Evaluation and Comparison of Reanalysis Precipitation Data in Iran

Ghasem Azizi*  
Associate Professor, Department of Physical Geography, Faculty of Geography, University of Tehran, Iran  
Taher Safarrad  
Assistant Professor, Faculty of Humanities and Social Sciences, University of Mazandaran, Iran  
Hosein Mohammadi  
Professor, Department of Physical Geography, Faculty of Geography, University of Tehran, Tehran, Iran  
Hasanali Faraji Sabokbar  
Associate Professor, Faculty of Geography, University of Tehran, Tehran, Iran

Received: 2 January 2015  Accepted: 6 February 2015

Extended Abstract
Introduction
There are significant differences in the spatial distribution of the Iran annual precipitation. This is resulted from spatial behavior of precipitation in the one hand and variation in the sources of precipitation on the other. The lack of adequate distribution of meteorological stations and the unavailability of long-term statistics of precipitation makes the analysis of precipitation more complicated. Precipitation data are constant inputs of research and the models related to water resources (e.g., climate, agriculture, hydrology, and environment). Most of research institutions are used to record the data and present it to different users. Different ways of interpolation of the data will cause different results. Therefore, it is a critical step to select the appropriate data based on research design. This study evaluates APHRODITE, GPCC and Delaware University precipitation data (UDel) based on precipitation stations using RMSE, R² and Taylor diagram techniques.

Materials and Methods
DATA¹
Daily gridded precipitation of APHRODITE (Asian Precipitation- Highly-Resolved Observational Data Integration Towards Evaluation) is the only long-term (1951 onward) continental-scale daily product that contains a dense network of daily precipitation-gauge data for Asia including the Himalayas, South and Southeast Asia and mountainous areas in the

* E-mail: ghazizi@ut.ac.ir  
¹. The description of the data have been driven from the following website: https://climatedataguide.ucar.edu
Middle East. The number of valid stations was between 5000 and 12,000; this represents 2.3 to 4.5 times the data available through the Global Telecommunication System network, which were used for most daily grid precipitation products. The products are available on a regional basis.

Key Strengths: High density and quality station network.
Key Limitations: Station network changes with time and season.

The Global Precipitation Climatology Centre (GPCC) has been established in 1989 on request of the World Meteorological Organization (WMO). It is operated by Deutscher Wetterdienst (DWD, National Meteorological Service of Germany) as a German contribution to the World Climate Research Programme (WCRP). The GPCC provides gridded gauge-analysis products derived from quality controlled station data. Two products are for climate: (a) the Full Data Reanalysis Product (1901-2010) is recommended for global and regional water balance studies, calibration/validation of remote sensing based precipitation estimations and verification of numerical models, and (b) the VASClimO 50-Year Data Set which is for climate variability and trend studies.

Key Strengths: Large number of stations used; gauge network extends beyond GHCN
Key Limitations: Variable number of stations per grid over time can be a major inhomogeneity source. Monitoring products are frequently updated but climate products are not.

The Global (land) precipitation, University of Delaware (UDel) is a series of gridded temperature and precipitation data sets. The data used are including station records served as bases for the Terrestrial Air Temperature, 1900-2010 Gridded Monthly Time Series (Version 3.01) and Terrestrial Precipitation, 1900-2010 Gridded Monthly Time Series (Version 3.02). These are used to help create new gridded climatologies of monthly and annual average air temperature (T) and total precipitation (P). These two sets of station time series were drawn primarily from recent versions of the Global Historical Climatology Network (GHCN version 2) and the Global Surface Summary of Day (GSOD) archive. Selected averages from Legates and Willmott’s (1990a and b) long-term station averages of monthly and annual T and P were also used to help produce this new gridded archive.

Key Strengths: relatively detailed global land surface temperature climatology; higher spatial resolution than comparable data sets
Key Limitations: Infrequent updates.

Materials and Methods
In order to evaluate the data, the closest point from the mentioned precipitation data to meteorological stations (max 40 km) were identified for the period 1961-2007. Then, we used RMSE, the coefficient of determination ($R^2$) and Taylor diagram to evaluate precipitation data. These methods are formulated as below:

$$
RMSE^2 = \sigma_{est}^2 + \sigma_{obs}^2 - 2\sigma_{est}\sigma_{obs} R
$$

$$
R = \frac{1}{N} \sum_{n=1}^{N} \left( \frac{p_{est}}{p_{obs}^i} - \frac{p_{obs}}{p_{est}} \right) \left( \frac{p_{est}^i - p_{obs}^i}{p_{obs}} \right)
$$

$$
\sigma_{est}^2 = \frac{1}{N} \sum_{n=1}^{N} \left( p_{est}^i - \frac{p_{obs}^i}{p_{est}} \right)^2
$$
\[ \sigma_{obs}^2 = \frac{1}{N} \sum_{n=1}^{N} \left( \frac{p_{obs}^{(i)} - \bar{p}_{obs}}{\sigma_{obs}} \right)^2 \]

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (p_{obs}^{(i)} - p_{est}^{(i)})^2}{n}} \]

\[ R^2 = \frac{n \left( \sum_{i=1}^{n} p_{est}^{(i)} \right)^2 - \left( \sum_{i=1}^{n} p_{obs}^{(i)} \right)^2}{n \sum_{i=1}^{n} p_{obs}^{(i)}^2 - \left( \sum_{i=1}^{n} p_{obs}^{(i)} \right)^2} \]

where \( p_{obs}^{(i)} \) and \( p_{est}^{(i)} \) are the precipitation values provided by instrumental data and precipitation data, respectively, \( \sigma_{obs}^2 \) and \( \sigma_{est}^2 \) are also the variance values of instrumental data and precipitation data, respectively; \( n \) indicates the number of stations.

Results and Discussion

Figure 1 shows the Taylor diagram, plotted by spatially averaged precipitation values. The diagram summarizes the relationship between testing and reference series standard deviations, correlation coefficient, and the RMSD (root mean square difference) as computed by series centered pattern, by means of a trigonometric similitude.

Fig. 1. Taylor diagram obtained from spatial averaged values plotted on the basis of standard deviation values, correlation coefficients between products and reference dataset, and root mean square differences of series-centered pattern, indicated as RMSD in the plot.
Aphrodite data are more accurate at, Khoi, Babolsar, Tehran and Yazd stations. GPCC data have better performance than other data at Zahedan and Bandar Abbas stations. For Shahr-e-Kord, Mashhad, and Zanjan stations, Aphrodite and GPCC data have similar RMSD. However, according to the stronger correlations between GPCC and instrumental data, the GPCC data are more appropriate than Aphrodite data.

Conclusion
Based on the long-term average annual precipitation, Aphrodite and GPCC data are more accurate than UDel data. Taylor diagram is based on the geometrical relationship between correlation coefficient, series standard deviation and centered mean square error. It is more useful than other uni-variable methods as RMSE and $R^2$. Aphrodite data are better to use for the Northwest, the southern Alborz and internal areas. The GPCC data will lead to better results in the West, South, Southeast and North East Iran. As UDel data consider spatial association of data with dependent variable, it estimates precipitation time series better than other data. This type of data is useful to analyze characteristics of precipitation in the areas with short-term time series.

Keywords: APHRODIATE, GPCC, RMSE, Taylor diagram, UDel.