UNIVERSIDAD DE CUENCA

FACULTAD DE INGENIERIA
MAESTRIA EN ECOHIDROLOGIA

Rainfall monitoring network design using conditioned Latin Hypercube Sampling and satellite precipitation estimates: An application in the Ecuadorian Amazon

Tesis previa a la obtención del título de
Magíster en Ecohidrología

Autor:
Juan José Contreras Silva
C.I: 0105688584

Directora:
Daniela Elisabet Ballari, PhD.
C.I: 1751894716

Codirector:
Esteban Patricio Samaniego Alvarado, PhD.
C.I: 0102052594

Cuenca – Ecuador
Junio 2018

Abstract

Rain gauge networks are crucial for enhancing the spatio-temporal characterization of precipitation. In tropical regions, scarcity of rain gauge data and climatic variability, make conventional approaches to design rain gauge networks inadequate and impractical. In this study, we propose the use of conditioned Latin Hypercube Sampling (cLHS) method with multi-temporal layers of remotely sensed precipitation measurements for capturing the spatio-temporal precipitation patterns in ungauged areas. The study was conducted in the Amazon Region of Ecuador, for which monthly precipitation averages were derived based on a 16-year period of Tropical Rainfall Measuring Mission (TRMM 3B43 V7) data which were used as prior information to select representative sampling points through cLHS. Two scenarios for the sampling design were considered and evaluated, one without and one with restrictions on accessible sites according to the proximity to roads and settlements. Results showed that both optimized networks captured the variability of precipitation according to the TRMM climatology. Furthermore, evaluation against an independent satellite precipitation dataset showed that the optimized networks support mapping precipitation based on ordinary kriging (OK). Comparison with regular and random sampling methods showed that, particularly when a practical scenario is considered, the optimized network provided more reliable results over time, highlighting the suitability of the network to capture temporal changes and map precipitation with high accuracy. The proposed approach could be easily adopted in other ungauged and poorly accessible regions for rain gauge network design as well as to the design of multi-objective monitoring networks.

Keywords: rain gauge network, spatio-temporal monitoring, conditioned Latin Hypercube Sampling, satellite precipitation data, ungauged areas, poorly accessible areas, Ecuadorian Amazon.

Resumen

Las redes de pluviómetros son cruciales para mejorar la caracterización espacio-temporal de la precipitación. En regiones tropicales, la escasez de datos pluviométricos y la variabilidad climática contrastante hacen que los enfoques convencionales para diseñar redes de pluviómetros sean inadecuados e imprácticos. En este estudio, proponemos el uso del método de muestreo por hipercubo latino condicionado (cLHS) con imágenes multitemporales de precipitación detectadas de forma remota para capturar los patrones espacio-temporales de precipitación en áreas no monitoreadas. El estudio se realizó en la región amazónica del Ecuador, cuyos promedios mensuales de precipitación se obtuvieron en base a un período de 16 años de datos de la Misión de Medición de Precipitaciones Tropicales (TRMM 3B43 V7) que se utilizaron como información previa para seleccionar puntos de muestreo representativos mediante cLHS. Se consideraron y evaluaron dos escenarios para el diseño del muestreo, uno sin y otro con restricciones en sitios accesibles de acuerdo con la proximidad a carreteras y asentamientos. Los resultados mostraron que ambas redes optimizadas capturaron la variabilidad de la precipitación de acuerdo con la climatología TRMM. Además, la evaluación frente a un conjunto de datos de precipitación por satélite independiente mostró que las redes optimizadas son adecuadas para el mapeo de la precipitación basado en el kriging ordinario (OK). La comparación con métodos de muestreo regular y aleatorio mostró que, particularmente cuando se considera un escenario práctico, la red optimizada proporcionó resultados más confiables a lo largo del tiempo, destacando la idoneidad de la red para capturar cambios temporales y cartografiar la precipitación con alta precisión. El enfoque propuesto podría adoptarse fácilmente en otras regiones no monitoreadas y de escasa accesibilidad para el diseño de redes de pluviómetros, así como para el diseño de redes de monitoreo multiobjetivo.

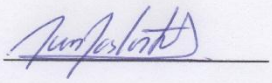
Palabras clave: red de pluviómetros, monitoreo espacio-temporal, muestreo condicionado por hipercubos latinos, datos de precipitación satelital, áreas no monitoreadas, áreas de difícil acceso, Amazonía ecuatoriana.

Table of contents

Cláusula de Propiedad Intelectual

Juan José Contreras Silva, autor del trabajo de titulación **"Rainfall monitoring network design using conditioned Latin Hypercube Sampling and satellite precipitation estimates: An application in the Ecuadorian Amazon"**, certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor.

Cuenca, junio de 2018



Juan José Contreras Silva

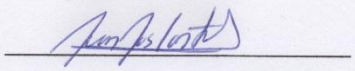
C.I: 0105688584

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Juan José Contreras Silva en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación **"Rainfall monitoring network design using conditioned Latin Hypercube Sampling and satellite precipitation estimates: An application in the Ecuadorian Amazon"**, de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, junio de 2018



Juan José Contreras Silva

C.I: 0105688584

1. Introduction

Precipitation is the most important component of the hydrological cycle and it has a fudamental role in different socio-economic activities. Accurate knowledge of the spatio-temporal variability of precipitation is essential for many scientific and management fields (Celleri, Willems, Buytaert, & Feyen, 2007; Michaelides et al., 2009; Padrón, Wilcox, Crespo, & Célleri, 2015; Shaghaghian & Abedini, 2013; Tapiador et al., 2012). Rain gauges are the most common instruments used to quantify the variability of precipitation (Michaelides et al., 2009; Tapiador et al., 2012). Despite advances in new technologies to estimate precipitation such as satellite remote sensing and weather radars, rain gauges remain to be the most accurate source of information and their measurements are essential for calibrating and validating predictions based on other sources (Zubieta, Getirana, Espinoza, & Lavado, 2015).

Many regions of the world, especially in tropical areas, remain ungauged owing to accessibility and budget constraints (Hobouchian, Salio, García Skabar, Vila, & Garreaud, 2017; Ochoa, Pineda, Crespo, & Willems, 2014; Ward, Buytaert, Peaver, & Wheater, 2011; Zubieta et al., 2015; Zulkafli et al., 2014). An ungauged region of particular importance is the Ecuadorian Amazon. This region is part of one of the richest biodiversity reserves of the planet (Ceballos & Ehrlich, 2006; Finer, Jenkins, Pimm, Keane, & Ross, 2008; Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000) and it is home for many indigenous ethnic groups, including some of the world's last uncontacted communities (Finer, Moncel, & Jenkins, 2010; Larrea & Warnars, 2009; Pappalardo, De Marchi, & Ferrarese, 2013). Hydrologically, the Ecuadorian Amazon basin supplies some of the major rivers within the Amazon, and contributes approximately 2.3% of the total discharge of the Amazon basin (Laraque, Ronchail, Cochonneau, Pombosa, & Guyot, 2007). The region has vast areas of intact tropical forest and according future projections, it has a high probability of stable climatic conditions regarding global warming (Killeen, Douglas, Consiglio, Jørgensen, & Mejia, 2007).

Despite this importance, little is known about the amount and variability of precipitation in the region since many areas remain ungauged. Several researchers have stressed the need of a well-developed monitoring network in the region with a representative spatio-temporal coverage that would support (1) an adequate characterization of precipitation variability (Morán-Tejeda et al., 2016), (2) the evaluation of satellite precipitation products and streamflow simulations (Manz et al., 2017; Zubieta, Getirana, Espinoza, Lavado-Casimiro, & Aragon, 2017) and (3) for a correct assessment of downscaling precipitation techniques (Ulloa, Ballari, Campozano, & Samaniego, 2017), among others. To design such monitoring network, it is required to take several things into account, such as the lack of prior information, the inherent spatio-temporal nature of precipitation and the inaccessibility that might make difficult the rain gauge network deployment and data retrieval.

Several methods have been proposed to accurately design rain gauge networks. These methods have been classified into two categories: measurement-free and measurement-based approaches (Chacon-Hurtado, Alfonso, & Solomatine, 2017). On one hand, measurement-free methods rely on technical guidelines or recommendations, which focus on the areal coverage per rain gauge (Baltas & Mimikou, 2009; K. Wang et al., 2015; WMO, 2008). They consider several physiographic characteristics of the area such as geomorphologic but they do not rely on existing precipitation measurements. On the other hand, measurement-based methods are based on prior information provided by existing rain gauge networks. These methods have shown to be the most robust and they are widely used for optimizing (augmentation, relocation and reduction) existing networks (Chacon-Hurtado et al., 2017). Such methods include geostatistics (e.g. Adhikary, Yilmaz, & Muttil, 2014; Cheng, Lin, & Liou, 2008; Pardo-Igúzquiza, 1998), cross-correlation (e.g. Nazaripour & Daneshvar, 2017), entropy theory (e.g. H. Xu et al., 2015; Yoo, Jung, & Lee, 2008), model output error (e.g. Volkmann, Lyon, Gupta, & Troch, 2010; H. Xu, Xu, Chen, Zhang, & Li, 2013) and hybrid methods (Shaghaghian & Abedini, 2013; P. Xu et al., 2018).

In ungauged or poorly gauged areas such as the Ecuadorian Amazon, the feasibility of measurement-based methods is limited by the deficiency of current rain gauge data. In addition, prevalent contrasting climatic regions turns difficult the extrapolation of ground-based information from other regions. For these reasons, approaches that incorporate auxiliary information sources to design rain gauge networks should be developed.

In ungauged areas, satellite predictions are the only source of information (Collischonn, Collischonn, & Morelli Tucci, 2008). Precipitation data provided by the quasi-global Tropical Rainfall Measuring Mission (TRMM) have been widely used for environmental monitoring applications (Du et al., 2013; Li, Christakos, Ding, & Wu, 2018; Moffitt, Hossain, Adler, Yilmaz, & Pierce, 2011; Su, Hong, & Lettenmaier, 2008; Xue et al., 2013), the study of climate trends and variability (Almazroui, Islam, Jones, Athar, & Rahman, 2012; Retalis, Katsanos, & Michaelides, 2016) as well as for the regionalization of precipitation (Ballari, Giraldo, Campozano, & Samaniego, 2018). Although satellite precipitation estimates are relatively inaccurate compared to rain gauge measurements, their exhaustive coverage provides valuable information to identify spatio-temporal features of precipitation over any given area (Libertino, Sharma, Lakshmi, & Claps, 2016). Previous studies in Ecuador showed the suitability of TRMM estimates to characterize long-term precipitation climatology (Ballari, Castro, & Campozano, 2016; Erazo et al., 2018; Ulloa et al., 2017). Dai et al. (2017) proposed a methodology for rain gauge network design based on remotely sensed precipitation measurements obtained from a weather radar. The method correctly captured average precipitation but temporal variability and accessibility restrictions were not included in the design approach. In addition, the availability of precipitation radar data is limited in developing countries.

Conditioned Latin Hypercube Sampling (cLHS) provides an approach for incorporating prior information from remote sensing instruments as well as accessibility restrictions in a sample design. cLHS is a multivariate stratified random strategy (Minasny & McBratney, 2006) that has been proven to be an efficient sampling method because it captures the marginal variability of several variables using a relatively small sample (Brungard & Boettinger, 2010; Domenech, Castro-Franco, Costa, & Amiotti, 2017; Ramirez-Lopez et al., 2014; Stumpf et al., 2016). Roudier et al. (2012), Mulder et al. (2013) and Yin et al. (2016) have shown the effectiveness of the method when restrictions of accessibility are taken into account. cLHS has been extensively used as a sampling strategy for digital soil mapping (e.g. Chu, Lin, Jang, & Chang, 2010; Clifford, Payne, Pringle, Searle, & Butler, 2014; Mulder et al., 2013; Pahlavan Rad et al., 2014; Rosemary, Vitharana, Indraratne, Weerasooriya, & Mishra, 2017; Vitharana, Mishra, Jastrow, Matamala, & Fan, 2017) but only few studies have considered other environmental fields. Yin et al. (2016) used cLHS with remote sensing images for validating leaf area index in mountainous areas. This work was extended to the multi-temporal domain and other biophysical properties in Yin et al. (2017).

The objective of this study was to apply and evaluate the cLHS technique with long-term TRMM climatology as a spatio-temporal sampling scheme for a rain gauge network design in ungauged areas. The approach was used to select the number and the locations of possible monitoring points in the Ecuadorian Amazon region. Due to the lack of coverage of ground-based precipitation measurements in the study area, the approach was evaluated through an independent satellite precipitation dataset, comparing the quality of monthly precipitation obtained by ordinary kriging (OK) with the optimized sampling points and alternative sampling schemes.

2. Study area

The study area is located in the eastern part of Ecuador. The Ecuadorian Amazon is one of the four biogeographic regions of Ecuador and it comprises almost half of the territory of the country. For the current study, the Amazon region of Ecuador was considered as the area located below the occidental foothills of the Andes covering an area of 83949 km^2 (Figure 1). The study area extends over an area of Tropical Cloud Forest and Tropical Rain Forest. The average temperature is close to 25˚C and annual rainfall exceeds 3000 mm (Laraque et al., 2007). Climate and seasonality are controlled by large-scale meteorological phenomena such as the Intertropical Convergence Zone (ITCZ) and the South American Monsoon System (SAMS) which cause convection and heavy rainfall in the northern parts of the Amazon (Espinoza Villar et al., 2009; Marengo et al., 2012). A bimodal regime is distinguished in the region, with two rainy seasons from March to July and another one from October to December (Ballari et al., 2018). According to the National Institute of Meteorology and Hydrology of Ecuador (INAMHI), a total of 20 rain gauges were reported for the Ecuadorian Amazon in 2015. The spatial distribution of the current network is uneven, as shown in Figure 1; where 15 gauges are located along the western border of the region.

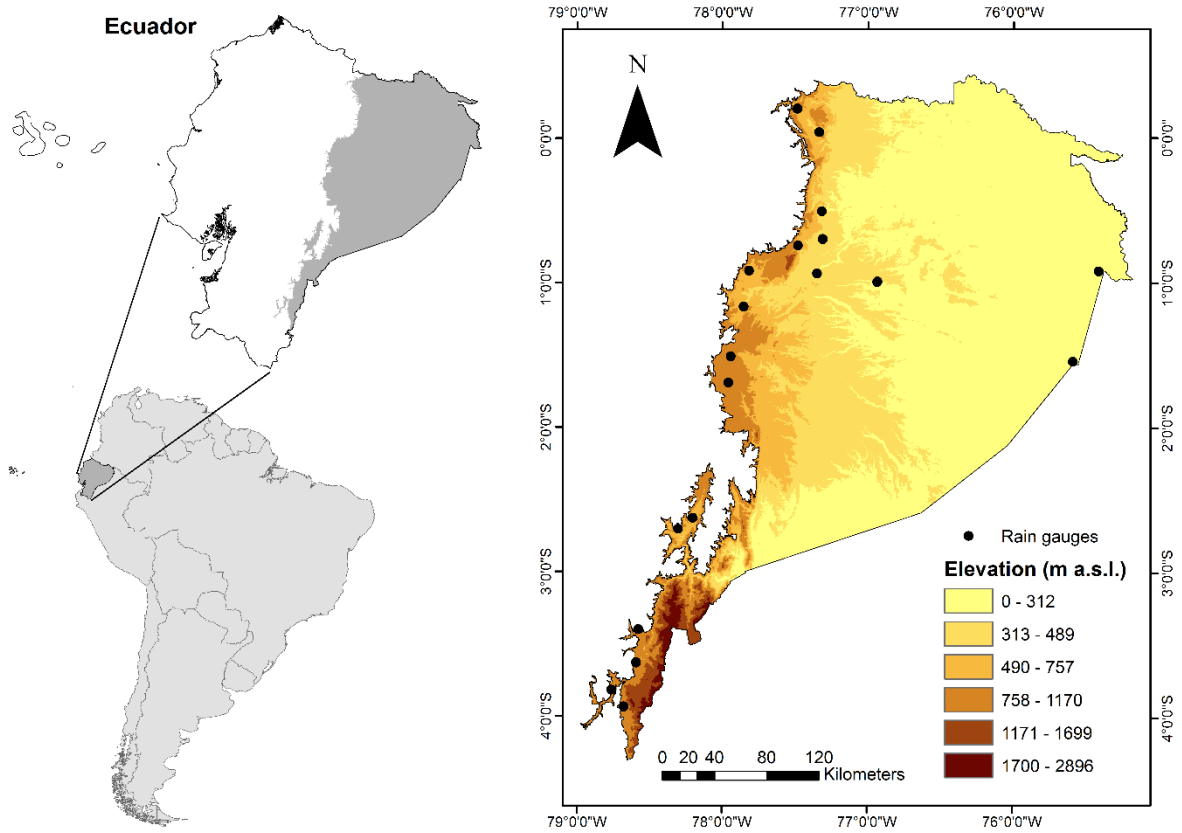


Figure 1. Study area and location of the rain gauge network.

3. Materials and methods

3.1. Satellite precipitation data

Two satellite-based precipitation datasets were used in our study: TRMM 3B43 Version 7 and GPM IMERG Level 3 Version 05. The first dataset was used as prior information to design the rain gauge network in the Ecuadorian Amazon, while the later was used for the evaluation of the optimized networks. Detailed information of the two datasets is presented below.

TRMM 3B43 Version 7

Monthly precipitation data from TRMM, called TRMM 3B43 were used in this study. The dataset was produced by the TRMM Multi-satellite Precipitation Analysis (TMPA) algorithm that combines all available precipitation datasets from different satellite sensors of TRMM and surface rain gauge data to correct bias and provide a best estimate of precipitation at relatively fine spatial resolution of 0.25° x 0.25° (~ 27 km x 27 km) covering 50°N–S areas (Huffman et al., 2007). TRMM 3B43 Version 7 was downloaded from the NASA database (https://pmm.nasa.gov/data-access/downloads/trmm) for the period January 1999 - December 2014. The year 1998 was excluded from our analysis due to the unusual occurrence of El Niño (ENSO)

event in front of the coast of Ecuador which could affect the construction of the precipitation climatology. Therefore, mean monthly TRMM precipitation (hereafter TRMM climatology) for the Ecuadorian Amazon was derived from the 16-year period comprising 192 satellite images. The original grid size of TRMM (~ 27 km) was resampled to 1 km by bilinear interpolation following the first procedure step of Ulloa et al. (2017). The resample procedure was done to take accessibility restrictions in the study area within the design scheme into account (see section 2.3.2; Step 3).

IMERG Level 3 Version 05

Satellite precipitation estimates from the Global Precipitation Measurement (GPM) mission were used as reference precipitation for the evaluation of the sampling scheme. GPM, which is the successor of TRMM, provides the next generation of precipitation products at a spatial resolution of 0.1° x 0.1° (~ 10 km x 10 km) since April 2014 (Hou et al., 2014). Early assessments of GPM products around the world have shown a better performance compared to TRMM products at different temporal scales (Prakash, Mitra, Pai, & AghaKouchak, 2016; Tang, Zeng, et al., 2016; Tang, Ma, Long, Zhong, & Hong, 2016; Wang, Lu, Zhao, Jiang, & Shi, 2017; Xu et al., 2017) and it also has better capabilities to identify local precipitation patterns caused by orographic effects (Manz et al., 2017; Mayor, Tereshchenko, Fonseca-Hernández, Pantoja, & Montes, 2017; Sharifi, Steinacker, & Saghafian, 2016). For the study, monthly estimates of the IMERG Level 3 Final Run research product Version 05 (hereafter IMERG L3) for 2015 were downloaded from the NASA database (https://pmm.nasa.gov/data-access/downloads/gpm). Similarly to the TRMM climatology data, IMERG L3 data were also resampled to 1 km resolution through the bilinear resample method.

3.2 Sampling design

3.2.1 Spatio-temporal sampling

In order to select representative spatio-temporal sampling points for monitoring precipitation and, at the same time, minimize costs related to monitoring and accessibility, cLHS with operational constraints was used (Roudier et al., 2012). The cLHS attempts to cover the multidimensional distribution corresponding to a set of predictor variables by using a stratified random sampling. Each of the marginal distributions of the covariate space is divided into equiprobable intervals that are each targeted to be sampled once. The steps of the cLHS algorithm are detailed below (Minasny & McBratney, 2006):

1. Divide the quantile distribution of *X* (being X the variables, i.e.12 images of TRMM climatology) into n strata (number of sampling points), and calculate the quantile distribution for each variable, $q_j^i, \dots, q_j^{i+1}$. Calculate the correlation matrix C for *X*.
2. Select n random sampling points from N (being N the total number of pixels of *X*), and calculate the correlation matrix T of *x* (being x a sub-sample of *X*).

3. Calculate the objective function (*OF*) and cost function (*CF*). Because precipitation is a continuous variable, the overall *OF* integrates two objective functions:

$$O1 = \sum_{i=1}^{n}\sum_{j=1}^{k}\left|\eta\left(q_j^i \le x_j \le q_j^{i+1}\right) - 1\right| \quad (1)$$

where k is the number of variables, $\eta\left(q_j^i \le x_j \le q_j^{i+1}\right)$ is the number of sampling points (x_j) that falls between quantiles q_j^i and q_j^{i+1}. To ensure that the correlation of the sampled variables will replicate the original data, the following objective function is added:

$$O2 = \sum_{i=1}^{k}\sum_{j=1}^{k}\left|c_{ij} - t_{ij}\right| \quad (2)$$

where *c* is an element of C, the correlation matrix of X; and t is the equivalent element of T, the correlation matrix of *x*. The overall *OF* is:

$$OF = w_1 O_1 + w_2 O_2 \quad (3)$$

where w is the weight given to each component of the objective function. Here w_1 and w_2 were both set to 1. The ideal value of the *OF* is zero which indicates that the sample is completely stratified and the correlation matrix is exactly reproduced.

Roudier et al. (2012) incorporated accessibility constraints and operational costs into the cLHS. They proposed the use of a second objective function called cost function (*CF*) equal to the sum of the costs over the sampling points. In the optimization process *CF* is evaluated along with the *OF* of standard cLHS. The *CF* is:

$$CF = \sum_{p=1}^{n} C_p \quad (4)$$

where C_p is the cost associated with sampling point p.

4. As cLHS becomes an optimization problem, Simulated Annealing (SA) is used to evaluate the quality of the objective function and the magnitude of the cost function to find the optimum solution. The SA is an iterative, combinatorial optimization algorithm in which a sequence of combinations is generated by deriving a new combination by randomly changing the previous combination (van Groenigen, Siderius, & Stein, 1999). Each time a new combination is generated, the quality of the *OF* (Eq. (3)) and the magnitude of the *CF* (Eq. (4)) are evaluated and compared with the value of the previous combination. The new combination is accepted if the quality of the *OF* or the magnitude of the *CF* has improved by the change. However, the annealing algorithm also accepts some changes that worsen the *OF* and *CF*. This is to avoid being trapped in a local optimum. The probability of accepting a worse sample in terms of *OF* is:

$$p_{OF} = e^{-(\Delta OF/T_i)} \quad (5)$$

with $T_i = T_{i-1} \times K$

where p_{OF} is the probability of accepting a solution that worsens the objective function, ΔOF is the change in the objective function, T_i is the temperature in iteration *i*, and K is a constant factor decreasing the temperature at each iteration. The probability of accepting a worse combination in terms of *CF* is:

$$p_{CF} = e^{-(\Delta CF/T_i)} \quad (6)$$

where p_{CF} is the probability of accepting a solution that worsens the cost function, and ΔCF is the change in the cost function. In order to configure the initial value of temperature and K several trials were performed previously to ensure the best solution. After multiple tests, the initial value for *T* was set to 2 and K was set to 0.99.

5. Run the SA algorithm until the *OF* and *CF* falls beyond a given stop criterion or a specific number of iterations.

For further details on the cLHS algorithm the reader is referred to Minasny & McBratney (2006) and Roudier et al. (2012).

3.2.2 Implementation

The current rain gauge network (Figure 1) was not included in the design scheme, instead a new network was optimized for the study area. For the design, two scenarios were defined:

Scenario 1, the entire study area without restrictions of accessibility was take into account for the location of the rain gauge network.

Scenario 2, the proximity to roads and human settlements as suitable areas for the location of the rain gauge network were taken into account.

Although scenario 1 could be considered ideal because it covers the whole study area, it is impractical in most real cases. Therefore, the scenario 1 was used to compare the theoretical impact of accessibility restrictions in the region on the spatio-temporal representativeness of the rain gauge network.

Regarding the variables, the TRMM climatology (12 images, from January to December) was used to capture the mean seasonal variation of precipitation throughout the year in the study area. In addition, to ensure a representative sampling in the geographic space and achieve a relatively uniform coverage of sampling sites in the study area, X-coordinate and Y-coordinate were used as additional variables.

A flowchart of the main steps of the implementation procedure is shown in Figure 2. The implementation of the cLHS algorithm was done in R with the *clhs* package (Roudier, 2017).

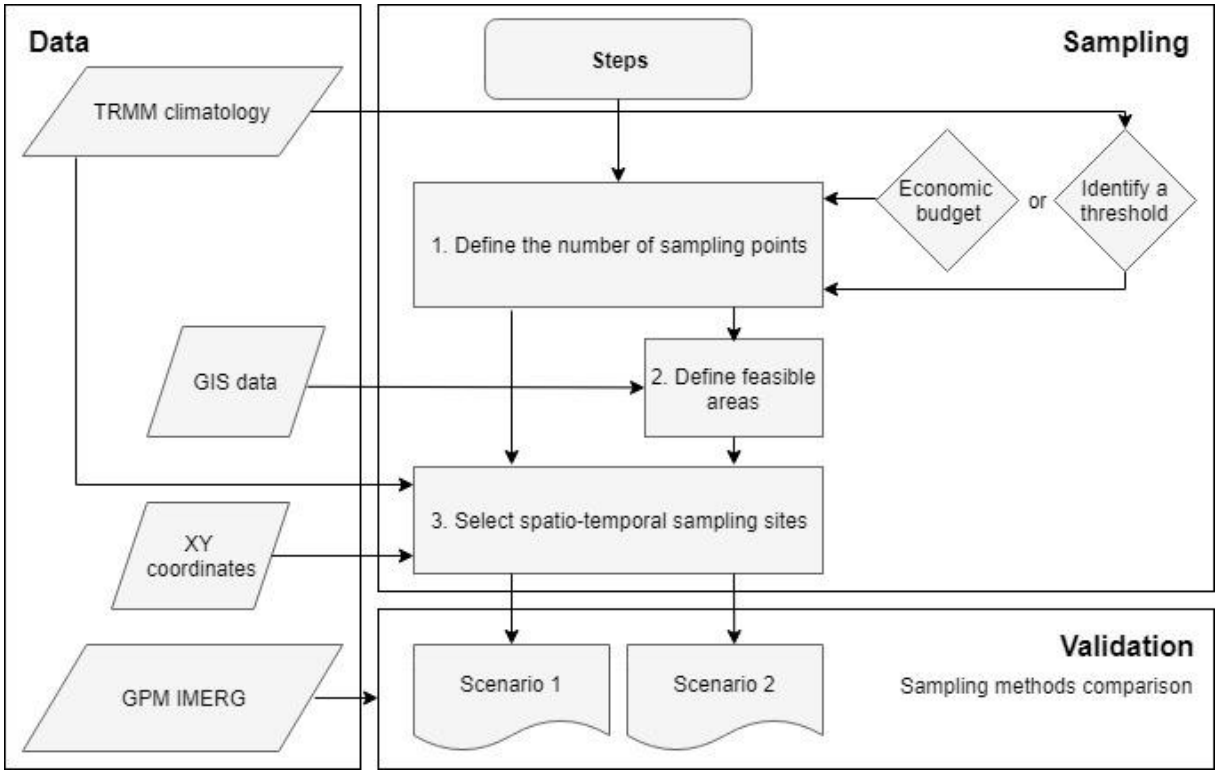


Figure 2. Flowchart of the implementation of the network design.

Step 1: Define the sample size

In many practical applications the number of available rain gauges may be defined by budget limitations. However, the selection of an adequate number can be based also on a threshold in the rate of increment in a given objective function. To select a representative number of sampling points which capture the variability of precipitation in the Ecuadorian Amazon, the absolute difference between the sample standard deviation obtained by cLHS (s_{TRMM}) and the original standard deviation of the TRMM image (σ_{TRMM}) was calculated for each month with different sample sizes. The evaluated number of sampling points ranged from 10 to 100 (in intervals of 5). For each climatology image and for each evaluated sampling size, the cLHS sampling was performed 100 times. The absolute difference between the standard deviation of sample TRMM and the original TRMM ($|s_{TRMM} - \sigma_{TRMM}|$) was computed as:

$$|s_{TRMM} - \sigma_{TRMM}| = \left|s(x_{ij}) - \sigma(x_i)\right|$$

where x_{ij} is the sample space of TRMM in month i (i.e. January, February… December) with sample size j (i.e. 10, 15 …100), x_i is the original space of TRMM in month i. Finally, the average of the monthly absolute standard deviation difference of the 100 runs was used to define the best trade-off between the sampling size and the representativeness of the network.

Step 2: Assess feasible areas for the location of monitoring points

For scenario 2, cLHS requires a cost map which represents the difficulty or the cost of reaching every place within the area. The cost layer is used in the optimisation process of the cLHS algorithm to penalise the points that are difficult or impossible to reach in the field, and guide the sampling process to schemes that are easier to implement operationally (Roudier et al., 2012). In our study area, sites easy to reach were those close to roads, trails and human settlements (Figure 3a). Although some navigable rivers could be used for transportation, especially to connect scattered settlements in the east of the region, these were not considered in our current design to avoid large travel times and high costs of boat transportation. Feasible areas to the allocation of rain gauges were defined by a buffer of 1 km around roads/trails and 3 km around settlements. We assumed that the difficulty of reaching any point in the region and the operational costs increase linearly with the increase of the distance from the feasible areas. Thus, distance from aforementioned sites were created using the Euclidean distance at a resolution of 1 km. Data were downloaded from the web page of the Military Geographic Institute of Ecuador (IGM) (http://www.geoportaligm.gob.ec/portal/) and the National Information System (SNI) (http://sni.gob.ec/accesibilidad-y-vialidad). The resulting cost map shows the difficulty of reaching every pixel within the area as a function of the distance from feasible areas. The cost map is shown in Figure 3b where higher costs are observed at the east border of the Ecuadorian Amazon where accessibility is more difficult given the lack of roads and human settlements.

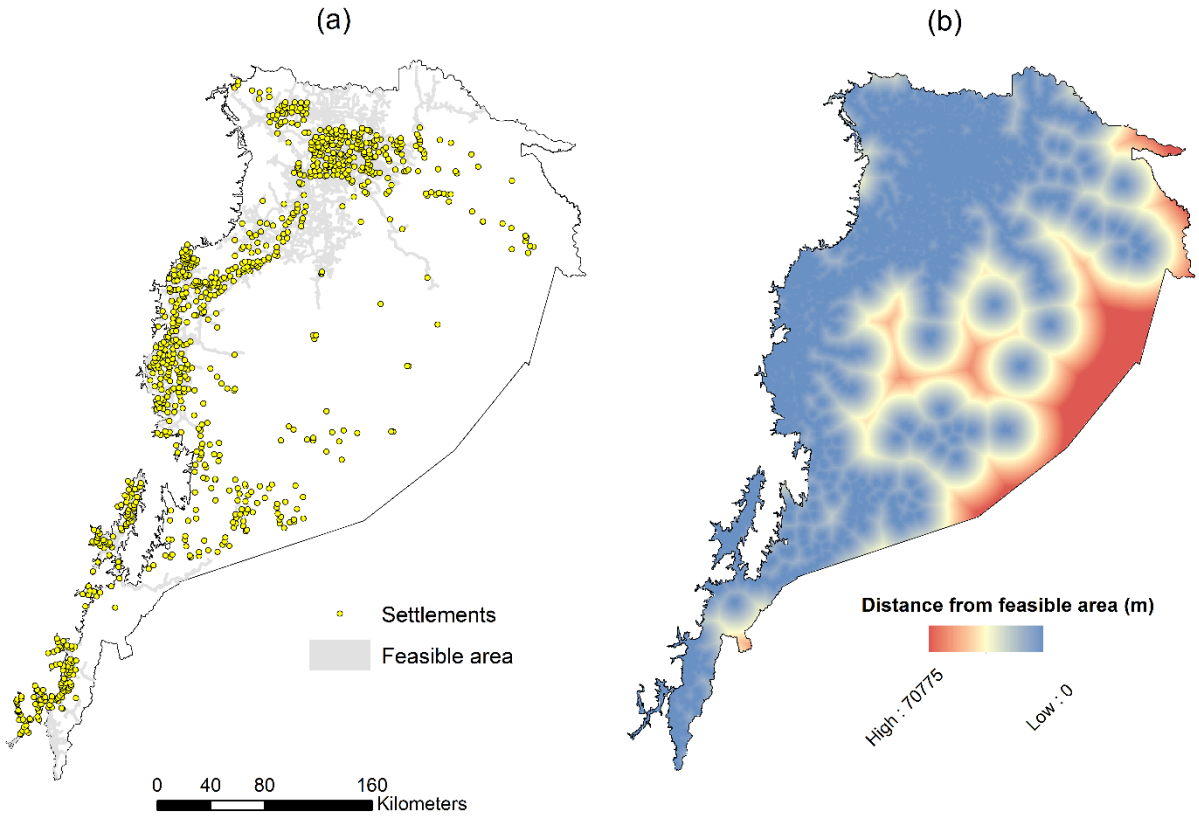


Figure 3. Feasible areas for the location of precipitation monitoring points. (a) Road/trail network, human settlements and defined feasible area in the Ecuadorian Amazon and (b) cost map representing the difficulty of reaching every place in the study area as a function of the distance from feasible sites.

Step 3: Selection of spatio-temporal sampling sites with cLHS

With the selected number of sampling sites, cLHS was run for both scenarios. The number of iterations for the SA algorithm was set to 1 x 10^7 to approach an optimum network output.

3.3 Evaluation of the optimized networks

A rain gauge network should provide a reasonably accurate precipitation estimate at any point in the study area trough spatial interpolation. Based on this criterion, the optimized networks were assessed in terms of the accuracy of mapping precipitation. We used two interpolation methods: ordinary kriging (OK) and inverse distance weighting (IDW). OK and IDW were used to generate monthly precipitation maps with the sampling points of the optimized networks. For the interpolations, precipitation values at each sampling point were extracted from the corresponding pixel of the monthly IMERG L3 image. Webster & Oliver (2007) recommend at least 50 to 100 points for satisfactory variogram estimates. However, having this number of observations might be a challenging issue in many climate studies due the scarcity of

ground measurements. Despite this limitation, several studies have successfully applied OK to interpolate precipitation with a limited number of samples (e.g. Adhikary, Muttil, & Yilmaz, 2017; Cruz-Roa, Olaya-Marín, & Barrios, 2017; Mair & Fares, 2011; S. Wang et al., 2014).

The *sp and gstat* R packages were used to define the variogram models and perform the interpolation with OK and IDW at 1 km resolution. The method of moments was used to fit variogram models to experimental variograms. The best variogram fit for each month was selected testing five variogram models: spherical, exponential, Gaussian, Matern and Matern-Stein. Resulting maps were evaluated pixel-to-pixel with the original IMERG L3 images for each month which were assumed as the "true" precipitation for the 2015 year. Percent bias (PBIAS), root mean squared error (RMSE) and Nash-Sutcliffe Efficiency coefficient (NSE) were used to assess the accuracy of the predictions. Correlation coefficient (r) was used to assess the degree of agreement between precipitation image and interpolation precipitation fields. All aforementioned indices were computed according to the formulas shown in Table 1.

Table 1. Statistical indices used to quantify the performance of mapping precipitation.

Statistical Metrics	Unit	Equation	Perfect Value	Equation Number
Percent Bias (PBIAS)	%	$PBIAS = \frac{1}{N}\sum_{i=1}^{N}\frac{(S_i - O_i)}{O_i} \times 100$	0	(7)
Root Mean Square Error (RMSE)	mm	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2}$	0	(8)
Nash-Sutcliffe Efficiency Coefficient (NSE)	NA	$NSE = 1 - \frac{\sum_{i=1}^{N}(S_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}$	1	(9)
Coefficient of Correlation (r)	NA	$CC = \frac{\sum_{i=1}^{N}(S_i - \bar{S})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{N}(S_i - \bar{S})\sum_{i=1}^{N}(O_i - \bar{O})}}$	1	(10)

where N is the total number of pixels of IMERG L3 and interpolation image; i is the ith of IMERG L3 and interpolation image; O_i means IMERG L3 observation and $\bar{O}$ the average of IMERG L3 observation. S_i and $\bar{S}$ are the interpolation estimates by OK and their average, respectively.

Finally, in order to evaluate the effectiveness of the proposed approach, the optimized networks were compared with two alternative sampling schemes: random and regular sampling. Although random and regular sampling methods differ from cLHS because they do not use prior information to guide the sampling, these are commonly alternative strategies for the allocation of sampling points (de Gruijter, Brus, Bierkens, & Knotters,

2006). The *sp* R package was used to generate 100 different networks for each sampling scheme. The monthly precipitation values provided by these networks were interpolated with OK for which variogram models were defined for each network. Resulting interpolation was evaluated according to the procedure mentioned above (Table 1). To assess the robustness of the networks, the median and the interquartile range (IQR) values of the statistical indices were selected as measures of central tendency and dispersion. Thus, the best sampling scheme should provide the lowest median value of PBIAS and RMSE, the highest value of NSE and r, and the lowest IQR for all statistical indices.

4. Results

4.1 Spatio-temporal patterns of TRMM climatology

Figure 4 shows the monthly spatio-temporal patterns of precipitation at 1 km captured by TRMM 3B43 V7 datasets and its distribution in the Ecuadorian Amazon during the study period. Although precipitation in the study area is present throughout the year (ranging from 70-480 mm per month), it shows a clearly different monthly spatial pattern (Figure 4a). A bimodal precipitation regime is observed with a relative dry month in August and two wet periods in the rest of the year, with two peaks in June and November. All year long, the highest levels of precipitation are present in the middle of the region close to the Andean Cordillera similar to those reported by Ballari et al. (2018). Figure 4b shows that precipitation distribution varies monthly and TRMM climatology resampled at 1 km conserve the overall distribution shape of the original data.

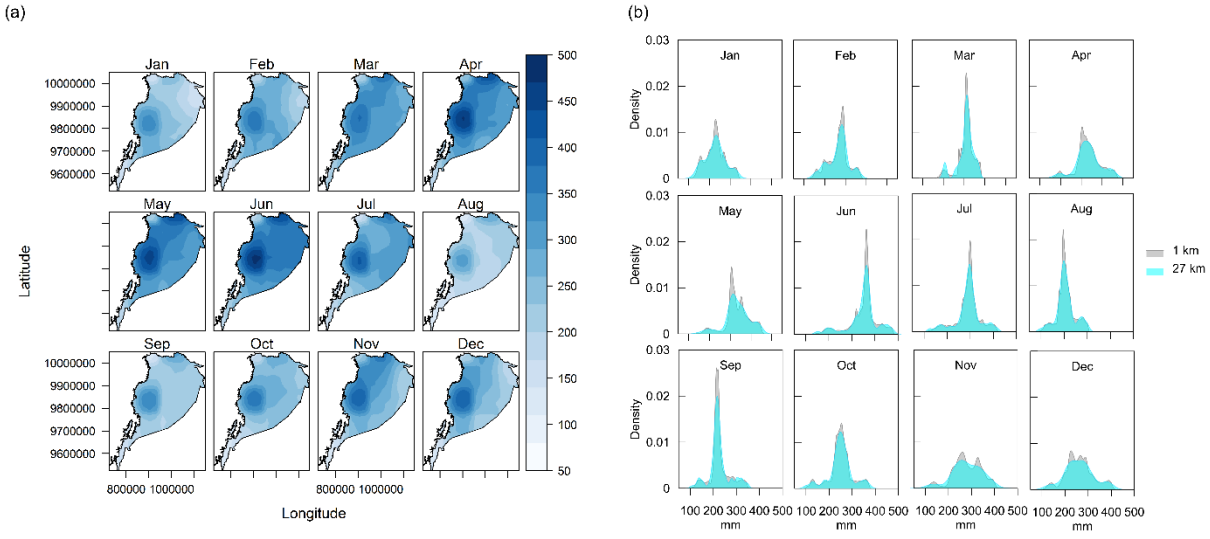


Figure 4. Precipitation in the Ecuadorian Amazon captured by TRMM 3B43 V7 dataset during 1999-2014. (a) Monthly spatial patterns and (b) Monthly distribution of original TRMM 27km and resample TRMM 1km.

4.2. Optimal sample size

As expected, Figure 5 shows that the absolute difference between the sample TRMM standard deviation and the entire TRMM standard deviation decreases when sample size increases. This trend indicates that representativeness will increase when sampling size increases. The difference between standard deviations showed larger dispersion for smaller sample sizes; however, this dispersion decreased gradually as the sample size increased. In fact, larger dispersions were found for samples smaller than 25 points (Figure 5). With the aim of selecting a number of sampling points in practical terms, 25 points were identified as the best compromise between the spatio-temporal representativeness of the network and the economic costs.

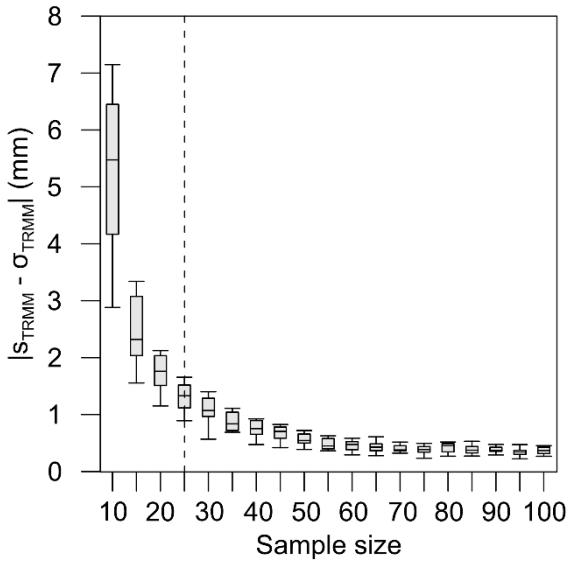


Figure 5. Relation between the monthly absolute difference between the sample TRMM and the original TRMM standard deviations and sample size. The red dash line indicates the selected optimal sample size.

4.3 Optimized networks

Visual inspection of the *OF* and *CF* traces revealed no improvements above 4.7 x 10^6 and 2 x 10^6 iterations respectively (Figure 6), suggesting that optimal (near) solutions were achieved. The *OF* reached a value of 60.50 and of 79.76 in scenario 1 and 2, respectively. In scenario 2, the reached value of the *CF* was 8211.10 meters.

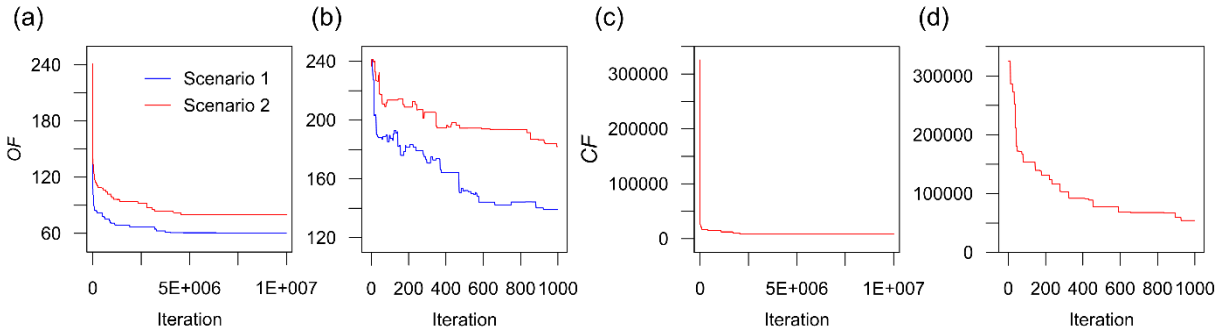


Figure 6. Evolution of the *OF* (a, b) and *CF* (c, d) with the number of iterations.

Figure 7 shows the optimized sampling sites for both scenarios as well as the location of the current rain gauge network. The optimized sampling sites for scenario 1 tended to be uniformly spread across the entire study area with only 3 sampling points located within the feasible areas (Figure 7a). On the other hand, in scenario 2, most of the sampling sites were spread within the feasible areas. Nevertheless, 2 sampling sites were located outside of these areas, one at the distance of 1000 m and another at 7211 m from the feasible areas (Figure 7b). Comparing the networks of both scenarios, we noted that the location of the farthest point is quite similar between the scenarios (note the biggest points of Figure 7). This suggests that this point contains unique information that is not found within the feasible area, and also that it was not possible to reallocate this point within this area without the loss of representativeness in the network.

Figure 8 shows the density distribution of the TRMM climatology and the distribution of the 25 sampling points from the optimized networks. Both networks capture reasonably well the original probability distribution of TRMM climatology in all months. The distributions of both networks do not show significant discrepancies between them. However, the optimized network for scenario 1, in general, outperformed the optimized network for scenario 2.

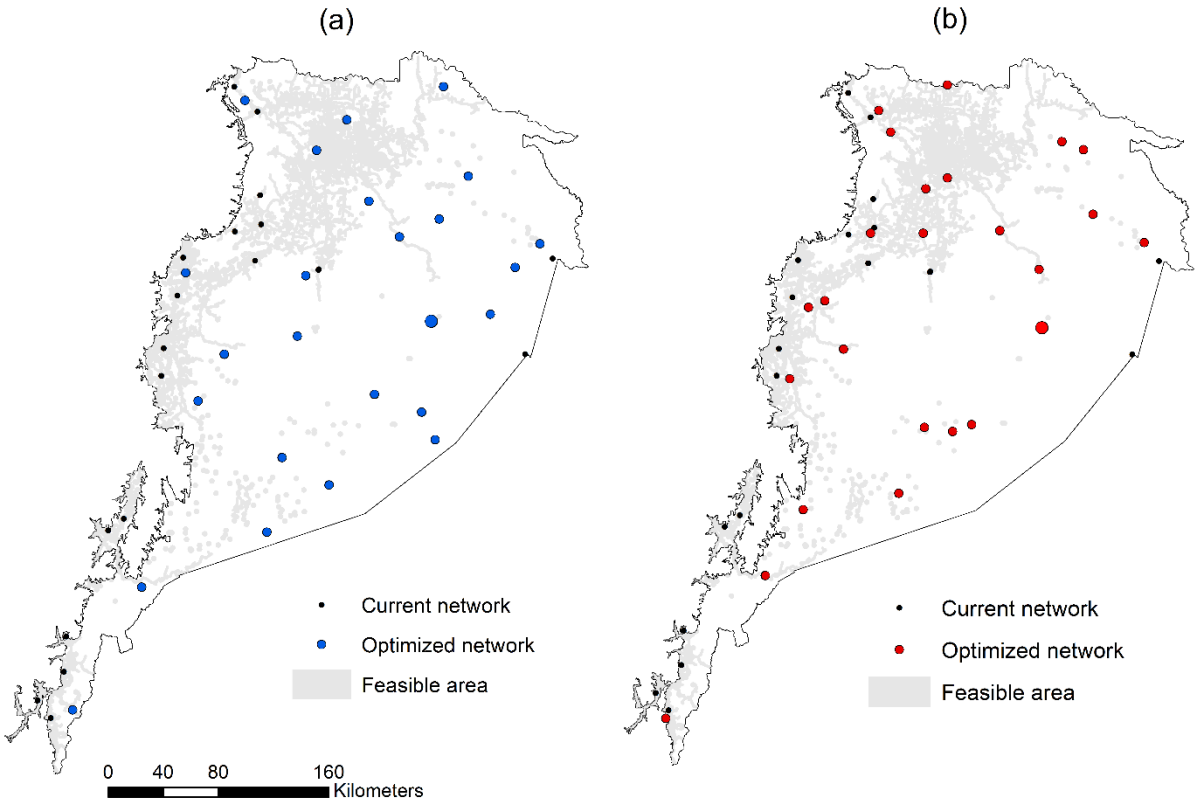


Figure 7. Location of the current and optimized rain gauge networks in (a) scenario 1 (considering entire study area) and (b) scenario 2 (considering proximity to accessible sites).

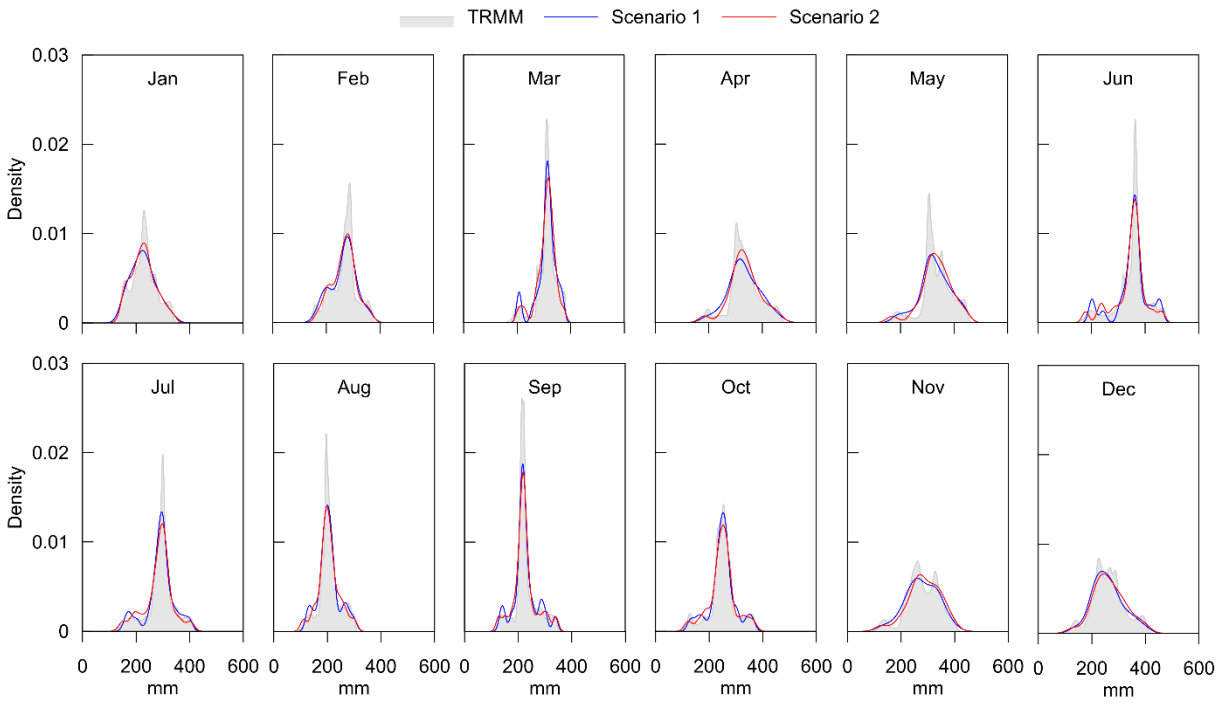


Figure 8. Distributions of TRMM climatology and the optimized networks in scenarios 1 and 2.

4.4 Evaluation of the optimized networks

In order to verify that OK provided optimum estimates of precipitation with the selected number of sampling points of both optimized networks, the average monthly performance of IDW and OK was compared. Table 2 shows that OK provided more accurate spatial predictions than IDW for both scenarios. OK provided a lower absolute PBIAS and a lower RMSE as well as a higher NSE and a higher r than IDW. This confirmed the suitability of OK to estimate precipitation with 25 sampling points in our study area. According to these results, the further assessments were based only on the use of the OK method.

Table 2. Average monthly statistics between the interpolated precipitation and IMERG L3 images using IDW and OK interpolation methods.

	Scenario 1				Scenario 2			
Interpolation Method	Absolute PBIAS (%)	RMSE (mm)	NSE	r	Absolute PBIAS (%)	RMSE (mm)	NSE	r
IDW	0.78	23.76	0.78	0.91	1.32	26.19	0.74	0.89
OK	0.29	17.40	0.87	0.93	0.70	18.27	0.87	0.94

4.4.1 Spatio-temporal performance of the optimized networks

No substantially differences were observed in the performance of the interpolation between the optimized networks. Figure 9 shows the monthly statistical indices between the interpolated precipitation and the IMERG L3 images for both optimized networks. High interpolation performances were obtained with both networks for all months, with PBIAS ranging from -1.7% to 1%, RMSE from 11.42 mm to 28.33 mm, NSE from 0.61 to 0.96, and r from 0.81 to 0.98. RMSE has its highest value in April (the maximum peak of precipitation), while NSE and r has their lowest values in August (the minimum peak of precipitation) in both networks. Furthermore, similar temporal performances of the networks were observed for all statistical indexes except for PBIAS (Figure 9). Regarding to PBIAS no distinctive pattern was observed in the monthly performance of the networks (Figure 9a). In average, the optimized network in scenario 1 slightly outperformed the network in scenario 2 according to most of the statistical indices (Table 2 and Figure 9). This indicates that no substantial difference was observed in the temporal performance of the optimized network when only accessible sites were considered (scenario 2) compared to an ideal network (scenario 1).

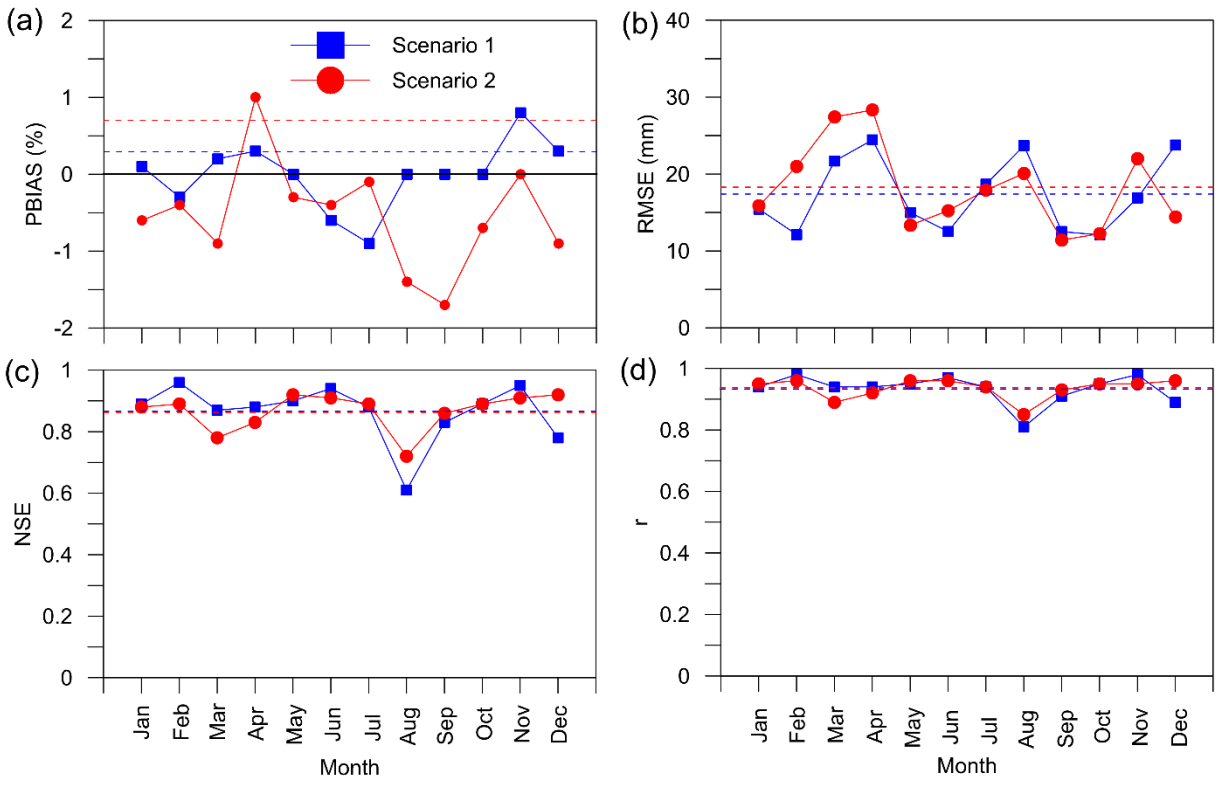


Figure 9. Monthly statistical indices of (a) PBIAS, (b) RMSE, (c) NSE and (d) r between precipitation obtained by OK and IMERG L3 images. Horizontal blue and green dash lines are the average of the performance of the optimized network in scenario 1 and 2.

Figure 10 shows the average monthly spatial distribution of statistical indices between the interpolated precipitation and the IMERG L3 precipitation images for both optimized networks. The interpolated precipitation with the network optimized in scenario 1 exhibited a better agreement with the IMERG L3 data than the optimized network in scenario 2. However, a quite similar spatial performance of the networks was observed with PBIAS, NSE and r, in contrast to RMSE values. In both networks, a relatively uniform distribution of errors was observed especially for PBIAS, NSE and r (Figure 10a, c, d). Low values of PBIAS (±12%) are observed for most parts of the study area. However, higher overestimation (around 20% to 65%) is observed along the west boarder of the Amazon over the foothills of the Andean Cordillera (Figure 10a). Similar spatial patterns were observed for NSE in both scenarios (Figure 10c), with high values of NSE (>0.75) in most of the region and lower values (<0.50) along the foothills of the Andes. On the other hand, high values of r (>0.8) are observed for the whole study area in both scenarios (Figure 10d). Regarding to RMSE, the lowest values are observed close to the sampling points in both scenarios. However, high values were found along the border of the Andean Cordillera, and in some disperse regions in the central and east parts of the Amazon in scenario 1, whereas in scenario 2 high RMSE values are observed in most central and east areas of the Amazon where no sampling points were located (Figure 10b).

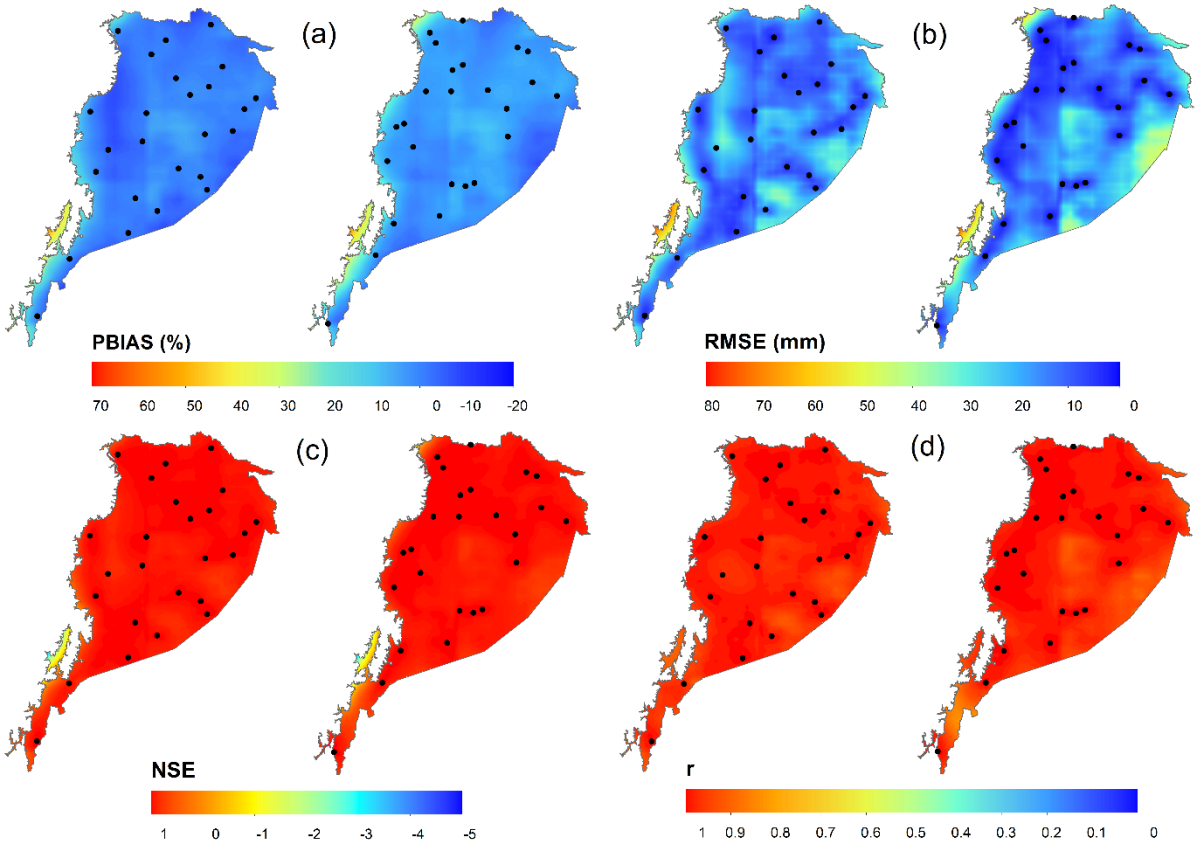


Figure 10. Average monthly spatial distribution of (a) PBIAS, (b) RMSE, (c) NSE and (d) r, between interpolated precipitation (OK) with the optimized networks and the IMERG L3 images. Black points are the location of the sampling points.

4.4.2. Comparison with alternative sampling methods

For a fair comparison of the sampling schemes in scenario 2, an additional buffer was established around the accessible areas to sample with random and regular methods. The buffer distance was set to 7211 meters, corresponding to the maximum distance of the optimized sampling points from the accessible areas (scenario 2).

The summary of monthly statistical values between the interpolated precipitation and the IMERG L3 obtained with the optimized networks and the 100 networks with each alternative sampling scheme are presented in Table 3 and Figure 11. In general, the monthly performance of the optimized networks obtained by means of the cLHS method outperformed the networks obtained by a random and a regular sampling, especially when a pragmatic scenario is considered.

Scenario 1 slightly overestimated median PBIAS while it was slightly underestimated in scenario 2. In both scenarios, the optimized network performed better than random and regular sampling and had the smallest IQR value compared to the other sampling methods (Table 3 and Figure 11a). Regarding to RMSE, median values indicated that the optimized network outperformed the alternative sampling schemes in both scenarios, however it is observed that the dispersion of errors depicted by the IQR is higher than a regular scheme for scenario 1. On the other hand, in scenario 2 the smallest IQR value was found with the optimized network. According to the median and IQR values of NSE and r indices, the optimized networks outperformed the alternative sampling schemes in both scenarios.

To a further detailed comparison of the performance of the alternative sampling schemes, the median and IQR values of each random and regular network were compared with the values of the optimized networks (Table 3). In general, no random and regular network outperformed the optimized networks in all statistical indices (Table 4). However, a greater number of regular networks compared to random networks slightly outperformed the optimized networks, especially in scenario 1 according to RMSE, NSE and r indices. In contrast to scenario 1, only two networks outperformed the optimized network according the RMSE in scenario 2. This indicates that overall, the cLHS is a better sampling scheme compared to random and regular sampling methods especially when accessibility restrictions are considered.

Table 3. Median and interquartile range (IQR) values of the monthly statistics for the optimized network and the 100 random and regular networks for scenario 1 and 2.

		PBIAS (%)		RMSE (mm)		NSE		r	
	Method	Median	IQR	Median	IQR	Median	IQR	Median	IQR
Scenario 1	Random	0.80	3.05	24.47	12.66	0.77	0.23	0.89	0.11
	Regular	0.70	1.30	18.74	7.98	0.86	0.10	0.93	0.04
	Optimized	0.00	0.40	16.17	10.18	0.89	0.07	0.94	0.04
Scenario 2	Random	-0.60	3.80	25.57	14.92	0.73	0.28	0.88	0.13
	Regular	-0.85	2.90	23.12	10.54	0.80	0.15	0.91	0.06
	Optimized	-0.50	0.70	16.89	7.61	0.89	0.07	0.95	0.04

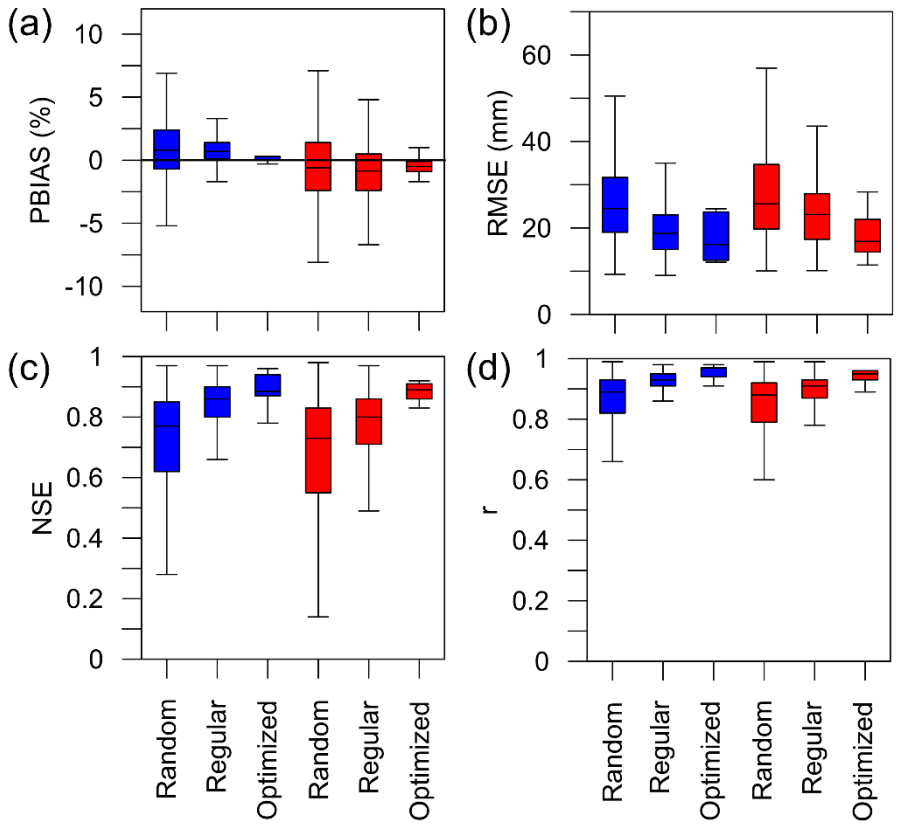


Figure 11. Boxplots of monthly statistical indices of the optimized networks and 100 random and regular sampling networks for scenario 1 (blue boxplots) and 2 (red boxplots).

Table 4. Number of random and regular networks that outperformed the optimized network in each scenario.

	Scenario 1				Scenario 2			
Method	PBIAS	RMSE	NSE	r	PBIAS	RMSE	NSE	r
Random	0	0	0	0	0	0	0	0
Regular	0	11	8	16	0	2	0	0

5. Discussion

In this study, TRMM 3B43 V7 satellite precipitation data and cLHS were coupled and presented as a spatio-temporal sampling scheme for monitoring precipitation in ungauged areas. The proposed method was applied to select the number and the location of a rain gauge network in the Ecuadorian Amazon. Furthermore, different performing evaluations supported the efficiency of the method. Despite the increase of satellite information, studies about the design of rain gauge networks based on the use of remote sensing measurements are scarce. Recently, Dai et al. (2017) provided an scheme for rain gauge network design using precipitation radar measurements. Although that network was designed to estimate the average precipitation for different events, we found similar results regarding the performance of the network. Unlike to Dai et al. (2017), we proposed a spatio-temporal sampling scheme to characterize the climatology of precipitation and, at the same time which guarantees the suitability of the sampling locations in poor accessible regions.

5.1. Representativeness of the optimized networks

Twenty-five sampling points were identified as the minimum number of monitoring points required to appropriately capture the spatio-temporal variability of precipitation in the Ecuadorian Amazon region. In order to evaluate the representativeness of the optimized network under accessibility restrictions, two networks were assessed. The spatial distribution of the optimized networks showed a quite different spatial arrangement mainly due the accessibility restrictions in the study area. Yet both networks spread the sampling points across the whole study area in scenario 1 and across the feasible areas in scenario 2. This is mainly because of the inclusion of geographic coordinates in the cLHS method. Results showed that optimized networks are representative of the whole study area and they well-capture the overall seasonality of precipitation in the region according to the TRMM climatology. This result agrees with other studies which highlight the ability of cLHS to capture the variability of multiple input covariates with a limited number of samples (Brungard and Boettinger, 2010; Domenech et al., 2017; Godinho Silva et al., 2014; Levi and Rasmussen, 2014; Mulder et al., 2013; Ramirez-Lopez et al., 2014; Yin et al., 2016, 2017). Although a less representative sampling is expected under accessibility restrictions (Godinho Silva et al., 2014; Roudier et al., 2012; Yin et al., 2016), no major difference were observed in the precipitation distribution captured by both networks. This indicates that the

accessibility restrictions in our study area did not significantly influence the spatio-temporal representativeness of the network. Nevertheless, it is important to note that these results might be influenced by the relative coarse spatial resolution of TRMM images and the fact that no further variability is gained with the bilinear resampling.

5.2. Prediction of precipitation with optimized networks and alternative sampling schemes

The evaluation with the GPM independent precipitation dataset (IMERG L3) indicated that the optimized networks were suitable to reproduce the monthly spatial surfaces of precipitation by means of OK. Both networks showed high agreement between the interpolated precipitation surfaces and the IMERG L3 images as well as a similar spatial and temporal performance. This result supports our previous statement that restrictions of accessibility no greatly influence the spatio-temporal representativeness of the network in the study area. Similar to our findings, Levi and Rasmussen (2014) reported that cLHS effectively captured the spatial variability of soils in southeastern Arizona (USA) and the method provided the base for the prediction of soil properties with OK and regression kriging. Rosemary et al. (2017) also used cLHS and OK to predict different soil properties in an Alfisol catena in Sri Lanka. Generated maps showed unbiased spatial predictions of all soil properties highlighting the effectiveness of cLHS to select sampling sites. Previous studies (e.g., Chu et al., 2014; Levi and Rasmussen, 2014; Rosemary et al., 2017; Vitharana et al., 2017) have reported that cLHS provide representative and dispersed sample locations with a range of short and long distance between sampling points, which is a mandatory requirement for variogram estimation (Bogaert and Russo, 1999). Although ideally a greater number of samples should be used to define a satisfactory variogram model, the better performance of OK over the IDW interpolation method confirmed the suitability of OK to provide reliable precipitation estimates even when a limited number of samples were used. The superiority of OK over IDW to estimate precipitation is in agreement with other studies in scarce monitoring regions (Adhikary et al., 2017; Delbari, Afrasiab, and Jahani 2013; Mair and Fares, 2011; Wang et al., 2014).

The use of cLHS and satellite precipitation data to identify representative spatio-temporal sampling points in the Ecuadorian Amazon showed to be a better sampling scheme compared to random and regular sampling methods. This result confirms the suitability of TRMM 3B43 V7 data as valuable prior information to identify representative sampling points for precipitation monitoring. The less performance variability depicted by different statistic indices in both optimized networks compared to those obtained by a random and a regular sampling showed the efficiency of the method to generate more stable spatial predictions of precipitation along the year. One interesting result of comparing all the regular networks with the optimized network is that when accessibility limitations were not considered, a good spatio-temporal representativeness could also be achieved with a regular sampling scheme. However, in an operational or realistic scenario, the network optimized with cLHS clearly outperformed a regular sampling; this result suggests that when accessibility is limited,

a more efficient sampling strategy which covers the entire feature space of precipitation is relevant to guarantee a good spatial and temporal prediction of precipitation.

5.3. Recommendations

In our study, precipitation climatology obtained by TRMM 3B43 V7 dataset was used as prior information for the selection of the most representative sites to locate a rain gauge network. However, data with a finer spatial resolution might be useful, especially for mountain regions where local factors such as topography and other meteorological variables driven precipitation which are not capture by coarse gridded satellite products. Although to-date no long-term GPM products are available (to provide a robust characterization of spatio-temporal variability of precipitation), these datasets could provide valuable insights for the location of rain gauge networks at smaller spatial scales (e.g. catchments).

In order to consider accessible sites, we took into account road/trail networks and human settlements as principal factors that influence the accessibility in the region. However, a more detailed cost map with a finer spatial resolution (meters) could be defined by taking into account other factors such as the slope, land cover and real economic costs associated to visit each sampling point.

The extensible nature of the cLHS scheme allows the inclusion of additional auxiliary variables that could be used to design multi-objective networks such as eco-hydrological networks. In this way, the multidimensional variable space could be reduced by the inclusion of pertinent information. Principal Component Analysis (PCA) could be applied to the climatology of precipitation and to the climatology of other environmental variables such as surface temperature, solar radiation, soil moisture, and other environmental variables. PCA methods have proven to be an appropriate method for selecting variables (Levi and Rasmussen, 2014) and to reduce dimensionality of data (Mulder et al., 2013).

6. Conclusions

In ungauged or poorly gauged regions, the increase of rain gauges are crucial to improve the spatial and temporal characterization of precipitation. As satellite estimates are the only available information in non-monitored sites, we proposed the use of monthly satellite-derived precipitation and cLHS method for the rain gauge network design in ungauged areas. Two networks were designed and evaluated: one considering homogeneous accessibility throughout the region and another -more realistic- scenario which accounted for difficult accessibility of remote locations.

The optimized networks composed of 25 sampling points (rain gauges) were representative for the study area and captured the overall seasonality of TRMM precipitation climatology. Results indicate that restrictions of accessibility in the region did no substantially affect the spatio-temporal representativeness of the network. Evaluation of the optimized networks showed that the networks in general adequately

estimate the spatio-temporal patterns of precipitation in the region by ordinary kriging. Comparison with other sampling schemes showed the efficiency of cLHS with monthly satellite precipitation information to produce more reliable interpolated results when a realistic scenario was considered. The proposed scheme may be applied in other ungauged regions because it only requires remotely sensed precipitation measurements commonly available at global scale. In addition, the proposed methodology could be used to design multi-objective monitoring networks such as eco-hydrological networks through the inclusion of other environmental variables generated by remote sensing instruments.

Acknowledgements

The current study has been financed by DIUC – Universidad de Cuenca through the project “Spatial sampling optimization of precipitation with multivariate geostatistics”. I want to express my sincere thanks to Daniela Ballari for her guidance and constant support in the development of this research work. In the same way I want to thank Esteban Samaniego who with his comments helped to improve the manuscript. My thanks in a special way to Professor Sytze de Bruin for his valuable guidance on the spatial sampling methods and for his comments that helped to improve the final manuscript version.

References

Adhikary, S. K., Muttil, N., & Yilmaz, A. G. (2017). Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological Processes*, *31*(12), 2143–2161. http://doi.org/10.1002/hyp.11163

Adhikary, S. K., Yilmaz, A. G., & Muttil, N. (2014). Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrological Processes*, *29*(11), 2582–2599. http://doi.org/10.1002/hyp.10389

Almazroui, M., Islam, M. N., Jones, P. D., Athar, H., & Rahman, M. A. (2012). Recent climate change in the Arabian Peninsula: Seasonal rainfall and temperature climatology of Saudi Arabia for 1979-2009. *Atmospheric Research*, *111*, 29–45. http://doi.org/10.1016/j.atmosres.2012.02.013

Ballari, D., Castro, E., & Campozano, L. (2016). Validation of satellite precipitation (TRMM 3B43) in Ecuadorian coastal plains, Andean highlands and Amazonian rainforest. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. XLI-B8, pp. 305–311). Prague, Czech Republic. http://doi.org/10.5194/isprs-archives-XLI-B8-305-2016, 2016

Ballari, D., Giraldo, R., Campozano, L., & Samaniego, E. (2018). Spatial functional data analysis for regionalizing precipitation seasonality and intensity in a sparsely monitored region: Unveiling the spatio- temporal dependencies of precipitation in Ecuador. *International Journal of Climatology*. http://doi.org/10.1002/joc.5504

Baltas, E. A., & Mimikou, M. A. (2009). GIS-based optimisation of the

hydrometeorological network in Greece. *International Journal of Digital Earth*, *2*(2), 171–185. http://doi.org/10.1080/17538940902818303

Bogaert, P., & Russo, D. (1999). Optimal spatial sampling design for the estimation of the variogram based on a least squares approach. *Water Resources Research*, *35*(4), 1275–1289. http://doi.org/10.1029/1998WR900078

Brungard, C. W., & Boettinger, J. L. (2010). *Conditioned Latin Hypercube Sampling: Optimal Sample Size for Digital Soil Mapping of Arid Rangelands in Utah, USA*. *Digital Soil Mapping*. http://doi.org/10.1007/978-90-481-8863-5_6

Ceballos, G., & Ehrlich, P. R. (2006). Global mammal distributions, biodiversity hotspots, and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(51), 19374–19379. http://doi.org/10.1073/pnas.0609334103

Celleri, R., Willems, P., Buytaert, W., & Feyen, J. (2007). Space-time rainfall variability in the Paute Basin, Ecuadorian Andes. *Hydrological Processes*, *21*(24), 3316–3327. http://doi.org/10.1002/hyp.6575

Chacon-Hurtado, J. C., Alfonso, L., & Solomatine, D. P. (2017). Rainfall and streamflow sensor network design: a review of applications, classification, and a proposed framework. *Hydrology and Earth System Sciences*, *21*(6), 3071–3091. http://doi.org/10.5194/hess-21-3071-2017

Cheng, K.-S., Lin, Y.-C., & Liou, J.-J. (2008). Rain-gauge network evaluation and augmentation using geostatistics. *Hydrological Processes*, *22*(14), 2554–2564. http://doi.org/10.1002/hyp.6851

Chu, H.-J., Chen, R.-A., Tseng, Y.-H., & Wang, C.-K. (2014). Identifying LiDAR sample uncertainty on terrain features from DEM simulation. *Geomorphology*, *204*, 325–333. http://doi.org/10.1016/j.geomorph.2013.08.016

Chu, H.-J., Lin, Y.-P., Jang, C.-S., & Chang, T.-K. (2010). Delineating the hazard zone of multiple soil pollutants by multivariate indicator kriging and conditioned Latin hypercube sampling. *Geoderma*, *158*(3–4), 242–251. http://doi.org/10.1016/j.geoderma.2010.05.003

Clifford, D., Payne, J. E., Pringle, M. J., Searle, R., & Butler, N. (2014). Pragmatic soil survey design using flexible Latin hypercube sampling. *Computers and Geosciences*, *67*, 62–68. http://doi.org/10.1016/j.cageo.2014.03.005

Collischonn, B., Collischonn, W., & Morelli Tucci, C. E. (2008). Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology*, *360*(1–4), 207–216. http://doi.org/10.1016/j.jhydrol.2008.07.032

Cruz-Roa, A. F., Olaya-Marín, E. J., & Barrios, M. I. (2017). Ground and satellite based assessment of meteorological droughts: The Coello river basin case study. *International Journal of Applied Earth Observation and Geoinformation*, *62*, 114–121. http://doi.org/10.1016/j.jag.2017.06.005

Dai, Q., Bray, M., Zhuo, L., Islam, T., & Han, D. (2017). A Scheme for Rain Gauge Network Design Based on Remotely Sensed Rainfall Measurements. *Journal of*

Hydrometeorology, *18*(2), 363–379. http://doi.org/10.1175/JHM-D-16-0136.1

de Gruijter, J., Brus, D., Bierkens, M., & Knotters, M. (2006). *Sampling for Natural Resource Monitoring*. *Springer Science & Business Media* (Springer). Berlin.

Delbari, M., Afrasiab, P., & Jahani, S. (2013). Spatial interpolation of monthly and annual rainfall in northeast of Iran. *Meteorology and Atmospheric Physics*, *122*(1–2), 103–113. http://doi.org/10.1007/s00703-013-0273-5

Domenech, M. B., Castro-Franco, M., Costa, J. L., & Amiotti, N. M. (2017). Sampling scheme optimization to map soil depth to petrocalcic horizon at field scale. *Geoderma*, *290*, 75–82. http://doi.org/10.1016/j.geoderma.2016.12.012

Du, L., Tian, Q., Yu, T., Meng, Q., Jancso, T., Udvardy, P., & Huang, Y. (2013). A comprehensive drought monitoring method integrating MODIS and TRMM data. *International Journal of Applied Earth Observation and Geoinformation*, *23*, 245–253. http://doi.org/10.1016/j.jag.2012.09.010

Erazo, B., Bourrel, L., Frappart, F., Chimborazo, O., Labat, D., Dominguez-Granada, L., … Mejia, R. (2018). Validation of Satellite Estimates (Tropical Rainfall Measuring Mission , TRMM) for Rainfall Variability over the Pacific Slope and Coast of Ecuador. *Water*, *10*(2), 213. http://doi.org/10.3390/w10020213

Espinoza Villar, J. C., Ronchail, J., Guyot, J. L., Cochonneau, G., Naziano, F., Lavado, W., … Vauchel, P. (2009). Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *International Journal of Climatology*, *29*(11), 1574–1594. http://doi.org/10.1002/joc.1791

Finer, M., Jenkins, C. N., Pimm, S. L., Keane, B., & Ross, C. (2008). Oil and Gas Projects in the Western Amazon: Threats to Wilderness, Biodiversity, and Indigenous Peoples. *PLoS ONE*, *3*(8), e2932. http://doi.org/10.1371/journal.pone.0002932

Finer, M., Moncel, R., & Jenkins, C. N. (2010). Leaving the Oil Under the Amazon: Ecuador's Yasuní-ITT Initiative. *Biotropica*, *42*(1), 63–66. http://doi.org/10.1111/j.1744-7429.2009.00587.x

Godinho Silva, S. H., Owens, P. R., Duarte de Menezes, M., Reis Santos, W. J., & Curi, N. (2014). A Technique for Low Cost Soil Mapping and Validation Using Expert Knowledge on a Watershed in Minas Gerais, Brazil. *Soil Science Society of America Journal*, *78*(4), 1310–1319. http://doi.org/doi:10.2136/sssaj2013.09.0382

Hobouchian, M. P., Salio, P., García Skabar, Y., Vila, D., & Garreaud, R. (2017). Assessment of satellite precipitation estimates over the slopes of the subtropical Andes. *Atmospheric Research*, *190*, 43–54. http://doi.org/10.1016/j.atmosres.2017.02.006

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., … Iguchi, T. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. http://doi.org/10.1175/BAMS-D-13-00164.1

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., …

Wolff, D. B. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *Journal of Hydrometeorology*, *8*(1), 38–55. http://doi.org/10.1175/JHM560.1

Killeen, T. J., Douglas, M., Consiglio, T., Jørgensen, P. M., & Mejia, J. (2007). Dry spots and wet spots in the Andean hotspot. *Journal of Biogeography*, *34*(8), 1357–1373. http://doi.org/10.1111/j.1365-2699.2006.01682.x

Laraque, A., Ronchail, J., Cochonneau, G., Pombosa, R., & Guyot, J. L. (2007). Heterogeneous Distribution of Rainfall and Discharge Regimes in the Ecuadorian Amazon Basin. *Journal of Hydrometeorology*, *8*(6), 1364–1381. http://doi.org/10.1175/2007JHM784.1

Larrea, C., & Warnars, L. (2009). Ecuador's Yasuni-ITT Initiative: Avoiding emissions by keeping petroleum underground. *Energy for Sustainable Development*, *13*(3), 219–223. http://doi.org/10.1016/j.esd.2009.08.003

Levi, M. R., & Rasmussen, C. (2014). Covariate selection with iterative principal component analysis for predicting physical soil properties. *Geoderma*, *219–220*, 46–57. http://doi.org/10.1016/j.geoderma.2013.12.013

Li, D., Christakos, G., Ding, X., & Wu, J. (2018). Adequacy of TRMM satellite rainfall data in driving the SWAT modeling of Tiaoxi catchment (Taihu lake basin, China). *Journal of Hydrology*, *556*, 1139–1152. http://doi.org/10.1016/j.jhydrol.2017.01.006

Libertino, A., Sharma, A., Lakshmi, V., & Claps, P. (2016). A global assessment of the timing of extreme rainfall from TRMM and GPM for improving hydrologic design. *Environmental Research Letters*, *11*(5), 54003. http://doi.org/doi:10.1088/1748-9326/11/5/054003

Mair, A., & Fares, A. (2011). Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island. *Journal of Hydrologic Engineering*, *16*(4), 371–383. http://doi.org/10.1061/%28ASCE%29HE.1943-5584.0000330

Manz, B., Páez-Bimos, S., Horna, N., Buytaert, W., Ochoa-Tocachi, B., Lavado-Casimiro, W., & Willems, B. (2017). Comparative Ground Validation of IMERG and TMPA at Variable Spatiotemporal Scales in the Tropical Andes. *Journal of Hydrometeorology*, *18*(9), 2469–2489. http://doi.org/10.1175/JHM-D-16-0277.1

Marengo, J. A., Liebmann, B., Grimm, A. M., Misra, V., Silva Dias, P. L., Cavalcanti, I. F. A., … Alves, L. M. (2012). Recent developments on the South American monsoon system. *International Journal of Climatology*, *32*(1), 1–21. http://doi.org/10.1002/joc.2254

Mayor, Y. G., Tereshchenko, I., Fonseca-Hernández, M., Pantoja, D. A., & Montes, J. M. (2017). Evaluation of Error in IMERG Precipitation Estimates under Different Topographic Conditions and Temporal Scales over Mexico. *Remote Sensing*, *9*(5), 503. http://doi.org/10.3390/rs9050503

Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation: Measurement, remote sensing, climatology and modeling.

Atmospheric Research, *94*(4), 512–533. http://doi.org/10.1016/j.atmosres.2009.08.017

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, *32*(9), 1378–1388. http://doi.org/10.1016/j.cageo.2005.12.009

Moffitt, C. B., Hossain, F., Adler, R. F., Yilmaz, K. K., & Pierce, H. F. (2011). Validation of a TRMM-based global Flood Detection System in Bangladesh. *International Journal of Applied Earth Observation and Geoinformation*, *13*(2), 165–177. http://doi.org/10.1016/j.jag.2010.11.003

Morán-Tejeda, E., Bazo, J., López-Moreno, J. I., Aguilar, E., Azorín-Molina, C., Sanchez-Lorenzo, A., … Vicente-Serrano, S. M. (2016). Climate trends and variability in Ecuador (1966-2011). *International Journal of Climatology*, *36*(11), 3839–3855. http://doi.org/10.1002/joc.4597

Mulder, V. L., de Bruin, S., & Schaepman, M. E. (2013). Representing major soil variability at regional scale by constrained Latin Hypercube Sampling of remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, *21*, 301–310. http://doi.org/10.1016/j.jag.2012.07.004

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. http://doi.org/10.1038/35002501

Nazaripour, H., & Daneshvar, M. R. M. (2017). Rain gauge network evaluation and optimal design using spatial correlation approach in arid and semi-arid regions of Iran. *Theoretical and Applied Climatology*, *129*(3–4), 1255–1261. http://doi.org/10.1007/s00704-016-1853-3

Ochoa, A., Pineda, L., Crespo, P., & Willems, P. (2014). Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific-Andean region of Ecuador and Peru. *Hydrology and Earth System Sciences*, *18*(8), 3179–3193. http://doi.org/doi:10.5194/hess-18-3179-2014

Padrón, R. S., Wilcox, B. P., Crespo, P., & Célleri, R. (2015). Rainfall in the Andean Páramo: New Insights from High-Resolution Monitoring in Southern Ecuador. *Journal of Hydrometeorology*, *16*(3), 985–996. http://doi.org/10.1175/JHM-D-14-0135.1

Pahlavan Rad, M. R., Toomanian, N., Khormali, F., Brungard, C. W., Komaki, C. B., & Bogaert, P. (2014). Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma*, *232–234*, 97–106. http://doi.org/10.1016/j.geoderma.2014.04.036

Pappalardo, S. E., De Marchi, M., & Ferrarese, F. (2013). Uncontacted Waorani in the Yasuní Biosphere Reserve: Geographical Validation of the Zona Intangible Tagaeri Taromenane (ZITT). *PLoS ONE*, *8*(6), e66293. http://doi.org/10.1371/journal.pone.0066293

Pardo-Igúzquiza, E. (1998). Optimal selection of number and location of rainfall gauges

for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology*, *210*(1–4), 206–220. http://doi.org/10.1016/S0022-1694(98)00188-7

Prakash, S., Mitra, A. K., Pai, D. S., & AghaKouchak, A. (2016). From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, *88*, 1–7. http://doi.org/10.1016/j.advwatres.2015.11.008

Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Demattê, J. A. M., & Scholten, T. (2014). Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma*, *226–227*(1), 140–150. http://doi.org/10.1016/j.geoderma.2014.02.002

Retalis, A., Katsanos, D., & Michaelides, S. (2016). Precipitation climatology over the Mediterranean Basin - Validation over Cyprus. *Atmospheric Research*, *169*(Part B), 449–458. http://doi.org/10.1016/j.atmosres.2015.01.012

Rosemary, F., Vitharana, U. W. A., Indraratne, S. P., Weerasooriya, R., & Mishra, U. (2017). Exploring the spatial variability of soil properties in an Alfisol soil catena. *Catena*, *150*, 53–61. http://doi.org/10.1016/j.catena.2016.10.017

Roudier, P. (2017). Package "clhs." *R Package*, 1–7. http://doi.org/10.1201/b12728-46>

Roudier, P., Hewitt, A. E., & Beaudette, D. E. (2012). A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In *Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping* (pp. 227–231). http://doi.org/10.1201/b12728-46

Shaghaghian, M. R., & Abedini, M. J. (2013). Rain gauge network design using coupled geostatistical and multivariate techniques. *Scientia Iranica*, *20*(2), 259–269. http://doi.org/10.1016/j.scient.2012.11.014

Sharifi, E., Steinacker, R., & Saghafian, B. (2016). Assessment of GPM-IMERG and Other Precipitation Products against Gauge Data under Different Topographic and Climatic Conditions in Iran: Preliminary Results. *Remote Sensing*, *8*(2), 135. http://doi.org/10.3390/rs8020135

Stumpf, F., Schmidt, K., Behrens, T., Schönbrodt-Stitt, S., Buzzo, G., Dumperth, C., … Scholten, T. (2016). Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. *Journal of Plant Nutrition and Soil Science*, *179*(4), 499–509. http://doi.org/10.1002/jpln.201500313

Su, F., Hong, Y., & Lettenmaier, D. P. (2008). Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and Its Utility in Hydrologic Prediction in the La Plata Basin. *Journal of Hydrometeorology*, *9*(4), 622–640. http://doi.org/10.1175/2007JHM944.1

Tang, G., Ma, Y., Long, D., Zhong, L., & Hong, Y. (2016). Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *Journal of Hydrology*, *533*, 152–167. http://doi.org/10.1016/j.jhydrol.2015.12.008

Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W., & Hong, Y. (2016).

Statistical and Hydrological Comparisons between TRMM and GPM Level-3 Products over a Midlatitude Basin: Is Day-1 IMERG a Good Successor for TMPA 3B42V7? *Journal of Hydrometeorology*, *17*(1), 121–137. http://doi.org/10.1175/JHM-D-15-0059.1

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., … de Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, *104–105*, 70–97. http://doi.org/10.1016/j.atmosres.2011.10.021

Ulloa, J., Ballari, D., Campozano, L., & Samaniego, E. (2017). Two-Step Downscaling of Trmm 3b43 V7 Precipitation in Contrasting Climatic Regions With Sparse Monitoring: The Case of Ecuador in Tropical South America. *Remote Sensing*, *9*(7), 758. http://doi.org/10.3390/rs9070758

van Groenigen, J. W., Siderius, W., & Stein, A. (1999). Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, *87*(3–4), 239–259. http://doi.org/10.1016/S0016-7061(98)00056-1

Vitharana, U. W. A., Mishra, U., Jastrow, J. D., Matamala, R., & Fan, Z. (2017). Observational needs for estimating Alaskan soil carbon stocks under current and future climate. *Journal of Geophysical Research: Biogeosciences*, *122*(2), 415–429. http://doi.org/10.1002/2016JG003421

Volkmann, T. H. M., Lyon, S. W., Gupta, H. V., & Troch, P. A. (2010). Multicriteria design of rain gauge networks for flash flood prediction in semiarid catchments with complex terrain. *Water Resources Research*, *46*(11). http://doi.org/10.1029/2010WR009145

Wang, K., Chen, N., Tong, D., Wang, K., Wang, W., & Gong, J. (2015). Optimizing precipitation station location: a case study of the Jinsha River Basin. *International Journal of Geographical Information Science*, *30*(6), 1207–1227. http://doi.org/10.1080/13658816.2015.1119280

Wang, S., Huang, G. H., Lin, Q. G., Li, Z., Zhang, H., & Fan, Y. R. (2014). Comparison of interpolation methods for estimating spatial distribution of precipitation in Ontario, Canada. *International Journal of Climatology*, *34*(14), 3745–3751. http://doi.org/10.1002/joc.3941

Wang, W., Lu, H., Zhao, T., Jiang, L., & Shi, J. (2017). Evaluation and Comparison of Daily Rainfall From Latest GPM and TRMM Products Over the Mekong River Basin. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(6), 2540–2549. http://doi.org/10.1109/JSTARS.2017.2672786

Ward, E., Buytaert, W., Peaver, L., & Wheater, H. (2011). Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Advances in Water Resources*, *34*(10), 1222–1231. http://doi.org/10.1016/j.advwatres.2011.05.007

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. *John Wiley & Sons*.

WMO. (2008). *Guide to Meteorological Instruments and Methods of observation*. *Guide to Meteorological Instruments and Methods of Observation* (Vol. I & II).

Xu, H., Xu, C.-Y., Chen, H., Zhang, Z., & Li, L. (2013). Assessing the influence of rain gauge density and distribution on hydrological model performance in a humid region of China. *Journal of Hydrology*, *505*, 1–12. http://doi.org/10.1016/j.jhydrol.2013.09.004

Xu, H., Xu, C.-Y., Sælthun, N. R., Xu, Y., Zhou, B., & Chen, H. (2015). Entropy theory based multi-criteria resampling of rain gauge networks for hydrological modelling - A case study of humid area in southern China. *Journal of Hydrology*, *525*, 138–151. http://doi.org/10.1016/j.jhydrol.2015.03.034

Xu, P., Wang, D., Singh, V. P., Wang, Y., Wu, J., Wang, L., … He, R. (2018). A kriging and entropy-based approach to raingauge network design. *Environmental Research*, *161*, 61–75. http://doi.org/10.1016/j.envres.2017.10.038

Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., & Hou, A. (2017). Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over southern Tibetan Plateau based on a high-density rain gauge network. *Journal of Geophysical Research: Atmospheres*, *122*(2), 910–924. http://doi.org/10.1002/2016JD025418

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., … Chen, S. (2013). Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? *Journal of Hydrology*, *499*, 91–99. http://doi.org/10.1016/j.jhydrol.2013.06.042

Yin, G., Li, A., & Verger, A. (2017). Spatiotemporally Representative and Cost-Efficient Sampling Design for Validation Activities in Wanglang Experimental Site. *Remote Sensing*, *9*(12), 1217. http://doi.org/10.3390/rs9121217

Yin, G., Li, A., Zeng, Y., Xu, B., Zhao, W., Nan, X., … Bian, J. (2016). A Cost-Constrained Sampling Strategy in Support of LAI Product Validation in Mountainous Areas. *Remote Sensing*, *8*(9), 704. http://doi.org/10.3390/rs8090704

Yoo, C., Jung, K., & Lee, J. (2008). Evaluation of Rain Gauge Network Using Entropy Theory: Comparison of Mixed and Continuous Distribution Function Applications. *Journal of Hydrologic Engineering*, *13*(4), 226–235. http://doi.org/10.1061/(ASCE)1084-0699(2008)13:4(226)

Zubieta, R., Getirana, A., Espinoza, J. C., Lavado-Casimiro, W., & Aragon, L. (2017). Hydrological modeling of the Peruvian-Ecuadorian Amazon Basin using GPM-IMERG satellite-based precipitation dataset. *Hydrology and Earth System Sciences*, *21*(7), 3543–3555. http://doi.org/10.5194/hess-21-3543-2017

Zubieta, R., Getirana, A., Espinoza, J. C., & Lavado, W. (2015). Impacts of satellite-based precipitation datasets on rainfall-runoff modeling of the Western Amazon basin of Peru and Ecuador. *Journal of Hydrology*, *528*, 599–612. http://doi.org/10.1016/j.jhydrol.2015.06.064

Zulkafli, Z., Buytaert, W., Onof, C., Manz, B., Tarnavsky, E., Lavado, W., & Guyot, J.-L.

(2014). A Comparative Performance Analysis of TRMM 3B42 (TMPA) Versions 6 and 7 for Hydrological Applications over Andean–Amazon River Basins. *Journal of Hydrometeorology*, *15*(2), 581–592. http://doi.org/10.1175/JHM-D-13-094.1