Daily precipitation concentration around the world according to several indices

Robert Monjo[1*] and Javier Martin-Vide[2]
[1] Climate Research Foundation (FIC), C/Gran Vía 22 (duplicado), 7º, 28013 Madrid, Spain (rma@ficlima.org)
[2] Climatology Group, Department of Physical Geography, University of Barcelona, C/Montalegre 6, 08001 Barcelona, Spain (jmartinvide@ub.edu)

IMPORTANT*: This is the pre-peer reviewed version of the following article:* Monjo R, Martin-Vide J. 2015. Daily precipitation concentration around the world according to several indices. *Int. J. Climatol.*, doi: 10.1002/joc.4596***, which has been published in final form at http://doi.org/10.1002/joc.4596. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.***

Abstract

The temporal concentration of precipitation may be characterised using several methods. For climate-scale precipitation, concentration measures are usually performed by means of dimensionless indices such as the Gini Index or the Theil Index. For the purposes of this paper, a set of 66,409 daily time series from around the world were analysed to estimate the climatic concentration of precipitation. To this end, some of the most widely used indices were tested, i.e. the Theil Index, the Gini Index, the Concentration Index, the classic *n* index and an ordered version of the *n* index. Results show a strong connection between several indices, mainly between the Gini Index and the ordered *n* index. The high correlation of these indices (R = 0.98) reflects a theoretical connection between the shape and integration of the Lorenz curve. As regards spatial distribution, the three main indices present the same relative areas of high and low concentration. The high temporal concentration of precipitation is generally linked to the rapid pace of physical processes such as convection in areas with a high degree of insolation and warm seas. The low temporal concentration of rainfall can be interpreted as a consequence of regular patterns (maritime flows or highly recurrent storms). A relationship between the number of rainy days and concentration indices was noted; however, their correlation depends on the region analysed.

Key words: daily precipitation, Gini Index, Concentration Index, *n* index, world

1. Introduction

In the context of climate change analysis or simply as a means to determine the current climate, irregular temporal patterns of precipitation prove of great interest. The temporal concentration of precipitation is key to monitoring the pace of physical processes in the atmosphere. Climatic regions with a more advective regime show regular rainfall patterns, e.g. a higher frequency. However, regions with mainly convective rainfall show greater irregularity in precipitation behaviour.

The temporal concentration of precipitation can be determined by means of several methods. Some authors use fractal theory to describe the self-similarity of rainfall, according to the box-counting method or the Hurst exponent (Feder, 1988; Olsson *et al.*, 1992; Taouti and Chettih, 2014). Rainfall irregularity in time is also

described using the Maximum Average Intensity (MAI) and the Intensity Duration Frequency (IDF) curves (Ghahraman and Hosseini, 2005). In particular, the rate of decline of the MAI curve can be analysed using the *n* index, which is related to the temporal distribution of precipitation (Monjo, 2015).

Other methods focus on the relationship between the accumulation of a variable and its cumulative frequency. This function is known as the concentration curve or Lorenz curve, widely used in many sectors (Shaw and Wheeler, 1994). The area under this curve is related to the Gini Index (*GI*) and can be modelled by theoretical probability distributions. If an exponential distribution is fitted, the resulting version of the Gini Index is referred to as the Concentration Index (*CI*) (Martin-Vide, 1994, 2004, 2013). The Concentration Index has been used to analyse the characteristics of rainfall in some countries such as Spain, Italy, Iran, China, Malaysia, Chile and Algeria (Alijani *et al.*, 2008; Zhang *et al.*, 2009; Suhaila and Jemain, 2012; Coscarelli and Caloiero, 2012; Sarricolea and Martin-Vide, 2014; Benhamrouche *et al*., 2015), as well as in Europe (Burgueño *et al.*, 2010; Cortesi *et al.*, 2012). In particular, the *CI* is related to the percentage of wet days, which collects most of the annual precipitation. This analysis is relevant for Mediterranean climates as it is an indicator of torrentiality and erosivity (Martín-Vide, 1994).

Some disciplines employ information entropy formulae such as the Theil Index (*TI*), derived from the Shannon Index (Cowell, 2000; Jost, 2006), and the Atkinson Index, or its related development of the Generalised Entropy Index (Atkinson, 1970; Ullah and Giles, 1998). The Theil Index is probably the simplest inequality indicator as it does not depend on any additional parameter; moreover, it is a particular Generalised Entropy Index case.

In a climate context, the Theil Index can be adapted to describe the temporal inequality of rainfall. Thus, the uneven distribution over time is another property for measuring the climatic concentration of precipitation.

This study seeks to analyse the main techniques for measuring the temporal concentration of daily precipitation in time series, and thereby to explore its spatial distribution around the world. For this purpose, several indices and their relationships are examined, i.e. the Gini Index, the Concentration Index, the Theil Index and a modified version of the *n* index. This paper therefore endeavours to further knowledge of how to measure the contribution of the wettest days to total precipitation. Thanks to this methodology, new pluviometric regions can be defined according to rainfall concentration characteristics, and the indices will prove useful for analysing changes in precipitation patterns and their possible influence on erosivity.

2. Data

A total of 66,409 daily time series are available for the period 1950-2014, taken from the Global Historical Climatology Network-Daily Database (GHCN-Daily) (Menne *et al.*, 2012).

The selected stations correspond to all those with at least 1,000 daily precipitation values and at least 70% temporal continuity. Of the 66,409 stations, slightly more than half had more than 8,000 records, and more than 75% exceeded 3,000 records with 98% continuity. Approximately 25% of the stations stand out as having more than 20,000 daily records with 99% continuity.

Quality control was performed using the Kolmogorov-Smirnov test to detect inhomogeneities (Monjo *et al.*, 2013). Only 5% of the time series presented some

inhomogeneities, and in these cases, the longest homogeneous sub-time series was selected.

On average, there is approximately one observatory for every 2,244 km^2 of Earth landmass. However, spatial cover is very uneven, depending on the country analysed. This especially depends on their economic development or their former colonial dependence (Fig. 1). The countries with greater station density include the USA, Brazil, India, Australia and the Netherlands. Africa presents a low density in virtually its entire area, except for the south. The deserts and the coldest areas of the planet are other regions with low station density, due to their small population.

3. Methodology

3.1. Previous concepts

The Lorenz curve is the function that relates the accumulation of a variable (y) and its cumulative frequency (x). Using this curve, the Gini coefficient or Gini Index (GI) is defined as (Haughton and Khandker, 2007):

$$GI = \frac{S}{S+A} = 1 - 2A \qquad \text{Eq. 1}$$

where S is the area compressed by the curve $y(x)$ and the equidistribution line $y = x$ (see Fig. 2). Therefore, S is equal to $1/2 - A$, where A is the area under the curve $y(x)$.Area A can be estimated using the trapezoidal rule to integrate $y(x)$, that is:

$$A = \sum_{i=2}^{N} \left(\frac{y_i + y_{i-1}}{2} \left(x_i - x_{i-1} \right) \right) \qquad \text{Eq. 2}$$

Probability distribution models can be used to simulate area A. According to Martin-Vide's definition of the Concentration Index (CI) (2004), the exponential distribution integral $y(x) = a \cdot \exp(b \cdot x)$ is commonly used for fitting and estimating this area in a theoretical manner:

$$A \approx \left[\frac{a}{b} e^{bx} \left(x - \frac{1}{b} \right) \right]_{x=0}^{x=1} \qquad \text{Eq. 3}$$

where a and b are fitted parameters of the curve $y(x)$.

3.2. Relative cumulative precipitation

Firstly, the N values of non-zero precipitation p are ordered from the lowest to the highest (p^L), or from the highest to the lowest (p^H). Hence, relative cumulative precipitation P_i (i.e. $P^L{}_i$ or $P^H{}_i$) on the position i is defined as:

$$P_i \equiv \frac{\sum_{k=1}^{i} f_k p_k}{\sum_{k=1}^{N'} f_k p_k} \qquad \text{Eq. 4}$$

where p_k is the k-th value of ordered precipitation (i.e. $p^L{}_k$ or $p^H{}_k$), f_k is its relative frequency and $N' \leq N$ is the number of classes or pairs (p_k, f_k). Precipitation p_k can take all values with repetitions (0.1, 0.1, …, 0.2, 0.2, …, and so $f_i = 1/N$ with $N' = N$), or it can take the unique values (0.1, 0.2, 0.3, 0.4, …, with $f_i \geq 1/N$), 0.1 being the minimum non-zero precipitation in millimetres. For both cases of p, the relative cumulative frequency F_i is defined as:

$$F_i = \sum_{k=1}^{i} f_k \qquad \text{Eq. 5}$$

The domain of F and P is between 0 and 1. In accordance with the order, F^L and F^H are defined as the cumulative relative frequencies of p^L and p^H respectively. Note that curve $P(F)$ is the same if one uses all values of p (with $f_i = 1/N$) or only the unique values (with $f_i \geq 1/N$).

3.3. Inequality-concentration indices

Besides the Concentration Index (*CI*) of Martin-Vide (Eq. 1 with 3), the Gini Index (*GI*) was also used in this study. From its definition (Eq. 1 and 2), it follows that variable x is F^L and y is P^L. The Gini Index can therefore be rewritten as (Haughton and Khandker, 2007):

$$GI = 1 - 2\sum_{i=2}^{N'} \left(\frac{P^L{}_i + P^L{}_{i-1}}{2} \left(F^L{}_i - F^L{}_{i-1} \right) \right) \qquad \text{Eq. 6}$$

The difference between the *CI* and the *GI* is that the first uses theoretical probability distribution representing the population of the rainfall (theoretical climate), while the second only employs the empirical curve, representing a sample of the rainfall (observed values).

On the other hand, the Theil Index adapted to precipitation is given by:

$$TI = \frac{1}{N \ln N} \sum_{i=1}^{N} \left(\frac{p_i}{\bar{p}} \cdot \ln \frac{p_i}{\bar{p}} \right) \qquad \text{Eq. 7}$$

where p_i is each one of the N non-zero precipitation values (with repetitions) and *p-bar* ($\bar{p}$) is its average. Note that in this case there is no need to sort the p values.

3.4. Ordered version of the n *index*

The n index is a dimensionless indicator of the temporal irregularity of rainfall (Monjo, 2015). It is defined as the exponent of the power law given by the Maximum Average Intensity (MAI) of the rainfall, according to its duration (averaging time) (Moncho *et al.*, 2009).

In order to compare climatic indices such as the *GI* with the n index, a reformulation of the latter was required. The n index depends on the natural time sequence of rainfall events, while the *GI* is estimated by breaking the natural sequence, i.e. by sorting all values from the minimum to the maximum. For this reason, an *ordered n index* (n_{or}) is defined using a MAI linked with the ordered daily precipitation records. Hence, the ordered MAI (I_i) is defined as:

$$I_i \equiv \frac{P^H{}_i}{i} \qquad \text{Eq. 8}$$

where $P^H{}_i$ is the relative cumulative precipitation (Eq. 4) and i is the time in days. Note that, generally speaking, ordered MAI is different from the real MAI. An example that shows these differences can be found in Table 1. Therefore, the definition of the ordered n index (n_{or}) is given by:

$$I_i = I_1 \left(\frac{1}{i} \right)^{n_{or}} \qquad \text{Eq. 9}$$

where I_1 and n_{or} are fitted parameters. Using Eq. 8 and Eq. 9, Eq. 10 can be rewritten as a Maximum Cumulative Precipitation (MCP) curve:

$$P^H{}_i = I_1 \cdot i^{1-n_{or}} \quad \text{Eq. 10}$$

Note that the ordered case of MCP is an absolute version of the Normalised Rainfall Curve (NRC) used by other authors (Soman and Kumar, 1990; Jolliffe and Hope, 1998; Burgueño *et al.*, 2010).

3.5. Statistical analysis

The n_{or} index was obtained by using the Profile Log-Likelihood (PLL) approach (Raue *et al.*, 2009). The climate average of the classic daily n (n_δ, Monjo, 2015) was also obtained to compare with the n_{or} index. All indices (n_δ, n_{or}, *CI*, *GI* and *TI*) were analysed to obtain a general intercomparison statistic and to determine their spatial distribution around the world. Finally, the relationship between the fraction of wet days (*WD*) and the concentration indices was explored.

In order to enhance the visual analysis, a spatial interpolation was performed using two-dimensional Thin Plate Splines (TPS). In particular, interpolation was applied to each point and its ten nearest neighbours, extrapolating in circles to the maximum distance of the neighbours. The resulting raster was obtained by averaging all contributions (66,409 interpolations), weighted by the inverse of the distance.

All statistical analysis was performed by using the R language (Muenchen and Hilbe, 2010; R Development Core Team, 2010). A very low level of significance (p-value < $2 \cdot 10^{-16}$) is considered for Pearson correlation between indices because it is the typical floating point error of R operations.

4. Results

4.1. Intercomparison of indices

A very high correspondence between the *CI* and the *GI* was found, with a Pearson correlation of $R = 0.994$ and a Mean Normalised Absolute Error (MNAE) of 0.8% (Fig. 3a). In 95% of cases, absolute error is less than 2%. Therefore, for practical purposes, they are indistinguishable, at least for daily precipitation.

The correlation between n_{or} and the *GI* is very high, $R = 0.98$ (Fig. 3b and Table 2). This result reflects a strong mathematical connection that can be seen in section 5.1. In the case of the classic n index (n_δ), its correlation with the *GI* is lower but also remarkable ($R = 0.71$, p-value < $2 \cdot 10^{-16}$). In fact, it is higher than the correlation between n_δ and n_{or} ($R = 0.63$, p-value < $2 \cdot 10^{-16}$). This result is surprising because the classic n_δ index results from individual calculation for each rainfall event ("weather perspective"), while the *GI* and n_{or} are calculated by sorting all the climatic values of each time series ("climate perspective") (see section 5.2).

The Theil Index does not use cumulative precipitation and therefore, a priori, it is the most different of the indices analysed. However, it shows a high correlation with n_{or} ($R = 0.76$) and with the *GI* ($R = 0.75$) (Fig. 3c and Table 2). The theoretical values of the *TI* varies between 0 and 1, but they barely reach the threshold of 0.25 (99.95^{th}-centile).

4.2. Spatial distribution of daily precipitation

In accordance with the mathematical relationships, the three main indices analysed (*GI*, n_{or} and *TI*) present a similar spatial distribution (Fig. 4 to 6). The location of the areas

with the highest/lowest concentration is practically the same (especially for n_{or} and *GI*), and therefore the three indices are analysed together.

Among the areas with higher climatic concentration of rainfall, two groups are noteworthy: (1) seasonally warm coastlines (Australia, Brazil, the Caribbean and the Mediterranean); and (2) areas near the world's major deserts, mainly in the western and northern sides. For example, mention can be made of the northern Gobi Desert (China and Mongolia), western Thar Desert (India), most of the Arabian and Sahara Deserts, western Sonora Desert (USA and Mexico), eastern Patagonian Steppe (Argentina) and Atacama Desert (Chile). According to Sarricolea and Martin-Vide (2014), Chile's highest concentration is found in the northern region with a Mediterranean climate (approximately 30°S), not far from the Atacama Desert.

As an extension of the group (1), other high concentration areas are the subtropical and tropical regions such as Madagascar, Southeast Asia and much of the Pacific Islands. In general, the characteristic heat of both groups leads to a faster life cycle for the precipitation systems, since the water cycle is amplified by higher temperatures (Roderick *et al.*, 2012; Durack *et al.*, 2012). East Asia is a special high concentration case due to the atmosphere dynamics: precipitation bands of the monsoon season move rapidly northward throughout late spring/early summer (Ding *et al.*, 2005). This rain belt, referred to as the Mei-Yu (China), Baiu (Japan) and Changma (Korea), reaches 45 degrees north (Bao and Yu 2014). The East Asia case contrasts with the Indian Summer Monsoon, which presents a more persistent precipitation regime and therefore rainfall concentration is moderate (values close to the median).

On the other hand, the regions with more regularity can be classified into three types: (1) arid/semi-arid areas with low convection (Kalahari Desert, eastern Sonora Desert extending to other areas of the Western United States and the Gobi Desert interior); (2) oceanic climate regions with a continuous maritime flow (e.g. New Zealand, Western France and some parts of Western Canada) and (3) rainforests or pluvisilva with a recurrent pattern of daily storms (e.g. interior of Brazil and Guinea).

The latitudinal profile (quartiles Q1, Q2 and Q3) of the indices is an interesting aspect of the spatial distribution. The *GI* profile is very similar to the n_{or} profile, differing only in terms of scale (mean value and width of the distribution) (Fig. 4b and 5b). However, the classic n_{δ} profile differs as regards the polar latitudes because these regions have a stratiform regime (low n_{δ}), but the rainfall is irregularly accumulated throughout the year (high n_{or}). The *TI* profile shows a great width (Q1-Q3) in the northern distribution (Fig 6b), probably due to the greater continentality of the Northern Hemisphere (with a higher temperature range). In fact, the Southern Hemisphere shows more regularity in rainfall accumulation according to the concentration indices (especially in 10ºS compared to 10ºN).

5. Discussion

5.1. Temporal data coherence

Data selection criteria were adopted according to one of the study's main objectives, namely, to obtain an approximate spatial distribution of the concentration indices around the world. However, some countries have a low station density (Fig. 1) and a short time-series length (less than 5,000 records), especially in some parts of Africa, Northwestern China and Oceania (Fig. 1). Therefore, relaxed criteria were chosen in order to provide sufficient spatial coverage, but with sufficient climate information: time-series with at least 1,000 records and with at least 70% temporal continuity.

This low threshold may seem insufficient to be compared with longer series, some of which cover the entire period (1951-2014). Nevertheless, an analysis was performed of the effects of considering one time period or another on the spatial distribution of the indices. In particular, a set of time series was chosen, all of which spanned at least 40 years, 20 years in the period 1951-1975 and 20 years in 1975-2000. The results showed that both spatial distributions are very similar, even to that obtained with the relaxed criteria. In fact, an exhaustive analysis of the *GI* trends was performed, taking in account all time series with at least 20 years' worth of data, but most of them did not reveal significant trends. Only 28% showed trends higher than ±0.1 century^{-1} at the level of $p \leq 0.05$, and 19% (included in the 28%) was significant at the level of $p \leq 0.01$. Therefore, we can conclude that the selected criteria do not affect the spatial analysis shown in this study.

5.2. Connections between the n_{or} and concentration-entropy indices

The results showed a strong relationship between the n_{or} index and the Gini Index; hence, a mathematical connection is expected. Firstly, it is easy to demonstrate that the MCP curve, $P^H = f(i)$, is the symmetrical one of the Lorenz curve, $P^L = f(F^L)$. Using a transformation of the time i in a relative time $F^H \equiv i/N$, it can be seen that the normalised rainfall curve $P^H = f(F^H)$ is the same as $1 - P^L = f(1 - F^L)$. For this reason, it is expected that the integration area of the Lorenz curve (related with *GI*) depends on its shape (related with n_{or}). In particular, the empirical relationship between both indices is a power law:

$$n_{or} \approx GI^{\gamma} \qquad \text{Eq. 11}$$

where $\gamma \approx 0.66 + 0.34 \cdot GI$, or simply $\gamma \approx 0.85$ (p-value < $2 \cdot 10^{-16}$).

However, the explicit theoretical relationship is not trivial at first glance. The Gini Index is directly related to the average of the relative cumulative precipitation (see Appendix A):

$$GI \approx 1 - 2\frac{1}{N}\sum_{i=1}^{N} P^L{}_i \qquad \text{Eq. 12}$$

While n_{or} is obtained as a fitted parameter of Eq. 9 or 10, this can be approximated by a weighted average of the P^H logarithm (see Appendix B):

$$n_{or} \approx 1 + \frac{1}{N}\sum_{i=1}^{N}\left(\alpha_N(i) \cdot \ln P^H{}_i\right) \qquad \text{Eq. 13}$$

and is very roughly approximated as:

$$n_{or} \approx 1 - \frac{2}{N}\sum_{i=1}^{N-1}\left(\ \alpha_N(N-i+1)\ \cdot\ P^L{}_i\right) \qquad \text{Eq. 14}$$

where $\alpha_N(i) \approx \ln(N) - \ln i - 1 + 0.5/\mathrm{N} \cdot \ln(2\pi\mathrm{N})$ is interpreted as a weight. Eq. 14 therefore presents a strong connection with Eq. 12, especially if the weight $\alpha_N(i)$ is replaced by 1. These weights, which are higher for lower positions ($i << N$), have been obtained for linear regression using least squares. Nevertheless, for a non-linear regression (e.g. PLL approach), the weights are closer to 1. Hence, the empirical relationship between n_{or} and *GI* (Eq. 11) has been theoretically justified by the connection between Eq. 12 and 14.

On the other hand, Eq. 13 has a structure close to information entropy formulae such as the Theil Index (Eq. 7) and one of the Atkinson indices (*AI*) (Eq. 15; Atkinson, 1970), or its equivalent version in the Generalised Entropy Index (*GE*) (Eq. 16; Ullah and Giles, 1998; Haughton and Khandker, 2007).

$$AI = 1 - \exp\left(1 + \frac{1}{N}\sum_{i=1}^{N}\left(\ln P^{L}{}_{i}\right)\right) \qquad \text{Eq. 15}$$

$$GE = -\frac{1}{g_o}\left(1 + \frac{1}{N}\sum_{i=1}^{N}\left(\ln P^{L}{}_{i}\right)\right) \qquad \text{Eq. 16}$$

where g_o is a normalisation constant (equal to $1 + \ln P^{L}{}_{1}$), which is found for the particular case of using a relative cumulative variable (i.e. $P^{L}{}_{i}$). Note that the relationship between both indices is $AI = 1 - \exp(-g_o \cdot GE)$.

Similarities between both entropy indices (*AI*, *GE*) and the tested indices (*GI*, n_{or}, *TI*) are shown in Table 2. In particular, it is found that the Pearson correlation between *GI* and *AI* is very high ($R = 0.91$).In fact, a relationship between both indices was found by other authors (Lambert *et al.*, 2003). The *GE* has a lower correlation than the *AI* for all indices except for *TI*, which shows $R = -0.92$. Overall, the Gini Index (or the *CI* version) is the most informative of all the indices (with an average correlation of 0.84). Note that the *CI* does not appear in Table 2 as it is considered that it provides redundant information regarding the *GI*, i.e. both have the same correlation with the rest of the indices.

5.3. Interpretation of the connections

The high correlation between the classic n_δ and the *GI* (0.71) is more difficult to explain than the connection between n_{or} and the *GI*, because n_δ is the average for all individual rainfall events (for a location), while the *GI* is estimated for the entire duration of a time series. In other words, a priori there is no mathematical connection between both climatic features. However, a physical connection is expected. In particular, it can be said that the time structure of a rainfall event is related to irregular precipitation distribution on wet days during the year: for climates with greater rainfall concentration, this concentration is also greater within the rainfall events. In other words, the time structure and irregularity of precipitation are reflected in a similar manner at low and high time scales; and this is because the pace of physical processes (evaporation + transport + condensation) is also interconnected at various time scales (the energy of the sensible and latent heat flux is propagated throughout the water cycle). In fact, the *n* index shows autosimilarity properties at several time scales, according to its connection with the fractal dimension of the rainfall events (Monjo, 2015). In addition and according to the current study, it could be said that n_{or} and the *GI* reflect characteristics of both the climatic concentration and information entropy of the precipitation, which are related to the physical processes of convection and transport.

It is hoped that the fractality of the rainfall concentration can be extended to the number of rainy days as an index of how precipitation is accumulated over a few or several days. In other words, it is expected for higher concentration to be observed in climates with fewer wet days. A negative correlation is therefore expected between the concentration indices and the fraction of wet days (*WD*).

However, the obtained Pearson correlation depends on the region analysed (Fig. 7). Regions with higher/lower rainfall concentration do not always correspond to the regions with the lowest/highest number of rainy days. That is, there may be cases where it rains few days per year, but very similar amounts are recorded over those days. Likewise, there may be areas where it rains almost every day, but the accumulation of rainfall is highly irregular for some days compared to others. This is observed throughout the positive or at least non-negative correlation of some regions of the world (Fig. 7), and somehow it is noted in the spatial distribution of the three main indices

(Fig. 4 to 6). The Kalahari Desert (Southwest Africa), the Gobi Desert interior (China and Mongolia) and interior areas of the Sonora-Mojave Desert (Mexico and the USA) do not show a large concentration, while oceanic climate regions (Cfb/Cfc according to the Köppen classification) such as Germany and the Netherlands do not show the uniformity that was expected according to their high number of wet days.

For most of Asia, northern Africa and the Mediterranean Basin, the correlation between *WD* and the *GI* is negative and statistically significant. Cortesi *et al.* (2012) found similar results for Europe ($R = -0.68$, p-value <0.01). In agreement with Burgueño *et al.* (2010), Europe presents a north-to-south spatial pattern, from the west maritime temperate climate (low *GI* and high *WD*) to the Mediterranean climate (high *GI* and low *WD*).

Differences in the patterns of India and Southeast Asia are also observed according to the sign of the correlation. The most persistent rainfall regime in India is framed within a shorter period (summer monsoon), i.e. yearly rainfall is accumulated as a truncated triangular (almost rectangular) function. Hence, rainfall concentration increases when the number of rainy days is extended (positive correlation). In contrast, the rain in Southeast Asia is irregularly recorded throughout the year as an entire triangular function. Therefore, a shorter rainfall period causes a higher concentration (negative correlation).

6. Conclusions

Strong connections were found between the indices analysed, and therefore just one can be used for most comparative analyses. In particular, we recommend using the Gini Index (*GI*) due to its simplest estimation and its highest correlation ($R > 0.7$) with the other indices. Taking this index as a basis for comparison, some mathematical features can be gleaned. According to its similarity with the Concentration Index (*CI*), the Lorenz curve can be simulated using an exponential function. In addition, on account of the mathematical relationship between the *GI* and the n_{or} index, it can be said that the n_{or} index is related to the shape of the Lorenz curve.

The high concentration of precipitation is linked to the rapid pace of physical processes such as convection. In fact, regions with the highest concentration correspond to areas with a high degree of insolation (arid or semi-arid regions) or are influenced by warm seas, such as the Gulf of Mexico, the Western Mediterranean and the Arabian Sea.

In general, low rainfall concentration can be interpreted as a consequence of regular patterns such as large-scale maritime flows (advective processes) or repetitive patterns such as a successive passage of lows or storms. Examples of the repetitive passage of lows are France and parts of Canada, whilst cases of recurrent storms occur in Brazil and Eastern Peru. Desert regions can be classified into two types according to their high or low precipitation concentration during the few days of rain. In particular, the highest concentration for deserts is observed in coastal areas (e.g. west of the Sonora, Atacama and Thar deserts), whilst the lowest concentration is located in inland areas (e.g. east of the Sonora, Thar and central Sahara deserts). The Asian monsoon also presents two different types: the Indian Summer Monsoon (with moderate rainfall concentration values) and the East Asian Monsoon (with very high concentration values), according to the higher or lower persistence of rainfall events.

The correlation between the number of rainy days and the concentration indices depends on the region analysed. For example, the regions with uneven rainfall (e.g. the Mediterranean Basin and South-Eastern China) present higher rainfall concentration

when less rainy days occur (negative correlation). However, the regions with regular rainfall patterns (e.g. Brazil and India) show a higher concentration when more rainy days are reported (positive correlation).

This study could serve to characterise the world's climates according to rainfall concentration and the physical mechanisms governing this behaviour. In fact, the concentration indices analysed prove useful for measuring climate changes in rainfall patterns and their influence on erosivity.

Acknowledgments

This work would not be possible without the contribution of the Global Historical Climatology Network-Daily Database (GHCN-Daily). In addition, we would like to thank the Spanish project CSO2014-55799-C2-1-R and the Group of Climatology (2014 SGR 300) for their support.

Appendix A. Obtaining Equation 12

Given the definitions of relative cumulative precipitation (P^L) and relative cumulative frequency (F^L) from Eq. 4 and 5, the Gini Index (*GI*) is defined as:

$$GI = \frac{S}{S+A} = 1 - 2A \qquad \text{Eq. A1}$$

where *S* is the area compressed by the curve $P^L(F^L)$and the equidistribution line $P^L = F^L$, while *A* is the area under the curve $P^L(F^L)$. Therefore *S* is equal to 1/2 – *A*. Using the trapezoidal rule for the integral of $P^L(F^L)$, area *A* is:

$$A = \sum_{i=2}^{N'} \left(\frac{P^L{}_i + P^L{}_{i-1}}{2} \left(F^L{}_i - F^L{}_{i-1} \right) \right) \qquad \text{Eq. A2}$$

Since $F^L{}_i - F^L{}_{i-1} = f^L{}_i$ (from Eq. 5), the Gini Index can be rewritten as:

$$GI = 1 - 2 \sum_{i=2}^{N'} \left(\frac{P^L{}_i + P^L{}_{i-1}}{2} f^L{}_i \right) \qquad \text{Eq. A3}$$

If all precipitation values (with repetitions) are taken (i.e., *N*'= *N*), frequency is replaced by $f^L{}_i = 1/N$. Then:

$$GI = 1 - 2 \sum_{i=2}^{N} \left(\frac{P^L{}_i + P^L{}_{i-1}}{2} \frac{1}{N} \right) \qquad \text{Eq. A4}$$

It is easy to verify that:

$$GI = 1 - 2 \sum_{i=2}^{N-1} \left(\frac{2 \cdot P^L{}_i}{2} \frac{1}{N} \right) - \frac{P^L{}_N + P^L{}_1}{N} \qquad \text{Eq. A5}$$

Finally, doubling the last term (which is negligible), a simplified version of the Gini Index is found:

$$GI \approx 1 - 2 \frac{1}{N} \sum_{i=1}^{N} P^L{}_i \qquad \text{Eq. A6}$$

Appendix B. Obtaining Equations 13 and 14

The ordered *n* index (n_{or}) is given by the relationship between relative cumulative precipitation ($P^H{}_i$) and the time *i*, which can be fitted with a power law (from Eq. 10):

$$P^H{}_i = I_1 \cdot i^{1 - n_{or}} \qquad \text{Eq. B1}$$

where I_1 and n_{or} are fitted parameters.

According to the linear regression using least squares, the n_{or} index can be obtained from:

$$n_{or} = 1 - \frac{N\cdot\sum_{i=1}^{N}\left(\ln i \cdot \ln P^{H}{}_{i}\right) - \sum_{i=1}^{N}\left(\ln i\right)\sum_{i=1}^{N}\left(\ln P^{H}{}_{i}\right)}{N\cdot\sum_{i=1}^{N}\left(\ln i \cdot \ln i\right) - \sum_{i=1}^{N}\left(\ln i\right)\sum_{i=1}^{N}\left(\ln i\right)} \qquad \text{Eq. B2}$$

The summation of ln$i$ can be rewritten using Stirling's formula:

$$\sum_{i=1}^{N}\left(\ln i\right) \approx N\cdot\ln(N) - N + \frac{1}{2}\ln(2\pi N) \qquad \text{Eq. B3}$$

and for N>100, the divisor of the Eq. B2 can be approximated by:

$$N\cdot\sum_{i=1}^{N}\left(\ln i \cdot \ln i\right) - \sum_{i=1}^{N}\left(\ln i\right)\sum_{i=1}^{N}\left(\ln i\right) \approx N^2 \qquad \text{Eq. B4}$$

with a less than 10% error. In particular, the bias of Eq. B4 is close to ½·(ln N)^2 + ln N. Using Eq. B3 and B4, Eq. B2 can be simplified as:

$$n_{or} \approx 1 + \frac{1}{N}\sum_{i=1}^{N}\left(\alpha_N(i)\cdot\ln P^{H}{}_{i}\right) \qquad \text{Eq. B5}$$

where $\alpha_N(i) \approx \ln(N) - \ln i - 1 + 0.5/\mathrm{N} \cdot \ln(2\pi\mathrm{N})$ is a positive function, whilst ln $P^{H}{}_{i}$ has a negative result because $P^{H}{}_{i}$ is less than 1.

Since $P^{H}{}_{N-i+1} = 1 - P^{L}{}_{i}$ for i between 1 and $N - 1$, when $P^{L}{}_{i}$ is between 0 and 0.99, the logarithm can be very roughly approximated by:

$$\ln P^{H}{}_{N-i+1} = \ln\left(1 - P^{L}{}_{i}\right) \approx -\left(P^{L}{}_{i}\right) - \frac{1}{2}\left(P^{L}{}_{i}\right)^2 - \frac{1}{3}\left(P^{L}{}_{i}\right)^3 - \ldots \approx -2\cdot P^{L}{}_{i} \qquad \text{Eq. B6}$$

Therefore, Eq. B5 is very roughly approximated as:

$$n_{or} \approx 1 - \frac{2}{N}\sum_{i=1}^{N-1}\left(\ \alpha_N(N-i+1)\ \cdot\ P^{L}{}_{i}\right) \qquad \text{Eq. B7}$$

Note that the sum is between 1 and $N - 1$ because the relationship $P^{H}{}_{N-i+1} = 1 - P^{L}{}_{i}$ is only valid for this interval.

References

Alijani B, O'Brien J, Yarnal B. 2008. Spatial analysis of precipitation intensity and concentration in Iran. *Theoretical and Applied Climatology* **94**: 107–124.

Atkinson AB. 1970. On the measurement of inequality. *Journal of Economic Theory* **2**: 244–263. doi:10.1016/0022-0531(70)90039-6.

Bao Q, Yu P. 2014. The Second Decadal Leading Mode of East Asian Summer Monsoon Rainfall. *Atmospheric and Oceanic Science Letters* **7**: 417–421.

Benhamrouche A, Boucherf D, Hamadache R, Bendahmane L, Martin-Vide J, d Teixeira Nery, J. 2015. Spatial distribution of the daily precipitation concentration index in Algeria. *Natural Hazards and Earth System Science* **15**: 617–625.

Burgueño A, Martínez M, Serra C, Lana X. 2010. Statistical distributions of daily rainfall regime in Europe for the period 1951–2000. *Theoretical and Applied Climatology* **102**: 213–226, doi:10.1007/s00704-010-0251-5.

Cortesi N, Gonzalez-Hidalgo JC, Brunetti M, Martin-Vide J. 2012. Daily precipitation concentration across Europe 1971-2010. *Natural Hazards and Earth System Science* **12**: 2799-2810, doi: 10.5194/nhess-12-2799-2012.

Coscarelli R, Caloiero T. 2012. Analysis of daily and monthly rainfall concentration in Southern Italy (Calabria region). *Journal of Hydrology* **416–417**: 145–156. doi:10.1016/j.jhydrol.2011.11.047.

Cowell FA. 2000. Measurement of inequality. In A. B. Atkinson and Ourguignon F. (Eds.), *Handbook of Income Distribution,* Chapter 2, pp. 87–166. Amsterdam: North Holland.

Ding Y, Wang H, Wang B. 2005. East Asian Summer Monsoon. In: Chang C-P, Wang B, Lau N-C G (Eds.), *The Global Monsoon System: Research and Forecast*, WMO/TD No. 1266, World Meteorological Organization, pp. 95–114.

Durack PJ, Wijffels SE, Matear RJ. 2012. Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. *Science* **336**: 455–458.

Feder J, 1988: Fractals, Plenum Press, New York.

Ghahraman B, Hosseini SM. 2005. A new investigation on the performance of rainfall IDF models. *Iranian Journal of Science and Technology. Transaction B* **29**: 333–342.

Haughton J, Khandker SR. 2007. Inequality Measures. In: Haughton J, Khandker SR. (Eds). Handbook on Poverty and Inequality. The World Bank, Washington D.C.

Jolliffe IT, Hope PB. 1998. Representation of daily rainfall distributions using normalized rainfall curves. *International Journal of Climatology* **16**: 1157–1163. doi:10.1002/(SICI)1097-0088(199610)16:10<1157::aid-joc71>.

Jost L. 2006. Entropy and diversity. *Oikos* 113, 363–375. doi:10.1111/j.2006.0030-1299.14714.x

Lambert PJ, Millimet DL, Slottje D. 2003. Inequality aversion and the natural rate of subjective inequality. *Journal of Public Economics* **87**: 1061–1090.

Martin-Vide J. 1994. Geographical factors in the pluviometry of Mediterranean Spain: drought and torrential rainfall. In U.S.–Spain Workshop on Natural Hazards. Iowa Institute of Hydraulic Research, The University of Iowa; 9–25.

Martín-Vide J. 2004. Spatial distribution of a daily precipitation concentration index in peninsular Spain. *International Journal of Climatology* **24**: 959–971.

Martín-Vide J. 2013. Daily concentration of precipitation in peninsular Spain. A map of torrential rainfall risk. In García-Legaz C, Valero F (Eds). *Adverse weather in Spain, WCRP Spanish Committee and CCS,* 149-161.

Menne MJ, Durre I, Vose RS, Gleason BE, Houston TG. 2012. An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology* **29**: 897–910. Data: ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/ (last access 5 April 2015).

Moncho R, Belda F, Caselles V. 2009. Climatic study of the exponent "n" in IDF curves: application for the Iberian Peninsula. *Tethys* **6**: 3–14. doi: 10.3369/tethys.2009.6.01.

Monjo R, Pótoles J, Ribalaygua J. 2013. Detection of inhomogeneities in daily data: a test based in the Kolmogorov-Smirnov goodness-of-fit test. 9th EUMETNET Data Management Workshop, El Escorial (Madrid), 6-8 November 2013.

Monjo R. 2015. Measure of rainfall time structure using the dimensionless n-index. *Climate Research* (accepted). doi: 10.3354/cr01359.

Muenchen RA, Hilbe JM, 2010. *R for Stata Users. Statistics and Computing*. Springer, ISBN: 978-1-4419-1317-3.

Olsson J, Niemczynowicz J, Berndtsson R, Larson M. 1992. An analysis of the rainfall time structure by box counting—some practical implications. *Journal of Hydrology* **137**, 261–277.

Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J. 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**: 1923–1929.

R Development Core Team, 2010. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: www.R-project.org.

Roderick M, Sun F, Farquhar G. 2012. Water Cycle Varies over Land and Sea. *Science* **336**: 1230–1231. doi: 10.1126/science.336.6086.1230-b.

Sarricolea P, Martín-Vide J. 2014. Spatial analysis of rainfall daily trends and concentration in Chile. *Investigaciones Geográficas* (Chile) **47**: 53–66.

Shaw G, Wheeler D. 1994. Statistical Techniques in Geographical Analysis. Halsted Press: New York.

Soman MK, Kumar KK. 1990. Some aspects of daily rainfall distribution over India during the south-west monsoon season. *International Journal of Climatology* **10**: 299–311. doi: 10.1002/joc.3370100307.

Suhaila J, Jemain AA. 2012. Spatial analysis of daily rainfall intensity and concentration index in Peninsular Malaysia. *Theoretical and Applied Climatology* **108**: 235–245. doi 10.1007/s00704-011-0529-2.

Ullah A, Giles DEA. 1998. Handbook of Applied Economic Statistics, CRC Press. ISBN 0-8247-0129-1

Taouti MB, Chettih M. 2014. Fractal and multifractal analyses of the temporal structure of daily rainfall in a Mediterranean climate in northern Algeria. *Tethys* **11**, 3–12. doi: 10.3369/tethys.2014.11.01.

Zhang Q, Xu CY, Gemmer M, Chen YQ, Liu CL. 2009. Changing properties of precipitation concentration in the Pearl River basin, China, *Environmental Research and Risk Assessment* **23**: 377–385.

TABLES

Table 1. Example of real and ordered MAI for a time series of eight days with precipitation equal to 1, 3, 2, 1, 6, 4, 2, 0 in millimetres. Note that only non-zero values are used.

Time (days) i	Real MAI P_i/i	Ordered MAI $P^H{}_i/i$
1	(6)/1	(6)/1
2	(6+4)/2	(6+4)/2
3	(6+4+2)/3	(6+4+3)/3
4	(6+4+2+1)/4	(6+4+3+2)/4
5	(6+4+2+1+2)/5	(6+4+3+2+2)/5
6	(6+4+2+1+2+3)/6	(6+4+3+2+2+1)/6
7	(6+4+2+1+2+3+1)/7	(6+4+3+2+2+1+1)/7

Table 2. Pearson correlation of the indices used.

Comparing	*GI*	n_{or}	n_{δ}	*TI*	*AI*	*–GE*
GI	-	**0.98**	0.71	0.76	0.91	0.71
n_{or}	**0.98**	-	0.63	0.76	0.81	0.65
n_{δ}	0.71	0.63	-	0.55	**0.71**	0.63
TI	0.75	**0.76**	0.55	-	0.61	0.92
AI	**0.91**	0.81	0.71	0.61	-	0.64
–GE	0.71	0.65	0.63	**0.92**	0.64	-
Average	0.81	0.77	0.65	0.72	0.74	0.71

FIGURES

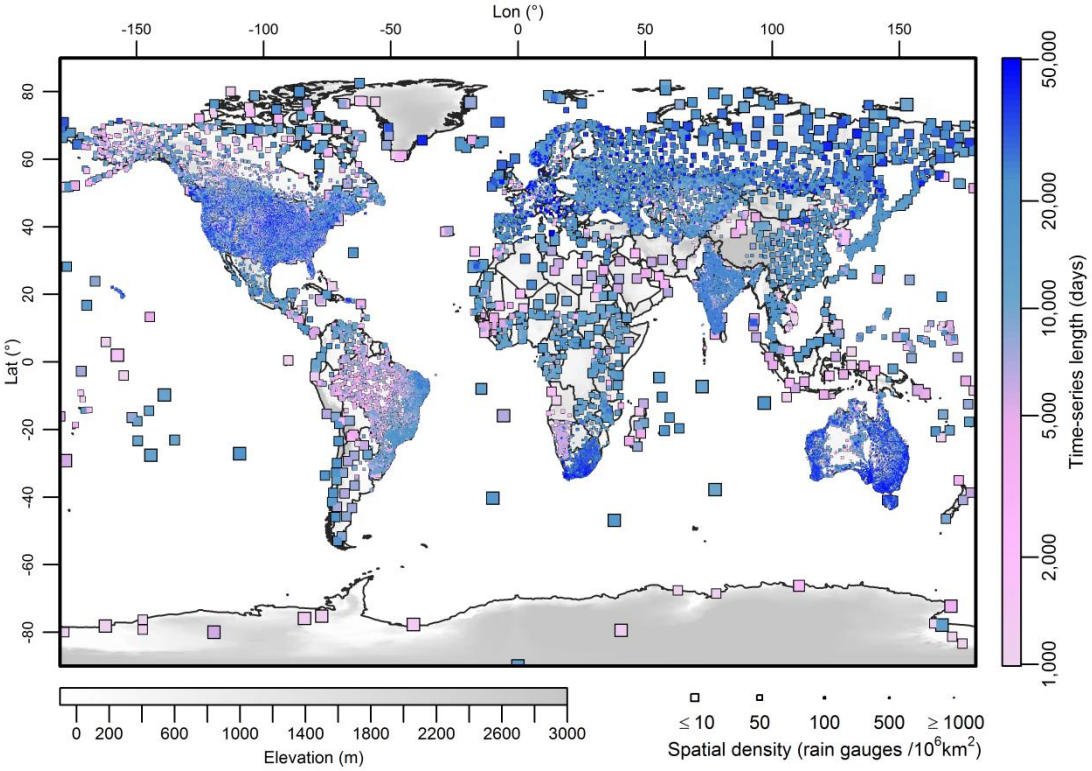

Fig. 1. Location of the rain gauges used, taken from the Global Historical Climatology Network-Daily Database. The colour indicates the duration of the time series (in days). Point size is inversely proportional to the point density in each area.

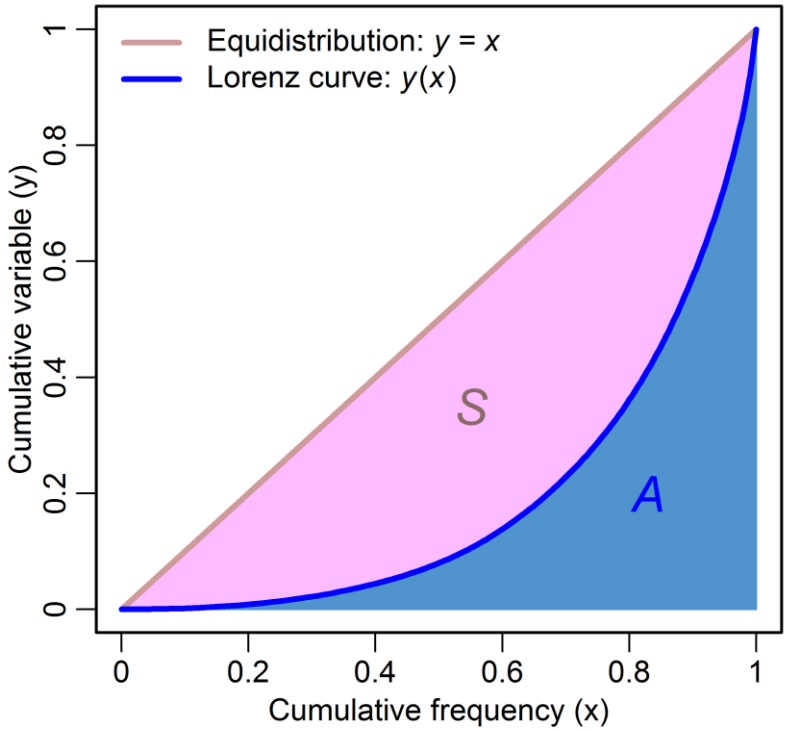

Fig. 2. Example of the Lorenz curve ($y(x)$) and its integrating areas (A and S) regarding the zero line ($y = 0$) and the equidistribution line ($y = x$). The Lorenz curve $y(x)$ is the polygonal line given by the empirical cumulative value of a variable according to its cumulative frequency.

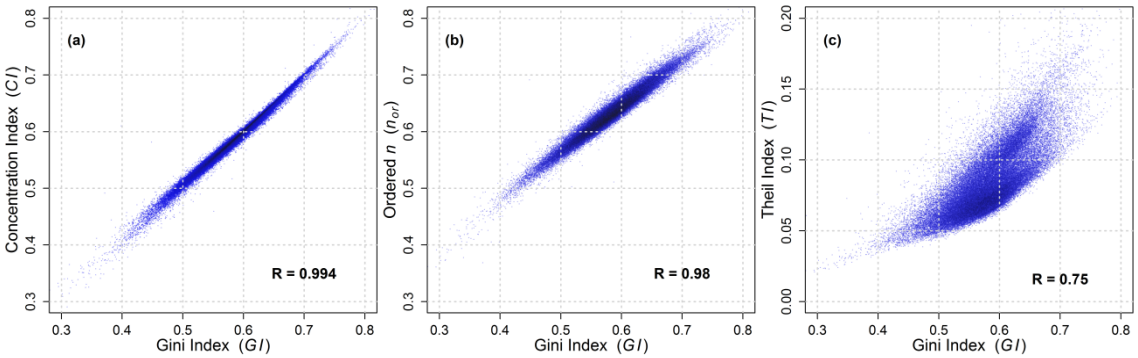

Fig. 3. Comparison between the *GI* and three indices estimated for all the time series: a) *CI*, b) n_{or} and c) *TI*.

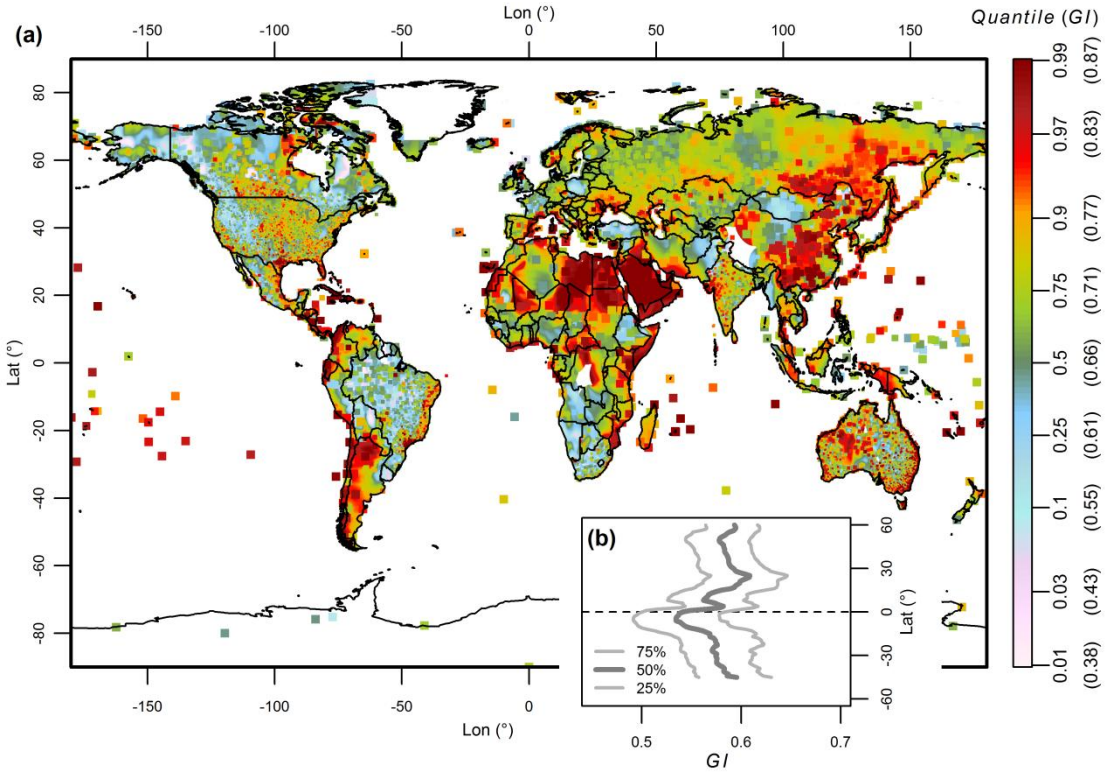

Fig. 4. a) Spatial distribution of the *GI*. b) Latitudinal profile of the *GI*, according to the centiles 25, 50 and 75.

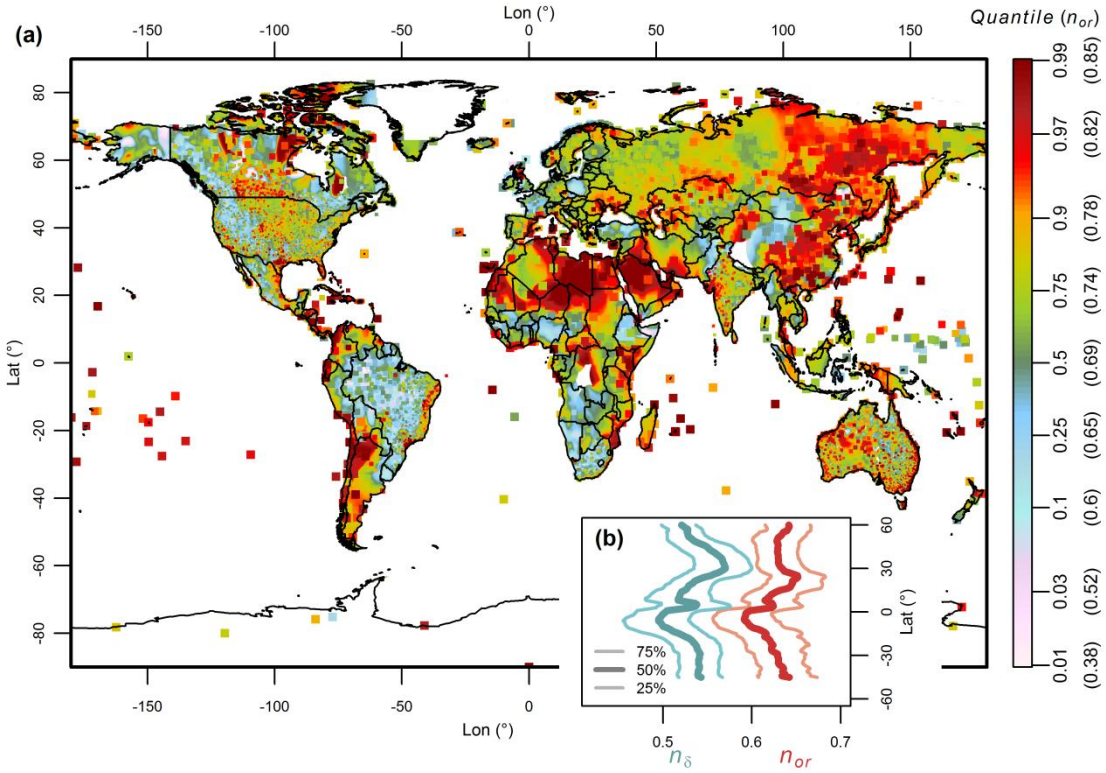

Fig. 5. a) Spatial distribution of n_{or}. b) Latitudinal profiles of n_δ (blue lines) and n_{or} (red lines), according to the centiles 25, 50 and 75.

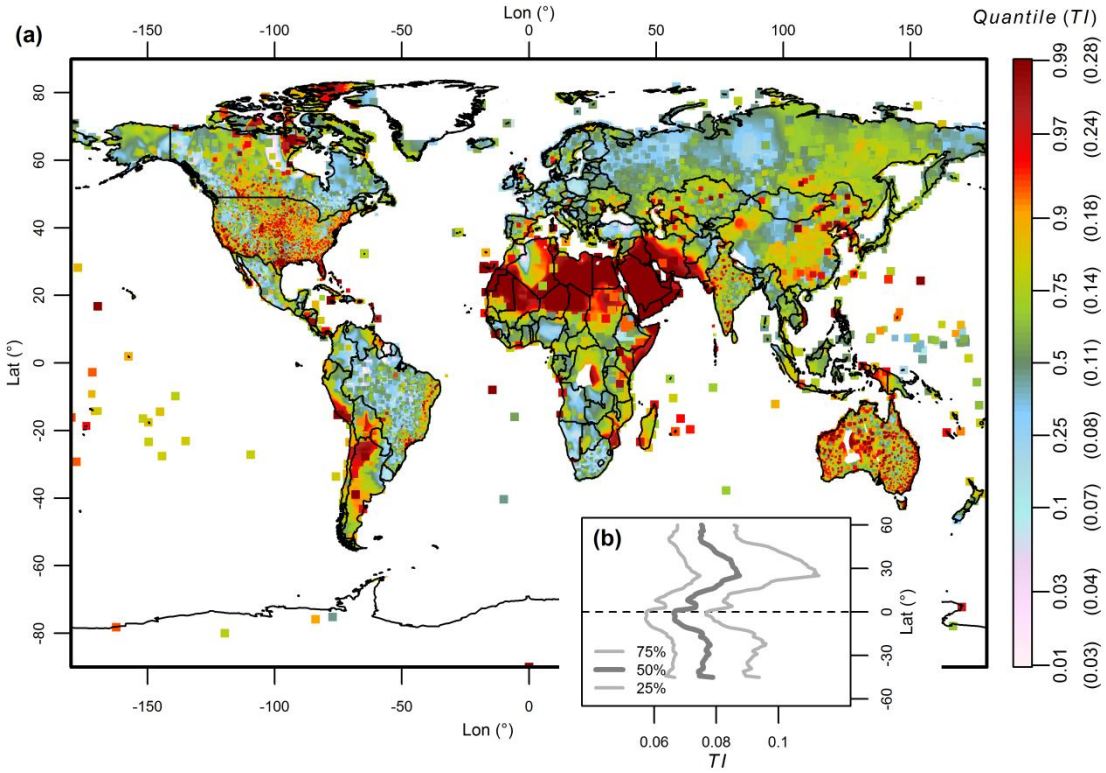

Fig. 6. a) Spatial distribution of the *TI*. b) Latitudinal profile of *TI*, according to the centiles 25, 50 and 75.

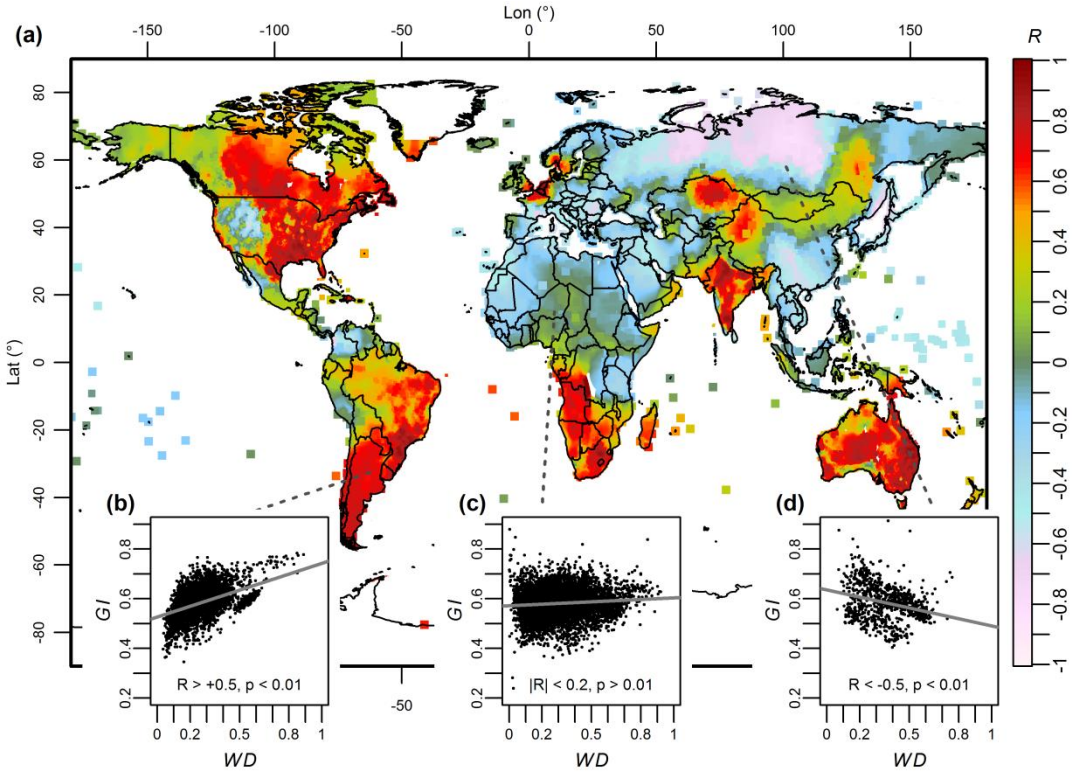

Fig. 7. a) Pearson correlation between the fraction of wet days (*WD*) and the Gini Index (*GI*) for each observatory and its 100 nearest neighbours. b) Sample with significant positive correlation $R > 0.5$. c) Sample with non-significant correlation. d) Sample with significant negative correlation $R < -0.5$.