An entropy-based approach for the optimization of rain gauge network using satellite and ground-based data

Claudia Bertini, Elena Ridolfi, Luiz Henrique Resende de Padua,
Fabio Russo, Francesco Napolitano and Leonardo Alfonso **IWA**

ABSTRACT

Accurate and precise rainfall records are crucial for hydrological applications and water resources management. The accuracy and continuity of ground-based time series rely on the density and distribution of rain gauges over territories. In the context of a decline of rain gauge distribution, how to optimize and design optimal networks is still an unsolved issue. In this work, we present a method to optimize a ground-based rainfall network using satellite-based observations, maximizing the information content of the network. We combine Climate Prediction Center MORPhing technique (CMORPH) observations at ungauged locations with an existing rain gauge network in the Rio das Velhas catchment, in Brazil. We use a greedy ranking algorithm to rank the potential locations to place new sensors, based on their contribution to the joint entropy of the network. Results show that the most informative locations in the catchment correspond to those areas with the highest rainfall variability and that satellite observations can be successfully employed to optimize rainfall monitoring networks.

Key words | CMORPH, information theory, rainfall, rain gauge network, satellite data, sensors' optimization

Claudia Bertini (corresponding author)
Fabio Russo
Francesco Napolitano
Department of Civil, Constructional and
Environmental Engineering,
La Sapienza University of Rome,
Via Eudossiana 18,
00184 Rome,
Italy
E-mail: *claudia.bertini@uniroma1.it*

Elena Ridolfi
Centre of Natural Hazards and Disaster Science,
CNDS,
75236 Uppsala,
Sweden

Luiz Henrique Resende de Padua
Universidade Federal de Minas Gerais,
Belo Horizonte,
MG,
Brazil

Leonardo Alfonso **IWA**
IHE Delft Institute for Water Education,
Delft,
The Netherlands

HIGHLIGHTS

- This study proposes a method to evaluate and optimize an existing rain gauge network using satellite observations.
- The most informative locations in the catchment are identified using a greedy ranking algorithm.
- The resulting optimal network provides higher joint entropy with fewer sensors.
- The most informative locations reflect those with highest variance.

INTRODUCTION

Quantification of precipitation is essential for improving knowledge about hydrological and water resources applications, including water allocation, water resources monitoring and risk assessment. Yet, the state of our

knowledge is subject to the density and distribution of rainfall monitoring networks over territories. For this reason, it is desirable to have dense rain gauge networks (Li *et al.* 2019). However, despite their crucial role, rainfall networks have been declining in the last decades due to their high maintenance and operating costs (Mishra & Coulibaly 2009; Dai *et al.* 2017) and data are scarce or lacking in some areas of the world (Walker *et al.* 2016). Although

many remote sensing products are now available, ground-based observations are still needed for their calibration, validation and bias removal (Li *et al.* 2019). In this context, many researchers have tried to answer the question on how to design optimal monitoring networks, which can guarantee accurate information and reduce uncertainty in precipitation (Chacon-Hurtado *et al.* 2017). The World Meteorological Organization (WMO) proposed minimum density requirements for hydrometric networks, based on the topography of the area and on the characteristics of the sensors adopted (WMO 2008).

Following the recent classification proposed by Chacon-Hurtado *et al.* (2017), methods for hydrometric network design can be distinguished as either statistics-based (e.g., Maddock 1974; Li *et al.* 2011), information theory-based, based on expert recommendations (WMO 2008) and based on the performance of hydrological models (Zeng *et al.* 2018).

Information theory (IT) (Shannon 1948) was first applied to water resources research by Amorocho & Espildora (1973) and then introduced by Caselton & Husain (1980) to design a rainfall network. From that moment on, many researchers have applied IT to solve the monitoring network design problem (Keum *et al.* 2017), for precipitation (e.g., Chen *et al.* 2008; Yoo *et al.* 2008), streamflow (e.g., Alfonso *et al.* 2013; Keum *et al.* 2019) and groundwater networks (e.g., Leach *et al.* 2016). The main principle behind all these studies is to maximize the information provided by the network, expressed through the concept of joint entropy (Shannon 1948). Many authors proposed multi-objective approaches, such as the minimization of network redundancy, expressed by total correlation (Alfonso *et al.* 2010b, 2013), mutual information (Chen *et al.* 2008; Li *et al.* 2012; Fahle *et al.* 2015) and directional information transfer (Yang & Burn 1994). Some studies adopted additional objective functions not related to IT, such as hydrological model efficiency (Xu *et al.* 2015), rainfall field interpolation accuracy (Xu *et al.* 2018) and spatiotemporality information (Huang *et al.* 2020).

A common issue for monitoring network design, in general, is that precipitation observations, and all the information we can derive from them, are available only at the locations where sensors are deployed. Thus, the question is how to decide which ungauged locations are the most convenient to place new sensors. Most authors either interpolate precipitation observations (e.g., Xu *et al.* 2018), for the

case of rainfall networks, or employ hydrological models to produce water level time series at ungauged locations (e.g., Werstuck & Coulibaly 2017). However, when interpolating precipitation, some biases are introduced and the accuracy of the resulting rainfall field depends on the specific interpolation technique adopted and on the characteristics of the area considered (Hofstra *et al.* 2008). This problem can be addressed using remote sensing data, which have been widely used in the last decade for many hydrological applications (Li *et al.* 2016; Mazzoleni *et al.* 2019; Bertini *et al.* 2020). Remote sensing products proved to better reflect spatial relationships among objects when compared to interpolated and simulated data (Li *et al.* 2016; Huang *et al.* 2020). Many satellite precipitation products have been developed in the last two decades, with different spatial coverage, going from 4 to 25 km, and temporal scale, from 30 min to monthly resolution. Among all the satellite precipitation products, the most popular in hydrological applications are probably the CMORPH (Climate Prediction Center MORPHing technique) (Joyce *et al.* 2004), PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) (Hsu *et al.* 1997), TMPA (Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis) (Huffman *et al.* 2007) and GPCP (Global Precipitation Climatology Project) (Adler *et al.* 2003).

A few authors have adopted satellite observations in network design; among them can be listed Contreras *et al.* (2019) and Huang *et al.* (2020). The former applied the conditioned Latin hypercube sampling method on a TMPA product to capture spatiotemporal precipitation in ungauged locations while the latter applied IT within a multi-objective optimization approach. However, methods that consider satellite information are very limited and still rely on prior data interpolation, without analysing the potential information content of satellite observations. As the main advantage of using a gridded dataset is the information contained at ungauged locations can be investigated without introducing interpolation biases, in this work we propose a method to use satellite precipitation estimates to evaluate and optimize an existing rain gauge network. The number and locations of new sensors are chosen in order to maximize the total information content of the network, given by its joint entropy, following the generally accepted fact that information content of time series can be taken as a

value of variability that can indicate where it is more appropriate to measure rainfall (e.g., Krstanovic & Singh 1992; Mishra & Coulibaly 2009). Entropy at ungauged locations is evaluated using version 1.0 of CMORPH rainfall product, which matches the requirements of a fine spatial scale and good performance. In contrast to the study of Huang *et al.* (2020), we do not take into account redundancy reduction, with the aim of obtaining a robust network and ensuring the capture of essential information even in case of a sensor failure. Indeed, we do not mean to change or remove the existing stations, as most hydrological applications need long time series; instead, we aim to increase the network density in order to have a better understanding of rainfall characteristics and an improvement of water resources assessment in the area.

This paper is organized as follows. First, we provide a background on IT; second, the case study and the dataset are introduced; then, details about the methodology adopted are provided. Finally, results and discussions are presented and conclusions of the study are drawn.

BACKGROUND

Information theory

The amount of information content and of redundancy given by a monitoring network can be measured using IT (Shannon 1948). Definitions of the IT-related quantities employed are presented below.

Given a set of n events, with known probabilities of occurrence $p_1, p_2, \ldots, p_n$, entropy is defined as the measure of uncertainty of the possible n outcomes. If more information about one of the events is obtained, then the uncertainty of the outcomes decreases. Information can be thus regarded as a decrease in uncertainty and entropy can be seen as a measure of information content. The concept of entropy can be extended to a random discrete variable X (Shannon & Weaver 1949), with discrete values $x_1, x_2, \ldots, x_n$ and corresponding probabilities $p(x_1), p(x_2), \ldots, p(x_n)$:

$$H(X) = - \sum_{i=1}^{n} p(x_i) \cdot \log p(x_i) \qquad (1)$$

where $H(X)$ is the entropy for the variable X, also called marginal entropy.

In a similar way, it is possible to evaluate the content of information from N multiple variables $X_1, X_2, \ldots, X_N$, introducing the concept of joint entropy:

$$JH(X_1, X_2, \ldots, X_N) = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \ldots \sum_{i_N=1}^{n_N} p(x_{i_1}, x_{i_2}, \ldots, x_{i_N}) \cdot \log p(x_{i_1}, x_{i_2}, \ldots, x_{i_N}) \qquad (2)$$

where $JH(X_1, X_2, \ldots, X_N)$ is the joint entropy of N random discrete variables and $p(x_{i_1}, x_{i_2}, \ldots, x_{i_N})$ is the joint probability of the $X_1, X_2, \ldots, X_N$ variables.

The logarithm in Equations (1) and (2) is base 2, therefore marginal entropy and joint entropy are measured in bits.

In monitoring network design and optimization problems, each precipitation time series recorded by a sensor can be regarded as a random discrete variable X, with marginal entropy $H(X)$. The information content provided by the whole network, made of N sensors, is given by the joint entropy $JH(X_1, X_2, \ldots, X_N)$.

Estimating joint entropy as defined in Equation (2) can be a complicated task, due to the difficulty in the computation of joint probability, especially for a large number of variables. To overcome this issue, the grouping property of mutual information (Kraskov *et al.* 2005) can be used. According to this property, joint entropy of a couple of variables X and Y is equal to the marginal entropy of a new variable Z, obtained agglomerating the original pair. The probability of occurrence of the new variable Z is then estimated by a histogram-based frequency analysis together with quantization, as applied by many researchers, e.g., Alfonso *et al.* (2013), Pádua *et al.* (2019), Ridolfi *et al.* (2012, 2014a, 2014b).

Quantization can be defined as the division of a quantity into a discrete number of smaller parts, often integral multiples of a common quantity (Gray & Neuhoff 1998). Its oldest version, which is rounding off, was already employed in 1898 for the estimation of densities by histogram (Sheppard 1897).

In this work, a normalized rounding off is adopted to convert a continuous signal, which is precipitation, into

discrete values, which are bins, filtering out the noise from the observed time series. The quantization here adopted rounds a value x to its nearest lowest integer x_q, which is a multiple of a predefined quantity a, following the rule:

$$x_q = k \cdot a \cdot \left\lfloor \frac{2x + a}{2a} \right\rfloor \tag{3}$$

where k is a constant value used to normalize the time series.

STUDY AREA AND DATA

Study area

The study is applied to the Rio das Velhas catchment, located in the central area of Minas Gerais state, Brazil. Rio das Velhas has a length of approximately 801 km and a drainage area of 29,173 km^2, 10% of which is occupied by the metropolitan area of the city of Belo Horizonte. The river is the major tributary of Rio Sao Francisco and its catchment belongs to the Alto Sao Francisco basin. As the main productive activities of the area, i.e., agriculture, cattle and mining, require high amounts of water every year, constant monitoring and assessment of water resources of the catchment are needed. To this aim, the existing network made of 28 rain gauges needs to be improved.

The precipitation regime is typical of the region with tropical climate, with wet periods during summer (October–March) and dry periods during winter (April–September) (Pinto 2005). The catchment is generally exposed to frequent drought cycles, especially in the urban area (Santos *et al.* 2019). The climate of the entire catchment is influenced by atmospheric large-scale processes which control the precipitation regime. The main systems governing precipitation in the area are the South Atlantic Subtropical Anticyclone (ASAS), the Lines of Instability (LI) and the South Atlantic Convergence Zone (ZCAS). The former is responsible for the high decrease in rainfall during the period June–August and for atmospheric instability during summer, while the latter generate long-term precipitation with large volumes in the months November–January. An important role in rainfall generation is also played by the presence of two mountain chains, Serra do Espinhaço and Serra da Mantiqueira, in the eastern and southern parts of the catchment, respectively (Figure 1). The two chains represent a natural barrier for the air masses moving from the ocean to inland and the other way around, generating high instability and frequent precipitation in the area, which is characterized by the highest yearly rainfall volumes of the catchment.

Precipitation datasets

Two rainfall datasets are considered, one from ground-based measurements and the other from satellite-based observations. The ground-based dataset consists of daily precipitation recorded by 28 rain gauges located throughout the catchment, with an average density of approximately 1,040 km^2 per sensor (Figure 1). The observations are provided by the Agência Nacional de Águas (ANA) and are available for the period January 1994–December 2014.

The satellite-based dataset employed is the CMORPH product, which provides precipitation estimates derived by the NOAA Climate Prediction Center (CPC) using the MORPhing technique. The CPC MORPhing technique uses geostationary satellite infrared (GEO IR) consecutive images, provided every 30 min, to derive cloud motion vectors via cross-correlation (Joyce *et al.* 2004). The cloud motion vectors are then employed to propagate in time the passive microwave (PMW) precipitation rate estimates both in the forward and backward directions, using temporal weights. This algorithm is used to produce the so-called CMORPH Raw (or version 0.X), which is provided in a spatial resolution of 8 km × 8 km, with quasi-global coverage (60° N–60° S), and in a 30 min temporal resolution. To improve the precipitation estimates, bias correction was performed on the raw CMORPH. More details about the bias-corrected CMORPH product, also known as version 1.0, can be found in Xie *et al.* (2017). CMORPH products are available from 1998 to the present. In this work, we adopted version 1.0 which ensures good performance at a fine spatial scale (Xie *et al.* 2007; Sapiano & Arkin 2009).

Data pre-processing

The optimization of the monitoring rainfall network is conducted using ground-based and satellite-based precipitation time series, both referring to the period 1998–2014, for a
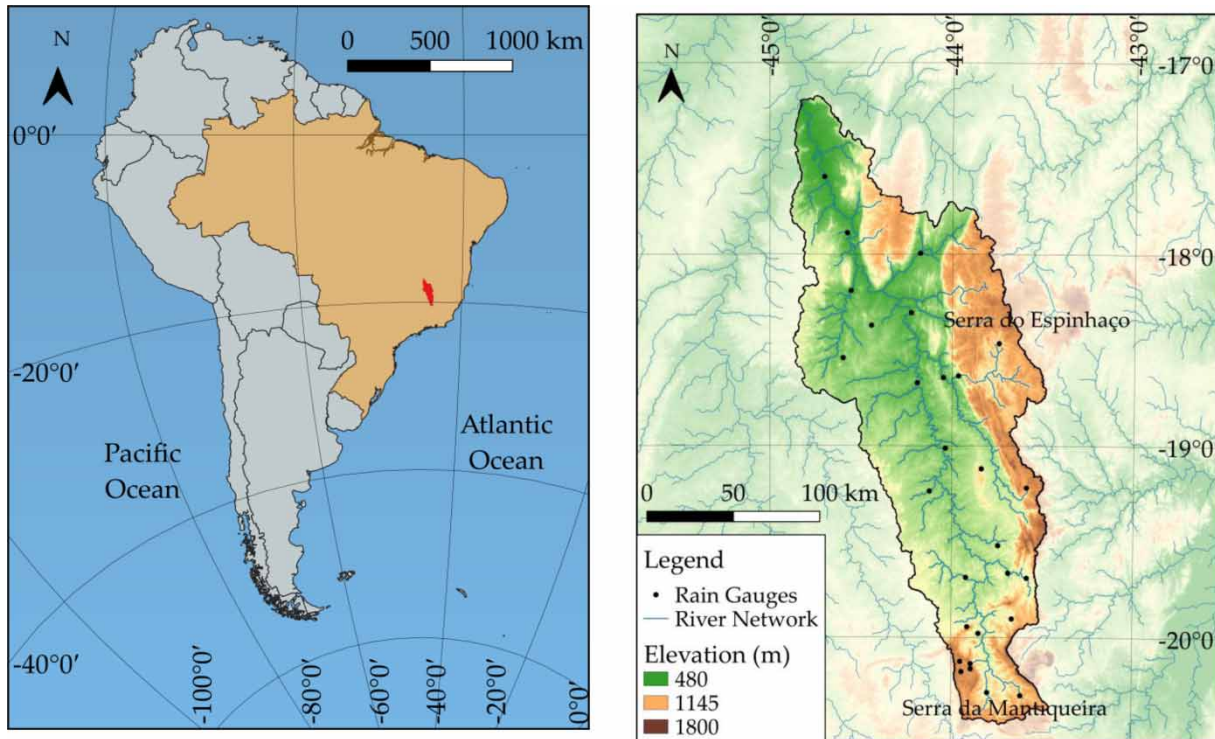
Figure 1 │ Left, localization of Rio das Velhas (red) within Brazil (orange); right, elevation map of the Rio das Velhas catchment. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2021.113.

total of 17 years of observations, meeting the requirement of a minimum 10 years of records when entropy is applied to network design (Keum & Coulibaly 2017).

The existing rain gauges provide daily precipitation depth estimates, while CMORPH gives 30 min rain intensity; therefore, we first pre-process satellite observations to make them comparable to the ground-based ones. First, we transform CMORPH intensity estimates into precipitation depth and we aggregate them to the daily scale, to have the same temporal resolution of gauge-based measurements. Finally, to match with the standard rain gauge minimum resolution, which is 0.1 mm for the case study, each record lower than 0.1 mm is set to 0 mm.

In both datasets missing records are removed from the daily time series with the following rule: if at time t missing information is detected in one of the rain gauges and/or in one of the satellite time series, then all the observations referring to time t, in both datasets, are removed.

The presence of zeros (i.e., dry days) in the precipitation records can affect the information content provided and, when there is a large number of zeros in the dataset, one

must isolate zero and non-zero values and deal with them separately (Gong *et al.* 2014). One option could be to use a delta function to deal with zeros, as proposed by Gong *et al.* (2014). As an alternative, dry days could be removed, considering only wet periods for the analysis, as done by Huang *et al.* (2020). We adopted the second strategy. With the purpose of isolating only the wet periods, we removed all dry days in both datasets with the following rule: if at time t no rain is detected from both CMORPH and rain gauges, then all precipitation records referring to time t are removed. At the end of the mentioned pre-processing procedures, both ground-based and remote sensing-based observations are reduced from 6,209 to 5,120 daily observations.

METHODOLOGY

The methodology we propose can be summarized in different steps: first, we perform experiments to rank the existing ground-based stations (experiment G) and the satellite cells (experiment S) based on their joint entropy, to identify the

most informative locations. Then, a sensitivity analysis on quantization parameters a and k is carried out to reduce uncertainty in estimation of IT-related quantities. Finally, the optimal monitoring network is found by solving an optimization problem, which defines the locations where new ground-based sensors should be located, using CMORPH observations (experiment GS). All three experiments are solved using a greedy algorithm (Li *et al.* 2012; Alfonso *et al.* 2013; Banik *et al.* 2017; Xu *et al.* 2018), aimed to maximize the joint entropy provided by the network. A summary of the undertaken experiments is provided in Table 1. Further details are given in the following subsections.

Ranking experiments

The existing stations and CMORPH cells are ranked to identify the most informative locations within the catchment. For experiment G, the set of candidate sensor locations is the set $g = \{g_1, g_2, \ldots, g_N\}$ of N existing rain gauges. The ranking procedure can be formulated as follows: from the set of candidate stations g, search for the station g^* such that $H(g^*)$ is the maximum (Krstanovic & Singh 1992; Ridolfi *et al.* 2011); when found, label it as g_1, store it in the set RG (ranked set g) and remove g^* from the original set g. Then, search for another station g^* among the candidates in the updated set g, such that $JH(g_1, g^*)$ is maximum; when found, label it as g_2, append it to RG and remove it from g. Repeat the procedure until the size of RG is N. The set RG, updated at each step of the algorithm, corresponds to a quasi-optimal set of stations. The mathematical procedure of the ranking problem is the following:

$$\begin{cases} \max\ [JH(g_1, g_2)]\ step\ 2 \\ \max\{JH\{\ \max\ [JH\ (g_1, g_2)],\ g_3\}\}\ step\ 3 \\ \max\{JH\{\max\ldots JH\{\ \max\ [JH\ (g_1, g_2)],\ g_3\}\ldots\},\ g_N\}\ step\ N \end{cases}$$
(4)

For experiment S, the set of candidate sensor locations is the set $s = \{s_1, s_2, \ldots, s_M\}$ of $M = 439$ available cells in which the catchment is spatially discretized. Similarly to experiment G, the ranking procedure is as follows. Search in set s for the cell s^* that maximize $H(s^*)$; when found, label it as s_1, store it in the set RS (ranked set s) and remove s^* from the original set s. Then, among the remaining candidates in the updated set s, search for the cell s^* such that $JH(s_1, s^*)$ is maximum; when found, label it as s_2, append it to RS and remove it from s. Repeat the procedure until the size of RS is M. The procedure can be mathematically expressed as:

$$\begin{cases} \max\ [JH(s_1, s_2)]\ step\ 2 \\ \max\{JH\{\ \max\ [JH\ (s_1, s_2)],\ s_3\}\}\ step\ 3 \\ \max\{JH\{\max\ldots JH\{\ \max\ [JH\ (s_1, s_2)],\ s_3\}\ldots\},\ s_M\}\ step\ M \end{cases}$$
(5)

Sensitivity analysis of quantization parameters

The estimation of both marginal and joint entropy requires the calculation of probabilities (see Equations (1) and (2)), which is done through frequency analysis of time series, previously filtered out with Equation (3), i.e., the quantization procedure (Alfonso *et al.* 2010a; Li *et al.* 2012; Huang *et al.* 2019). The filtering process requires the a priori assumption of the quantization parameters k and a. It means that obtained entropy-related quantities are influenced by the values of parameters k and a, which thus, in turn, have implications for the final layout of the optimized network. However, Alfonso *et al.* (2014) demonstrated that, when varying the quantization parameters, for each quasi-optimal set of stations, there exist most probable values of joint entropy, i.e., JH values that can be obtained using different combinations of a and k. These JH values will be referred to from now on as the most frequently selected or most

Table 1 | Experiment summary

Experiment name	Decision variables (description)	Decision variables (symbol)	Criterion/Objective function
G	Existing rain gauges	$g = \{g_1, g_2, \ldots, g_N\}$	$\max\{JH\{\max\ldots JH\{\max\ [JH\ (g_1, g_2)],\ g_3\}\ldots\},\ g_N\}$
S	CMORPH cells	$s = \{s_1, s_2, \ldots, s_M\}$	$\max\{JH\{\max\ldots JH\{\max\ [JH\ (s_1, s_2)],\ s_3\}\ldots\},\ s_M\}$
GS	Existing rain gauges and CMORPH cells	$gs = \{g_1, g_2, \ldots, g_m, s_1, s_2, \ldots, s_{M-m}\}$	$\max\{JH\{\max\ldots JH\{\max\ [JH\ (g_1, g_2, \ldots, g_m)],\ s_1\}\ldots\},\ s_n\}$

probable joint entropy values. Following the procedure proposed by the same authors, we performed a sensitivity analysis on quantization parameters to identify the couple of values providing the most probable values of joint entropy for each quasi-optimal set of stations. The analysis is performed both on the ground-based and satellite-based observations. It can be summarized in the following steps:

(a) Assume initial values for parameters k and a in Equation (3).
(b) Rank stations based on their joint entropy and determine quasi-optimal sets of sensors (RG or RS).
(c) Estimate joint entropy for each set identified in point b), varying parameters k and a in the ranges $k = 1, 2, \ldots, 50$ and $a = 1, 2, \ldots, 50$.
(d) Select the final values for parameters k and a which provide the most probable joint entropy values for each set of quasi-optimal stations.

Monitoring network optimization

The optimal layout of sensors' locations is defined with an optimization procedure (experiment GS) that takes into account the information provided by both ground-based and satellite-based datasets. The optimal network is defined in two steps: first, select a set of quasi-optimal rain gauges from the existing network, i.e., the first m sensors from set RG obtained in experiment G; second, identify n new sensor locations to complement these m stations, choosing them among all the CMORPH cells available over the catchment, which are directly treated as a candidate for a new station location (Chen *et al.* 2008; Yeh *et al.* 2017). The final layout is then defined by the optimal combination of a subset of ranked existing stations and a set of new locations, which together provide the highest information content.

The optimization problem is solved by the greedy algorithm explained in the previous section. Its mathematical formulation is given by:

$$\max \{JH\{\max \ldots JH\{\max [JH(g_1, g_2, \ldots, g_m)], s_1\} \ldots\}, s_n\}$$
(6)

where the set $\{g_1, g_2, \ldots, g_m\}$ is a subset of the first m stations in set RG. Moreover, the set $\{s_1, s_2, \ldots, s_n\}$ is the same set s from experiment S, but excluding the cells where the ground stations in set $\{g_1, g_2, \ldots, g_m\}$ are located. Therefore, the number of cells used as candidates in this experiment is $M - m$.

RESULTS

First, results of the ranking experiments of the datasets are shown, followed by our findings on the sensitivity analysis and by the optimal network obtained.

Ranking experiments

A map of the ranked existing rain gauges is shown in Figure 2, while the values of joint entropy obtained with an increasing number of stations is presented in Figure 3(b). The stations in the graph of Figure 3(b) increase according to the ranking obtained, so for instance, *JH* obtained with two stations is the one provided by the first two ranked stations, *JH* obtained with three sensors is the one given by the first three ranked stations and so on. Observing Figure 3(b), we can notice that, as the number of sensors increases, the joint entropy tends to converge to a stable value, in agreement with several previous works (Chen *et al.* 2008; Wei *et al.* 2014; Yeh *et al.* 2017). In other words, above a certain number of stations, adding new sensors does not provide a significant improvement in the information given by the entire network. The total amount of information given by the existing network is $JH_G = 8.7$ bits and already the first ten ranked sensors give almost 90% of JH_G.

The results of the ranking of CMORPH cells are presented in Figure 3(a) and 3(b). In both cases, only the first 28 cells ranked are presented, to facilitate the comparison with the existing rain gauge network, which is made up of 28 stations. At the catchment boundary, only the cells with more than 50% of their area falling inside the catchment are considered in the ranking process. Some boundary cells were therefore excluded, since they provide information mainly referring to an area that is outside our case study. To build the map of Figure 3(a), the identifiers of the satellite cells selected are placed in the centre of the selected cell. This is the reason why,
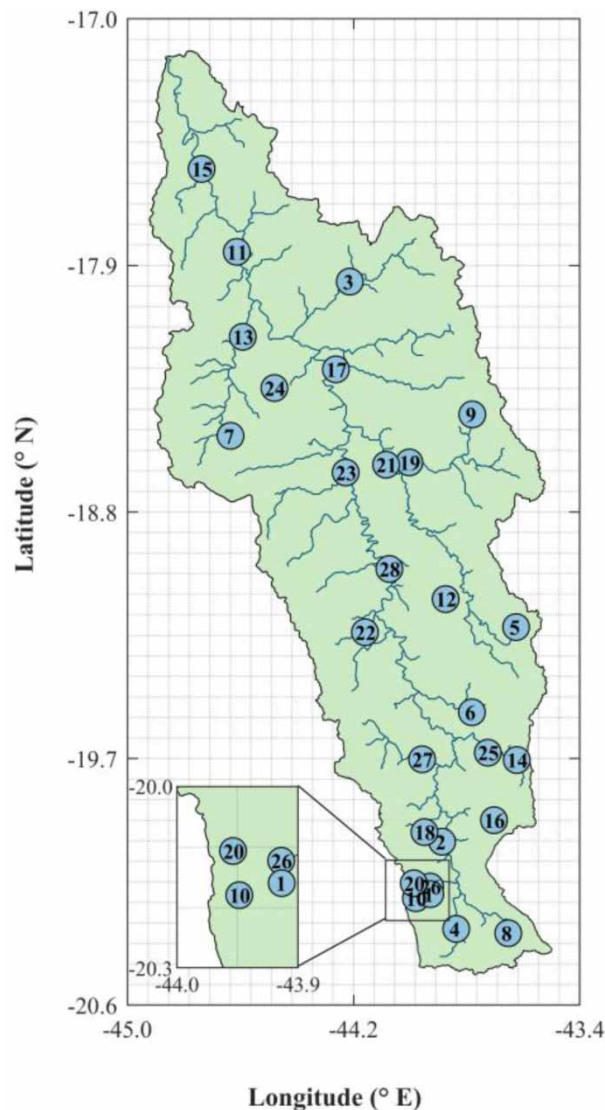
Figure 2 | Map of the existing rain gauges with the corresponding ranking (case G) –best rank is 1.

when a boundary cell is selected, the corresponding identifier could appear to be outside the catchment boundaries.

It is interesting to note that none of the existing rain gauge locations is selected. However, we can notice that both the total amount of information provided by a satellite-based network, which is $JH_S = 8.6$ bits, and the trend of joint entropy with an increasing number of stations are very similar to the corresponding obtained for rain gauge observations (Figure 3(b)).

Sensitivity analysis of quantization parameters

Bins' width and probabilities of occurrence evaluated with quantization are influenced by the parameters k and a of Equation (3) and, therefore, a sensitivity analysis on those parameters is performed. The quantization parameters k and a are first both assumed equal to 1, as suggested in Huang *et al.* (2019) and Keum & Coulibaly (2017). These values are adopted to transform both rain gauges and CMORPH time series, which are, in turn, employed to solve Equations (4) and (5). The stations ranked at each step of the greedy algorithm correspond to a set of quasi-optimal stations. For each set and each data source we then compute joint entropy with parameters k and a both varying in the ranges $k = 1, 2, \ldots, 50$ and $a = 1, 2, \ldots, 50$, for a total of 2,500 possible combinations.

The two-dimensional frequency distribution of joint entropy with an increasing number of stations is presented in Figure 4. For an increasing number of sensors (x-axis), joint entropy values (y-axis) are obtained for each combination of parameters k and a. Our findings confirmed those of Alfonso *et al.* (2014), i.e., that there exist JH values that can be obtained using different combinations of a and k, i.e., the most probable joint entropy values. Figure 4(a) shows that for rain gauge observations there is a well-defined interval of most probable JH values, while for CMORPH time series the same interval is still present but more dispersed (Figure 4(b)).

The main idea of this sensitivity analysis is to find the combination of parameter k and a which leads to the most probable values of JH, to reduce the uncertainty related to the evaluation of joint entropy. Since parameter values we first assumed ($a = 1$, $k = 1$) is one of the combinations providing the most frequent JH values, for both CMORPH and rain gauges, we decided to adopt $k = 1$ and $a = 1$ as the final values to perform quantization. These values are in agreement with those suggested by Huang *et al.* (2019) and Keum *et al.* (2017).

Monitoring network optimization

The optimal monitoring network is defined with an optimization problem that combines information provided by
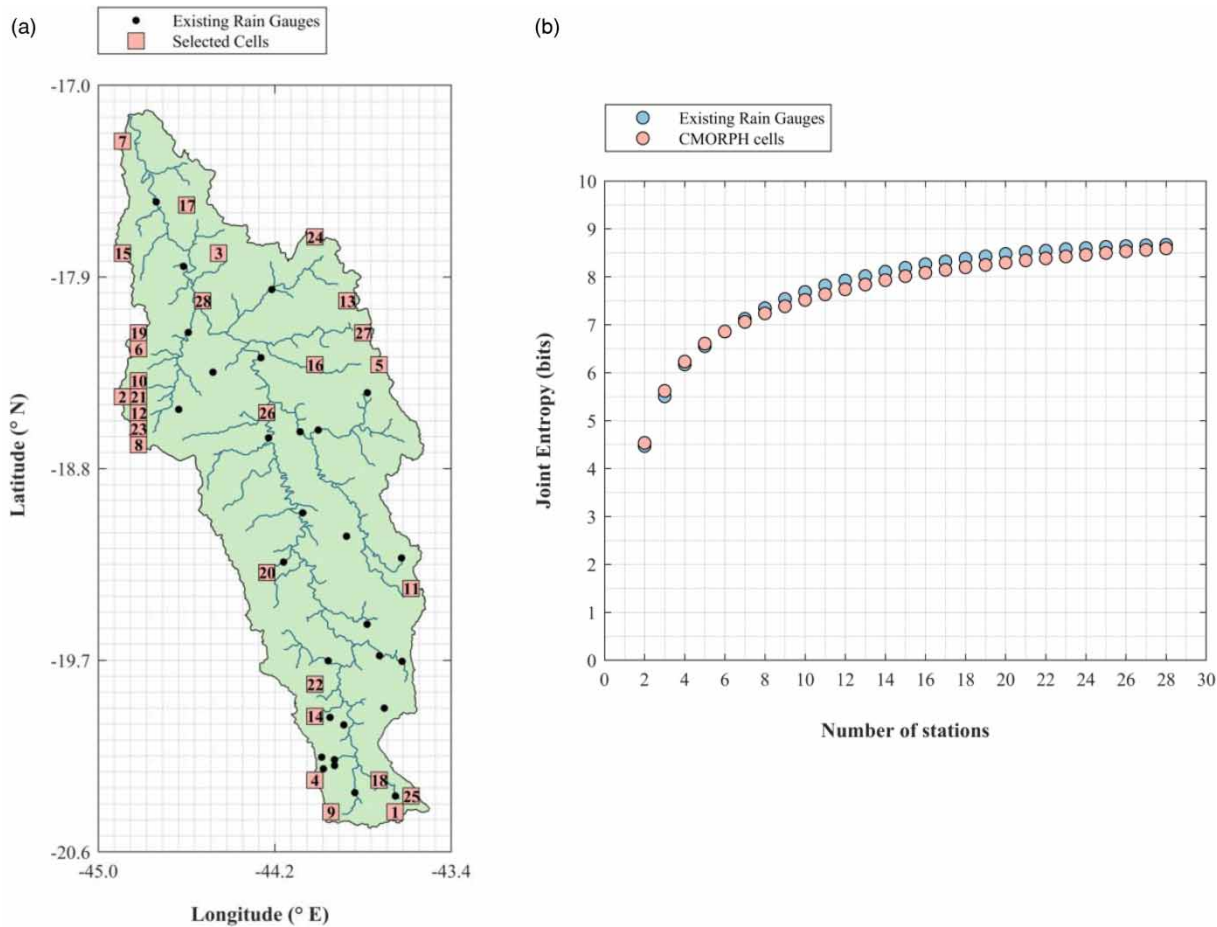
Figure 3 | (a) Map of the first 28 ranked CMORPH cells (case S); (b) joint entropy (JH) variation with increasing number of locations selected, both for existing rain gauges (light blue) and CMORPH cells (pink). Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2021.113.

ground-based and satellite-based observations. We take a subset made of m optimal rain gauges and complement it with locations chosen among CMORPH cells, with the aim of maximizing the information provided by the final network, as expressed by Equation (6). The network is completed when adding one more station would provide an increase in total joint entropy, in principle, lower than 1%.

The initial set of quasi-optimal rain gauges should be defined so that it provides a high amount of information, which is also lower than the maximum value of joint entropy given by the existing network. In other words, if we refer to the graph of Figure 3(b), the point representing the initial subset should be located in the ascending part, before it reaches the stable value of JH. In this way, we ensure that the high information content of the existing

network is preserved but that, at the same time, the network can be improved. To this end, we take $m = 8$ rain gauges, e.g., the first eight ranked.

The results we obtained are presented in Figure 5. To complete the optimal network only eight sensors are needed, four of which are placed in the west, two in the south and the remaining two in the north-east. Looking at Figure 5(b), it can be observed that combining a set of quasi-optimal rain gauges and a set of quasi-optimal satellite cells we obtain a total amount of information $JH_{GS} = 9$ bits, higher than the total information provided by both the rain gauges and CMORPH networks. This value is achieved with fewer sensors. This is probably due to the fact that the observations from the two datasets are less correlated than those coming from the same data source.
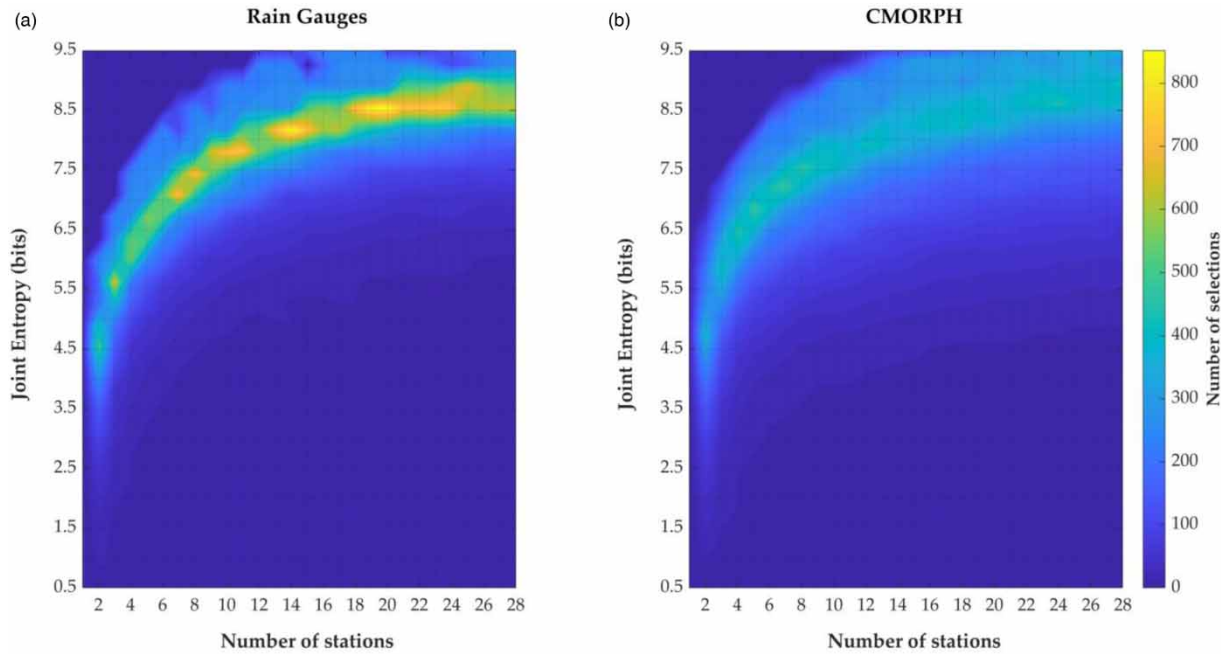
Figure 4 | Two-dimensional frequency distribution of joint entropy with increasing numbers of stations obtained for (a) rain gauges and (b) CMORPH observations.
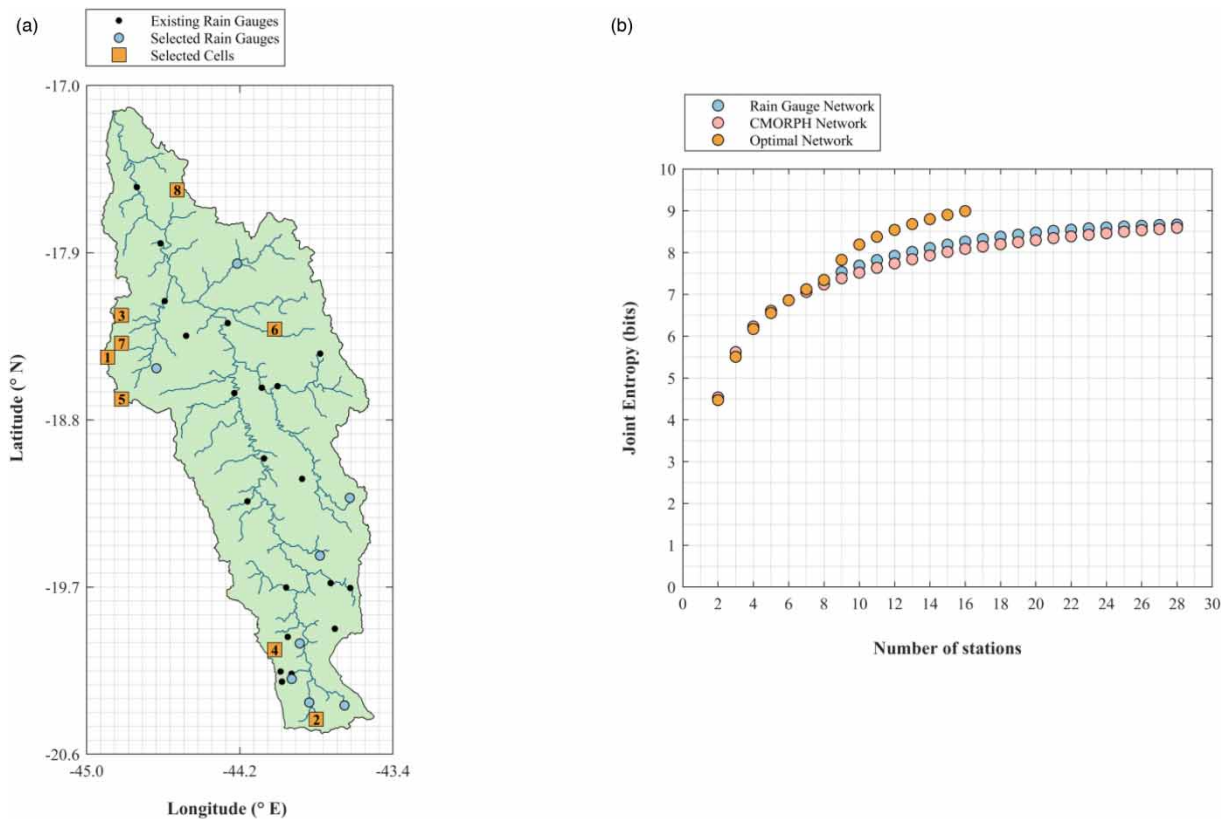
Figure 5 | (a) Map of the locations of the optimal network (case GS). The light blue circles stand for the initial rain gauges' subset, orange squares represent the selected CMORPH cells and black circles represent the existing rain gauges left out from optimization; (b) variation of joint entropy values with increasing number of stations obtained for rain gauges (light blue), CMORPH (pink) and optimal network (orange). Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2021.113.

DISCUSSION

To verify whether the ground-based and satellite-based data-sets are capturing rainfall variability in the same way, their information content was analysed using the concept of variance.

For satellite-based observations, we computed the variance of time series recorded in each cell of the grid within the catchment, while for ground-based observations we used the time series recorded by the existing network, applying linear interpolation to derive variance values in ungauged locations. The results are presented in Figure 6. Comparing the two maps, it emerges that time series recorded from rain gauges exhibit generally higher variance than those coming from satellite. Despite the difference in the absolute values, the two datasets exhibit similar spatial patterns, with the highest values localized in the western, southern and south-eastern parts of the catchment. The high variance in the last two of those areas is due to the presence of the mountainous chains that block air masses moving from the ocean to inland and the other way around, generating high instability in the precipitation.

Comparing the map of the variance (Figure 6) with the location of the first eight ranked sensors (Figure 2), the sensors themselves are located in the areas with highest variance. Similar results can be observed for experiment S (Figure 3(a)). These findings confirm that entropy is mainly driven by precipitation variability and, therefore, time series variance (Alfonso *et al.* 2016) and that CMORPH precipitation estimates have high capability of capturing rainfall variability (Xie *et al.* 2007).

Our study gives insights into the information content of satellite data, both from a statistical and an IT perspective. It emerged that, when merging observations provided by rain gauges and satellite, an improvement in terms of information amount is obtained. However, more research in this direction is needed, to verify whether this result is due either to the optimal layout defined or to the combination
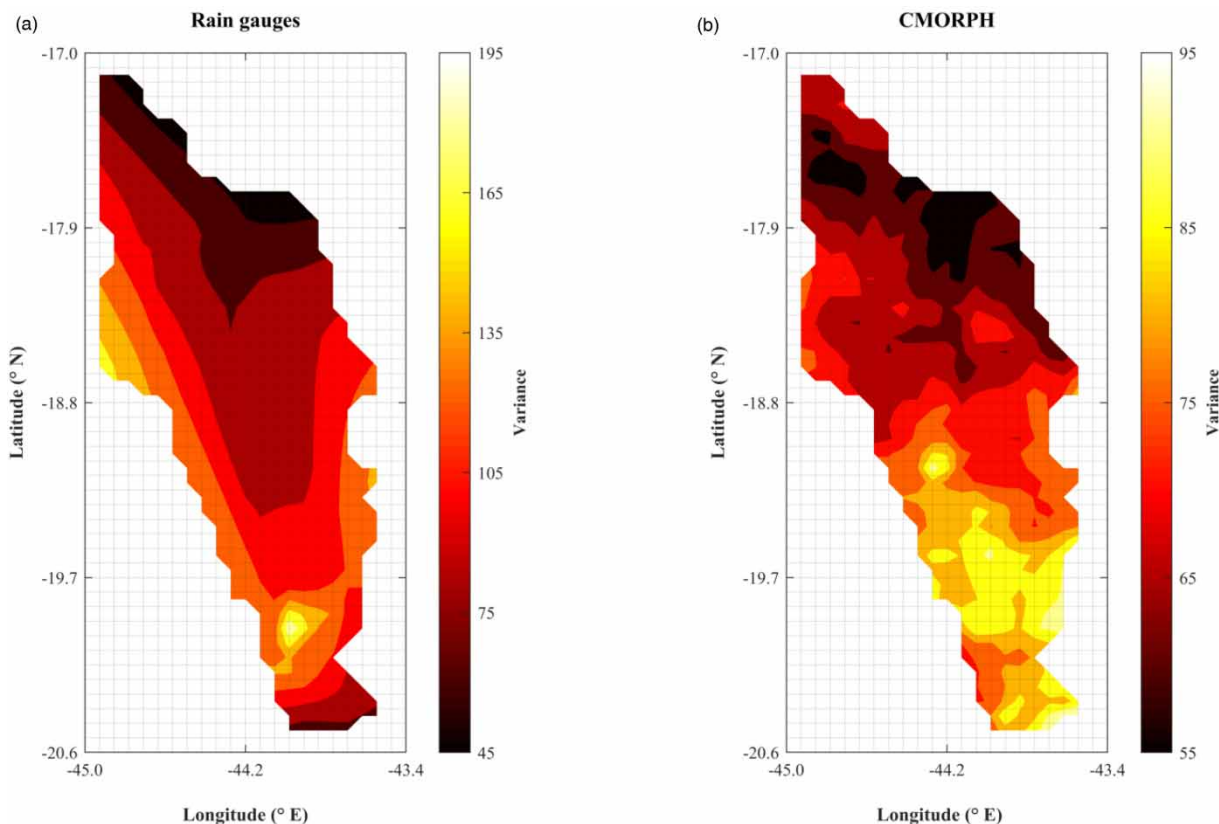
Figure 6 │ Variance of time series evaluated from (a) rain gauge observations and (b) CMORPH observations.

of two less dependent data sources. To this aim, future developments should test the efficiency of the optimal network, for example, by comparing the catchment response in terms of discharge by means of a hydrological model. Similar research is currently ongoing.

It is noteworthy that using satellite observations to derive data at ungauged locations avoids introducing additional bias related to rain gauge data interpolation, which are to be added to the systematic bias of ground measurements, as for instance under-catch due to wind effects (Pollock *et al.* 2018), evaporation and blowing snow (WMO 2008). On the other hand, using satellite data to identify locations for new sensors has the disadvantage of working with a coarser resolution with respect to ground-based data. Therefore, additional considerations, such as areas' accessibility, should be done to properly locate new stations.

Even though the optimal network does not contain some of the existing stations located in the inner part of the catchment (Figure 5(b)), these stations should be kept operating, both to improve rainfall knowledge in the area and to meet minimum density requirements suggested by the WMO (2008). It is worth noting that accuracy, understood as the deviation of the measurement from the real rainfall value, is not included in our analysis. We are aware that rainfall, as for instance, measured by gauge-radar comparison, can have an average difference of ±8% (Vieux & Vieux 2005), and that similar situations can happen with satellite data. We are also aware that accuracy of rainfall estimates depends on rainfall intensity, topography and climatic conditions of the area and that the use of remote sensing products adds even more uncertainties to these estimates (Yang & Luo 2014). Further considerations in this direction could be addressed in future research, e.g., adding to each cell time series black noise within a pre-defined range and repeating the optimization procedure, obtaining a family of Pareto fronts whose probabilistic distribution can be analysed. Finally, some limitations may arise when applying the methods in small catchments, as some of them could be characterized by localized convective rainfall events, while some others could have more uniformly distributed precipitation, depending on the climate and topography of the study area.

CONCLUSIONS

In this paper we present a method to optimize rain gauge networks using satellite observations with an entropy-based approach. The main idea is to use CMORPH precipitation records to derive information at ungauged locations and identify the most suitable locations to place new ground-based sensors, based on their information content.

To quantify the information achievable by rainfall records we applied the concept of joint entropy (JH). To identify the most informative locations within the catchment, we ranked the existing rain gauges and satellite cells based on their joint entropy, employing a greedy ranking algorithm. The results show that, in both cases, the first eight ranked stations are located in the areas with the highest variance of time series, confirming what had emerged in previous literature works, i.e., that entropy is mainly driven by precipitation variability. Although the locations selected in the two optimizations were not matching, the two networks, rain gauge-based and satellite-based, provide a similar amount of information from the IT perspective.

Finally, we combined the information coming from the two datasets to define the optimal network layout. The optimal network configuration is made of the first eight ranked existing rain gauges, complemented with eight locations chosen from CMORPH observations through an optimization problem, based on the maximization of the joint entropy. The number of existing rain gauges, i.e., eight, is chosen in order to preserve the high information content of the original network, while the addition of satellite cells is stopped when the increment in joint entropy is very limited, i.e., lower than 1%. The total amount of information provided by the optimal network is higher than the corresponding value obtained in the two previous networks considering the same number of sensors. Also, in this case, the most informative locations were found to be in the areas of the catchment with the highest variance. It is important to note that, although the optimal layout does not include all the existing rain gauges, we intend to preserve also the stations excluded, in order to maintain time series accuracy and length and to obtain a robust network.

An investigation of the variance distribution over the catchment from the two datasets was also conducted, to

check whether the two data sources were capturing the same rainfall variability. It emerged that, despite a difference in the absolute value, with higher values for ground-based observations, the spatial pattern of variance is the same for both data sources. The south and south-east areas and the western part of the catchment have the highest variance, probably due to the influence of topography.

In conclusion, satellite observations proved to be a powerful tool to derive rainfall records at ungauged locations to solve the network optimization problem. However, this result should be interpreted carefully and more research is needed to verify whether the significant information obtained is due either to the combination of two different and less dependent data sources or to the specific spatial configuration of sensors. Furthermore, the spatial scale of CMORPH product, even if finer than that of other satellite-based products, still remains coarser than the resolution of rain gauges. Other considerations, such as areas' accessibility, are needed to precisely locate the sensors within the identified cells.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories. https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00948#; http://www.snirh.gov.br/hidroweb.

REFERENCES

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P. P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P. & Nelkin, E. 2003 The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *Journal of Hydrometeorology* **4** (6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP > 2.0.CO;2.

Alfonso, L., Lobbrecht, A. & Price, R. 2010a Information theory-based approach for location of monitoring water level gauges in polders. *Water Resources Research* **46** (10). https://doi.org/10.1029/2009WR008101.

Alfonso, L., Lobbrecht, A. & Price, R. 2010b Optimization of water level monitoring network in polder systems using information theory. *Water Resources Research* **46** (12). https://doi.org/10.1029/2009WR008953.

Alfonso, L., He, L., Lobbrecht, A. & Price, R. 2013 Information theory applied to evaluate the discharge monitoring network of the Magdalena River. *Journal of Hydroinformatics* **15** (1), 211–228. https://doi.org/10.2166/hydro.2012.066.

Alfonso, L., Ridolfi, E., Gaytan-Aguilar, S., Napolitano, F. & Russo, F. 2014 Ensemble entropy for monitoring network design. *Entropy* **16** (3), 1365–1375. https://doi.org/10.3390/e16031365.

Alfonso, L., Mazzoleni, M., Chacon-Hurtado, J. C. & Solomatine, D. P. 2016 Optimal Design of Hydrometric Monitoring Networks with Dynamic Components Based on Information Theory. In: *12th International Conference on Hydroinformatics*, Incheon, South Korea, 21–26 August.

Amorocho, J. & Espildora, B. 1973 Entropy in the assessment of uncertainty in hydrologic systems and models. *Water Resources Research*. https://doi.org/10.1029/WR009i006p01511.

Banik, B. K., Alfonso, L., Di Cristo, C. & Leopardi, A. 2017 Greedy algorithms for sensor location in sewer systems. *Water (Switzerland)* **9** (11). https://doi.org/10.3390/w9110856.

Bertini, C., Buonora, L., Ridolfi, E., Russo, F. & Napolitano, F. 2020 On the use of satellite rainfall data to design a dam in an ungauged site. *Water* **12** (11), 3028. https://doi.org/10.3390/w12113028.

Caselton, W. F. & Husain, T. 1980 Hydrologic networks: information transmission. *Journal of the Water Resources Planning and Management Division, ASCE* **106**, 503–520.

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. https://doi.org/10.5194/hess-21-3071-2017.

Chen, Y. C., Wei, C. & Yeh, H. C. 2008 Rainfall network design using kriging and entropy. *Hydrological Processes* **22** (3), 340–346. https://doi.org/10.1002/hyp.6292.

Contreras, J., Ballari, D., de Bruin, S. & Samaniego, E. 2019 Rainfall monitoring network design using conditioned Latin hypercube sampling and satellite precipitation estimates: an application in the ungauged Ecuadorian Amazon. *International Journal of Climatology* **39** (4), 2209–2226. https://doi.org/10.1002/joc.5946.

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. https://doi.org/10.1175/jhm-d-16-0136.1.

Fahle, M., Hohenbrink, T. L., Dietrich, O. & Lischeid, G. 2015 Temporal variability of the optimal monitoring setup assessed using information theory. *Water Resources Research* **51** (9), 7723–7743. https://doi.org/10.1002/2015WR017137.

Gong, W., Yang, D., Gupta, H. V. & Nearing, G. 2014 Estimating information entropy for hydrological data: one-dimensional case. *Water Resources Research* **50** (6), 5003–5018. https://doi.org/10.1002/2014WR015874.

Gray, R. M. & Neuhoff, D. L. 1998 Quantization. *IEEE Transactions on Information Theory* **44** (6), 2325–2383.

Hofstra, N., Haylock, M., New, M., Jones, P. & Frei, C. 2008 Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research Atmospheres* **113** (D21). https://doi.org/10.1029/2008JD010100.

Hsu, K. L., Gao, X., Sorooshian, S. & Gupta, H. V. 1997 Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology* **36** (9), 1176–1190. https://doi.org/10.1175/1520-0450(1997)036<1176:PEFRSI > 2.0.CO;2.

Huang, Y., Zhao, H., Jiang, Y., Lu, X., Hao, Z. & Duan, H. 2019 Comparison and analysis of different discrete methods and entropy-based methods in rain gauge network design. *Water (Switzerland)* **11** (7). https://doi.org/10.3390/w11071357.

Huang, Y., Zhao, H., Jiang, Y. & Lu, X. 2020 A method for the optimized design of a rain gauge network combined with satellite remote sensing data. *Remote Sensing* **12** (1), 194. https://doi.org/10.3390/RS12010194.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Gu, G., Nelkin, E. J., Bowman, K. P., Hong, Y., Stocker, E. F. & 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. https://doi.org/10.1175/JHM560.1.

Joyce, R. J., Janowiak, J. E., Arkin, P. A. & Xie, P. 2004 CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology* **5** (3), 487–503. https://doi.org/10.1175/1525-7541(2004) 005<0487:CAMTPG > 2.0.CO;2.

Keum, J. & Coulibaly, P. 2017 Information theory-based decision support system for integrated design of multivariable hydrometric networks. *Water Resources Research* **53** (7), 6239–6259. https://doi.org/10.1002/2016WR019981.

Keum, J., Kornelsen, K. C., Leach, J. M. & Coulibaly, P. 2017 Entropy applications to water monitoring network design: a review. *Entropy* **19** (11). https://doi.org/10.3390/e19110613.

Keum, J., Awol, F. S., Ursulak, J. & Coulibaly, P. 2019 Introducing the ensemble-based dual entropy and multiobjective optimization for hydrometric network design problems: EnDEMO. *Entropy* **21** (10), 947. https://doi.org/10.3390/e21100947.

Kraskov, A., Stögbauer, H., Andrzejak, R. G. & Grassberger, P. 2005 Hierarchical clustering using mutual information. *Europhysics Letters* **70** (2), 278–284. https://doi.org/10.1209/epl/i2004-10483-y.

Krstanovic, P. F. & Singh, V. P. 1992 Evaluation of rainfall networks using entropy: II. Application. *Water Resources Management* **6** (4), 295–314. https://doi.org/10.1007/BF00872282.

Leach, J. M., Coulibaly, P. & Guo, Y. 2016 Entropy based groundwater monitoring network design considering spatial distribution of annual recharge. *Advances in Water Resources* **96**, 108–119. https://doi.org/10.1016/j.advwatres.2016.07.006.

Li, J., Bárdossy, A., Guenni, L. & Liu, M. 2011 A Copula based observation network design approach. *Environmental Modelling and Software* **26** (11), 1349–1357. https://doi.org/10.1016/j.envsoft.2011.05.001.

Li, C., Singh, V. P. & Mishra, A. K. 2012 Entropy theory-based criterion for hydrometric network evaluation and design: maximum information minimum redundancy. *Water Resources Research* **48** (5). https://doi.org/10.1029/2011WR011251.

Li, Y., Grimaldi, S., Walker, J. P. & Pauwels, V. R. N. 2016 Application of remote sensing data to constrain operational rainfall-driven flood forecasting: a review. *Remote Sensing* **8** (6), 456. https://doi.org/10.3390/rs8060456.

Li, S., Heng, S., Siev, S., Yoshimura, C., Saavedra, O. & Ly, S. 2019 Multivariate interpolation and information entropy for optimizing raingauge network in the Mekong River Basin. *Hydrological Sciences Journal* **64** (12), 1439–1452. https://doi.org/10.1080/02626667.2019.1646426.

Maddock, T. 1974 An optimum reduction of gauges to meet data program constraints. *Hydrological Sciences Bulletin* **19** (3), 337–345. https://doi.org/10.1080/02626667409493920.

Mazzoleni, M., Brandimarte, L. & Amaranto, A. 2019 Evaluating precipitation datasets for large-scale distributed hydrological modelling. *Journal of Hydrology* **578**, 124076. https://doi.org/10.1016/j.jhydrol.2019.124076.

Mishra, A. K. & Coulibaly, P. 2009 Developments in hydrometric network design: a review. *Reviews of Geophysics* **47** (2). https://doi.org/10.1029/2007RG000243.

Pádua, L. H. R., Nascimento, N. D. O., Silva, F. E. O. E. & Alfonso, L. 2019 Analysis of the fluviometric network of rio das velhas using entropy. *Revista Brasileira de Recursos Hidricos* **24**. https://doi.org/10.1590/2318-0331.241920180188.

Pinto, E. J. D. A. 2005 *Estudo De Indicadores Climáticos Para a Previsão De Longo Termo De Vazões Na Bacia Do Alto São Francisco (Study of Climate Indicators for Long Term Flow Forecasting in the Upper São Francisco Basin)*. PhD thesis, Departamento de Engenharia Sanitária e Ambiental, Departamento de Engenharia Hidráulica e Recursos Hídricos, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M. E., Colli, M., Stagnaro, M., Lanza, L. G., Lewis, E., Kilsby, C. G. & O'Connell, P. E. 2018 Quantifying and mitigating wind-induced undercatch in rainfall measurements. *Water Resources Research* **54** (6). https://doi.org/10.1029/2017WR022421.

Ridolfi, E., Montesarchio, V., Russo, F. & Napolitano, F. 2011 An entropy approach for evaluating the maximum information content achievable by an urban rainfall network. *Natural Hazards and Earth System Science* **11** (7), 2075–2083. https://doi.org/10.5194/nhess-11-2075-2011.

Ridolfi, E., Yan, K., Alfonso, L., Di Baldassarre, G., Napolitano, F., Russo, F. & Bates, P. D. 2012 An entropy method for floodplain monitoring network design. *AIP Conference Proceedings* **1479** (1), 1780–1783. https://doi.org/10.1063/1.4756522.

Ridolfi, E., Alfonso, L., Di Baldassarre, G., Dottori, F., Russo, F. & Napolitano, F. 2014a An entropy approach for the optimization of cross-section spacing for river modelling. *Hydrological Sciences Journal* **59**, 126–137. https://doi.org/10.1080/02626667.2013.822640.

Ridolfi, E., Servili, F., Magini, R., Napolitano, F., Russo, F. & Alfonso, L. 2014b Artificial Neural Networks and entropy-based methods to determine pressure distribution in water distribution systems. *Procedia Engineering* **89**, 648–655. https://doi.org/10.1016/j.proeng.2014.11.490.

Santos, M. S., Costa, V. A. F., Fernandes, W. D. S. & de Paes, R. P. 2019 Time-space characterization of droughts in the São Francisco river catchment using the Standard Precipitation Index and continuous wavelet transform. *Revista Brasileira de Recursos Hidricos* **24**. https://doi.org/10.1590/2318-0331.241920180092.

Sapiano, M. R. P. & Arkin, P. A. 2009 An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *Journal of Hydrometeorology* **10** (1), 149–166. https://doi.org/10.1175/2008JHM1052.1.

Shannon, C. E. 1948 A mathematical theory of communication. *Bell System Technical Journal* **27** (4), 623–656. https://doi.org/10.1002/j.1538-7305.1948.tb00917.x.

Shannon, C. E. & Weaver, W. 1949 *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, USA.

Sheppard, W. F. 1897 On the calculation of the most probable values of frequency-constants, for data arranged according to equidistant division of a scale. *Proceedings of the London Mathematical Society* **s1–29** (1), 353–380. https://doi.org/10.1112/plms/s1-29.1.353.

Vieux, B. & Vieux, J. 2005 Rainfall accuracy considerations using radar and rain gauge networks for rainfall-runoff monitoring. *Journal of Water Management Modeling* **13**. https://doi.org/10.14796/JWMM.R223-17.

Walker, D., Forsythe, N., Parkin, G. & Gowing, J. 2016 Filling the observational void: scientific value and quantitative validation of hydrometeorological data from a community-based monitoring programme. *Journal of Hydrology* **538**, 713–725. https://doi.org/10.1016/j.jhydrol.2016.04.062.

Wei, C., Yeh, H. C. & Chen, Y. C. 2014 Spatiotemporal scaling effect on rainfall network design using entropy. *Entropy* **16** (8), 4626–4647. https://doi.org/10.3390/e16084626.

Werstuck, C. & Coulibaly, P. 2017 Hydrometric network design using dual entropy multi-objective optimization in the Ottawa River basin. *Hydrology Research* **48** (6), 1639–1651. https://doi.org/10.2166/nh.2016.344.

WMO 2008 *Guide to Hydrological Practices. Volume I: Hydrology–from Measurement to Hydrological Information*. World Meteorological Organization, Geneva, Switzerland.

Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C. & Yang, S. 2007 A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology* **8** (3), 607–626. https://doi.org/10.1175/JHM583.1.

Xie, P., Joyce, R., Wu, S., Yoo, S. H., Yarosh, Y., Sun, F. & Lin, R. 2017 Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *Journal of Hydrometeorology* **18** (6), 1617–1641. https://doi.org/10.1175/JHM-D-16-0168.1.

Xie, P., Joyce, R., Wu, S., Yoo, S.-H., Yarosh, Y., Sun, F. & Lin, R. 2019 *NOAA CDR Program (2019): NOAA Climate Data Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates, Version 1 [a]*. NOAA National Centers for Environmental Information. https://doi.org/10.25921/w9va-q159.

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. https://doi.org/10.1016/j.jhydrol.2015.03.034.

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

Yang, Y. & Burn, D. H. 1994 An entropy approach to data collection network design. *Journal of Hydrology* **157**, 307–324. https://doi.org/10.1016/0022-1694(94)90111-2.

Yang, Y. & Luo, Y. 2014 Evaluating the performance of remote sensing precipitation products CMORPH, PERSIANN, and

TMPA, in the arid region of northwest China. *Theoretical and Applied Climatology* **118** (3). https://doi.org/10.1007/s00704-013-1072-0.

Yeh, H. C., Chen, Y. C., Chang, C. H., Ho, C. H. & Wei, C. 2017 Rainfall network optimization using radar and entropy. *Entropy* **19** (10), 1–14. https://doi.org/10.3390/e19100553.

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). https://doi.org/10.1061/(ASCE)1084-0699(2008)13:4(226).

Zeng, Q., Chen, H., Xu, C. Y., Jie, M. X., Chen, J., Guo, S. L. & Liu, J. 2018 The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. *Journal of Hydrology* **563**, 106–122. https://doi.org/10.1016/j.jhydrol.2018.05.058.