Proceedings of 14th IOE Graduate Conference
Peer Reviewed
Year: 2023 Month: December Volume: 14
ISSN: 2350-8914 (Online), 2350-8906 (Print)

Performance of the ERA5 data in the Snowfed region : A case study in Langtang River Catchment

Ashish Devkota [a], Pawan Kumar Bhattarai [a], Susen Shrestha [b], Asha Bhatta [c]

[a] *Department of Civil Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[b] *Climate Change and Transformation, EURAC Research, Bolzano, Italy*
[c] *Faculty of Science and Technology, Pokhara University*
[a] 078mswre002.ashish@pcampus.edu.np, [a] nepal.pawan@gmail.com, [b] sshrestha@eurac.edu, [c] bhattaasha94@gmail.com

Abstract
Hydro-meteorological observation remains a significant challenge in any mountainous catchment due to its difficult terrain. The Hindu Kush Himalaya region is no exception, and the limited number of stations in Nepal further increases the challenges associated with data acquisition. These limitations hinder comprehensive studies related to climate and hydrological science for multiple purposes in Nepal. Hence, the primary objective of this study is to introduce and assess the state-of-the-art fifth-generation ERA-5 reanalysis dataset for runoff reproduction in the high-altitude catchment of the Langtang River basin. In the first part of the study, we focused on the characteristics of ERA5 precipitation and temperature, which are presented here. ERA5 precipitation exhibited an overall positive bias of 1.68 primarily due to the overprediction of low-intensity rainfall. While, ERA5 temperature showed a cold bias, which could be attributed to its spatial scale. The rapidly changing orography of this region makes it challenging for ERA5 to accurately capture these features.The closest ERA5 station with the Kyanging exhibited a precipitation bias of 0.0379 during the calibration period(2008 to 2010), while a bias of 0.078 was observed for the validation period (2011 to 2013). Similarly, for temperature, differences of -0.008 and -1.01 were noted between the calibration and validation periods. Nevertheless, an acceptable temperature lapse rate of 0.0550℃/m was observed, considering the extreme ERA5 grids. In case of temperature, a monthly additive corrections was evaluated based on the calibration period and were consistently applied throughout the whole timeseries. Following these corrections, the precipitation bias was reduced to 0.037 and 0.078 for the calibration and validation period but the KGE is underperforming for both the cases. While the temperature performance was also enhanced with temperature difference diminishing to -0.008(Calibration) and -1.01(validation).The performance of the reanalysis precipitation data shows that it is unable to capture the feature even the bias is reduced significantly even after the correction in high altitude region and the temperature data is acceptable for application in the data scarce region after correction.

Keywords
ERA5, lapse rate, reanalysis

1. Introduction

Despite the importance of observed data for hydrological sciences, the historical climatology network of monthly temperature datasets has seen a net decrease in the number of stations since the turn of the twenty-first century[1]. The promise of relevant hydrometeorological data being made available globally, across vast swaths of land, and in places with sparse or nonexistent observational networks has long been fulfilled by remote sensing datasets. Remote sensing datasets have long held the promise of delivering relevant hydrometeorological data across vast areas of land, up to the global scale, and over areas where observational networks are absent or sparse[2]. The European Centre for Medium-Range Weather Forecasts (ECMWF) has released the ERA5 dataset, the most complex reanalysis output to date. It was developed using methods that gave it several advantages over the previous release, the ERA-Interim reanalysis tool. Notably, it incorporates more data sources, uses a more complex assimilation system, has a finer spatial resolution, and is stored at the hourly time step.Reanalysis incorporates a wide range of measurable and remotely sensed data into a dynamically and physically connected numerical model. They use the analytical component of a weather forecasting model, which is driven by data assimilation to the closest possible representation of the current state of the atmosphere. A reanalysis is a retrospective examination of historical data that employs more recent iterations of numerical models and assimilation schemes, as well as ever-increasing computational resources.[3].This is thought to be caused in part by reanalysis, which has a rather coarse spatial resolution.

Numerous studies have shown that the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis, ERA-Interim, is one of the best-performing reanalysis products[4]. A glacier mass balance-runoff model was used for the reconstruction using daily air temperature and precipitation fields from ERA5 reanalysis files. Modules for temperature-index, accumulation, and rain were included in this model[5].The majority of this is explained by the large temporal and spatial gaps in the global watershed discharge observing network. In many parts of the world, there are simply not enough long-term river discharge observations at high enough spatial densities, and the vast majority of countries lack access to real-time hydrometric data[6].Reanalysis has the advantage of producing a large number of variables at various atmospheric levels, rather than just at the land surface. The majority of the data assimilated in a reanalysis come from the atmosphere and ocean; surface data, including weather station data, are rarely used. As a result, reanalysis results can potentially provide surface variables in areas with little to no surface coverage and are not dependent on the density of surface observational networks. Several modeling organizations now provide reanalysis in a variety of spatial and temporal dimensions[7, 8].The ERA5 reanalysis can be used as a potential reference dataset by substituting the precipitation and temperature data from ERA5 for the observations made during the hydrological modeling process[9].ERA-I introduced a sophisticated four-dimensional variational analysis assimilation technique with a 12-hour time step. It calculates 60 vertical levels from the surface up to 0.1 hPa. It has a horizontal resolution of about 80 km. Temperature and precipitation data from a 12-hour time step were aggregated to the daily scale in this study. Because ERA-production I's will end in August 2019, it will cover the time period from January 1, 1999 to that month[10].

The lack of observation sites makes it difficult to obtain high-quality data for mountainous basins like Langtang[11].The performance of gridded precipitation products over mountainous areas is especially important due to the lack of observations[12].So for the data scarce region with the very high spatial gap for the meteorological station the gridded datasets could be crucial to understand the hydrometeorological behavior of the catchment. The main objective of the study is to compare the ERA5 reanalysis product with the observed data in the high altitude snow fed region, correcting it and check the quality of the corrected ERA5 products for the future application. Also, the objective is to check whether the ERA5 reanalysis product is worth to understand the meteorological behavior in the data scarce region.

2. Study area and Methods

2.1 Study area

The Rasuwa district, in the heart of the Nepalese Himalaya, is home to the Langtang River Catchment, which is located 60 kilometers north of Kathmandu. The altitude varies from 3647 to 7178 meters above sea level. In our work, we primarily focus on the Langtang basin, also referred to as the catchment for the Langtang River. 352.6 square kilometers make up the Langtang River's watershed. 36% (126.91 sq km) of the basin's total area is made up of glaciers (debris-covered and clean ice), while the remaining 64% (225.05 sq km) is made up of rock and vegetation. The hydrological station for the Langtang River catchment is situated at 3650 masl. On the high plateaus and the steep hillsides, there are trees and boulders[13].Due in part to the sporadic passage of westerly troughs, the post-monsoon (October to November) and winter (December to February) have relatively little precipitation. Precipitation for the pre-monsoon is provided by isolated convective precipitation episodes[14].

2.2 Data Preparation

An important part of facilitating the entire research process is data preparation. The outcome must be produced using a trustworthy data set. The main focus of this phase is on making data available and getting them ready for model input. The first step is to choose the research topic, assess the data availability, prepare the data, and input the data into the model. Data from various hydrological and meteorological stations are gathered from the catchment area's close-by stations. Data pre-processing is done. Any missing data are interpolated using the appropriate interpolation method. Data from the ERA5 reanalysis is obtained from https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset, compared with the observed temperature and precipitation data, and then the ERA5 data is corrected as needed. GIS is used to define the catchment area.The following stations' data were gathered from ICIMOD and DHM. The hydrological stations at

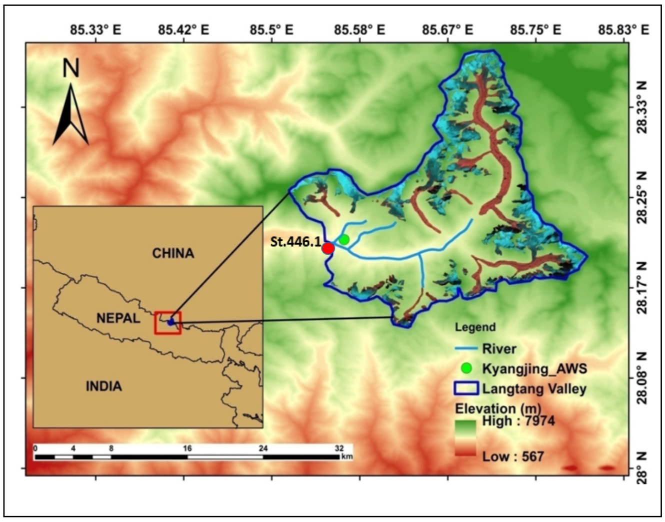

Figure 1: Langtang Catchment representing hydrometeorological station

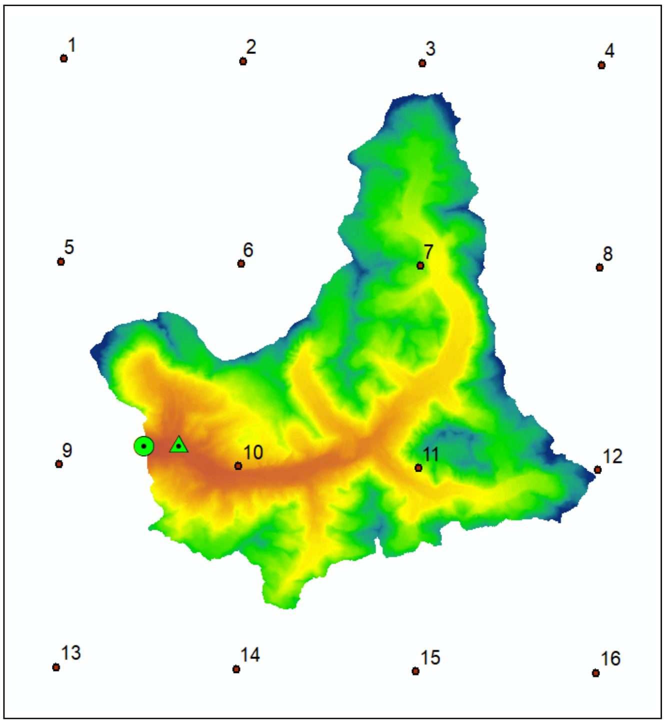

Figure 2: ERA grids covering the catchment

Langtang River (28° 12’ 34" N, 85° 32’ 50.31" E, and 3661m elevation) were used to collect the hydrological data. Data on the weather were gathered from the Kyanging meteorological station (28°12’39.66" N, 85°34’1.63" E, 3842m elevation). From `https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset`, hourly ERA5 Land data were downloaded. 2m temperature (hourly) and total precipitation are the data sets for ERA5 Land that were downloaded. Total precipitation information is the sum of all 24-hour precipitation information. The hourly 2m temperature data were in Kelvin and were converted to centigrade by taking away 273.15. from the given data. The daily data for that day is taken from the 24-hour precipitation data. The mean daily temperature is calculated by averaging the 24-hour temperature data. For all grids, daily temperature and precipitation data were plotted to see if there was any anomaly in the data.

2.3 Data correction

Data correction for precipitation was accomplished using two techniques. The first approach omits the smaller precipitation events. Data on precipitation took into account ignoring precipitation events that were less than 0.001mm,0.01mm,0.1mm,1mm,1.1mm, and 1.14mm. For each of these events, the average annual precipitation is calculated, and the best-performing event is chosen for analysis. To evaluate its effectiveness, the BIAS and Klings Gupta Efficiency (KGE) are calculated. Quantile mapping is the second technique used. The ERA data and observed data from the years 2008 to 2010 are used to calculate the monthly quantile and the monthly correction factor is applied to the ERA data sets.For the correction of the temperature data the ERA grid comprising the ERA station is selected for the analysis.The monthly correction factor is calculated for the year 2008 to 2010 and same factor is used to correct the temperature data for the year 2011 to 2013.The performance of the ERA grid temperature is evaluated based on the Pbias. The lapse rate for the Catchment is compared with the lapse rate of the ERA temperature.

3. Result and Discussion

3.1 Temperature and Precipitation

While the observed average precipitation for the basin is 755.8 mm, the average annual precipitation of the ERA data comprising the station is 2030.1 mm. This demonstrates that the precipitation estimate made using ERA data was off by 168%. The average precipitation of the catchment considered by using all the contributing ERA grid to the catchment is 1811.193mm which is overestimated by the 139%.Similarly, the station’s ERA temperature for the average annual temperature is -4.767 C while the observed temperature data is 4.101 C. Based on the DEM’s coverage, the ERA contribution to the catchment is calculated. 4.9 km along the longitude and 5.5 km along the latitude make up the coverage area. The table 1 makes it obvious that the grids 1,2,4,8,13,15, and 16 do not affect the catchment.

Table 1: ERA fractional coverage

ERA grids	Fractional Coverage in percentage
1	0
2	0
3	0.051
4	0
5	0.017
6	0.051
7	0.255
8	0
9	0.016
10	0.261
11	0.267
12	0.061
13	0
14	0.016
15	0
16	0

3.2 ERA and DEM elevation

In Table 2,the elevation of the ERA grid considered by the ERA data is based on the geo potential of that point.This values is divided by the value of acceleration due to gravity (g) to obtain the elevation of the ERA grid.Also the ERA grid from the DEM is calculated based on the coverage of the ERA grid on the DEM.The extent of the coverage is half the distance between the two grid on either sides.The elevation of the ERA grid comprising the station is 5011m. The elevation of the grid comprising the station based on the DEM is 4011m.

Table 2: ERA and Dem elevation(m)

Grid no.	Elev(DEM)	Elev(ERA)	diff
1	–	4434	–
2	–	5365	–
3	5868.44	5813	-55.43
4	–	5785	–
5	5709.97	4170	-1539.97
6	5631.23	5158	-473.22
7	5418.05	5607	188.94
8	–	5562	–
9	4916.70	4435	-481.70
10	4677.80	5011	333.20
11	5122.29	5153	30.70
12	5780.32	5082	-698.32
13	–	4009	–
14	5147.90	4119	-1028.90
15	5319.82	4109	-1210.82

The observed precipitation is overestimated by the ERA data. The higher precipitation event for the ERA data is upto 80 mm per day while for the observed data is 36mm per day.The pattern of the precipitation is almost better represented by the ERA grids based on the seasons.

3.3 Temperature Lapse rate

The ERA temperature underestimates the observed temperature.This is because the elevation consideration by the ERA grid is overestimated than the ERA grid coverage of the DEM. The Lapse rate for the entire valley floor and upper valley floor are 0.005°C/m and 0.0054 °C/m[15]. The lapse rate calculated for ERA grid 14(lowest elevation) and ERA grid 3 (highest elevation) is 0.00546°C/m which is similar.

3.4 Correction of Temperature data

The ERA temperature's lapse rate falls within the range of the Langtang Catchment's observed lapse rate. The correction factor is applied to the ERA grid containing the station about the observed station data and other station because the lapse rate of the temperature data is within the range of the observed data. The monthly factor is used to correct the data. The average catchment temperature is calculated after applying the same correction factor to all the other contributing grids. The data (figure 4) for the years 2011 to 2013 are validated using the correction factor calculated for the years 2008 to 2010 (figure 3). Bias is 0.008% during calibration and 24.71% during validation. This represents that the correction factor is good enough to correct the data for the rest of the years.

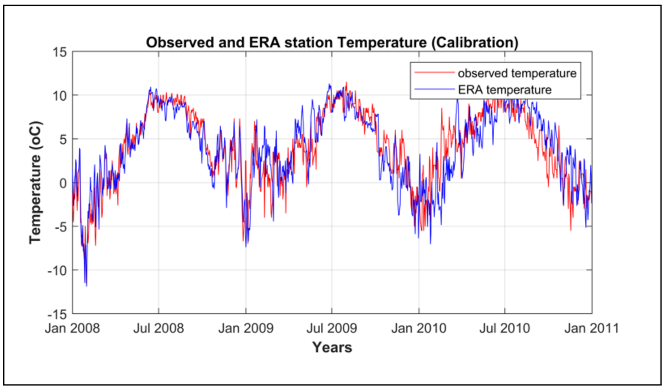

Figure 3: Calibration for the ERA temp

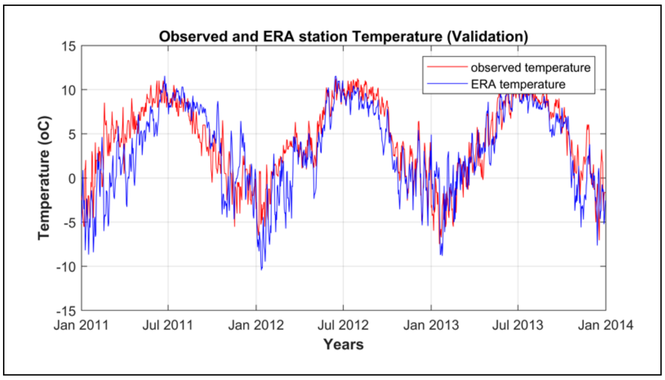

Figure 4: Validation for the ERA temp

3.5 Correction of precipitation neglecting precipitation events

The Quantiles for the ERA grid data and the observed data is calculated.The precipitation events are neglected based on the Q-Q plot for the data.

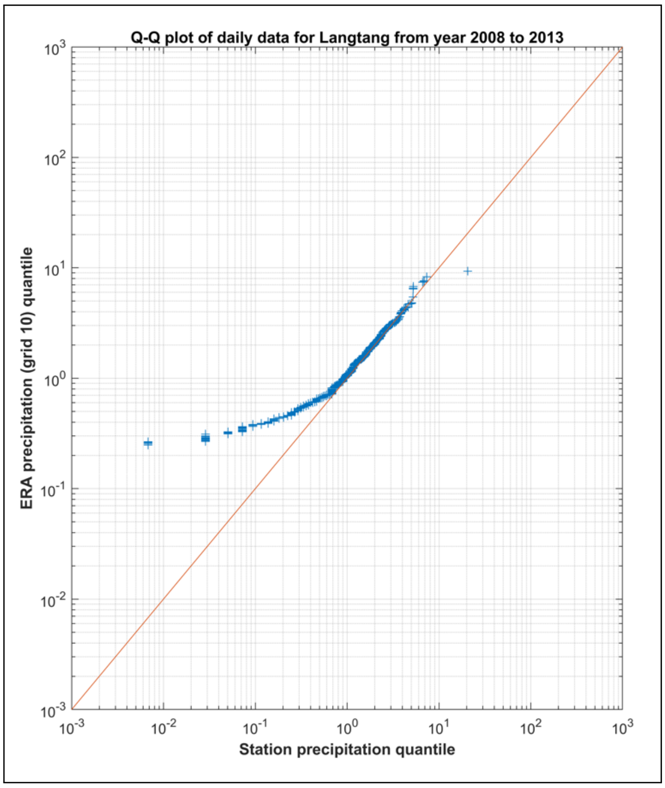

Figure 5: Q-Q plot for ERA grid 10 Precipitation and obs precipitation

Figure 5 shows the Q-Q plot for the ERA grid 10 precipitation and observed precipitation at the meteorological station at Kyanging.The Q-Q plot shows that the lower precipitation events are overestimated than the higher precipitation event. The higher precipitation event follows the similar distribution.The better option to correct the precipitatation data is neglecting the lower precipitaion events as the higher precipitation events are well predicted by the ERA data. The precipitation data is corrected neglecting the lowewr precipitation event of 0.001mm, 0.01mm, 0.1mm, 1mm, 1.1mm, 1.2mm, 1.3mm, 1.4mm, 1.5mm daily precipitation event and the data is compared with the observed precipitation data from the year 2008 to 2013.

The correction of the precipitation data is best for the precipitation event neglecting 1.4mm on hourly precipitation. The Pbias and Klings Gupta Efficiency(KGE) for the data neglecting this event is 6.21% and 0.067 respectively.This indicates that the bias has been significantly reduced but the performance is unsatisfactory.The timing of the event is not captured properly.

3.6 Correction using quantile mapping

Another popular method is to correct the data is based on the quantile mapping.The data is mapped based on the quantiles of the observed data using the monthly correction factor.

The precipitation ERA data is corrected with observed station data using quantile mapping. It is challenging to validate the gridded data at the higher elevation because there is only one hydrological and meteorological station located within the catchment. The observed station precipitation is transferred to all contributing grids using the precipitation gradient to get around this. To transfer precipitation to a grid other than the grid containing the station, a precipitation gradient of 42% per 1000m elevation is used[16]. For quantile mapping, the monthly quantile of the monthly precipitation of the observed data and the same for the ERA grid 10 data during the calibration period (2008 to 2010) is used to calculate the monthly correction factor. The monthly correction factor calculated is used to validate the data for the validation period (2011 to 2013). A similar analysis is done to the grid and the catchment average precipitation is calculated. To know the variation same process is done to correct the ERA grid data without applying the precipitation gradient.

The average catchment precipitation using the same correction factor to the every contributing grids in the catchment is 689.272mm while the observed precipitation is 755.83mm.Here the precipitation after the bias correction is lower than the observed precipitation.The elevation the meteorological station is at the elevation of 3842m but the extent of the highest elevation is 7178m.As the precipitation event goes on increasing on increasing the elevation,the gridded datsets should be corrected based on their elevation.As the catchment has a single observed station.The data is corrected based on the precipitation gradient. The observed data is transferred to the era grid data and ERA data is corrected in each grid.The corrected data is used to determine the catchment average precipitation based on the ERA fraction.The average catchment precipitation is 930mm.

Figure 6 and 7 shows the calibration and validation of the precipitation of grid 10.The Bias for the calibration period for all the grids is less than 0.005 with the better performance of the correction at grid 12.The KGE for the calibration period is 0.35 for all grids.This means the precipitation pattern in similar in each grids covering the catchment.The Pbias for for the validation period is less than 0.01 with the better KGE of 0.38 at grid 14 and KGE of 0.33 at grid 3.The KGE ranges from 0.33 to 0.38 for both the periods.The correction of ERA precipitation data at the higher elevation is challenging and should be given a special care while using in a hydrometeorological analysis process.

The variability ratio for the data on all grid during the calibration and validation period is less than 1.This indicates the standard deviation is greater in ERA data than in observed data.Also the mean bias ratio is greater than 1 in all cases.This indicated the ERA data is overpredicting even after correction.The correlation ranges from 34% to 42% in all cases. This means the timing of the events is captured only 34 to 42% of the time.

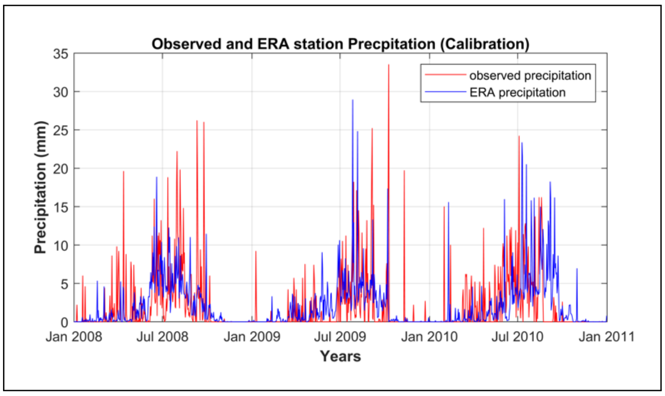

Figure 6: Calibration for the ERA precipitation

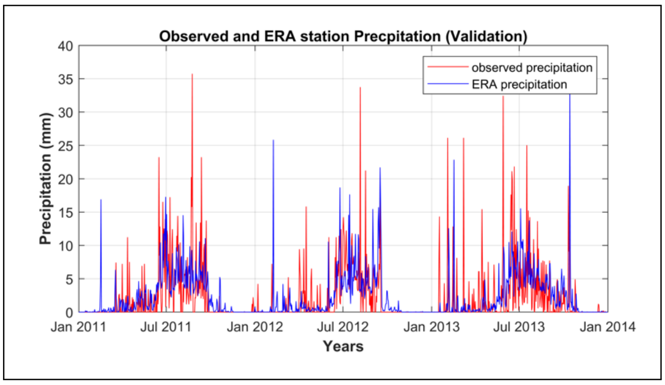

Figure 7: Validation for the ERA precipitation

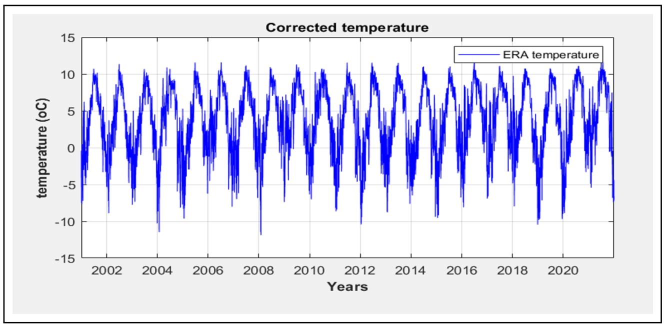

Figure 8: Average temperature considering ERA grid contribution

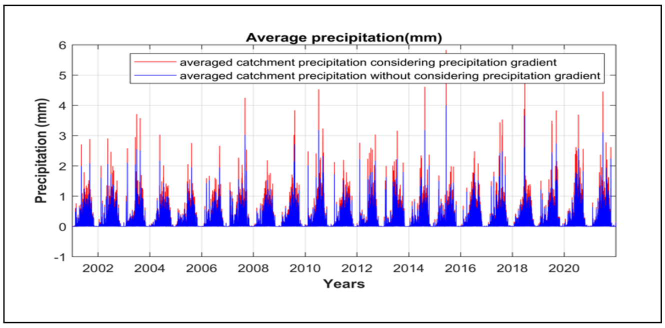

Figure 9: Average Precipitation considering ERA grid contribution

Figure 8 shows the average Era temperature after the correction of all ERA grids temperature is 3.68°C .The average temperature is underestimated by 9.5%The average catchment temperature is low as the temperature decreases with the increase in elevation. Similarly, Figure 9 show the average catchment precipitation considering the precipitation gradient and without considering the precipitation gradient.

4. Conclusion

In conclusion there exist the difference in the elevation considered by the ERA as compared to the average elevation based on the DEM.The temperature is underestimated as compared to the observed station.The temperature events are bias generating due to the overestimation of the elevation by the ERA.The precipitation data are bias generated due to the overestimation of the lower precipitation events. But on neglecting the lower precipitation event the ERA data are even bias orientated and difficult to correct. It is difficult to adjust the ERA data using the observed data. These difficulties may result from the under- or overestimation of the grid elevation in comparison to the actual elevation, from the ERA's estimation of precipitation focusing only on one season, or from the ERA's overestimation of precipitation during the dry season. When compared to the ERA precipitation, the ERA temperature performs well. The unsatisfactory performance of the precipitation data may be caused by the reanalysis data's inability to adequately capture the event in the topography at higher elevations. Despite this, the ERA5 land can be a useful substitute to roughly comprehend the meteorological process of the region where the data is lacking. For precipitation, the Pbias has been reduced to less than 10%, and the KGE for all grids is currently around 0.35. The data for each grid that contributes to the catchment can be corrected using the temperature lapse rate alone.

The Himalayas steep terrain makes accurate weather and climate prediction difficult.To understand the usability of the ERA product, it must be studied in multiple locations.This study has been carried out in the region which lacks the Long term observed data and multiple observational stations. Due to the lack of multiple observation stations, the data are corrected based on the lapse rate and the precipitation gradient of the observed data.

Acknowledgments

The authors are thankful to Dr. Rijan Bhakta Kayastha for his valuable inputs on accomplishing the research.

References

[1] Matthew J Menne, Claude N Williams, Byron E Gleason, J Jared Rennie, and Jay H Lawrimore. The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate*, 31(24):9835–9854, 2018.

[2] Dennis P Lettenmaier, Doug Alsdorf, Jeff Dozier, George J Huffman, Ming Pan, and Eric F Wood. Inroads of remote sensing into hydrologic science during the wrr era. *Water Resources Research*, 51(9):7309–7342, 2015.

[3] Mostafa Tarek, François P Brissette, and Richard Arsenault. Evaluation of the era5 reanalysis as a potential reference dataset for hydrological modelling over north america. *Hydrology and Earth System Sciences*, 24(5):2527–2544, 2020.

[4] Qiaohong Sun, Chiyuan Miao, Qingyun Duan, Hamed Ashouri, Soroosh Sorooshian, and Kuo-Lin Hsu. A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, 56(1):79–107, 2018.

[5] Mohd Farooq Azam and Smriti Srivastava. Mass balance and runoff modelling of partially debris-covered dokriani glacier in monsoon-dominated himalaya using era5 data since 1979. *Journal of Hydrology*, 590:125432, 2020.

[6] David A Lavers, Shaun Harrigan, Erik Andersson, David S Richardson, Christel Prudhomme, and Florian Pappenberger. A vision for improving global flood forecasting. *Environmental Research Letters*, 14(12):121002, 2019.

[7] Ayan H Chaudhuri, Rui M Ponte, and An T Nguyen. A comparison of atmospheric reanalysis products for the arctic ocean and implications for uncertainties in air–sea fluxes. *Journal of Climate*, 27(14):5411–5421, 2014.

[8] Ron Lindsay, Mark Wensnahan, A Schweiger, and J Zhang. Evaluation of seven different atmospheric reanalysis products in the arctic. *Journal of Climate*, 27(7):2588–2606, 2014.

[9] Susen Shrestha, Mattia Zaramella, Mattia Callegari, Felix Greifeneder, and Marco Borga. Evaluation of the era5 reanalysis as a reference dataset for fine-scale hydrological modelling over alpine basins. In *EGU General Assembly Conference Abstracts*, pages EGU21–14265, 2021.

[10] Copernicus climate change service (c3s). 2017.

[11] Samit Thapa, Bo Li, Donglei Fu, Xiaofei Shi, Bo Tang, Hong Qi, and Kun Wang. Trend analysis of climatic variables and their relation to snow cover and water availability in the central himalayas: a case study of langtang basin, nepal. *Theoretical and Applied Climatology*, 140:891–903, 2020.

[12] Jie Chen, François P Brissette, and Hua Chen. Using reanalysis-driven regional climate model outputs for hydrology modelling. *Hydrological processes*, 32(19):3019–3031, 2018.

[13] SM Uppala, PW Kållberg, AJ Simmons, U Andrae, V Da Costa Bechtold, M Fiorino, JK Gibson, J Haseler, A Hernandez, GA Kelly, et al. The era-40 re-analysis, qj roy. *Meteor. Soc*, 131(612):2961–3012, 2006.

[14] Akiko Sakai, Koji Fujita, and Jumpei Kubota. Evaporation and percolation effect on melting at debris-covered lirung glacier, nepal himalayas, 1996. *Bulletin of glaciological research*, 21:9–16, 2004.

[15] Prashant Baral, Rijan B Kayastha, Walter W Immerzeel, Niraj S Pradhananga, Bikas C Bhattarai, Sonika Shahi, Stephan Galos, Claudia Springer, Sharad P Joshi, and Pradeep K Mool. Preliminary results of mass-balance observations of yala glacier and analysis of temperature and precipitation gradients in langtang valley, nepal. *Annals of glaciology*, 55(66):9–14, 2014.

[16] Walter W Immerzeel, LPH Van Beek, M Konz, AB Shrestha, and MFP Bierkens. Hydrological response to climate change in a glacierized catchment in the himalayas. *Climatic change*, 110:721–736, 2012.