Proceedings of 12th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2022 Month: October Volume: 12

Quality Assessment of Temporal and Spatial Input Data for Water Resources Modelling

Koshish Raj Maharjan [a], Utsav Bhattarai [b], Pawan Kumar Bhattarai [a]

[a] *Department of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[b] *University of Southern Queensland, Toowoomba, Queensland, Australia*
[a] koshishrajmaharjan@gmail.com

Abstract

Performance of hydrological, hydraulic and other water resources models is largely dependent on the quality of the temporal and spatial input data. That is why quality assessment of these data during the pre-processing is extremely important. In this study, we applied the Standard Normal Homogeneity Test and Pettitt's Test to assess the homogeneity of the precipitation and temperature observations from the climate stations in the Kaligandaki River Basin as a case. The precipitation series for the selected stations were found to be homogeneous in most of the cases except for Musikot. On the other hand, only half of the selected temperature stations were found to be homogeneous. The Digital Elevation Model (DEM) is the most important spatial data required for modelling. Therefore, vertical accuracy of five freely available space-borne DEMs (ASTER, SRTM, Copernicus, AW3D30 and NASA) were assessed by comparison with the DEM generated using contour and spot heights from the Department of Survey, Government of Nepal. Based on visual comparison and statistical analysis, AW3D30 DEM was found to be the most accurate in the Kaligandaki River Basin. The applicability of this study is beyond the selected area and shall be useful for data quality control for hydro-meteorological modelling in Nepal.

Keywords

Quality assessment, Standard Normal Homogeneity Test, Pettitt's Test, Vertical Accuracy, Digital Elevation Model

1. Introduction

"Garbage in, garbage out" is a common phenomenon in modelling studies. The accuracy and reliability of the model is highly dependent on the quality of data used [1]. Modelling the hydrology, hydro-meteorological extremes such as floods and droughts, and climate change is no different.

Modern hydro-meteorological models generally require spatial and temporal data. The variation in an ideal homogeneous climate series, is solely due to an influence of the climate variation [1]. The quality of temporal data can be assessed in terms of homogeneity in their series. These records may contain inhomogeneities because of station relocation, equipment alterations, and changes in the data collection method, among others [2]. As a result, the observed data series are laden with faulty or missing data which lead to high chances of erroneous model outputs. There are different methods such as Normal Ratio, Multiple Regression, Gridded data, etc. to correct such data inadequacies.

The homogeneity can be assessed either by an absolute or a relative approach for a time series [1, 3]. In absolute method, the test is based only on the time series of the single station, whereas in relative method, the contribution of neighboring stations is considered, to isolate non-climatic influence [4, 5, 6]. There are various methods to implement a homogeneity test of such climatic records. Mihajlović [7] tested the homogeneity of monthly precipitation over the Pannonian part of the Croatia, using the Standard Normal Homogeneity Test (SNHT). Wijngaard et al. [4] used the SNHT, Pettitt Test, Buishand Range Test and Von Neumann Ratio, to test the homogeneity of European daily precipitation and temperature series. Firat et al. [1] used SNHT, Swed-Eisenhart Runs and Pettitt tests to assess the homogeneity in the temperature series of the stations of Turkey, using the annual mean temperatures.

As with temporal data, spatial data may contain inaccuracies arising from the data acquisition method, level of ground truthing and the degree of correction applied. These also need to be corrected using possible rectification methods before using in models. It is extremely important in data such as Digital Elevation Models (DEMs) which form the core of hydrological and hydrodynamic analysis [8]. DEMs are usually applied to calculate the slope, aspect, or delineate a watershed for hydrological studies, and the quality of results depends upon their accuracy. These DEMs can be generated via techniques such as photogrammetry, interferometry, laser scanning, aerial stereo images and topographic surveys [9]. DEMs from ground based topographic surveys, drone-based photogrammetric or air-borne laser scans, are the best for detailed studies but, they are infeasible to be used in developing countries which generally rely on freely available ones [8]. When multiple data sources are available, the best approach would be to choose the source which best fits the observed data. Some freely available DEMs include Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER), Shuttle Radar Topography Mission (SRTM), Copernicus, NASA DEM, Advanced Land Observing Satellite (ALOS) World 3D (AW3D30) among many others. Talchabhadel et al. [8] assessed AW3D30, ASTER Global DEM Version 3, SRTM-30, and Carto DEM v3 in the West Rapti river basin and concluded AW3D30 DEM to be the better choice. Gautam et al. [10] evaluated the differences in performances of a hydrological model for Kaligandaki and Bagmati basin, under different elevation sources: Contour of Nepal, ASTER Global DEM v2 and SRTM v4.1 and assessed the performance of these elevation sources.

The overall objective of this study is to carry out quality assessment of the temporal and spatial data taking the case of the Kaligandaki Basin. The specific objectives are:

- To assess the homogeneity of the climate variables: precipitation and temperature (maximum and minimum) and
- To assess the most applicable space-borne DEM among a set of freely available alternatives (ASTER, SRTM, Copernicus, AW3D30 and NASA DEM) by comparing their vertical accuracies.

This study could act as a basis for data quality control among hydro-meteorological modelers which can be conveniently replicated in other basins of Nepal.

2. Study Area

This study is carried out in the Kaligandaki River Basin (27.75°N to 29.33°N latitudes; 82.83°E to 84.25°E longitudes) with a drainage area of 10,590 km^2 at the outlet - Department of Hydrology and Meteorological (DHM) hydrological station 419.1, Ansing. The basin mainly spreads over Nepalese districts: Mustang, Myagdi, Baglung, Gulmi, Syangja, Parbat, Kaski and Palpa, along with minor percentage of area in Tibet. The elevation in the catchment ranges from 336 to 8188 meters above mean sea level. The annual precipitation ranges from 114 mm to 5527 mm, with the lowest rainfall being at the higher elevations [11]. The minimum temperature in the higher mountains may even drop to below -25°C while the maximum temperature in the lower parts of the basin might even exceed 35°C in the summer [12].

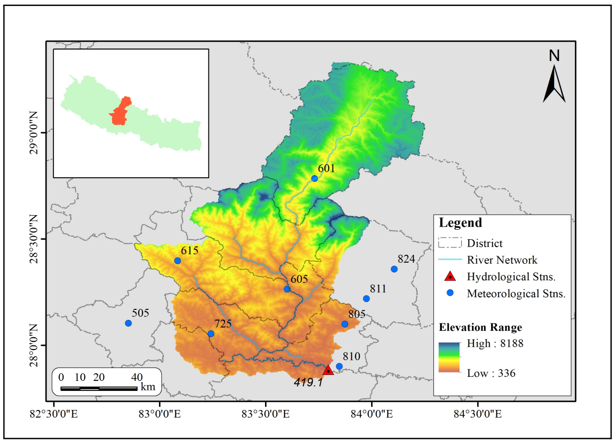

Figure 1: Location map of the Kaligandaki River Basin

3. Data and Methodology

3.1 Data

The observed times series data for daily precipitation, minimum and maximum temperature are obtained from DHM. A period of 30 years (1985 to 2014) is taken as the baseline. Among the available stations, only those are considered which have less than 15% missing data during the period of analysis and are listed in Table 1 and Table 2.

The freely available space-borne DEMs of 30m x 30m resolution, considered for an assessment in the study along with their download source, is listed in Table 3.

Table 1: Precipitation stations with their missing values

DHM Stn. No.	Name	Lat. (°N)	Lon. (°E)	Missing %
505	Bijuwatar	28.10	82.87	1.96
601	Jomsom	28.78	83.72	0.36
605	Baglung	28.23	83.60	10.74
615	Bobang	28.40	83.10	4.44
722	Musikot	28.17	83.27	0.27
810	Chapkot	27.83	83.82	0.25
824	Siklesh	28.37	84.10	0.00

Table 2: Temperature stations with their missing values

				Missing %	
DHM Stn. No.	Name	Lat. (°N)	Lon. (°E)	Tmax	Tmin
601	Jomsom	28.78	83.72	3.86	0.00
725	Tamghas	28.07	83.25	10.88	10.22
805	Syangja	28.10	83.88	1.38	3.71
810	Chapkot	27.88	83.82	1.07	9.35
811	Malepani	28.12	84.12	5.58	9.86

Table 3: List of DEMs assessed in this study

SN	DEM	Download Source
1	SRTM-30	OpenTopography (https://portal.open topography.org/datasets)
2	NASA DEM-30m	
3	ALOS World 3D -30m (AW3D30)	
4	Copernicus GLO-30	
5	ASTER GDEM v3	Earth Explorer-USGS (https://earth explorer.usgs.gov/)

3.2 Homogeneity Tests of the Climatic Time Series

The data gaps of the selected stations are filled using the gridded APHRODITE (Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation) datasets. The filled climate series are then tested for homogeneity at an annual scale. The two methods used in this study to test an absolute homogeneity are briefly discussed subsequently. The test results are classified as done by Wijngaard et al. [4]: (Useful) if the series satisfies the null hypothesis of both tests, (Doubtful) if the series satisfies the null hypothesis of only one test, and (Suspect) if the null hypothesis of both tests are rejected.

3.2.1 Standard Normal Homogeneity Test (SNHT)

To carry out the SNHT, a statistic T(k) is considered which describes the mean of first 'k' years of the record with the last '(n-k)' years of record [13].

$$T(k) = k\bar{z}_1^2 + (n-k)\bar{z}_2^2, \quad k = 1,2,....,n \tag{1}$$

where,

$$\bar{z}_1 = \frac{1}{k}\sum_{i=1}^{k}\frac{(Y_i - \bar{Y})}{s} \tag{2}$$

$$\bar{z}_2 = \frac{1}{(n-k)}\sum_{i=k+1}^{n}\frac{(Y_i - \bar{Y})}{s} \tag{3}$$

$$s = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{4}$$

The value of the statistic T(k) is maximum at the year 'k' if the break is in that year and the test statistic T_0 is given by maximum of the statistic T(k) [4, 13].

$$T_0 = \max_{1 \le k \le n} T(k) \tag{5}$$

The SNHT test indicates the inhomogeneity if this test statistic T_0 is greater than the critical value, which depends upon the sample size. The critical value at 99% significance level [2], is 10.153 for a sample size of n=30.

3.2.2 Pettitt's Test

Pettitt's method is a non-parametric rank test in which the ranks r_i of the Y_i are used to calculate the statistic X_k [4, 5].

$$X_k = 2\sum_{i=1}^{k}(r_i - k(n+1)), \quad k = 1,2,....,n \tag{6}$$

The statistic X_k is maximum at the year 'k' when the break occurs and the test statistic X_E is given by:

$$X_E = \max_{1 \le k \le n}|X_K| \tag{7}$$

The Pettit's method indicates inhomogeneity in the time series if the value of test-statistic is greater than the critical value, which depends upon the sample size. The critical value at 99% significance level [4], is 133 for a sample size of n=30.

3.3 Vertical Accuracy Assessment of DEMs

Vertical accuracy assessment is conducted comparing the elevations of the selected space-borne DEMs at defined points with the DEM generated using contours and spot heights published by the Department of Survey (DoS), Government of Nepal. A total of 1025 sample points are taken along the river longitudinally at an interval of 5 km and also across at 10 km interval along the Kaligandaki River and its major tributaries (Myagdi River, Modi River and Badi Gad River) (Figure 2). The cross sections were taken up to a width of 500 m. To examine how the DEMs compared with the actual values of DoS, two evaluation metrics, namely, Root Mean Square error and Mean error is computed. The DEM with the lowest errors is considered as the best DEM for future application in the study area.

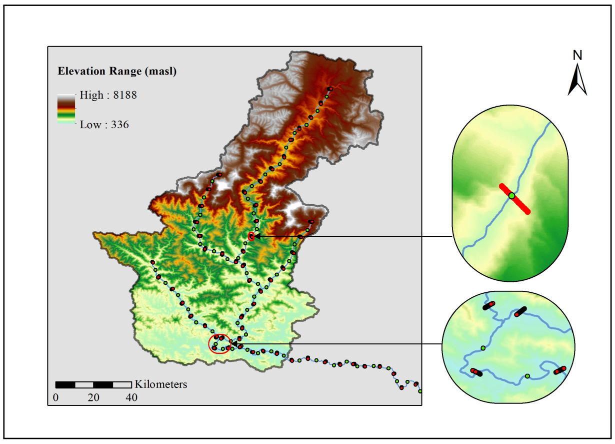

Figure 2: Sample points in the Kaligandaki Basin considered for DEM assessment

4. Results and Discussion

4.1 Homogeneity Tests of the Climatic Records

The homogeneity test results for the annual precipitation and temperature (maximum and minimum) of the stations listed in Table 1 and Table 2, are given in Table 4, Table 5 and Table 6. The **bold** figures represent the inhomogeneous series at 99% significance level and the year in parenthesis is the detected break years. Based on the results for each of the test, the series for each station is classified as mentioned under the heading 3.2.

In case of precipitation series, all of the considered stations are homogenous for both tests, except for the station: 722 at Musikot, which has been classified to be suspect.

Table 4: Homogeneity test results of precipitation stations

DHM Stn. No.	SNHT	Pettitt's Test	Remarks
505	3.34	73	Useful
601	6.92	116	Useful
605	2.52	65	Useful
615	5.01	105	Useful
722	**14.89 (1994)**	**184 (1994)**	**Suspect**
810	2.60	85	Useful
824	1.49	58	Useful

The changes in the mean annual precipitation of the 'suspect' station 722 at Musikot as per SNHT, is shown in Figure 3. As the precipitation station 722 has been classified as 'suspect', it would be unwise to use this station for the hydro-meteorological studies.

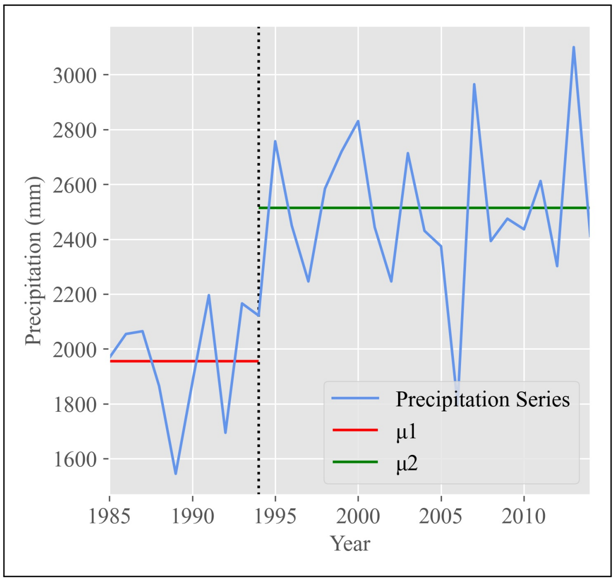

Figure 3: Changes in mean annual precipitation at station 722, Musikot: μ_1=1954.99 mm and μ_2=2514.28 mm

For the maximum temperature, only two stations: 601 and 810 are classified as useful, as only two of these succeed both SNHT and Pettitt's test while the other three stations were either doubtful or suspect. The changes in the mean annual maximum temperature of 'suspect' station 725 at Tamghas, as depicted by SNHT is shown in Figure 4.

Table 5: Homogeneity test results of maximum temperature stations

DHM Stn. No.	SNHT	Pettitt's Test	Remarks
601	5.35	105	Useful
725	**14.05 (2004)**	**181 (2003)**	**Suspect**
805	**12.85 (2013)**	75	*Doubtful*
810	5.2	114	Useful
811	**16.42 (2003)**	**195 (2003)**	**Suspect**

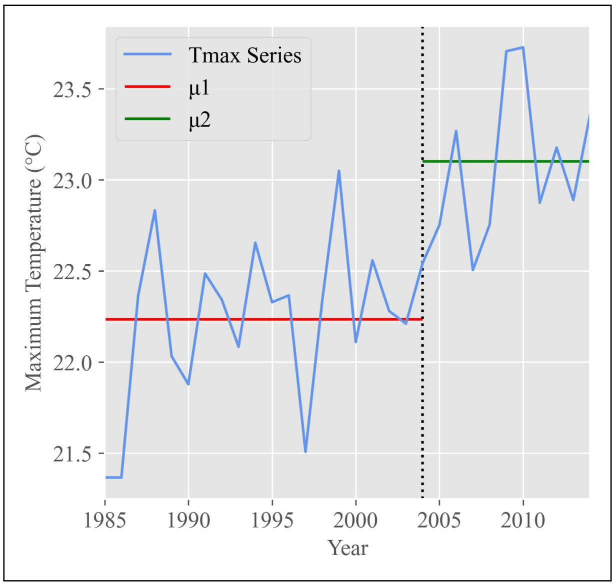

Figure 4: Changes in mean annual maximum temperature at station 725, Tamghas: μ_1=22.24 °C and μ_2=23.10 °C

Likewise, the minimum temperature series of the stations: 725 and 810 are classified as useful as these are homogeneous under both tests while rest of the stations are classified as suspect according to their homogeneity results. The changes in the mean annual minimum temperature of the 'suspect' station 601 at Jomsom, as depicted by SNHT is shown in Figure 5.

The use of these 'Suspect' and 'Doubtful' stations should be avoided as far as possible for the accuracy and reliability of the work. Even though the study on homogeneity assessment has been more focused on the precipitation and temperature variables, similar methods of assessment can be also conducted for other sets of climate and flow variables.

Table 6: Homogeneity test results of minimum temperature stations

DHM Stn. No.	SNHT	Pettitt's Test	Remarks
601	**14.02 (1994)**	**168 (1994)**	**Suspect**
725	5.06	99	Useful
805	**11.78 (2001)**	**165 (2001)**	**Suspect**
810	5.94	108	Useful
811	**14.56 (1997)**	**191 (1997)**	**Suspect**

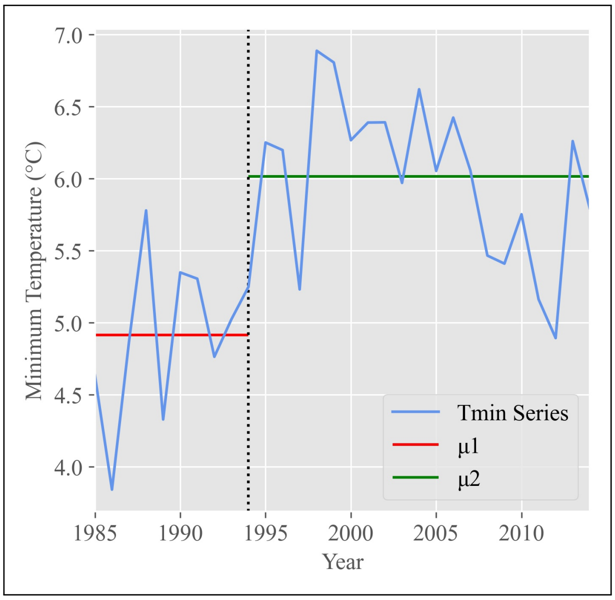

Figure 5: Changes in mean annual minimum temperature at station 601, Jomsom: μ_1=4.91 °C and μ_2=6.01 °C

4.2 Vertical Accuracy Assessment of DEMs

The summary results after the assessment of all DEMs is listed in Table 7. All space-borne DEMs under study shows a quality correlation with the DEM generated from the topo map, as shown in Figure 6.

Although the mean error for ASTER DEM is the least, it had the highest RMSE of 23.1m and standard deviation of 23.1m. The AW3D performed better than others, in terms of both RMSE (13.0m) and standard deviation (12.3m). In addition to this, the mean error of -4.2m for AW3D was the second best among the DEMs. In case of the SRTM DEM, it underestimated the elevation, especially in case of low regions of the study area while the NASA DEM and Copernicus DEM overestimated the elevation in those regions. The RMSE and standard deviation of the latter three DEMs are in the similar spectrum, but aren't the best one. Based on the error statistics listed in Table 7, the AW3D30 DEM is declared as the best DEM from the pool for the study area.

Table 7: Error Statistics (in metres) of vertical accurac -y of DEMs with respect to Topo maps (n=1072)

	DEM				
	SRTM	AST.	COP.	AW3D	NASA
Min	-44.7	-61.6	-49.2	-38.7	-41.4
Max	27.9	58.4	32.9	28.9	25.9
$\bar{x}$	7.1	0.5	-7.1	-4.2	-6.7
σ	12.4	23.1	14.7	12.3	12.4
RM SE	14.3	23.1	16.4	13.0	14.1
R^2	99.90	99.98	99.99	99.99	99.99

AST.=ASTER and COP.=Copernicus

Here $\bar{x}$, σ, RMSE and R^2 represents mean error, standard deviation of error, Root Mean Square Error and coefficient of determination respectively.

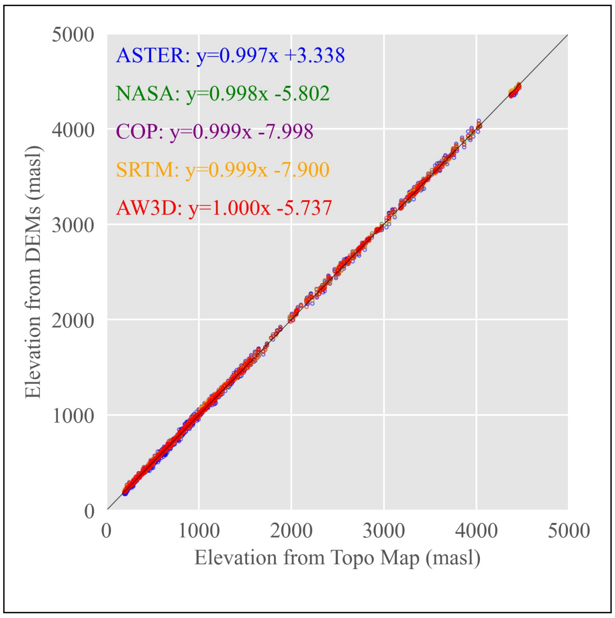

Figure 6: Comparison of space-borne DEMs with the elevation data from topo map

The differences in elevation of the DEMs can be visualized with a sample cross-section of Kaligandaki river, given in the Figure 7 and hints at the 'potential' difference in the results of a study with the variation of DEMs.

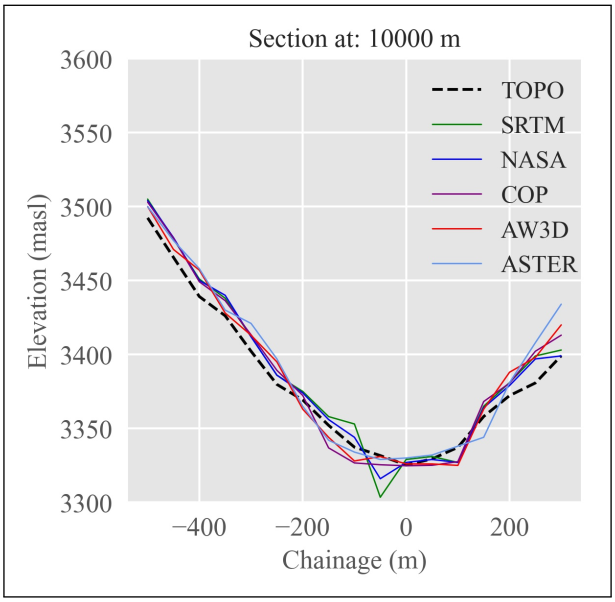

Figure 7: Sample cross-section at the Kaligandaki river using DEMs

5. Conclusion

The homogeneity is tested for the precipitation and temperature series (maximum and minimum) of the meteorological stations considered for the Kaligandaki river basin over the period 1985-2014, using Standard Normal Homogeneity Test and Pettitt's Test at an annual scale. The results of the study revealed that the precipitation series across the basin were homogeneous in most of the cases. While in case of temperature series, there was an inconsistency among the stations. Even when the record for the maximum temperature is homogeneous, the minimum temperature series at the same station is inhomogeneous. The year breaks in each of the inhomogeneous series were detected and its graphical abstract also verified the inhomogeneity. If the historical metadata of the inhomogeneous stations are available, the break points can be verified and the series can also be corrected. One of the objectives of this study is to encourage the use of such simple and effective methods to validate the quality of the temporal data by quantifying the homogeneity.

Upon the assessment of several space-borne DEMs (ASTER GDEM v3, SRTM-30, Copernicus GLO-30, ALOS World 3D-30m (AW3D30) and NASA DEM-30m) for the Kaligandaki river basin, the ALOS World 3D DEM (AW3D30) showcased itself as the suitable option for the topographic data in the study area with the least RMSE, mean error and the better correlation to the national topographic database. The difference in the ratings of these DEMs shows that it is essential to consider the vertical accuracy assessment before the topographic data from the DEMs are used for any analysis.

References

[1] M Firat, F Dikbas, A Koc, and M Gungor. Analysis of temperature series: estimation of missing data and homogeneity test. *Met. Apps*, pages 397–406, 2012.

[2] M Khaliq and TB Ouarda. On the critical values of the standard normal homogeneity test (SNHT). *Int. J. Climatol*, 27(5):681–687, 2007.

[3] NH Ahmad and SM Deni. Homogeneity test on daily rainfall series for Malaysia. *MATEMATIKA*, 29(1c):141–150, 2013.

[4] J Wijngaard, AK Tank, and G Konnen. Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology*, 23:679–692, 2003.

[5] K Ahmed, S Shahid, T Ismail, N Nawaz, and XJ Wang. Absolute homogeneity assessment of precipitation time series in an arid region of Pakistan. *Atmósfera*, 31(3):301–316, 2018.

[6] AG Dhorde and M Zarenistanak. Three-way approach to test data homogeneity: An analysis of temperature and precipitation series over southwestern Islamic Republic of Iran. *J. Ind. Geophys. Union*, 17(3):233–242, 2013.

[7] D Mihajlović. Monitoring the 2003–2004 Meteorlogical drought over Pannonian part of Croatia. *International Journal of Climatology*, 26:2213–2225, 2006.

[8] R Talchabhadel, H Nakagawa, K Kawaike, K Yamanoi, and BR Thapa. Assessment of vertical accuracy of open source 30m resolution space-borne digital elevation models. *Geomatics, Natural Harards and Risk*, 12(1):939–960, 2021.

[9] A Mondal, D Khare, S Kundu, S Mukherjee, A Mukhopadhyay, and S Mondal. Uncertainty of soil erosion modelling using open source high resolution and aggregated DEMs. *Geosci Front*, 8(3):425–436, 2017.

[10] S Gautam, V Dahal, and R Bhattarai. Impacts of Dem Source, Resolution and Area Threshold Values on SWAT Generated Stream Network and Streamflow in Two Distinct Nepalese Catchments. *Environmental Processes*, 2019.

[11] M Shrestha, SC Acharya, and PK Shrestha. Bias correction of climate models for hydrological modelling – are simple methods still useful? *Meteorological Applications*, 24(3):531–539, 2017.

[12] B Mishra, MS Babel, and NK Tripathi. Analysis of climatic variability and snow cover in the Kaligandaki River Basin, Himalaya, Nepal. *Theoretical and Applied Climatology*, 116:681–694, 2014.

[13] H Alexandersson. A homogeneity test applied to precipitation data. *Journal of Climatology*, 6(6):661–675, 1986.