Bias correction of the rainfall CMORPH satellite product using observed data from Bogotá, Colombia

Correction des biais des mesures de pluie du satellite CMORPH à l'aide de données observées sur Bogotá, Colombie

Oscar Manuel Baez Villanueva[1], Juan Diego Giraldo Osorio[2], Luis Arturo González Ortiz[1] and Jackson Roehrig[3]

[1]Universidad Autónoma de San Luis Potosí (México); oscarmbaez89@hotmail.com; lagonzalez@uaslp.mx
[2]Pontificia Universidad Javeriana (Colombia); j.giraldoo@javeriana.edu.co
[3]Institute for Technology and Resources Management in the Tropics and Subtropics (Germany); jackson.roehrig@th-koeln.de

RÉSUMÉ

Des stations pluviométriques installées à Bogotá (Colombie) ont été utilisées pour corriger un biais non-uniforme présent dans le produit satellite CMORPH à résolution temporaire de 3 heures, en utilisant une procédure de régression gaussienne. La correction des données satellite a été effectuée sur une base mensuelle et annuelle, de manière à observer l'influence du caractère saisonnier dans le processus de correction. Cette approche a montré une réduction significative des biais présents dans les données satellite, en calant correctement les chutes de pluie les plus fortes. Les résultats ont également présenté des limites du fait que la procédure a une tendance à sur-ajuster les chutes de pluie de faible intensité.

ABSTRACT

Rain gauge stations from the urban area of Bogotá (Colombia) were used to correct the nonuniform bias in the CMORPH satellite product, which has a temporal resolution of 3 hours, using a Gaussian Process Regression. The correction of the satellite data was performed in both monthly and yearly basis, to observe the influence of seasonality in the bias correction process. This approach showed that the satellite biases were reduced significantly and the heavy rain events were well adjusted. The results also showed limitations because it tends to overadjust the light rain events.

KEYWORDS

Bias correction, CMORPH, daily precipitation, Gaussian Process Regression, satellite rainfall data

1 INTRODUCTION

Due to the satellite development, it is possible to identify and quantify in an approximated way the rainfall in a constant time scale. Rainfall satellite data is available worldwide and it reduces or matches the timescale of the rain gauge stations. The objective of the satellite measured rainfall is to generate information of the occurrence, quantity and spatial distribution of precipitation (V. Levizzani, R. Amorati, & F. Meneguzzo, 2002). CMORPH (CPC Morphing technique) is one of the existing satellite precipitation products. The entire CMORPH record (December 2002- present) has a global spatial coverage at 0.25°x0.25° (NOAA, 2015).

The advantages of satellite based precipitation datasets are the good spatial coverage, spatial measurement and the reduced time resolution compared to gauge stations (C. Kidd, V. Levizzani, & R. Ferraro, 2009). The importance of the application of the satellite rainfall datasets in hydrology is due of the high temporal distribution of the measurement and in some places, the observations from spaceborne instrumentation currently produce the only measured data for a large part of the world (T. Cohen Liechti, J. P. Matos, J.-L. Boliat, & A. J. Schleiss, 2012). For these reasons, satellite datasets are commonly used for hydrological purposes but they must be calibrated to overcome biases of the precipitation measurement.

2 METHODOLOGY

Three automatic rain gauge stations, located in Bogotá were used to perform the correction of the satellite product. The used stations are located in the same pixel of CMORPH intentionally to prove the effectiveness of the spatial rainfall assessment of the product. The pixel area represents 760 km^2 in the studied zone.

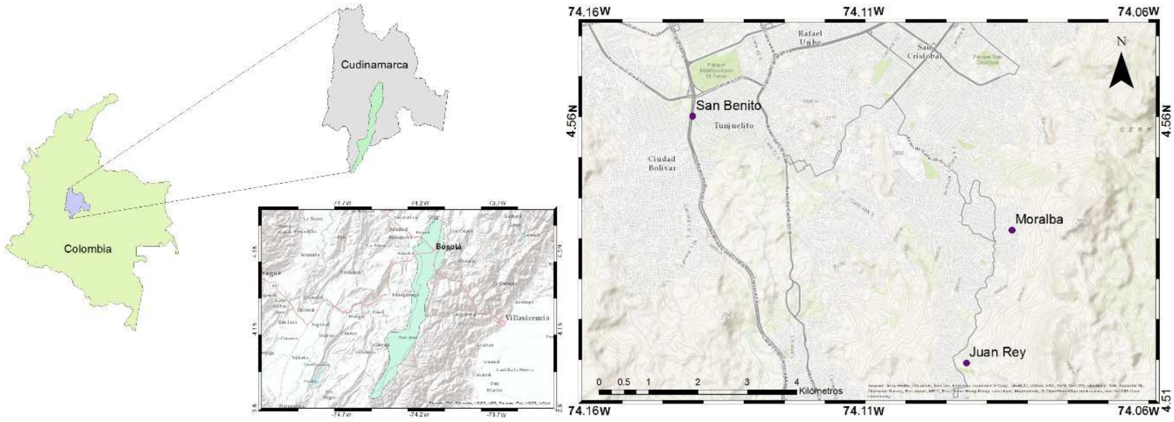

Figure 1: Macro and micro localization of the rain gauge stations.

2.1 Gaussian Process Regression

The advent of kernel machines, such as Support Vector Machines and Gaussian Processes, has opened the possibility of flexible models which are practical to work with (Schölkopf & J. Smola, 2002). A Gaussian Process is a collection of random variables of any finite number of which have a joint Gaussian distributions. (C. Edward Rasmussen, 2002)

In this method is assumed that the observed variables are normal, and the coupling between them takes place by means of the covariance matrix of a normal distribution. The kernel matrix is used as the covariance matrix which is a convenient way of extending Bayesian modelling of linear estimators to nonlinear situations.

A linear kernel function was used to perform the bias correction for CMORPH using the average daily value of the rain gauge stations located in the same area where the satellite product time serie was obtained. This kernel is useful especially dealing with large sparse data vectors as is usually the case of precipitation.

2.2 Adjustment of CMORPH to observed data

The time window of the rain gauge stations used to correct the satellite product CMORPH time serie was from January of 2008 to December of 2013. The satellite product time serie was obtained for the area of interest using a bash script, which extracted the 3 hourly intensity value of the 17,536 raster maps that conform the mentioned time window.

The intensity values were converted to precipitation depth, then aggregated in daily values, to be compared with the average of the rain gauge stations. Two different approaches were made to correct the bias of the CMORPH time serie; the non-split bias correction and the monthly bias correction.

2.2.1 Non-split bias correction

Non-split bias correction, is the correction performed with the data of the analyzed time window.The cumulative distribution function (CDF) was constructed for the average of the rain gauge stations and compared with the CDF of CMORPH for all the dataset in the studied years. The Gaussian Processes Regression model with the linear kernel was applied to the daily data to adjust CMORPH to the gauge stations data.

2.2.2 Monthly bias correction

In this approach, the data of each month were taken separately from the dataset, and the comparative of the CDF´s (CMORPH product and gauge stations) were built for each month. The CDF´s of one month of the dry and wet season, February and November respectively, are shown in the results (the figures of the other months were omitted).The same linear kernel in the Gaussian Processes Regression was used for all the months separately to adjust the satellite product.

2.2.3 Adjustment of the dry days data

It was detected in both approaches, that in the observed dry days in the rain gauge stations, the corrected CMORPH time serie, had a constant value of precipitation (in all cases less than 1 mm). The excess value was subtracted of the corrected CMORPH daily values to correct the bias of the days with no rain.

3 RESULTS

It was observed that the satellite product data represents the average rain in each time step (3 hours) by using the three rain gauge stations located in the study area.

3.1 Non-split bias correction

The CMORPH product bias was corrected using the linear kernel. In figure 2, is showed the comparison of the averaged rain gauge stations CDF and the corrected CMORPH CDF. The correction of the satellite product fits well to observed data including the values of the high quantiles that represent the high rain events.

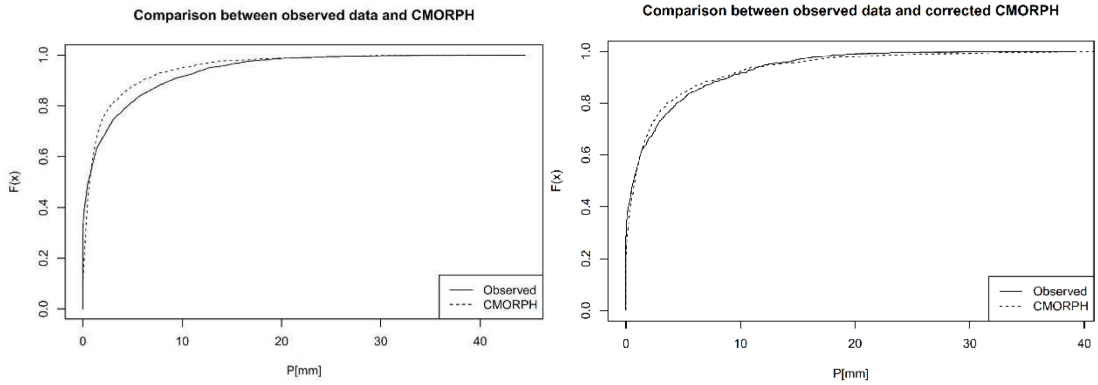

Figure 2: Comparison between the average of the rain gauge stations, CMORPH and corrected CMORPH.

To compare the correction, the root mean square deviation (RMSD) was calculated between the observed data and the uncorrected CMORPH product (observed RSMD) and between the observed data and the corrected CMORPH values (corrected RSMD). The value of the observed RMSD is 26.56, bigger than the corrected RSMD with a value of 16.94, the reduction in this value shows a better fit to observed data with the corrected satellite data.

3.2 Monthly bias correction

The CMORPH product was corrected monthly with the linear kernel. In figure 3 and 4, is showed the comparison of the dry month of February and the wet month of November respectively, between the observed values and the CMORPH satellite product observed and corrected.

The RMSD was applied to compare the correction, as in the first approach the observed RMSD was calculated between the observed data and the uncorrected CMORPH product (the value is the same in both approaches) and the corrected RMSD, between the observed data and the monthly corrected CMORPH values. The corrected RMSD value is 1.83, it is visible that the monthly correction is a better approach to correct the bias of the mentioned satellite product.

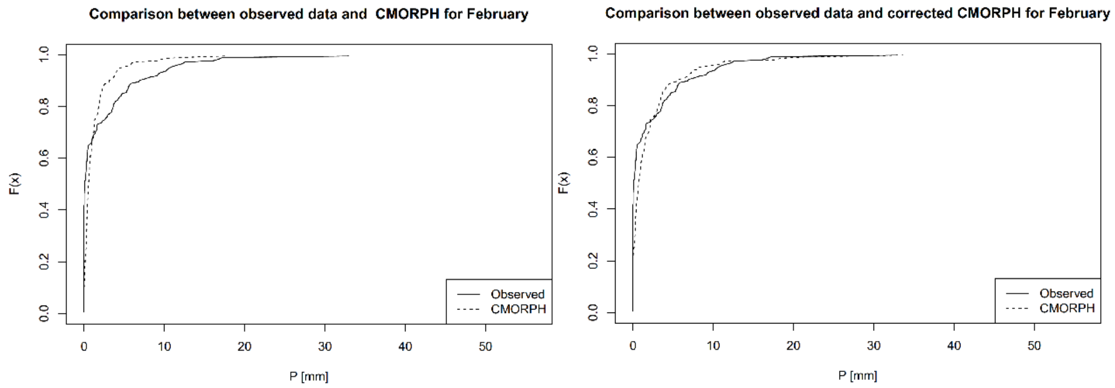

Figure 3: Comparison between the average of the rain gauge stations, CMORPH and corrected CMORPH in the dry month of February.

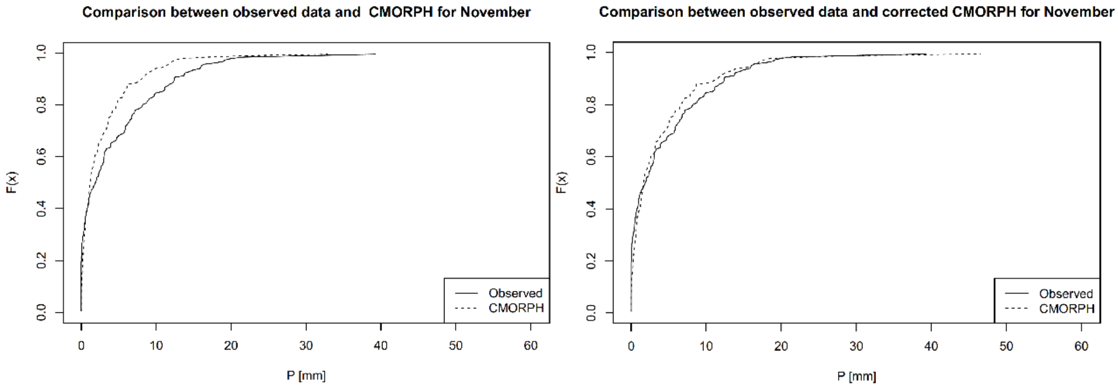

Figure 4: Comparison between the average of the rain gauge stations, CMORPH and corrected CMORPH in the wet month of November.

4 CONCLUSION

The rainfall satellite data is an important tool for the assessment of precipitation. The time scale of satellite data has a shorter time resolution than the common gauge station which measures daily values of precipitation depths; the satellite adjusted time series can be used to determine past event durations in flood analysis without the need of short time resolution gauge stations placed on the study area, problem frequently found in praxis.

The Gaussian Process Regression is a good approximation to adjust CMORPH satellite product data and the linear kernel has shown a good correlation, mainly on heavy rain events. The monthly bias correction has the best adjustment of both approaches, this is due to the inclusion of the seasonality in the correction process. Anyhow the light rain events are overadjusted.

Other kernels and functions can be applied to reduce the corrected RMSD value as the multiparameter gamma which takes in account the zeros (days with no precipitation) in the series to have a better correlation of the data. This approach can be done applying bootstrapping theory to locate the intervals of the different gamma parameters to obtain the best fitted function to correct the CMORPH satellite precipitation product.

The precipitation data of satellite products is recommended in areas with low density of rain gauge stations and when more time resolution in precipitation events is needed, this in order to obtain more information about the spacial variability in the precipitation process. Although rain gauge stations are needed for the satellite bias correction and the low density of stations can be used to develop a statistical model to remove the bias of satellite data.

LIST OF REFERENCES

Kidd, C., Levizzani, V., & Ferraro, R. (2009). *“Satellite Precipitation Measurements for Water Resource Monitoring”,* Journal of the American Water Resources Association, Vol. 45(3), 567–579.

Levizzani, V., Amorati, R. and Meneguzzo, F. (2002). *A Review of Satellite-based Rainfall Estimation Methods*. Istituto di Scienze dell’Atmosfera e del Clima. Bologna, Italia

Liechti, T. Cohen. Matos, J.P., Boliat, J.L. and Schleiss, J. (2012). *Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River Basin,* Hydrology and Earth System Sciences 16, 489-500.

NOAA. (2015). *NOAA Satellite Information System (NOAASIS).* Retrieved in October 06-2015, from http://noaasis.noaa.gov/NOAASIS/ml/genlsatl.html

Rasmussen, C. Edward. (2002). *Gaussian Processes in Machine Learning.* Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

Schölkopf, B., & Smola, J. A. (2002). *“Learning with kernels”*. Cambridge, Massachusetts, The MIT Press.