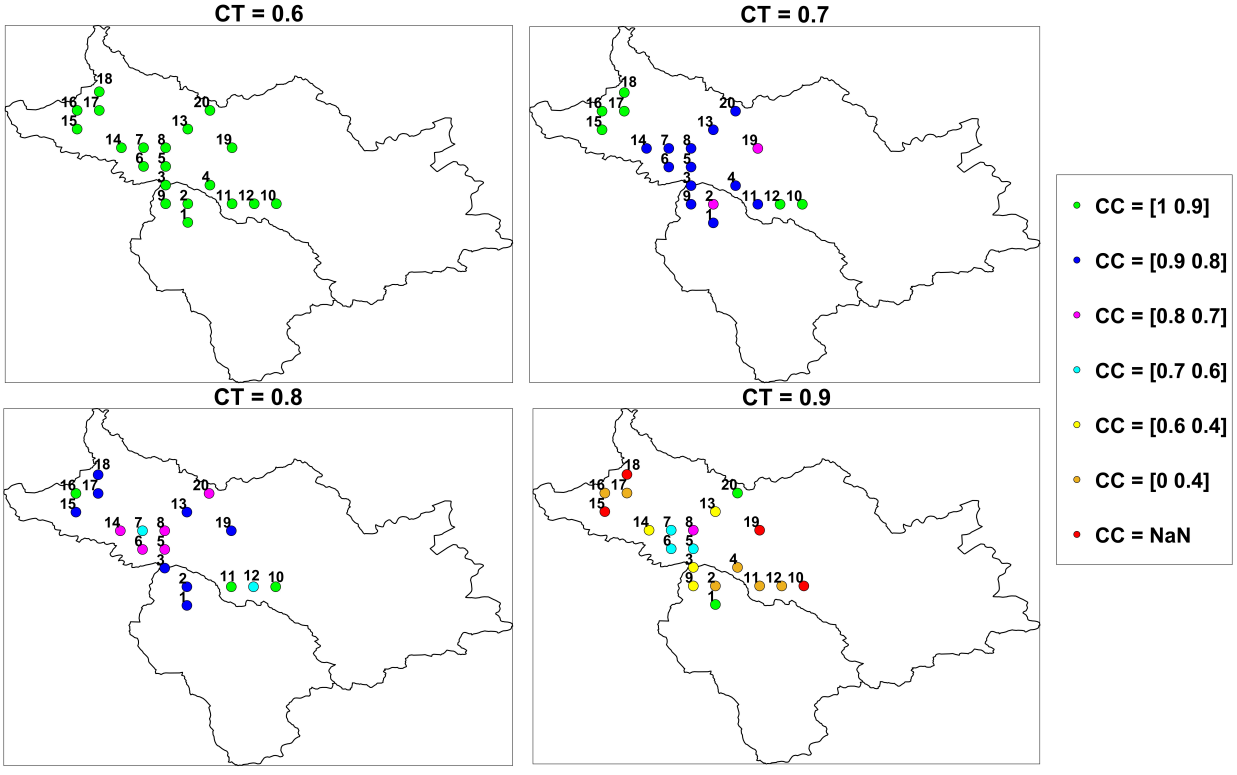

Figure 10. Clusters of sites obtained using clustering coefficient measure of complex network with CT value 0.8. The three clusters are represented in green (CC = 0.9 to 1), blue (CC = 0.8 to 0.9), and pink (CC = 0.7 to 0.8).

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

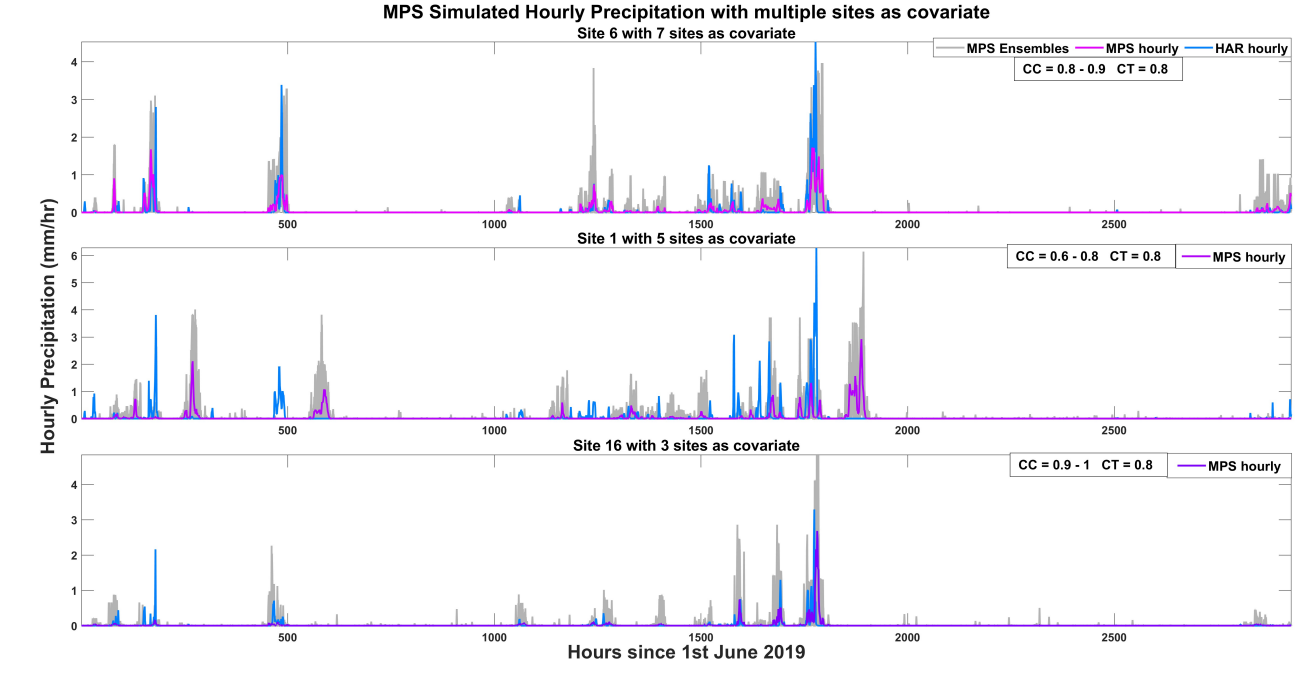

Figure 11. Hourly simulated precipitation for the monsoon season under experimental run E3_R7, E3_R9, and E3_R11 for avalanche site_6, site_1, and site_16, respectively. Each subplot represents time series of the HAR hourly precipitation, MPS hourly precipitation with 50 ensembles and mean of the ensembles for a time period of 1st June 2019 to 30th September 2019.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

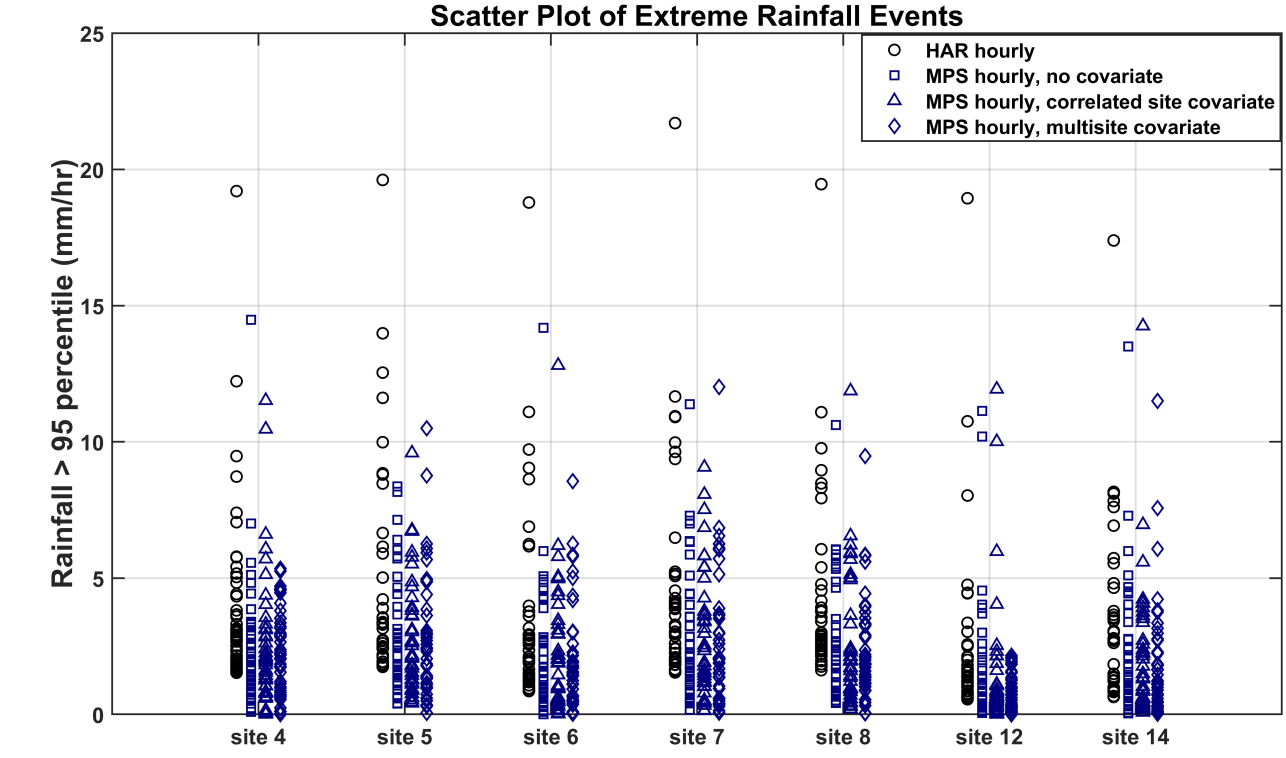

Figure 12. Scatter plots of rainfall events $\geq$ 95% in the years 2019 and 2020 comparing results from experiments E1_R1, E2_R5, and E3_R11.

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Table 1. Details of the fixed parameters used in the DS method.

Fixed parameters	
Simulation method	Direct Sampling
Number of Realizations	50
Maximum search distance	72 1 1
Anisotropy ratios in the search window	1 1 1 (isotrope window)
Transformations	0 (no transformations)
Path type	0 (random path)
Type of variable	0 for categorical,1 for continuous
The exponent of search distance	0 (mean absolute error) (isotrope window)
Syn-processing parameters	0 0 0 0
Initial seed	0
Parallelization	1
Parameters reduction	1

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Table 2. Details of varied parameters used in the DS method.

Varied parameters		
Name	**Default**	**Range**
Distance threshold (t)	0.05	Between 0 and 1
Max fraction of TI to scan (f)	0.25	0.05 to 1
Max number of points in the neighbourhood (n)	48	1-72

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

Table 3. Details of the experimental runs followed in GDHPM. Here, the runs are based on the moving average technique and the experiments are based on the covariate selection.

Runs	Experiments	Covariate selection method	Covariates	Moving-mean on daily data
E1_R1	E1 (NO Covariate)	NA	NA	3-hr
E1_R2				6-hr
E1_R3	E2 (One Site Covariate)	Station proximity	1 site	3-hr
E1_R4				6-hr
E1_R5	E2 (One Site Covariate)	Data Correlation	1 site	3-hr
E1_R6				6-hr
E1_R7	E3 (Multiple Site Covariate)	Clustering coefficient	3 sites	3-hr
E1_R8				6-hr
E1_R9	E3 (Multiple Site Covariate)	Clustering coefficient	5 sites	3-hr
E1_R10				6-hr
E1_R11	E3 (Multiple Site Covariate)	Clustering coefficient	7 sites	3-hr
E1_R12				6-hr

https://doi.org/10.5194/hess-2024-155
Preprint. Discussion started: 14 June 2024

Hydrology and Earth System Sciences
Discussions

Table 4. Mean RMSE, MAE, and bias values of the simulated hourly precipitation obtained from each experimental runs of experiment 3.

Runs	Run property	RMSE (mm/hr)	MAE(mm/hr)	BIAS(%)
E3_R7	3_hour, 7_site	0.1651	0.0209	-5.12
E3_R8	6_hour, 7_site	0.1822	0.0347	-5.99
E3_R9	3_hour, 5_site	0.2174	0.0404	-13.08
E3_R10	6_hour, 5_site	0.2170	0.0401	-2.84
E3_R11	3_hour, 3_site	0.1367	0.0205	-11.12
E3_R12	6_hour, 3_site	0.1282	0.0209	72.43