Comparison of Gridded Precipitation Time Series Data in APHRODITE and Asfazari Databases within Iran’s Territory

Esmail Nasrabadi¹, Seyyed Abolfazl Masoodian², Hossein Asakereh³

¹Department of Physical Geography, Faculty of Geographical Sciences and Planning, University of Isfahan, Isfahan, Iran
²Department of Physical Geography, University of Isfahan, Isfahan, Iran
³Department of Physical Geography, University of Zanjan, Zanjan, Iran

Email: esmailnasrabaadi@gmail.com

Received December 27, 2012; revised January 30, 2013; accepted February 7, 2013

Copyright © 2013 Esmail Nasrabadi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

The V1003R1 version of the monthly, seasonal and annual precipitation time series of the Middle East APHRODITE database and Asfazari database within Iranian territory in the time interval 1961 and 2004 were compared with each other. The monthly, seasonal and annual time series of the both databases in most cases show a random behavior and the time series follow a similar pattern with a significant autocorrelation of both databases. Studying cross-correlation between the time series of the two bases indicate that the zero lag significance in the monthly, seasonal and annual time series statistically confirm the coincidence of the peak and fall in the time series of both bases. Wilcoxon Test with 95% confidence confirms significance of the mean difference of the two series. However, there is not enough evidence to confirm the null hypothesis suggesting lack of difference between means of the two series. The statistical-objective analysis of the two bases’ time series indicates that although the series follow a similar course, the estimated precipitation quantity in monthly, seasonal and annual time series of APHRODITE base, except in several monthly time series and in trivial quantities, has been less than the estimated precipitation of Asfazari bases, but the amount of the difference was not constant and did not follow a regular pattern, although this difference has been narrowed in recent years.

Keywords: Precipitation Database; Cross-Correlation; Wilcoxon Test; APHRODITE; Asfazari

1. Introduction

Precipitation data at international, regional and national scale are prepared by different research and commercial institutes and scientists. The main criteria for study of precipitation’s diverse features include trend, frequency change, increase or decrease of rainy days and they give us insight by enriching our knowledge on behavior of climatic processes in the past and present. Therefore, analysis, comparison, evaluation, specification of their weak points and eventual errors are indispensable. Given the availability of diverse evaluation methods and technical qualitative control applied to the data in the centers, comparison of the data quality from different bases in temporal and spatial dimensions will significantly help enrich the knowledge on the data’s common aspects, differences and validity.

The signal output of many physical and financial systems is often characterized by variables which can be both autocorrelated and cross-correlated [1]. The analysis in the temporal domain using such methods as autocorrelation and cross-correlation in time series of different variables as well as in diverse areas is one of the useful and conventional tools in this regard. For example, precipitation’s spatial behavior in the Baden-Wuerttemberg region of Germany from 1997 through to 2004 using Doppler Radar Combined Images data in networks of 4 × 4 km² and the data of 101 rain gauging stations using cross-correlation were compared. This approach shows that cross correlation varies depending on the chosen quantile. In the lower quantiles, the correlation is very similar in rain gauge and radar data [2]. As another example, the analysis of autocorrelation and cross-correlation between time series of air temperature, soil temperature, rainfall, relative humidity, and Diaprepes root weevil across a period of 30 month in Florida suggested an association of temperature and precipitation with time distribution of Diaprepes root weevil [3]. Analysis of
changes in precipitation temperature time series with plant cover (products and forest) during 1988-2005 in the area of Oreto watershed (Italy) indicated a relationship between plant cover, precipitation and temperature together with time lag value which varies from 4 to 8 months lag [4]. Application of correlation, cross-correlation and Spectral Density Function in studying time changes in hydrological processes (such as atmospheric pressure, rainfall and groundwater levels) in “Wunju” of Korea has also confirmed competence of these methods in the study of hydrological time series [5]. These methods are especially useful in identification of short term changes of hydrological systems. This application has been demonstrated in a peat basin in Malaysia by examining cross-correlation and cross spectral correlation between precipitation, water flow and level of aquifer [6]. The analyzed characterization of the transformation between the input rainfall and the output discharge of two karstic system during 1984 to 1992 in central Italy using autocorrelation, cross-correlation and spectral analysis gave reasonable results [7]. Used time series of meteorological parameters such as pressure, temperature, rainfall, relative humidity, and wind speed in the statistical period 1990-2005 in nine meteorological stations of Saudi Arabia, For coastal to costal pair of stations, pressure time series was found to be strongly correlated. In general, the temperature data were found to be strongly correlated for all pairs of stations and the rainfall data the least [8]. In a study of the relationship between stream discharge and the concentration of different solutes in a small forested watershed near Montreal (Quebec) in two years of 1995 and 1996, [9] and in a study of the relationship between salinity degree and water discharge in Apalachicola Bay of Florida, [10] cross correlation method were used and the correlation degree was investigated and analyzed in different time delays.

As for comparison of precipitation over Iranian territory, Javanmard et al. (2010) was compared to evaluate the satellite rainfall estimates of Tropical Rain Measurement Mission (TRMM) level 3 output (3B42) with high-resolution gridded precipitation datasets (0.25° × 0.25° latitude/longitude) based on rain gauges (Iran Synoptic gauges Version 0902 (ISO902)). Spatial distribution of mean annual and mean seasonal rainfall in both ISO902 and TRMM 3B42 from 1998 to 2006 shows two main rainfall patterns along the Caspian Sea and over the Zagros Mountains. However, for the entire country, the Caspian Sea region and the Zagros Mountains, TRMM 3B42 underestimate mean annual precipitation by 0.17, 0.39, and 0.15 mm·day⁻¹, respectively [11]. Among other comparative studies of different precipitation databases, it can be referred to the comparison of spatial pattern of the data 0.50 × 0.50 degree gridded daily rainfall for the period 1980-2002 in India obtained from 6000 precipitation gauging stations with those obtained from APHRODITE project [12] as compared with India Meteorological Department gridded database with Variability Analysis of Surface Climate Observations database during 1951-1995 and calculation of correlation, similarity and difference between the features obtained from these databases [13], and the estimates rainfall over South Asia by combining rain gauge and merged satellite observations are validated using Automatic Weather Station (AWS) rain gauge data and other available rainfall products such as those of the APHRODITE (V1003) with 0.25 × 0.25 resolution degree over India [14].

In most areas, with equal number of input gauging stations for the months (January 1951-December 2007), APHRO_0101 estimates less precipitation than the Global Precipitation Climatology Centre (GPCC) product. Hence, the difference seems to be because of 1) quality control and 2) different interpolation methods [15].

Substantial progress has been made in the last two decades in quantitatively documentation global precipitation. Surface gauge observations have been collected, digitalized, and quality controlled by the centers in several countries [16]. Hence, this research intends to identify the structure and features of these data and to investigate similarities and differences between time series of the two reliable daily gridded precipitation databases for Iran, while the eventual weaknesses and strengths in the data are identified, the choice of suitable and goal-directed data is made possible. This research is the first study in which the data of Afzazari National database are compared with APHRODITE database for Iran.

2. Data

The data of the Afzazari national database were formed through using the precipitation data of 1437 synoptic, climatic and rain-gauging stations interpolated by Kriging Method in pixels to the dimension of 15 × 15 kg in the time interval 1961-2004. The daily gridded precipitation data of this base are prepared in an array of 15,992 × 7187 (15,992 days over row and 7187 cells over columns). The coordinate system of this database is Lambert Conformal Conical.

APHRODITE database has been launched in 2006 by the research foundation of Japan Meteorological Agency with membership of several other countries. This database has been formed based on precipitation gauging stations of such sources as local meteorological and hydrological organizations, regional researchers, Global Historical Climatology Network (GHCN), Carbon Dioxide Information Analysis Center (CDIAC), National Center for Atmospheric Research, Data Archive (NCAR-DS), National Climatic Data Center (NCDC), Global Telecom-
munication System (GTS) [17]. The APHRODITE database daily gridded precipitation data for the Middle East region has been prepared with 0.5 × 0.5 and 0.25 × 0.25 resolution for the time interval 1951-2007.

In this research, the version V1003R1 of APHRODITE database in the Middle East on the daily gridded precipitation data has been utilized for the region of Iran with 0.25 × 0.25 resolution degree of geographical longitude/latitude and the daily gridded precipitation data of Asfarazi in the common time interval of 1961-2004 [18].

3. Methods

By programming in MATLAB software environment, the monthly, seasonal and annual time series were extracted from the two bases. First, autocorrelation of the monthly, seasonal and annual time series in different lags was calculated for each database separately. Next, cross-correlation between the time series of the two bases was examined and the results were analyzed and interpreted. Given that application of parametric tests to the time series didn’t meet the required conditions, Wilcoxon non-parametric test was utilized in SPSS software and the test of null- and alternative hypotheses was examined. In the following, by drawing the time series graph and their difference against each other, we have tried to provide an objective view to the time series of the two databases.

3.1. Autocorrelation

When studying time series, autocorrelation coefficient is frequently used. The prefix “auto” signifies the relationship of a variable with itself [19]. Considering that this coefficient measures correlation between successive observations, it is also known as “serial correlation”. If correlation between the observations with k distance from each other is found, this correlation is called “correlation coefficient in k lag” [20]. Suppose N observations (x₁, x₂, …, xₙ) is the discrete time series, then we can construct N-1 observation pairs (Equation (1)).

\[
(x_1, x_2), (x_2, x_3), \ldots, (x_{N-1}, x_N)
\]

If the first observation in each pair is considered as the first variable and the second observation as the second variable [21], Equation (2) will be obtained which represents the autocorrelation between the observations with k distance [22].

\[
r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N} (x_t - \bar{x})^2}
\]  (2)

For the interpretation of autocorrelation, “autocorrelo-

gram” is used in which autocorrelation coefficient is drawn against lags (lags). Although interpretation of this diagram is not easy, the following general states can be considered in this respect. If a time series is completely random, for large amounts of N and all non-zero values of k, the value of autocorrelation coefficient is equal to zero (r_k = 0). Static time series often show short term correlation in which a relatively large amount of r_k with two or three time series, significantly greater than zero, consecutively get smaller and for longer lags, r_k values tend to zero. If a time series on two sides of the mean tends to sequence, correlogram will tend to sequence as well, and if a time series has a trend, r_k values do not drop, except for very large values of lag. From this correlogram cannot be much inferred since other aspects are surrounded by trend [20]. Correlogram represents a time series with seasonal changes (fluctuation in the same frequency). If correlogram in these series is drawn after elimination of seasonal change from data, we are provided with more information. With regard to time series and outlying observations, first, the outliers should be adjusted and then correlogram can be drawn so as it won’t be affected by these outliers. In addition, by correlogram and series partial correlogram, Order ARIMA Model [23] or the series randomness is determined. In this study, the data randomness was examined by calculation of autocorrelation and drawing correlograms.

3.2. Cross-Correlation

In case observations of two time series are available and the relationship between them is to be found, two states can be distinguished. In the first state, two series occur in “equal condition” the correlation of which is to be found. In the second case, which is more important, two series are “randomly correlated” so as one series is regarded as the input of a linear system and another series as the output [20]. In this research, precipitation time series of the two bases are in similar condition. Equation (3) is for calculation of cross-correlation between the two time series.

\[
r_{xy}(k) = r_{y|x}(k) = r_{y|x}(0) + \sum_{t=1}^{\infty} \rho_{xy}(k) r_{y|x}(0)
\]  (3)

This function measures correlation between X(t) and Y(t+k). In the Equation (3), cross-correlation is calculated through covariance of the series. The symbols used in the equation are as follows:

\[
COV(X, Y_{t+k}) = r_{x|y}(k)
\]  (4-1)

\[
COV(Y, X_{t+k}) = r_{y|x}(k)
\]  (4-2)

\[
COV(Y, Y_{t+k}) = r_{y|y}(k)
\]  (4-3)

[20]. Application of cross-correlation is not recommend-
ed for the condition in which two time series do not follow an identical trend, since it may produce undesirable results.

### 3.3. Wilcoxon Test

To compare the mean of two time series, numerous parametric and non-parametric techniques can be employed if the required conditions are met. In cases where data do not meet the condition for performance of parametric tests, equivalent non-parametric tests can be applied. The quantile-quantile plot indicates that the data do not have a normal distribution. Hence, Wilcoxon non-parametric test which is equivalent to the two-sample parametric \( t \)-test was used which is a good substitute to it in case \( t \)-test conditions are not met. Mathematical form of this test, like \( t \)-test with a paired sample (correlated) is as the following Equation (3).

\[
z = \frac{\frac{T}{2} - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \tag{5}
\]

This test examines both direction and degree of difference between peer groups. Therefore, it answers the question as which part of pairs is greater than other and ranks the differences in the order of their absolute value, i.e. by means of this test it can be judged what part is greater than the other [24].

### 4. Research Findings

#### 4.1. Study of Time Series Autocorrelation

Autocorrelation of the annual time series in none of the lags outside the upper and lower limit is significant and the amount of autocorrelation can be supposed equal to zero so as randomness of the annual time series can be confirmed, although APHRODITE time series is closer to a perfect randomness (Figures 1 and 2).

Table 1 provides the results of autocorrelation application for a month lag in the monthly time series and a season lag for the seasonal time series. The autocorrelation close to zero is confirmed in the monthly time series of April-May, September-October, October-November, January-February, February-March, and in the seasonal time series of spring and autumn. In other time series, autocorrelation in the mentioned lags is significant in the table. The time series of July-August lies farthest from the condition of a random time series. In addition, autocorrelation coefficient estimation indicates two points: first, in none of the monthly and seasonal time series there is a significant autocorrelation coefficient. Secondly, despite the difference in value of autocorrelation coefficient in some time series of the databases, the amount of this difference is trivial and statistically insignificant and the monthly and seasonal precipitation time series of the two databases have an identical autocorrelation in the month and season lags. After interpreting figures in Table 1, it can be claimed that in March-April time series in both bases indicate autocorrelation of precipitation in March-April 5 and 15 years before and after, whereas in the time series of April-May, the precipitation is not significant in any lag. In other words, April-May precipitation has no significant correlation with April-May precipitation the years before and after.

![Figure 1. Correlogram database APHRODITE in the lag year.](image1)

![Figure 2. Correlogram database Asfazari in the lag year.](image2)
Table 1. Time series autocorrelation significance in the month lag for the monthly time series and the season lag for seasonal time series.

<table>
<thead>
<tr>
<th>Database series</th>
<th>Autocorrelation</th>
<th>Database series</th>
<th>Autocorrelation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asfazari</td>
<td>APHRODITE</td>
<td>Asfazari</td>
<td>APHRODITE</td>
</tr>
<tr>
<td>Jun</td>
<td>-</td>
<td>Sep</td>
<td>2</td>
</tr>
<tr>
<td>Feb</td>
<td>-</td>
<td>Oct</td>
<td>-</td>
</tr>
<tr>
<td>Mar</td>
<td>16</td>
<td>Nov</td>
<td>-</td>
</tr>
<tr>
<td>Apr</td>
<td>5 - 15</td>
<td>Dec</td>
<td>13</td>
</tr>
<tr>
<td>May</td>
<td>-</td>
<td>Winter</td>
<td>9</td>
</tr>
<tr>
<td>Jun</td>
<td>7 - 12</td>
<td>Spring</td>
<td>-</td>
</tr>
<tr>
<td>Jul</td>
<td>12</td>
<td>Summer</td>
<td>8</td>
</tr>
<tr>
<td>Aug</td>
<td>9-11-14</td>
<td>Autumn</td>
<td>-</td>
</tr>
</tbody>
</table>

*: Signal to autocorrelation insignificance.

4.2. Studying Time Series Cross-Correlation Fluctuation Patterns

In Figure 3, the cross-correlation between the annual time series of APHRODITE and Asfazari databases indicates that correlation between the two series is significant in zero lag, i.e. maximum value of the cross-correlation in zero lag signifies simultaneity of the two series’ fluctuation peak [25], and as expected, despite the difference in precipitation estimation value, since the two bases represent one area, they have simultaneous peak and fall. For the monthly and seasonal time series, cross-correlation between time series of the two bases is significant in zero lag, indicating simultaneous peak and fall of monthly and seasonal time series. Only March-April and August-September time series have different situation so as in March-April time series, in addition to significance of cross-correlation in zero lag, cross-correlation in +5 and −5 lag is significant as well (Figure 4). In other words, the precipitation peak in March-April occurs with 5 days after and before. And in August-September time series, in addition to significance of cross-correlation in zero lag, cross-correlation in +2 and −2 lag is significant as well (Figure 5).

4.3. Means Difference Significance Test

In quantile-quantile plot, in low precipitation quantities, the observed precipitation in both bases is less than the expected normal precipitation. In APHRODITE data along with increasing precipitation, the observed and normal precipitation get closer to each other (Figure 6), whereas within the same area, the observed Asfazari precipitation is higher than the expected normal quantity (Figure 7), and in high precipitation quantities, in both bases, the observed amounts are less than the expected normal precipitation. However, in APHRODITE base, the difference between the observed and expected precipitation is greater. Considering that application of the parametric tests by time series did not meet the required conditions, Wilcoxon non-parametric test was used as compared with the means, in which the null hypothesis suggests “there is no difference between the means of the two time series’ data” and the alternative hypothesis proposes “there is a difference between the means of the two time series’ data”. Based on the results, since the number and the mean of Asfazari ranks (positive) are greater than the number and the mean of APHRODITE ranks (negative), the mean Asfazari precipitation is greater than the mean of APHRODITE precipitation. This led to a negative value for Z-test statistic based on which at 0.05 chance of error
there is no sufficient reason to confirm the null hypothesis and at 95% confidence interval the difference between the mean precipitation of Asfara zi and APHRODITE is statistically significant and the alternative hypothesis is confirmed (Table 2).

The results of Wilcoxon test for the data’s mean in the monthly and seasonal time series are similar to those obtained from the test of annual time series. In the time series of March-April, May-June, July-August, August-September and December-January, although the value of

**Table 2. Wilcoxon ranks and sign test of APHRODITE-Asfarazi annual precipitation.**

<table>
<thead>
<tr>
<th>Annual precipitation</th>
<th>Number ranks</th>
<th>Total ranks</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative ranks</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Positive ranks</td>
<td>43</td>
<td>22</td>
<td>946</td>
</tr>
<tr>
<td>Ties</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of ranks</td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td>-5.711</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asymp. sig. (2-tailed)</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The graphs depict the annual precipitation data from 1961 to 2001 for different months:

1. **Precipitation Jun. (mm):** The data shows variability with peaks in 1976 and 1991. There is a general increase towards the end of the period.

2. **Precipitation Feb. (mm):** Similar trends to Jun., with peaks in 1976 and 1991. The data shows a gradual increase over the period.

3. **Precipitation Mar. (mm):** The pattern resembles that of Jun. and Feb., with notable peaks in 1976 and 1991. The data trend shows an increasing pattern over the years.


The data suggests a pattern of increased precipitation towards the end of the observed period, with notable peaks in specific years.
Figure 8. Monthly precipitation time series plot (mm).
Figure 9. Seasonal precipitation time series plot (mm).

Figure 10. Annual precipitation time series (mm).
Figure 11. Difference of monthly precipitation time series database APHRODITE and Asfazari (mm).

Figure 12. Difference of monthly precipitation time series of APHRODITE and Asfazari databases (mm).

Figure 13. Difference of monthly precipitation time series of APHRODITE and Asfazari databases (mm).
Figure 14. Difference of monthly precipitation time series of APHRODITE and Asfazari databases (mm).

Figure 15. Difference of seasonal precipitation time series of APHRODITE and Asfazari databases (mm).

Figure 16. Difference of annual precipitation time series of APHRODITE and Asfazari databases (mm).
the test statistic somewhat becomes smaller, this drop in value is not statistically significant and still the mean difference of the time series is significant. In the time series of April-May, August-September, September-October, October-November, November-December, January-February and February-March and all the seasonal series, the negative ranks become zero. In these series, mean Asfarazi precipitation is greater than mean APHRODITE precipitation. In the monthly time series of March-April, May-June, July-August, August-September and December-January, the means of negative ranks are 1, 2.5, 7, 1 and 1, respectively. In other words, in these time series, there are cases where the mean monthly precipitation of APHRODITE is greater than that of Asfarazi.

4.4. Comparison of Time Series Plots

In Figures 8 to 10, APHRODITE time series are represented with continuous line and those of Asfarazi with dotted line at three monthly, seasonal and annual levels. The estimated precipitation quantity of Asfarazi base is greater than that of APHRODITE base. However, in March 1962, May 1964, 2002, June 2003, July 1991 and December 1965, precipitation quantity of APHRODITE base is found to be greater than that of Asfarazi, but this difference is so insignificant that even in the maximum case it reaches 1.3 mm which can be ignored. Yet, fluctuation in both series plots follows an almost similar pattern indicating an identical trend in them. The zero precipitation on APHRODITE base in three months of September-October, October-November and November-December of 1962 seems abnormal for which no data construction took place for explanation of precipitation features based on the data of this base. In Figures 11 to 16 respectively the difference of APHRODITE monthly (in 3 figures), seasonal and annual precipitation time series from those of Asfarazi (accept a few dispensable cases) is positive. In other words, precipitation quantity of Asfarazi base is greater than that of APHRODITE base but does not follow a particular pattern and is not regular and has fluctuations. For example, while precipitation difference of the two bases reaches 18 mm in January 2001, in January of the next year this difference decreases to 5 mm and in 2002 this difference becomes the least (1.5 mm). In general, the monthly and seasonal time series have decreased in recent decades. Yet, this is a general conclusion and some fluctuation cases can be noticed as well. Maximum difference between annual precipitation of the two bases is observed in 1962 (79 mm), minimum difference in 1999 (24 mm) and average difference is 50 mm. Comparison of the annual data difference indicates a decreasing trend in the difference between the two bases in recent years which most probably is due to the increase in number of the used gauging stations in APHRODITE bases.

5. Conclusions

Despite availability of diverse qualitative assessment and control methods applied in data supply and processing centers, for many reasons these data may be erroneous and comparison of data from different bases is a suitable method in assessment of data’s reliability and accuracy.

In this study, the time series of APHRODITE and Asfarazi bases were compared. The research results indicate that the estimated precipitation quantity in the monthly, seasonal and annual time series of APHRODITE base (except a few monthly series with insignificant quantities) has been smaller than that of Asfarazi base. However, the amount of this difference has not been constant and has had fluctuation. The data of Asfarazi database which has used a greater number of gauging stations in interpolation are more valid compared to those of APHRODITE base. However, in such studies, long time period and data extending beyond Iran’s political borders are required so as this base can be used considering the lower estimation of precipitation data in the latter base.

The monthly, seasonal and annual time series in both bases show a random behavior and in the series with autocorrelation follow a similar pattern. Studying cross-correlation between series of the two bases indicates a statistically significant lag in zero lag in the monthly, seasonal and annual series, confirming simultaneity of the peak and fall in the two series. Wilcoxon test at 95% confidence confirms significance of the mean difference between the two series, and there is no sufficient evidence to confirm the null hypothesis suggesting absence of a difference between the means of the two series.

Finally, it seems that narrowing difference among the estimates of the two bases in recent years is due to the greater number of gauging stations used by APHRODITE database.

REFERENCES


ily, Italy)," European Water 23/24, 2008, pp. 133-145.


