



Determining Relative Errors of Satellite Precipitation Data over The Netherlands

Qingyu Wang¹, Yijian Zeng²*, Chris M. Mannaerts³ and Vahid Rahimpour Golroudbary³

¹ Power China Northwest Engineering Corporation Limited, Xi'an, China; 837995672@qq.com, qingyuwang12@gmail.com

² Faculty of Geo-Information Science and Earth Observation (ITC), Department of Water Resources, University of Twente, Enschede, The Netherlands; yijianzeng@gmail.com, y.zeng@utwente.nl

³ Faculty of Geo-Information Science and Earth Observation (ITC), Department of Water Resources, University of Twente, Enschede, The Netherlands; c.m.m.mannaerts@utwente.nl (C. M.M.); v.rahimpourgolroudbary@utwente.nl (V.R.G.)

* Correspondence: yijianzeng@gmail.com, y.zeng@utwente.nl; Tel.: +31534874498

Published:

Abstract: Satellite precipitation data are widely used for a variety of studies. However, satellite precipitation estimation is inevitably followed with errors which are caused by different factors. Therefore it is essential to evaluate the relative errors of satellite precipitation data. A realizable method which can be used to quantify the relative errors in large-scale datasets is triple collocation. This method can objectively obtain the relative errors for at least three or more independent products. But before estimation of relative errors, the bias of the products relative to each other should be reduced or removed. This study tests the cumulative distribution function (CDF) matching approach which aims to reduce the bias among three precipitation products over the Netherlands. Afterwards, the triple collocation technique is applied to determine the relative errors of these precipitation products. The three precipitation datasets are, the Climate Prediction Center morphing method (CMORPH), the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) and the gridded rain gauge data interpolated from in situ rain gauge measurement data provided by the Royal Netherlands Meteorological Institute (KNMI). For the relative errors among the three sets of precipitation data, it is found that the relative error of CMORPH is lower than the other two products', KNMI data is at the medium while PERSIANN is the highest one.

Keywords: precipitation products; bias correction; triple collocation; relative errors.

1. Introduction

The study of surface precipitation is important for society and people's livelihood, because inaccurate measurements and forecasts can mean risk to crops, livestock, property and even live [1]. Therefore, obtaining reliable and accurate precipitation data is crucial for local, regional and global agriculture management and hydrologic prediction, like urban flood early warning system. In addition, precipitation has a more direct impact on human life than other atmospheric phenomena, such as heavy rain events and flash floods [2]. In order to understand such disaster and for its reduction, it is necessary not only to improve urban drainage systems, but also to estimate the precipitation in advance and make an early warning system to accommodate extremely rapid response times. Traditionally precipitation is usually measured with rain gauges, but variant instruments have been deployed until now. Most representative instruments are satellites, ground-based radar,

distrometers, microwave links, and in-situ rain gauges. As the most common measurement, the most important advantage for rain gauges is giving direct measurement of rain accumulation. However, there are several drawbacks of rain gauges, such as poor spatial coverage, suffered from wind effects and other resources of errors [1]. Besides, gauges are limited to land regions and islands, thus they are unable to verify oceanic rainfall estimations [3].

Satellite precipitation estimates are widely used to measure global rainfall on near real-time and monthly timescales. In addition, satellites provide insight into the synoptic scale precipitation and are able to obtain an estimate of precipitation in areas where are too remote for ground-based instruments. However, satellite estimates are often affected by instrument noise, semitransparent clouds, and uncertainty in surface emission modelling [4]. In addition, the images from satellites are lack of the details and also usually have larger quantitative errors than ground-based instruments [1]. Therefore, similar to any observation data, it is crucial to investigate their accuracy, internal variability and error structures. This investigation can be done by verifying the satellite estimates against independent data from rain gauges measurements [5].

2. Materials and Methods

2.1. Study Area

The study area is selected as the Netherlands. The country covers an areas of 41543 square kilometers, the geographic coordinates is 5.45° E and 50.30° N. Due to the proximity of the ocean and the effect of the north Atlantic Gulf Stream, it belongs to the temperature zone climate with small climatological variations. The mean annual rainfall changes from 725mm to 925mm [6].

2.2. Precipitation data

There are four kinds of precipitation data which have been used in the Netherlands. As we need a long-term (from 2003 to 2013) precipitation data and also with high spatial and temporal resolution, the precipitation products from the Climate Prediction Center morphing method (CMORPH) and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) are the appropriate choices [7,8].

2.3. Methodology

2.3.1. Create sub maps

Getting the Netherlands sub map from the global CMORPH and PERSIANN precipitation map can be achieved by the ILWIS. As the CMORPH and PERSIANN data needed for this study is 11 years from 2003 to 2013 with 3-hourly temporal resolution and 25 kilometer spatial resolution. So there is a large number of maps need to be edited using the same method. To improve the efficiency, a script was written to create the sub maps for a whole month. Thus, all the maps over the Netherlands can be created month by month.

2.3.2. Create TIFF file and raster map

For subsequently processing the raster map using ArcGIS, the map should be converted to raster type, which is the TIFF file. Raster progress is processed to define the original coordinate system of the sub map, transform to the local coordinate system and extract the Netherlands out using the shape file. In this study, an ArcGIS model is built to achieve this objective .

2.3.3. Interpolation of rain gauge data

Select the hourly precipitation amount data (RH) of 32 stations from KNMI website and sum to 3 hourly. A MATLAB code is developed to integrate the complex data to clear 3 hourly data for everyday over 11 years with 32 rain gauge stations' number on the leftmost row and the stations' name on the rightmost row. As the data collected from in situ rain gauges is discrete and random, therefore spatial interpolation is necessary for creating a continuous dataset. In this study, the

ordinary kriging is selected to interpolate the in situ rain gauge data using R Scripts.

2.3.4. CDF Matching

As precipitation data set derived from satellite is characterized by its specific value and dynamical range. Therefore, satellite data always require scaled before their actual use within hydrological or meteorological models [9]. In this paper, the cumulative distribution function (CDF) matching technique is used to adjust two satellite observations against the interpolation precipitation products and applied for each 25km pixel individually.

2.3.5. Error estimation using triple collocation

Triple collocation can be used to estimate the random error variance in three collocated datasets of the same geophysical variable [10]. Triple collocation assumes the following error model for each time series:

$$R = \alpha + \beta R_t + \epsilon \quad (\text{eq. 1})$$

Assume R_t is the true value of precipitation, α and β are additive and multiplicative biases of the data and ϵ is the relative errors which we want to estimate. In order to Estimate the relative error ϵ , it is necessary to scale or calibrate the datasets to the reference dataset (removing α and β) and calculating the relative error based on these datasets.

3. Result and Discussion

3.1. Statistic difference among different precipitation products

The statistic difference is investigated by calculating the correlation coefficient and the root mean square error (RMSE). Generally, the value of a correlation coefficient can range between -1 and 1 and the weakest linear relationship is indicated by a correlation coefficient equal to 0. The greater the absolute value of a correlation coefficient, the stronger the linear relationship between two variables. The average correlation coefficient of CMORPH vs. PERSIANN is 0.352, the value of CMORPH vs. interpolation is 0.355, and of PERSIANN vs. interpolation is 0.185. It indicates the correlation between CMORPH and interpolation datasets is higher than the other two pairs, while the correlation between PERSIANN and interpolation datasets is the weakest one. The RMSE represents the sample standard deviation of the differences between predicted values and observed values. The average RMSE of CMORPH vs. interpolation is 3.35, while for PERSIANN vs. interpolation the value is 4.13.

3.2. Data histogram

In order to further observe the three precipitation products, a pixel which from row 5 and column 8 from each products is selected out to do the analysis. As it is shown in histogram figure 1, for CMORPH data, the frequency of 0 to 5 mm precipitation close to 1700, for PERSIANN data the frequency is almost the same with CMORPH, but for interpolation data, the frequency is close to 14000. There appears a great difference between two satellite estimation and interpolation estimation because of the satellite's unsuccessful retrievals of precipitation for the relatively low precipitation amount. In addition, the histogram also provides the information that in the Netherlands, the precipitation amount mostly concentrates on 0-5mm for 3hourly products.

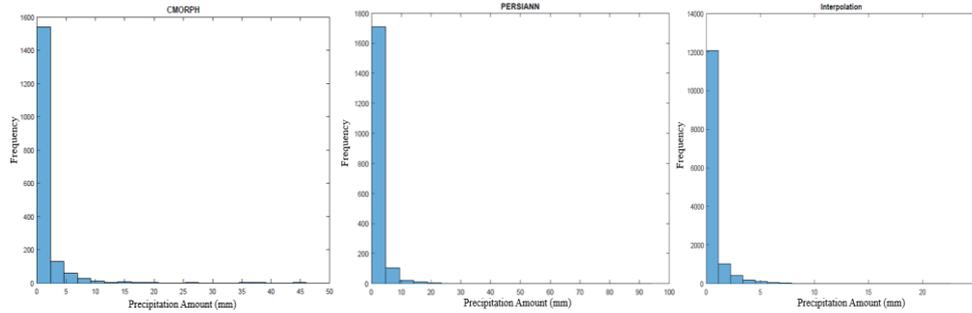


Figure 1. Histogram of one pixel from 3 hourly CMORPH, PERSIANN and interpolation precipitation products.

3.3. Bias correction (CDF matching)

The figure 2 shows results of the correlation coefficient of CMORPH and PERSIANN products versus the interpolation product after CDF matching. From the figure we can see, the average correlation coefficient of CMORPH vs. interpolation is 0.386 and the value of PERSIANN vs. interpolation is 0.221. Compared with the values before CDF matching, the average correlation coefficient of CMORPH vs. interpolation improved from 0.352 to 0.386 and the value of PERSIANN vs. interpolation is improved from 0.185 to 0.221. The improvement of average correlation coefficient is because the CDF matching approach reduced the systematic differences between the satellite datasets and interpolation datasets.

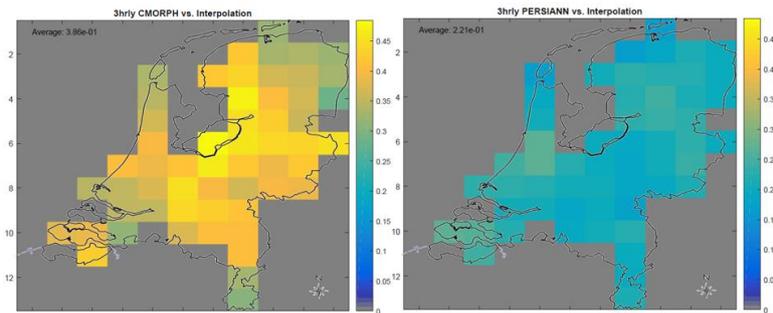


Figure 2. The correlation coefficient of CDF matched 3 hourly CMORPH and PERSIANN products.

The figure 3 shows the lower root mean squared error (RMSE) of CDF matched CMORPH and PERSIANN versus interpolation. The average RMSE of CMORPH vs. interpolation is 3.14, while for PERSIANN vs. interpolation the value is 2.80. Both of them are lower than before CDF matching' values.

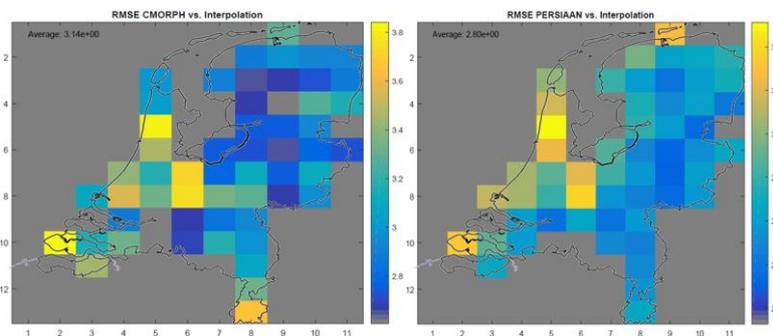


Figure 3. RMSE of CDF matched 3 hourly CMORPH and PERSIANN reference to interpolation.

The pixel from row 5 and column 8 was chosen from the two bias corrected products to draw the histogram, and shown in figure 4. According to the statistics, the frequency of precipitation amount range from 0-5mm, for CMORPH it is about 880, for PERSIANN is 910, while the frequency of precipitation amount range from 0-10mm, for CMORPH it is about 950, for PERSIANN is 930. The statistical result illustrates that the CDF Matching bias correction provided a much better correlation among these two satellite precipitation products.

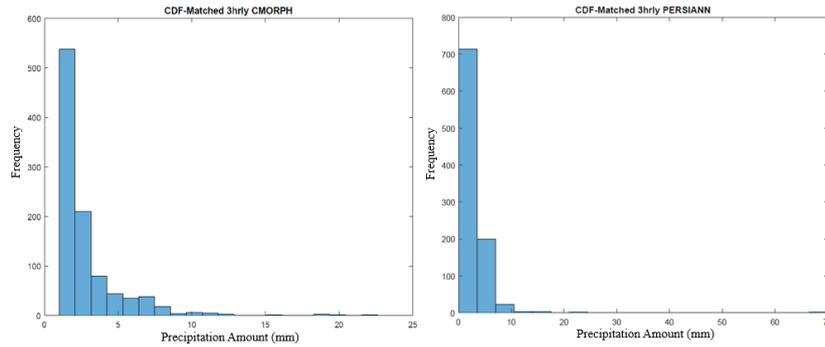


Figure 4. Histogram of one pixel from CDF matched 3 hourly CMORPH, PERSIANN precipitation products.

3.4. Triple collocation

In this section, we present the results of triple collocation analysis. The data used for this analysis are the CMORPH, PERSIANN and interpolation precipitation products, the time scale is 3 hourly and daily respectively.

Firstly, the 3 hourly and daily scale’s triplet number were respectively shown in the figure 5 below. The triplet number in this paper is defined as the number of estimations who are collocated to each other among these three precipitation datasets. From the figure we can see, the daily datasets have much greater triplet numbers than 3hourly. This indicates that at high temporal resolution (e.g. 3hrly in this study) satellite data cannot accurately predict the occurrences of precipitation. It seems that the daily datasets are preferable to be processed using triple collocation technique.

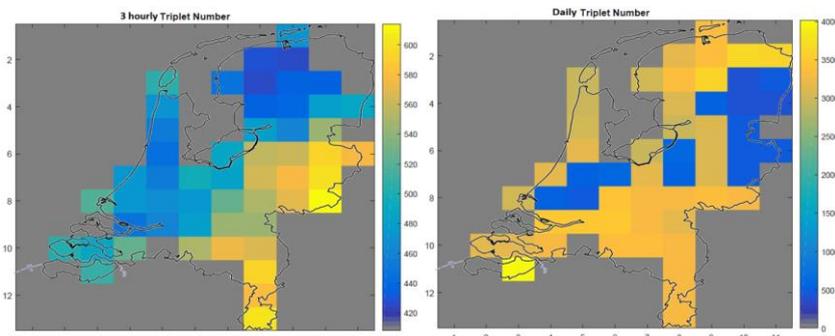


Figure 5. Triplet number of 3 hourly and daily scale respectively.

The results of triple collocation process is showed below as figure 6 and figure 7 for the 3 hourly and daily scale respectively. Comparing these figures carefully, it is not difficult to find that, for 3 hourly scale, the average relative error of CMORPH is 0.58, PERSIANN is 3.64 while interpolation is 2.68. For daily scale, the average relative error of CMORPH is 1.93, PERSIANN is 5.47 while interpolation 4.31. Therefore, the conclusion can be summarized that the relative error of CMORPH is the lowest among these three products and interpolation is at the medium while PERSIANN is the highest one.

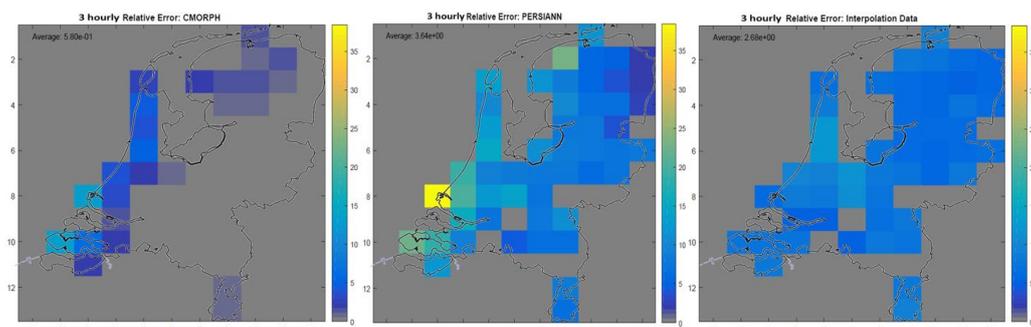


Figure 6. Relative errors of CMORPH, PERSIANN and interpolation products in 3hourly scale.

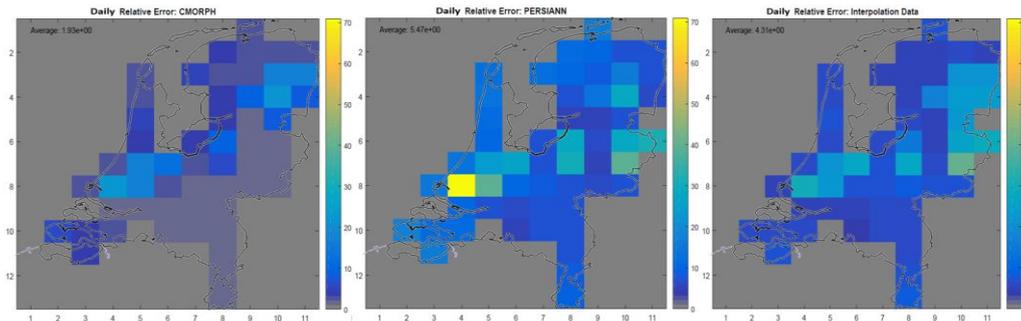


Figure 7. Relative errors of CMORPH, PERSIANN and interpolation products in daily scale

4. Conclusion and Prospect

Based on the research above, the following conclusions can be drawn:

1. The correlation between CMORPH and interpolation rain gauge data is the strongest, two satellite precipitation products (CMORPH and PERSIANN) is medium while PERSIANN and interpolation rain gauge data is the weakest one.
2. CPMORPH product's behavior is better than PERSIANN's when they are correlated to the interpolation products.
3. For the low precipitation amount like 0-5mm, the two satellites provide a relatively weak retrieval.
4. The relative error of CMORPH is lower than the other two products', interpolation is the medium while PERSIANN is the highest one.

The research can be referenced to the bias correction and triple collocation of the precipitation products over the Netherlands. The results of this paper can be useful for further determination of the relative weights of these precipitation products and obtain a merged precipitation product.

References

1. Beek, R. van de. (2013, June 10). Rainfall variability in the Netherlands from radars, rain gauges, and disdrometers. s.n.]. Retrieved from <http://edepot.wur.nl/259986>.
2. Levizzani, V., Amorati, R., & Meneguzzo, F. (2002). A Review of Satellite-based Rainfall Estimation Methods. Retrieved from http://www.researchgate.net/publication/252272255_A_Review_of_Satellite-based_Rainfall_Estimation_Methods.
3. Ebert, E. ., & McBride, J. . (2000). Verification of precipitation in weather systems: determination of systematic errors. *Journal of Hydrology*, 239(1-4), 179–202. [http://doi.org/10.1016/S0022-1694\(00\)00343-7](http://doi.org/10.1016/S0022-1694(00)00343-7).
4. Miralles, D. G., Crow, W. T., & Cosh, M. H. (2010). Estimating Spatial Sampling Errors in Coarse-Scale Soil Moisture Estimates Derived from Point-Scale Observations. *Journal of Hydrometeorology*, 11(6), 1423–1429. <http://doi.org/10.1175/2010JHM1285.1>.
5. Levizzani, V., Bauer, P., & Turk, F. J. (2007). *Measuring Precipitation from Space: EURAINSAT and the Future*. Springer Science & Business Media. Retrieved from <https://books.google.com/books?id=uks14jNh8bwC&pgis=1>.
6. Attema, J., & Lenderink, G. (2011). Mean precipitation changes and coastal effects in the ENSEMBLES regional climate model simulations. *KNMI Scientific Report 2011-03*, 42 pp.
7. Joyce, R. J., Janowiak, J. E., Arkin, P. a., & Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *Journal of Hydrometeorology*, 5(3), 487–503. [http://doi.org/10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](http://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
8. Satellite Precipitation - CHRS. (n.d.). Retrieved August 28, 2015, from http://chrs.web.uci.edu/research/satellite_precipitation/activities00.html.
9. Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., & Wagner, W. (2008). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters*, 35(24), L24403. <http://doi.org/10.1029/2008GL035599>.
10. Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research*, 103(C4), 7755. <http://doi.org/10.1029/97JC03180>.



© 2018 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).