

Simulation of Weather Data in Brahmaputra Basin Using K-Nearest Neighbour Model

Azhar Husain*, Mohammed Sharif

Department of Civil Engineering, Jamia Millia Islamia (Central University), New Delhi

*Corresponding author: ahusain3@jmi.ac.in

Received June 08, 2018; Revised July 15, 2018; Published August 08, 2018

Abstract This paper describes the application of a K-nearest neighbor weather-generating model that allows resampling with perturbation of the observed data to simulate (1) duration of extreme wet spells, and (2) duration of extreme dry spells in the Brahmaputra River Basin. The results of the simulation of extreme events carried out by the K-NN model clearly indicated that the model generated unprecedented extreme events that were not seen in the observed record. Several extreme wet spells were simulated by the KNN model, which are a critical input to flood management models. The model generated several extreme dry spells that are important for evaluation of effective drought management policies for the basin. The analysis conducted herein has the potential for providing valuable aid in developing efficient flood and drought management strategies for the Brahmaputra basin because of the ability of the model to simulate extreme dry and wet spells. It may be concluded that the utility of flood prediction models in estimating the probability of extreme events may be greatly enhanced if their performance is evaluated based on synthetic sequences generated in the present research.

Keywords: Brahmaputra, weather, K-NN, model, precipitation, extreme

Cite This Article: Azhar Husain, and Mohammed Sharif, "Simulation of Weather Data in Brahmaputra Basin Using K-Nearest Neighbour Model." *American Journal of Water Resources*, vol. 6, no. 3 (2018): 137-142. doi: 10.12691/ajwr-6-3-4.

1. Introduction

Global climate is expected to change significantly due to the continually increasing levels of carbon dioxide and other greenhouse gases. By the year 2056 the CO₂ concentration in the atmosphere is likely to double [1]. The future projections of climate change indicate a global average warming of between 1.5° to 4.5°C, greater surface warming at high latitudes in winter, but less during the summer. An increase of 3 to 15 % in global precipitation is expected, mainly due to globally increasing temperature, which causes greater evaporation of sea surface water. A year-round increase in precipitation in high-latitude regions is expected, whilst some tropical areas may experience small decreases. One of the most important and immediate effects of increased greenhouse gas emissions is the rise in long-term average temperatures.

Warming of the atmospheric system is expected to cause adverse impacts on many aspects of the natural environment, including the water resources. Temperature changes are accompanied by changes in precipitation and runoff amounts. Consequently, the increase in temperature will accelerate hydrologic cycle, altering the rainfall and eventually the runoff. One of the most important and immediate effects of global warming would be the changes in local and regional water availability [2]. Changes in extreme precipitation events will likely alter

runoff patterns, and would lead to increase in the frequency and magnitude of extreme events [3]. As a result, hydrological systems are anticipated to experience not only the changes in the average availability of water but also changes in the extremes [2,4]. Therefore, simulation of weather data under plausible scenarios of climate change is required for evaluation of hydrologic impacts of climate change at basin scale.

2. Literature Review

Recently, weather generators have been employed for simulation of weather data. Weather generators are stochastic models capable of simulating historical and future climatic conditions on a daily time scale either at a single or multiple sites [5,6]. An important advantage of weather generators is that they allow simulation of synthetic series of meteorological variables that are long enough to be used in the assessment of risk in hydrological or agricultural applications. Weather generators have been employed in climate change impact studies to generate scenarios with high temporal and spatial resolutions based on the output from GCMs (e.g., [7,8]). An important class of nonparametric weather generators are those based on the K - Nearest neighbor (KNN) resembling approach. Successful applications of K-NN weather generators to simulation of weather data have been described by Rajagopalan and Lall [9], Buishand and Brandsma [10]

and Yates et al. [11] among others. Sharif and Burn [12] describes an improved KNN weather generator for simulating plausible climate change scenarios in the Upper Thames River basin, Canada.

Muluye [13] describes the application of six variations in a nearest neighbor resampling approach to downscale station daily precipitation and minimum and maximum temperature fields for the Chute-du-Diable meteorological station in northeastern Canada. Gangopadhyay et al. [14] proposed a new K-NN algorithm that incorporated principal component analysis [15]. Most of these applications have focussed on resampling the observed data without simulating extreme events not observed in the historical record. A major limitation of these models is that they merely reshuffle the historical data to generate synthetic weather data without producing new values. Use of synthetic sequences simulated by such weather generators, in conjunction with hydrological models, to catchment response evaluation could lead to under-exploration of the possible effects of climate variability. To overcome this problem, Sharif and Burn [12] developed an improved K-NN model that can simulate large number of unprecedented values of variables and extreme events not seen in the observed record while preserving important statistical properties of the observed data.

3. Study Area and Data

The Brahmaputra basin spreads over Tibet (China), Bhutan, India and Bangladesh having a total area of 5,80,000 Sq.km. In India, it spreads over states of Arunachal Pradesh, Assam, West Bengal, Meghalaya, Nagaland and Sikkim and lies between $88^{\circ}11'$ to $96^{\circ}57'$ east longitudes and $24^{\circ}44'$ to $30^{\circ}3'$ north latitudes and extends over an area of 1,94,413 Sq.km which is nearly 5.9% of the total geographical area of the country. It is bounded by the Himalayas on the north, by the Patkari range of hills on the east running along the India-Myanmar border, by the Assam range of hills on the south and by the Himalayas and the ridge separating it from Ganga basin on the west. The Brahmaputra River originates in the north from Kailash ranges of Himalayas at an elevation of 5,150 m just south of the lake called

Konggyu Tsho and flows for about a total length of 2,900 km. In India, it flows for 916 km. The principal tributaries of the river joining from right are the Lohit, the Dibang, the Subansiri, the Jiabharali, the Dhansiri, the Manas, the Torsa, the Sankosh and the Teesta whereas the Burhidihing, the Desang, the Dikhow, the Dhansiri and the Kopili joins it from left. The major part of basin is covered with forest accounting to 55.48% of the total area and 5.79% of the basin is covered by water bodies. The basin spreads over 22 parliamentary constituencies (2009) comprising 12 of Assam, 4 of West Bengal, 2 of Arunachal Pradesh, 2 of Meghalaya, 1 of Sikkim and 1 of Nagaland. The Figure 1 shows the Brahmaputra basin and Figure 2 shows the location of ten stations as given below:

3.1. Climate Data

Hydrological observations in the sub-basin are carried out by the Central and State Governments. The Central Water Commission maintains 108 H.O sites in the basin. In addition, gauge data at 80 sites, gauge-discharge data at 15 sites and gauge, discharge and sediment data at 25 sites, maintained by the State Governments and the Brahmaputra Board, are also available. The Central Water Commission operates 27 flood forecasting stations in the sub-basin. The daily maximum, minimum temperature and daily precipitation data was obtained from IMD through IIT, Guwahati. The data for monthly average for temperature and total monthly precipitation at Cherapunji and Guwahati climate stations was obtained from the website of IMD (www.imd.gov.in).

3.2. GIS Data Layers

The data derived using Shuttle Radar Topography Mission (SRTM) has been utilized for this research [16]. These digital elevation data are available for download from <http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1>. The SRTM consisted of a specially designed radar system that flew onboard the Space Shuttle Endeavor during an 11-day mission in February of 2000. The land use data were obtained from the database of University of Maryland Global Land Cover Facility with one km grid cell.

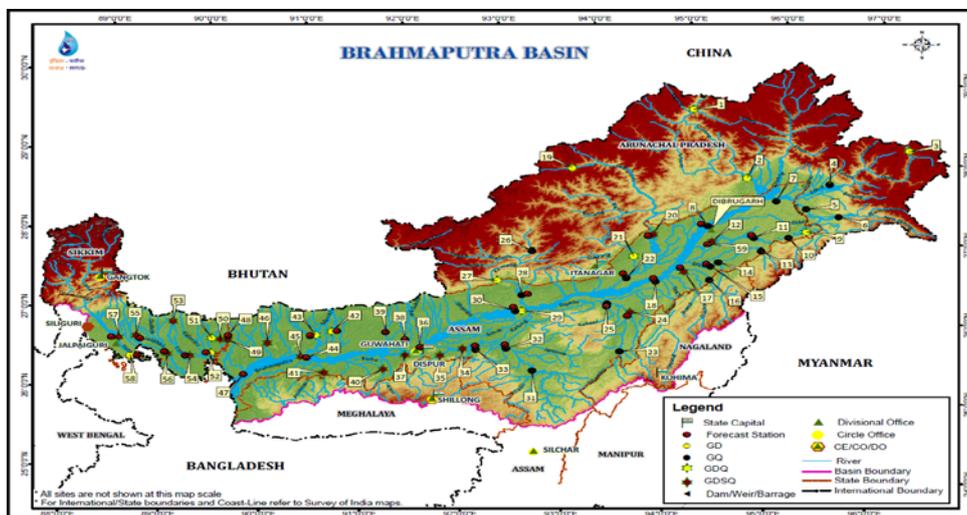


Figure 1. The Brahmaputra Basin (Source: India-WRIS)

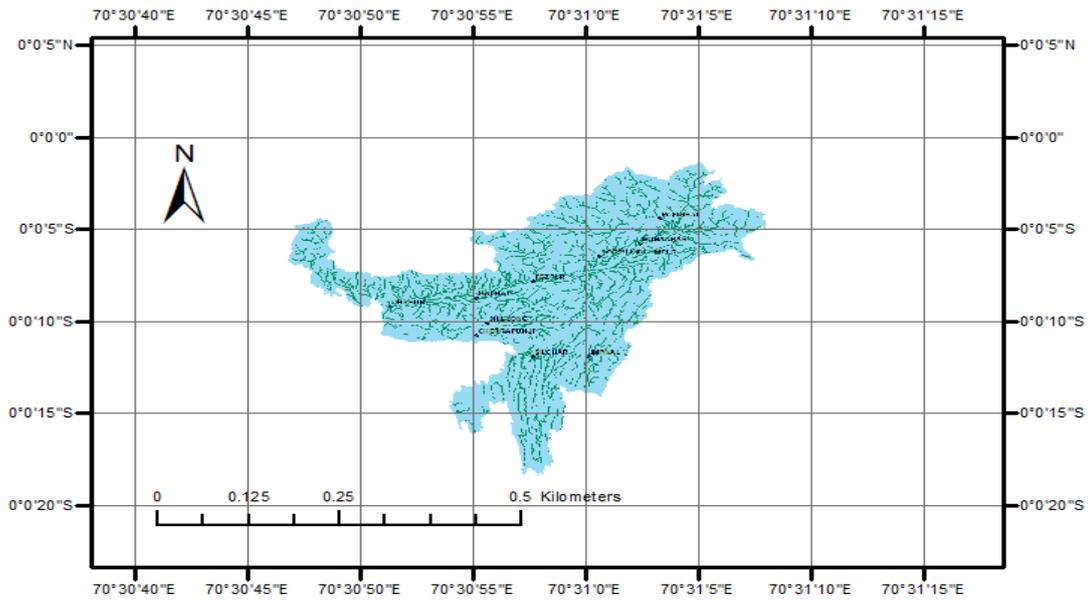


Figure 2. Location of Stations in Brahmaputra Basin

4. Research Objectives

The major objective of the present research is to simulate weather data at two locations in Brahmaputra River Basin using an improved K-NN model developed by Sharif and Burn [12]. The intent was to employ the improved K-Nearest Neighbour (KNN) to produce a variety of synthetic weather sequences that can be used as an input into hydrological models. Particular emphasis is laid on the simulation of weather sequences that model unprecedented precipitation events in the basin as the simulation of extreme precipitation is crucial for modelling flood generation mechanism in the basin.

5. Methodology

The improved K-NN weather generating model [12] was applied to perform two different types of simulations at two climate stations, namely Cheerapinji and Guwahati in the Brahmaputra basin. A K-NN algorithm typically involves selecting a specified number of days similar in characteristics to the day of interest. One of these days is randomly resampled to represent the weather of the next day in the simulation period. Despite their inherent simplicity, nearest neighbour algorithms are considered versatile and robust. These methods have been intensively investigated in the field of statistics and in pattern recognition procedures that aim at distinguishing between different patterns. The nearest neighbour approach involves simultaneous sampling of the weather variables, such as precipitation and temperature. The sampling is carried out from the observed data, with replacement. To simulate weather variables for a new day $t+1$, days with similar characteristics as those simulated for day t are first selected from the historical record. One of these nearest neighbours is then selected according to a defined probability distribution or kernel and the observed values for the day subsequent to that nearest neighbour are adopted as the simulated values for day $t+1$. A series of simulation of weather data in the basin were performed to

produce extreme precipitation events at different stations in the basin. Model runs were carried out to simulate 800 years of synthetic data based on the driving data set for the model.

6. Results and Discussion

Two different types of simulation were conducted in this research. The duration of wet spells and duration of dry spells were simulated. For each simulation box plots have been used to present the statistics of interest. Box plots are a favored method of data analysis in many hydrological applications as they show the range of variation in statistics of simulations and provide a straightforward method of comparing the statistics of simulations with historical data. The bottom and top horizontal lines in the box in a box plot indicate the 25th and 75th percentile, respectively, of the statistics computed from the simulated data. The horizontal line within the box represents the median. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. The whisker extends to the most extreme data value within 1.5 times the inter-quartile range of the data. The values beyond the ends of the whiskers are called outliers and are shown by dots. The statistics of the historical record are represented by dots and joined by solid lines.

6.1. Wet Spells Simulation

Analysis of statistics of wet days is important as it gives an indication about the ability of the model to reproduce the persistence structure of the underlying data. For each year in the historical and simulated series, the most extreme spell of wet days is determined. Thus, boxplots were created using 32 values for the historical data and 800 values for the simulated data. Figure 5 shows the distribution of the extreme wet spells for the historical as well as the simulated data at Cheerapinji. For the historical data, the median of wet spells duration was

around 65 days, whereas it was around 100 days for the simulated data. As expected, the median of the simulated data is substantially higher than the median of the historical data. A single extreme wet spell with a duration of around 180 days was simulated. This extreme spell lies beyond the whiskers which are at 1.5 times the inter-quartile range of the simulated data. Such extreme spells are particularly crucial for the analysis of flooding events in the basin. For Guwahati, which has a substantially smaller annual average precipitation and wet days compared to Cheerapunji, the simulations

produced relatively smaller median. The median for the historical data was of the order of 30 days for the historical data, whereas it was around 40 days for the simulated. This was expected as the improved model tends to perturb the data points to produce values that are not present in the observed record. The boxplots for the duration of extreme annual wet spells at Guwahati are shown in Figure 4. At Guwahati, several events beyond the whiskers were produced by the KNN model with the most extreme event being of the order of 100 days.

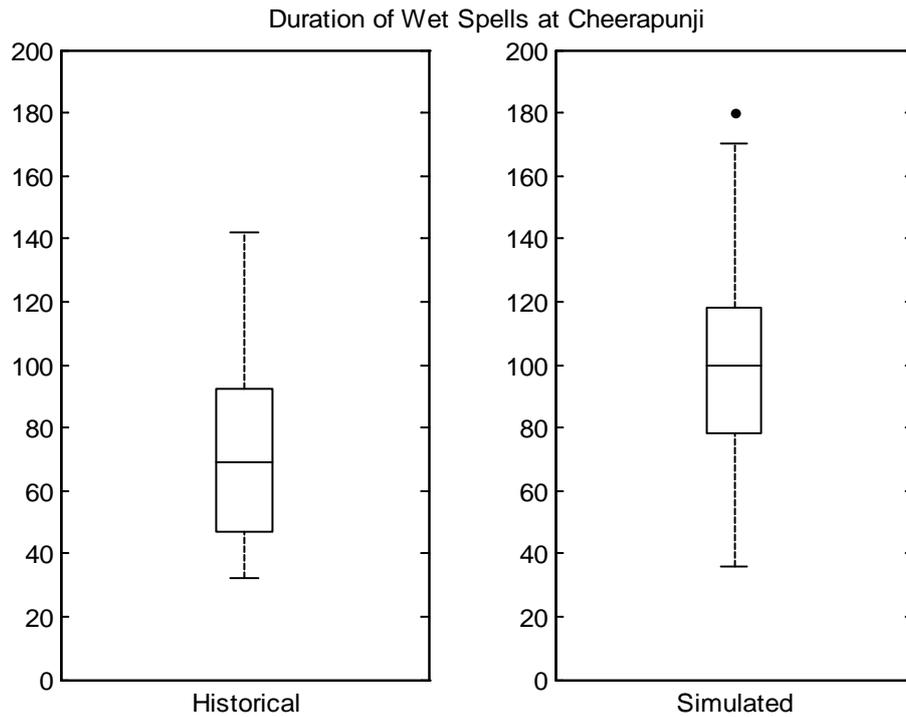


Figure 3. Box plots of extreme wet spells at Cheerapunji

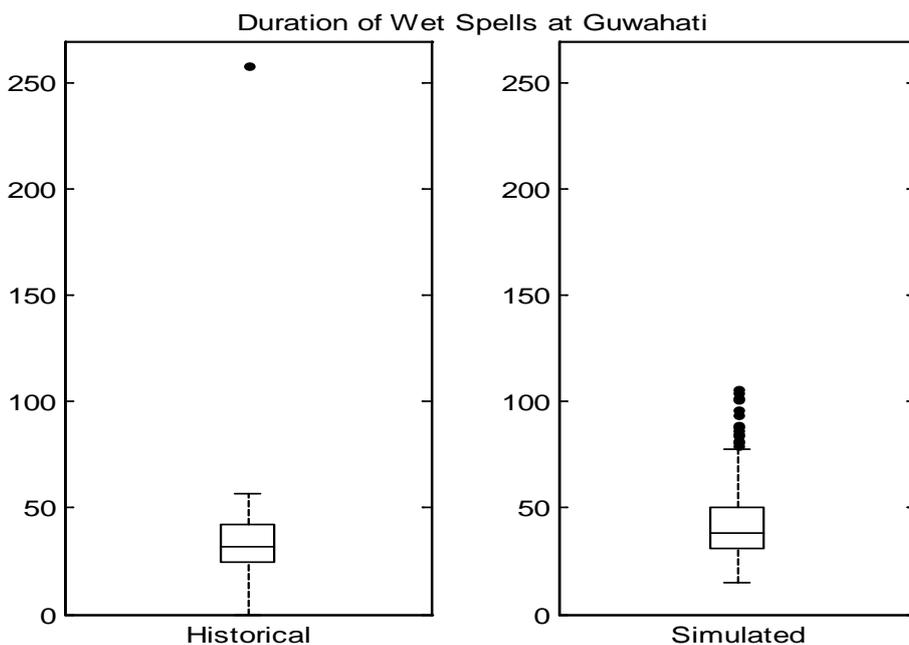


Figure 4. Box plots of extreme wet spells at Guwahati

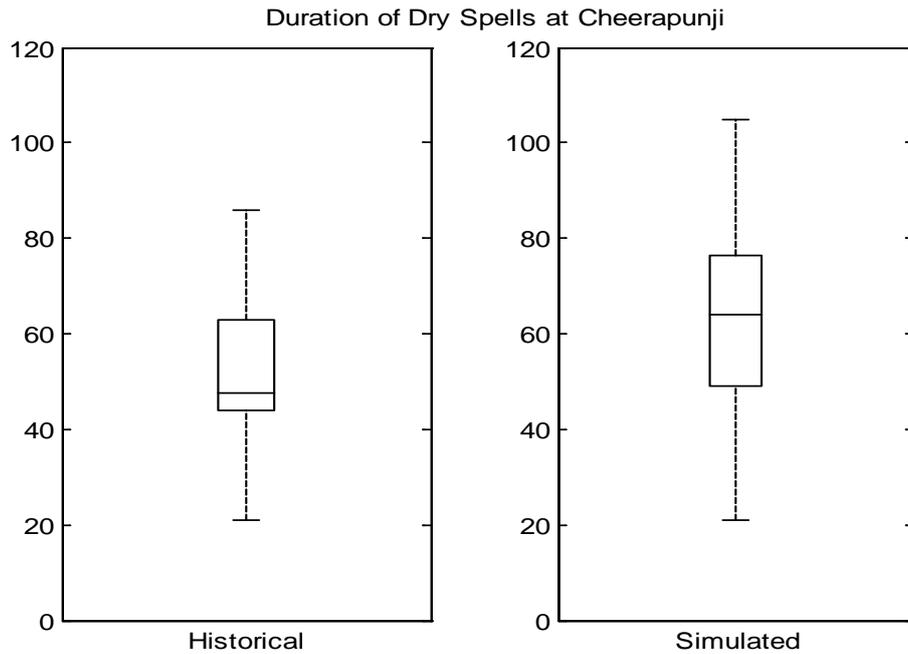


Figure 5. Box plots of extreme dry spells at Cheerapunji

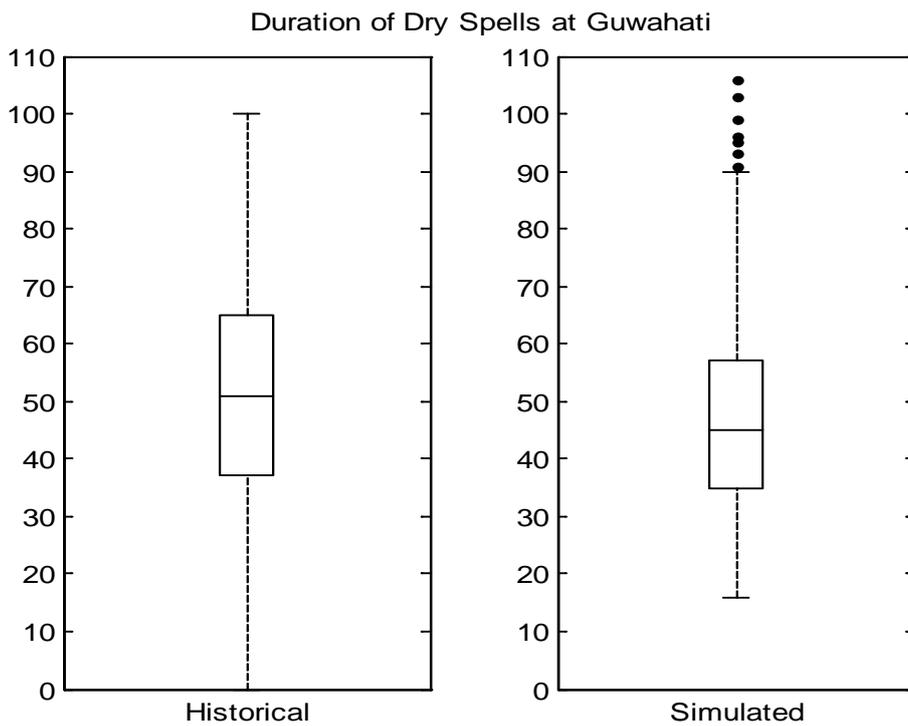


Figure 6. Box plots of extreme dry spells at Guwahati

6.2. Dry Spells Simulation

It is important to determine dry spell characteristics of the simulated data in order to assess the risks associated with drought in the basin under future climatic conditions. Figure 5 presents box plots of the total number of days during extreme dry spells in each year of the observed as well as simulated record. The first box plot in Figure 5 presents the distribution of extreme dry spells computed from 32 years of observed data, whereas the second boxplot is based on 800 years of simulated data. As can be seen from the box plot in Figure 5, the median dry spell

duration for the observed data is 46 and the corresponding statistic in the simulated data is 65. The higher median value produced by the KNN model may be attributed to the nature of the improved model, which tends to produce more severe events than observed in the historical data. However, unlike the wet spells no dry spells with durations exceeding 1.5 times the inter-quartile range were simulated by the model. The boxplots of dry spells at Guwahati for the historical and simulated data are shown in Figure 6. For the historical data, the median dry spell duration was around 50 days, whereas it was slightly less than 50 days for the simulated data. The KNN model was

able to simulate several events with durations exceeding 90 days. Use of weather sequences simulated by the KNN model would lead to better reliability in assessing the vulnerability of basin to drought events. An encouraging aspect of the model used herein is that extreme unprecedented events, both low precipitation and high precipitation, can be simulated. This allows for evaluation of the response of rainfall–runoff models for a wide variety of simulated extremes.

7. Conclusions

With the improved K-NN model, a series of unprecedented dry spells that were not seen in the historical record were produced. The incorporation of such extreme spells as an input to the hydrological model would increase the reliability of simulation of flooding events in the basin. As in the case of wet spells, the model simulated several events that are more severe than present in the observed data, thus providing a wider range of events as input to a hydrologic model. These results clearly indicate that the model produced extreme dry spells more severe than those observed in the historical record. It is clear from the results of simulation of dry spells that greater variability is associated with sustained periods of precipitation and dry days than is present in the observed data. It can be concluded that the use of weather sequences simulated by the KNN model would lead to better reliability in assessing the vulnerability of the basin to drought events. An encouraging aspect of the model used herein is that extreme unprecedented events, both low precipitation and high precipitation, can be simulated. This allows for evaluation of the response of rainfall–runoff models for a wide variety of simulated extremes.

References

- [1] Bardossy, and A. Plate E.J. (1991). Modelling daily rainfall using a semi-Markov representation of circular pattern occurrence. *J. of Hydrol.* 122: 33-47.
- [2] Buishand, T.A., Brandsma. T., (2001). Multisite simulation of daily precipitation and temperature in the Rhine Basin by nearest-neighbor resampling. *Water Resources Research* 37(11), 2761-2776.
- [3] Davis, J., (1986). *Statistics and Data Analysis in Geology*, John Wiley, Hoboken, N. J.
- [4] Dubrovsky M, Nemesova I, Kalvova J. (2005). Uncertainties in climate change scenarios for the Czech Republic. *Clim Res* 29: 139-156.
- [5] Gangopadhyay, S, Rajagopalan, B, Clark, M. (2005). Statistical downscaling using K-nearest neighbors, *Water Resources Research* Volume 41, Issue 2, Article first published online: 18 FEB.
- [6] Hutchinson, M.F. (1995). Stochastic space time weather models from ground based data. *Agric. Forest Meteorol.*, 73: 237-264.
- [7] IPCC. (2014). The Synthesis Report (SYR), constituting the final product of the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC), is published under the title *Climate Change*.
- [8] Jenkins, G. J. & Derwint, R. G. (1990). Climate consequences of emissions. In *Climate Change. The IPCC scientific assessment* (ed. by J. T. Houghton, G. J. Jenkins & J. J. Ephraums). WMO, UNEP, Cambridge University Press, UK.
- [9] Jiang, et al. (2007). Comparison of hydrologic impacts of climate change simulated by six hydrologic models in the Dongjiang Basin, South China. *J. Hydrol.* 336, 316-333.
- [10] Muluye Getnet Y. (2011). Deriving meteorological variables from numerical weather prediction model output: A nearest neighbor approach. *Water Resources Research* 47:7, n/a-n/a. online publication date: 1-Jul-2011.
- [11] Rabus B, Eineder M, Roth A, Bamler R. (2003). The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *Journal of Photogrammetry and Remote Sensing* 57: 241-262.
- [12] Rajagopalan, B., Lall, and U., (1999). A k-nearest neighbour simulator for daily precipitation and other variables. *Water Resources Research* 35(10), 3089-3101.
- [13] Semenov, M. A. (2008). Simulation of extreme weather events by a stochastic weather generator, *Clim Res*, 35: 203-212.
- [14] Sharif M, Burn DH (2007). Improved K-nearest neighbor weather generating model. *J Hydrologic Eng., ASCE*, 12(1): 42-51.
- [15] Simonovic, S.P., Li, L., (2003). Methodology for Assessment of Climate Change Impacts on Large-Scale Flood Protection System. *ASCE Journal of Water Resources Planning and Management* 129(5), 361-372.
- [16] Yates et al. (2003). *Water Resources Research*, VOL. 39, NO. 7, 1199.