Generation of Cressman Interpolated High-resolution Gauge-based Gridded Observations (CIHGGO) for climatic variables using in-situ data over Pakistan

Burhan Ahmad
Syed Ahsan Ali Bukhari
Sohail Babar Cheema

Pakistan Meteorological Department, Islamabad, Pakistan

Abstract. Scarcity of in-situ observations over an expanse of 881,913 km^2 (including northern sides of high altitude regions) of Pakistan invites generation of gridded datasets at World Meteorological Organization's (WMO's) recommended spatial resolutions. Normally used for forecasts, projections and impacts assessments, gauge-based gridded datasets over particular regions have deficiencies either in spatial expanse of incorporated in-situ data or in robustness of long-term temporal scales over particular regions. This work has addressed these issues by incorporating quality controlled and strategically adopted 31 stations daily temperature and daily precipitation data in to an algorithm of designing a Cressman Interpolated High-resolution Gauge-based Gridded Observations (CIHGGO) at 0.45° (49.95 Km approximately) over a temporal scale of 39 years (1980-2018) for Pakistan region. The CIHGGO is available in NetCDF compression format for ease of use in decision support systems, research and applications.

Key words: Cressman interpolation, gauge-based gridded dataset, Pakistan.

Introduction

Development of streamline programming languages and state-of-the-art computing facilities have provided an opportunity for climate scientists to estimate climatic variables and develop reliable data over long periods of time, mainly for areas where data is scarcely available. These datasets are generally categorized into reanalyses, satellite products, gauge-based observations, and sometimes their compendiums. Gauge-based gridded temperature and precipitation datasets are regularly utilized by the climate scientists due to their availability over longer spatial and temporal scale (see e.g., Eum et al. (2014: 4250-4271) and Yin et al. (2015: 2809-2827). Several gauge-based gridded temperature and precipitation datasets have been developed in the last decade where improvement in emulation of real climate have remained a challenge (see e.g., Herrera et al., (2012: 74-85) and Herrera et al., (2016: 900-908).

In pursuit of constructing a gridded climate dataset CRU TS3.10, monthly observations at meteorological stations across the world's land areas have been interpolated on a 0.5° latitude/longitude grid cells covering the global land surface (Harris et al., 2014: 623-642). Another database known as Global Historical Climatology Network (GHCN)-Daily described by Menne et al. (2012: 897-910) has been designed to fulfil need for daily climate data over global land areas. On comparable resolution, still another gauge-based analysis of daily precipitation has been constructed on a 0.5° latitude–longitude grid over East Asia (5°–60°N, 65°–155°E) for a 26-yr period from 1978 to 2003 using gauge observations as collected from several individual sources (Xie et al., 2007: 607-626).

As seen in already cited research, gridded products have been mostly developed at regional and global scales. The regional products were developed for specific regions or countries, for example, a gauge-based gridded climatic dataset was developed by Yatagai et al. (2009: 137-140) for Asia and was used by several (see e.g., Ahmad et al.,

(2015) over high altitude regions of Pakistan. However, attributed to scarcity of gauging instruments and to the long term historical records country-scale gridded observations are still rare. The current study has, therefore, addressed such gaps by generating state-of-the-art Cressman Interpolated High-resolution Gauge-based Gridded Observations (CIHGGO) at country-scale for ease of use in formulation of forecasts and performing impacts assessments.

Study area and general climatology

Pakistan lies between 20-40°N and 60-80°E on western side of South Asia (Fig. 1). In much of Pakistan, the climate is tropical or subtropical, semi-arid or desert, but in the north the area is mountainous which is quite snowy and frigid on peaks of Himalayas. In the cold half of the year, from late autumn to early spring, the north is reached by weather fronts of Mediterranean origin, which cause rainfall in lowlands and snowfall over the mountains. In spring (i.e. March and April), the clash between air masses can cause thunderstorms and strong winds. In summer, from July to mid-September, the country is reached by an offshoot of the Indian monsoon, yet in most of the country it is not able to bring heavy rains, while it doesn't arrive at all in the western part. However, the warmest months are those that precede the monsoon, especially June, which is very hot in plains and hills, and up to quite high altitudes.

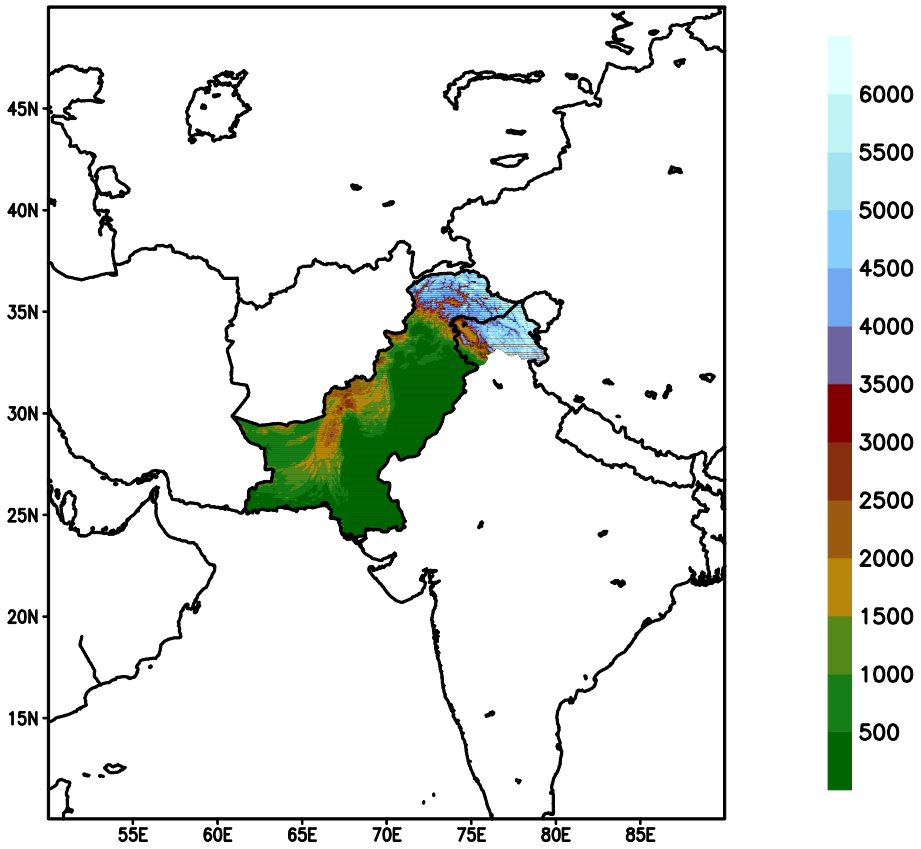

Fig. 1. Coordinate location of Pakistan with overlaid topography (m)

In the mountainous areas of the north and west, the climate is continental, with a wide temperature range between winter and summer, and often also between night and day. The temperature naturally decreases with altitude. The northern area, as well as being the coldest at equal altitude, is more prone to cold fronts brought by the westerly

winds of the middle latitudes from December to May. But not all areas receive a lot of precipitation: it depends on slope exposure. The southern side (the mountains north of Peshawar and Islamabad) is much rainier than the northern one.

The monsoon has an irregular pattern; during some years, it may have an unusual force, generating floods, while in other years, it doesn't even arrive. Rivers may overflow even at a great distance from the area where the heaviest rainfall occurred, which typically happen in the north. So, the great valley of the Indus and its tributaries may also be affected by widespread flooding in the southern area, where normally it doesn't rain much. The cycle called ENSO can affect the monsoon's performance; in La Niña years, rainfall is heavier than normal, while El Niño brings drought.

Methodology

Selection of ground stations

For 31 ground stations listed in Table 1, acquisition of daily maximum temperature, daily minimum temperature and total precipitation from Pakistan Meteorological Department was done. The ground station selection included general siting (urban, rural, snow–cover, agro–climate, latitudinal location, proximity to ocean, reach of monsoon and westerly systems etc.), and topographic settings since these factors attributed high variability in both the temperature and the precipitation values. Since snow cover can have a major impact on temperature change (especially by its freeze and thaw process) in its vicinity, therefore special attention was paid over to select stations located in those snow cover areas. Keeping in view these dynamics, the ground station density in higher latitudes above 31°N (below 31°N) was 16(15). Zonally, 3 ground stations were selected in the snow–covered regions of Gilgit-Baltistan, 2 ground stations in the hills of Azad Kashmir, 6 ground stations in hills and plains of Khyber Pakhtunkhwa, 7 ground stations in the plains of Punjab, 5 ground stations in deserts and ocean bordering plains of Sindh, and 8 ground stations in barren hills and coastal line of Balochistan.

Table 1. Attributes of ground stations selected for generation of CIHGGO over Pakistan

Sr. No.	Station	LON	LAT	Climate Characteristic	Province
1	Badin	68.54	24.38	desert/coastal	Sindh
2	Bahawalnagar	73.15	29.57	hot desert	Punjab
3	Barkhan	69.43	29.53	semi arid	Balochistan
4	Cherat	71.53	33.49	hill station	Khyber Pakhtunkhwa
5	Chhor	69.47	25.31	hot desert	Sindh
6	Dalbandin	64.24	28.53	hot desert	Balochistan
7	D.I.Khan	70.55	31.49	hot desert	Khyber Pakhtunkhwa
8	Dir	71.51	35.12	humid subtropical	Khyber Pakhtunkhwa
9	Gilgit	74.20	35.55	cold desert	Gilgit-Baltistan
10	Gupis	73.24	36.10	subarctic	Gilgit-Baltistan
11	Islamabad	73.05	33.42	humid subtropical	Punjab
12	Jiwani	61.48	25.04	hot desert	Balochistan
13	Karachi	67.08	24.54	arid/coastal	Sindh

14	Khanpur	70.41	28.39	hot desert	Khyber Pakhtunkhwa
15	Khuzdar	66.38	27.50	arid	Balochistan
16	Kotli	73.54	33.31	monsoon-influenced humid subtropical	Azad Kashmir
17	Lahore	74.24	31.31	semi arid	Punjab
18	Mianwali	71.33	32.33	hot desert	Punjab
19	Multan	71.26	30.12	arid	Punjab
20	Muzaffarabad	73.29	34.22	humid subtropical	Azad Kashmir
21	Nawabshah	68.22	26.15	hot desert	Sindh
22	Nokkundi	62.45	28.49	hot desert	Balochistan
23	Panjgur	64.06	26.58	hot desert	Balochistan
24	Parachinar	70.05	33.52	humid subtropical	Khyber Pakhtunkhwa
25	Peshawar	71.35	34.01	hot semi arid	Khyber Pakhtunkhwa
26	Quetta	66.53	30.15	cold semi arid	Balochistan
27	Rohri	68.54	27.42	hot desert	Sindh
28	Sargodha	72.40	32.03	hot semi arid	Punjab
29	Sailkot	74.32	32.30	monsoon-influenced humid subtropical	Punjab
30	Skardu	75.41	35.18	cold semi arid	Gilgit-Baltistan
31	Zhob	69.28	31.21	semi arid	Balochistan

Quality control

After identification of the weather stations that best represented diverse climatic regions, acquisition of data for 1980-2018 historical period (daily precipitation, maximum and minimum temperatures) was done. The raw data was then put to quality control which identified outlying (+/- 3 standard deviations) and questionable entries that might be corrupted. Spatial and temporal inspection was carried out on the dataset elements to see if anything looked amiss. The data was further checked to see if it were plausible physically, temporally (e.g., questionable value supported by preceding or following values), or spatially (e.g., questionable value supported by neighbouring stations).

Spatial aggregation

In final step the observations are spatially aggregated with optimal coverage at 0.45°×0.45° over a 39-year temporal scale using Cressman (1959: 367-374) interpolation method (see, e.g., Goodin et al., 1979). Firstly, the algorithm bins station observations into grid cells based on their location. If more than one station is located within a grid box, the observations are averaged together to produce analysed value. Secondly the method interpolates site data to a user-demarcated latitude-longitude grid. Several passes are made through the grid at successively tighter radii of influence to upsurge precision. The radius of influence is the maximum radius from a grid point to a site by which observation can be weighted to assess the computation at the grid. Sites outside the radius of influence have no relevance on a grid point assessment. For every iteration, a different result is computed for each grid site established on its adjustment factor. This adjustment factor is resolved by evaluating each site within the radius of influence. For each such location, an anomaly is described as the tiff between the site assessment and an assessment attained by interpolation from the grid to that site. A distance-weighted

method as in equation (1) and in equation (2) is later used in all such anomalies contained by the radius of influence of the grid to reach at an adjustment for that grid. The adjustment factors are imposed over all grids before the subsequent iteration is executed. Observations with highest proximity to the grid hold chief weights. As the proximity reduces, the observations bear recessed weight.

$$z = \frac{S^2 - s^2}{S^2 + s^2} \; for \; s^2 \leq S^2 \quad (1)$$

$$z = 0 \; for \; s^2 > S^2 \quad (2)$$

Here S is influence radius, s is distance between the site and the grid and z is the weighting function.

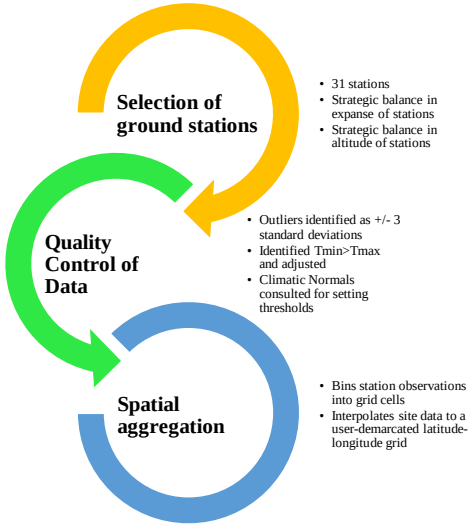

Fig. 2. Flow chart of CIHGGO algorithm designed for this study

Results

Post processed output of the CIHGGO from 31 stations (Fig. 2), replicating some extreme events over the historical data length is shown in Fig. 3. The Cressman objective analysis performed on the station data to arrive at a gridded result has presented robust results. The deployed technique has intricately interpolated station data to the defined latitude-longitude grid. Multiple passes made through the grid has increasingly rendered smaller radii of influence (10, 7, 6, 2 and 1 in our case). At each pass, a new value is depicted for each grid point based on the calculated correction factor that is determined by looking at each station within the radius of influence.

For each such station, an error is found as the difference between the station value and a value arrived by interpolation from the grid to that station. The correction factor is seen to balance distance weights applied to all such errors within the radius of influence. The correction factors are further seen to satisfactorily correct each grid point before the next pass is made. All the grid boxes that did not have stations within the third specified radius are set to the missing data value. The actual values of the gridded expression have been ignored in processing the data since the grids themselves act as templates to perform the analysis. Scaling of the grid is linear in latitude-longitude axes.

The CIHGGO algorithm was time-efficient, which mostly depended on our modestly chosen grid resolution and station data densities. The CIHGGO algorithm becomes highly unstable if the grid density is substantially higher than the station data density (i.e, far more grid points than station data points). In such cases, the analysis can produce

extrema in the grid values that are not realistic. It is thus suggested to examine the results and compare them to the station data to insure they meet satisfactory standards. One such comparison is performed by Ahmad and Mahmood (2017: 1) where an interpolated data over Pakistan is seen in analogy with a previously published data by Chaudhry et al. (2009: 367-374) over the same region.

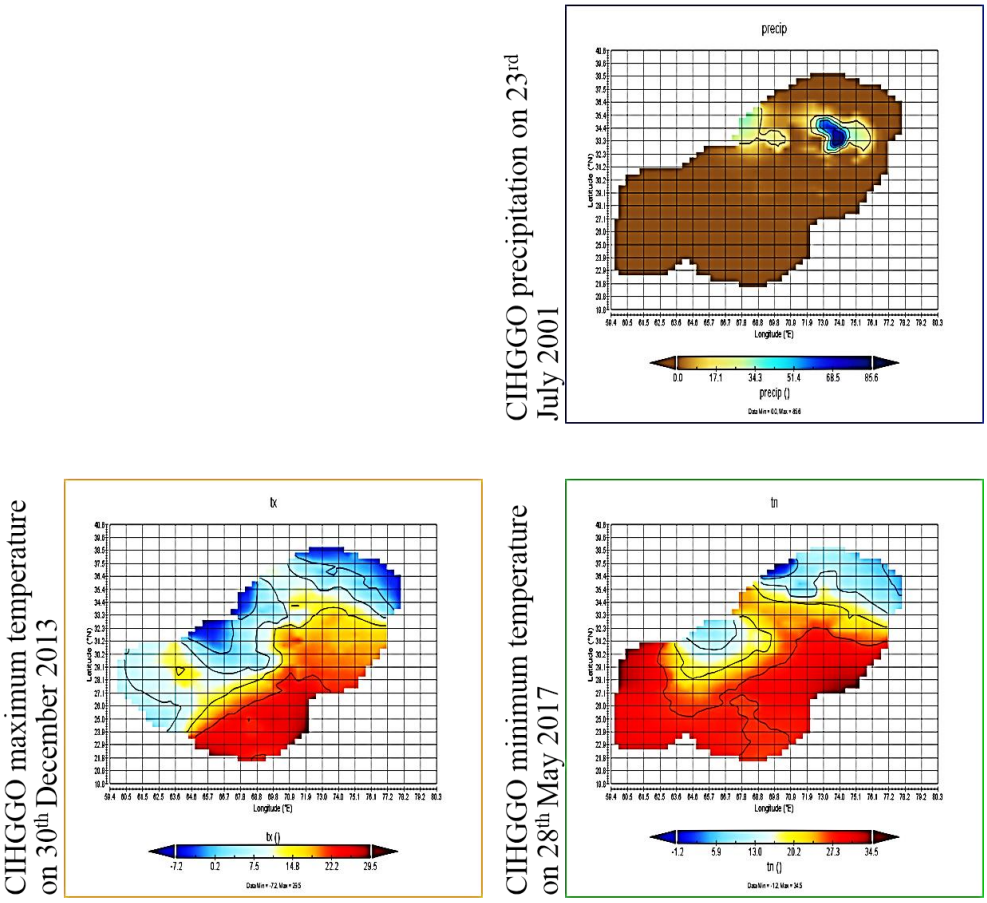

Fig. 3. Post processed results of CIHGGO emulating some extreme events of the past

Conclusion

Global temperature datasets have different spatio-temporal resolutions and durations. Quality of gridded datasets is governed by the data assimilation system, its optimization and the quality of observation data used. Differences/biases in stations can introduce specious variability and trends into gauge-based gridded datasets. Furthermore, there are uncertainties associated with gridded datasets, because of large natural, spatial and temporal variability, lack of observation data over large areas, changes in station networks, rain gauge types, and observational practices. Therefore, reliability of gauge-based gridded temperature data varies with time and regional climate. Gridding and interpolation of in-situ datasets using CIHGGO algorithm has established robust results. The correction factors are seen to satisfactorily correct each grid point of the constructed data. The CIHGGO is therefore recommended for use in applications considering all uncertainties mentioned in this study.

References

Ahmad, B., Mahmood, Sh. (2017). Observed, Simulated and Projected Extreme Climate Indices over Pakistan. Hamburg: Diplomica Publishing GmbH. Available at: http://www.anchor-publishing.com/ebook/371678/observed-simulated-and-projected-extreme-climate-indices-over-pakistan. ISBN 978-3-96067-672-0

Ahmad, B., Syed, A., Rasul, Gh., Shreshtha, Ab., Shea, Jm. (2015). Generation of high-resolution gridded climate fields for the upper Indus River Basin by downscaling CMIP5 outputs. Journal of Earth Science & Climatic Change, 6(2), 1.

Chaudhry, Q., Mahmood, A., Rasul, G., Afzaal, M. (2009). Climate Change Indicators of Pakistan. Pakistan Meteorological Department. PMD-22/2009.

Cressman, G.P. (1959). An Operational Objective Analysis System. Monthly Weather Review, 87, 367-374.

Eum, H.-Il, Dibike, Yo., Prowse, T., Bonsal, B. (2014). Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca Watershed, Canada. Hydrological Processes, 28(14), 4250-4271. https://doi.org/10.1002/hyp.10236

Harris, I.P.D.J., Jones, P.D., Osborn, T.J., Lister, D.H. (2014). Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. International journal of climatology, 34(3), 623-642. https://doi.org/10.1002/joc.3711

Herrera, S., Fernández, J., Gutiérrez, J.M. (2016). Update of the Spain02 gridded observational dataset for EURO-CORDEX evaluation: assessing the effect of the interpolation methodology. International Journal of Climatology, 36(2), 900-908. https://doi.org/10.1002/joc.4391

Herrera, S., Gutiérrez, J.M., Ancell, R., Pons, M.R., Frías M.D., Fernández, J. Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). International Journal of Climatology, 32(1), 74-85. https://doi.org/10.1002/joc.2256

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G. (2012). An overview of the global historical climatology network-daily database. Journal of Atmospheric and Oceanic Technology, 29(7), 897-910.

Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., Yang, S. (2007). A gauge-based analysis of daily precipitation over East Asia. Journal of Hydrometeorology, 8(3), 607-626. https://doi.org/10.1175/JTECH-D-11-00103.1

Yatagai, A., Arakawa, O., Kamiguchi, K., Kawamoto, H., Nodzu, M., Hamada, A. (2009). A 44Year Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. Sola, 5, 137-140. https://doi.org/10.2151/sola.2009-035

Yin, H., Donat, M.G., Alexander, L.V., Sun, Yi. (2015). Multi-dataset comparison of gridded observed temperature and precipitation extremes over China. International Journal of Climatology, 35(10), 2809-2827. https://doi.org/10.1002/joc.4174