Manuscript prepared for Hydrol. Earth Syst. Sci.
with version 5.0 of the LaTeX class copernicus.cls.
Date: 11 August 2016

Describing the inter-annual variability of precipitation with the derived distribution approach: Effects of record length and resolution

Claudio I. Meier[1], J. Sebastián Moraga[2], Geri Pranzini[2], and Peter Molnar[3]

[1]Department of Civil Engineering, University of Memphis, USA
[2]Departamento de Ingeniería Civil, Universidad de Concepción, Chile
[3]Institute of Environmental Engineering, ETH Zürich, Switzerland

Correspondence to: C. I. Meier (cimeier@memphis.edu), P. Molnar (molnar@ifu.baug.ethz.ch)

Abstract.

Inter-annual variability of precipitation is traditionally described by fitting a probability model to yearly precipitation totals. There are three potential problems with this approach: a long record (at least 25∼30 yrs) is required in order to fit the model, years with missing rainfall data cannot be used, and the data need to be homogeneous, i.e., one has to assume stationarity. To overcome some of these limitations, we test an alternative methodology proposed by Eagleson (1978), based on the derived distribution (DD) approach. It allows estimation of the probability density function (pdf) of annual rainfall without requiring long records, provided that continuously gauged precipitation data are available to derive external storm properties. The DD approach combines marginal pdfs for storm depths and inter-arrival times to obtain an analytical formulation of the distribution of annual precipitation, under the simplifying assumptions of independence between events and independence between storm depth and time to the next storm. Because it is based on information about storms and not on annual totals, the DD can make use of information from years with incomplete data; more importantly, only a few years of rainfall measurements should suffice to estimate the parameters of the marginal pdfs, at least at locations where it rains with some regularity.

For two temperate locations in different climates (Concepción, Chile, and Lugano, Switzerland), we randomly resample shortened time series to evaluate in detail the effects of record length on the DD, comparing the results with the traditional approach of fitting a Normal (or Lognormal) distribution. Then, at the same two stations, we assess the biases introduced in the DD when using daily, totalized rainfall, instead of continuously gauged data. Finally, for randomly selected periods between 3 and 15 years in length, we conduct full blind tests at 52 high-quality gauging stations in Switzerland, analyzing the ability of the DD to estimate the long-term standard deviation of annual rainfall, as compared to direct computation from the sample of annual totals.

Our results show that, as compared to the fitting of a Normal or Lognormal distribution (or equivalently, direct estimation of the sample moments), the DD approach reduces the uncertainty in annual precipitation estimates (especially inter-annual variability) when only short records (below 6∼8 years) are available. In such cases, it also reduces the bias in annual precipitation quantiles with high return periods. We demonstrate that using precipitation data aggregated every 24 h, as commonly available at most weather stations, introduces a noticeable bias in the DD. These results point to the tangible benefits of installing high-resolution (hourly, at least) precipitation gauges, next to the customary, manual rain-measuring instrument, at previously ungauged locations. We propose that the DD approach is a suitable tool for the statistical description and study of annual rainfall, not just when only short records are available, but also when dealing with non-stationary time series of precipitation. Finally, to avert any misinterpretation of the presented method, we should like to emphasize that it only applies for climatic analyses of annual precipitation totals; even though storm data are used, there is no relation to the study of extreme rainfall intensities needed for engineering design.

1 Introduction

Total annual precipitation and its variability between years are important climatic variables for water balance studies, developing regional climatologies, planning and management of water resources, and assessing water stress in general. Inter-annual variability in rainfall results from many factors such as long term multi-year atmospheric anomalies (ENSO, NAO, etc.; see, e.g., Higgins et al., 1999; Barlow et al., 2001), the strength and persistence of seasonality (e.g., Fatichi et al., 2012), and stochasticity in weather and precipitation formation. Inter-annual variation in precipitation is an important descriptor of the climatic environment which directly impacts the occurrence of droughts (e.g., Dai et al., 2004; Dai, 2011), vegetation productivity in water-limited ecosystems (e.g., Knapp and Smith, 2001; Reyer et al., 2013; Fatichi and Ivanov, 2014), as well as the distribution of rainfall extremes (e.g., Groisman et al., 2005).

A traditional statistical analysis of annual precipitation typically consists of estimating key statistics (mean, variance, skewness, etc.) and fitting a probability distribution model to the annual (or seasonal) data. According to Markovic (1965) and Linsley et al. (1982), in temperate zones this would typically be a Normal or a Lognormal distribution, fitted to a sample of at least 25$\sim$30 years of data. However, this approach is often impractical, because at many locations only a few years of precipitation data are available and many records are incomplete. With short records, the estimated statistics and parameters of the fitted probability model are highly uncertain. Moreover, natural fluctuations in climate over decadal or longer time scales, now accentuated by anthropogenic change, imply that most long climate records are not statistically homogeneous and stationary (Milly et al., 2008). This leads to the problem that while long records are required to accurately estimate the statistics and probability distribution of annual rainfall, precipitation itself might in fact be non-stationary over such long periods. Thus, an approach is needed that would allow for a better estimation of the probability distribution of annual precipitation without requiring long records.

Eagleson (1978) developed such a methodology by deriving the distribution of annual precipitation from the properties of the individual storms making up the yearly totals. Given independent storm arrivals and using prescribed models for the marginal probability distributions of storm inter-arrival times and storm depths, the probability density function (pdf) of annual precipitation can be derived analytically. Under this derived distribution (DD) approach, only a few years of continuously gauged precipitation data, from which storm arrivals and depths can be extracted and their distributions estimated, are necessary to estimate the probability distribution of annual precipitation for a site. Even though Eagleson's (1978) original paper has a large number of citations, most of these relate to ecohydrological modelling of soil moisture and vegetation dynamics (e.g., Dufrêne et al., 2005; Ivanov et al., 2008), derived distributions of runoff and flood frequency (e.g., Freeze, 1980; Díaz-Granados et al., 1984), rainfall modelling (for example, Onof et al., 1998; Willems, 2001), or morphological evolution of drainage basins (Tucker and Bras, 2000). We are not aware of any previous attempt at applying Eagleson's DD approach to the study of the inter-annual variability of precipitation, even though the method seems particularly well-suited to deal with locations with short records, as well as to account for non-stationarities introduced by a changing climate.

The main aim of this work is to investigate the performance of the DD approach for describing inter-annual variability of rainfall. We do this by comparing it with traditional procedures based on long series of annual precipitation totals. This paper specifically addresses two questions in detail: (a) what is the effect of record length on the estimates of annual precipitation (mean, deviation, and quantiles) obtained with the different methods? (b) What is the effect of rainfall temporal resolution (sampling time-step) on the results? The latter question is important for sites where only daily rainfall data are available from manually-read instruments, so that the accuracy of storm statistics for the DD is reduced.

2 Methods

2.1 Study sites and data

The DD and Normal/Lognormal probability distributions were fitted to precipitation data for two temperate locations with dissimilar climate. For the first site in Concepción, Chile, the data come from the Bellavista Research Weather Station, which was operated by Universidad de Concepción. They consist of 19 years of daily and 6 years of weekly pluviograms (paper rain charts), continuously recorded over the period 1975–1999 with a Lambrecht float-recording and siphoning rain gauge. For the second site in Lugano, Switzerland, we used a 32 year long precipitation record (1981–2012) available from MeteoSwiss (the Swiss Federal Office of Meteorology and Climatology). The full blind tests were conducted on 52 weather stations of the MeteoSwiss network, including that in Lugano. These are all 10 min precipitation depths, with 0.1 mm resolution, measured over the exact same period (1981-2012) with the same Lambrecht tipping-bucket instruments, with standardized calibration and maintenance. These rainfall data are of very good quality and have recently been used in other studies of storm properties in Switzerland (e.g., Gaál et al., 2014; Molnar et al., 2015).

2.2 Event definition

Using derived distributions first requires defining independent storms in the record, in order to obtain the parameters needed for the marginal distributions of storm depth and inter-arrival time. Although there are many different approaches for selecting a criterion for event independence (e.g., Dunkerley, 2008), we chose to discriminate successive independent storms based on a Minimum Inter-event Time, MIT (Restrepo-Posada and Eagleson, 1982; Driscoll et al., 1989; Gaál et al., 2014). With this scheme, any dry spell (i.e., between recorded precipitation) longer than the MIT defines two independent storms. Conversely, if a gap without precipitation is shorter than the MIT, then we assign both precipitation pulses to the same storm event (see Figure 1). For each independent storm, we obtained the following external storm properties from the data: storm depth H, rainfall event duration T_r, time elapsed between the end of the storm and the beginning of the next storm T_b, and time between the beginning of successive storms T_a (inter arrival time). These variables are shown in Figure 1, where independent storms are simplified into rectangular pulses in the lower panel.

2.3 Derived distribution of annual precipitation

Eagleson (1978) defined annual precipitation (P_a) as the sum of precipitation depths over the finite number of events that occur throughout a year (or wet season). P_a can thus be considered a compound variable which depends on the number ν of storm events in a given year (wet season), as well as on the storm depths H_j contributed by each storm:

$$P_a(\nu) = \sum_{j=1}^{\nu} h_j \qquad (1)$$

Both ν and H_j, are random variables with probability distributions that can be estimated on the basis of available, continuously gauged precipitation data. We are interested in obtaining the probabilistic behavior of the compound variable P_a, knowing the pdfs of the external properties of independent storm events.

In his work, Eagleson (1978) assumed that both the inter-arrival time T_a and the rainfall depth per storm H_j are identically and independently distributed (iid) variables, which are in turn mutually independent. The first assumption (“identically distributed”) means that the probabilistic behaviour of these variables is time-invariant, i.e., storms behave similarly in terms of their frequency and rainfall depth, every year, and throughout the year. Although weather disturbances are much more frequent and intense in certain seasons, both in Concepción (central south, temperate part of Chile) and in Switzerland, there are no clear limits between dry and wet seasons (rainfall events occur all year round). Thus, it should be fine to assume that P_a corresponds to an integration of the precipitation process at the yearly scale. Instead, homogeneity could also be assumed at the seasonal scale, if there were evidence for this in the data. In turn, the “independently distributed” assumption implies that the characteristics of a given storm are not affected by previous rainfall events. Finally, the assumption of mutual independence entails that storm depths are not affected by the time elapsed since the previous event, and vice-versa. Under these three assumptions, the distribution of annual rainfall is given by:

$$f_{P_a}(y) = \sum_{\nu=1}^{\infty} f_{P_a(\nu)}(y) \cdot P_\theta(\nu) \qquad (2)$$

where $f_{P_a}(y)$ is the probability density corresponding to an annual rainfall of exactly y mm; $f_{P_a(\nu)}(y)$ is the probability density corresponding to a rain depth of y mm occurring in ν storms; and $P_\theta(\nu)$ is the discrete probability mass of having exactly ν storms in a given year.

Equation 2 represents the probability density that the sum of the rainfall depths contributed by ν annual storms adds up to exactly y mm, weighted by the discrete probability of having ν storms in that year. We followed Eagleson (1978) in modelling the occurrence of storm events as a Poisson process in order to determine the discrete probability (or probability mass) of having ν storm events over a period of length t:

$$P_\theta(\nu) = \frac{(\omega t)^{\nu} \cdot e^{-\omega t}}{\nu!} \qquad \nu = 0,1,2,... \qquad (3)$$

The single parameter in this distribution ω represents the average rate of arrival or occurrence of events, whilst its inverse ω^{-1} corresponds to the average time elapsed between the beginning of two consecutive events, i.e., the mean inter-arrival time. As explained above, in our analysis t is a whole calendar year, but in semi-arid and arid climates it could represent the duration of the wet season within the year, as described in Eagleson (1978). Note that mathematically, the above choice is equivalent to fitting an Exponential distribution with parameter ω to the sample of inter-arrival times.

To obtain $f_{P_a(\nu)}(y)$ in Equation 2 it is necessary to prescribe the probability distribution of precipitation depths of the iid events. For this, Eagleson (1978) chose the Gamma distribution with two parameters λ and κ, because of its versatility and its regenerative property. The latter means that the sum of n iid Gamma(λ,κ) variables also has a Gamma distribution with parameters ($\lambda, n\kappa$) such that the mean storm depth is then $m_H = \kappa/\lambda$ and its variance $\sigma_H^2 = \kappa/\lambda^2$. The density function of total precipitation y from ν storms, $f_{P_a(\nu)}(y)$, can then be expressed as:

$$f_{P_a(\nu)}(y) = \frac{\lambda \cdot (\lambda y)^{\nu\kappa-1} \cdot e^{-\lambda y}}{\Gamma(\nu\kappa)} \qquad y > 0 \qquad (4)$$

where $\Gamma(x)$ is the gamma function.

Replacing the expressions for $P_\theta(\nu)$ and $f_{P_a(\nu)}(y)$ in Equation 2 yields the probability density function of annual precipitation as (Eagleson, 1978):

$$f_{P_a}(y) = \sum_{\nu=1}^{\infty} \frac{\lambda \cdot (\lambda y)^{\nu\kappa-1} \cdot e^{-\lambda y}}{\Gamma(\nu\kappa)} \cdot \frac{(\omega t)^{\nu} \cdot e^{-\omega t}}{\nu!} \qquad y > 0$$

(5)

Integrating Equation 5 results in the cumulative distribution function (cdf) for annual precipitation (Eagleson, 1978):

$$F_{P_a}(y) = e^{-\omega t} + \sum_{\nu=1}^{\infty} \frac{(\omega t)^{\nu} \cdot e^{-\omega t}}{\nu!} P[\nu\kappa, \lambda y] \qquad y \geq 0 \quad (6)$$

where $P[\nu\kappa, \lambda y]$ is Pearson's incomplete Gamma function.

In concluding this section, we should briefly discuss the choice of the models for the marginal distributions. Having rainfall data at a given location for a specific period, one could certainly use the pdfs of best fit, instead of prescribing the Exponential and Gamma distributions for the storm inter-arrival times and depths, respectively. In such case, though, it would be highly improbable that a closed-form solution for $f_{P_a}(y)$ could be found, so that numerical methods would be needed. As we are primarily interested here in testing through comparisons the general ability of the DD approach for describing the inter-annual variability of precipitation, we stick to Eagleson's (1978) formulation, as described above.

2.4 Performance of the DDA

If long records are available, e.g., at least 25 $\sim$ 30 yrs, then key statistics as well as the distribution of annual precipitation totals can be estimated from such data. For example, in temperate, humid areas, annual precipitation typically follows a Normal or a Lognormal distribution (e.g., Linsley et al., 1982; Markovic, 1965), which can be fitted to the yearly totals. On the other hand, the DD approach summarized above allows for the description of annual precipitation based only on storm statistics (with parameters representing the mean number of storms in a year, as well as the mean and variance of event depth). At locations with sufficient storms, such statistics can be adequately estimated from much shorter records, provided these have a temporal resolution that is detailed enough to accurately describe event properties. We assess the performance of the DD method vs. traditional model fitting to annual precipitation data with two important effects in mind.

First is the effect of record length, i.e., the uncertainty and bias which come from using short precipitation records both for the traditional and derived distribution approaches. We address this issue using two different methods:

(i) After testing for one-year lag independence, we randomly subsample shorter records from the original series in Concepción and Lugano, without replacement, to which we apply both the DD and the traditional fitting of a distribution. Instead of using independent continuous records (as done by Pranzini, 2000; Pranzini and Meier, 2001), which would yield only a handful of subsamples, the analyses are carried out after assembling 200 n-year-long resampled records, where n is the number of different, randomly picked years (n = 3, 5, 7, 10, 15 years). For example, one of the 200 5-yr-long resampled datasets for Concepción was assembled with the rainfall data for the years 1988-1991-1977-1981-1994. At each of the two locations we thus consider 200 shorter, resampled records for each one of the five record durations n, on top of the full (25 or 32 year-long) original time-series. Next, for each one of the 1001 different records at each site, we identify all independent storm events in the series, extract their inter-arrival times and total precipitation depths, and then fit the respective Exponential and Gamma distributions, thus obtaining the DD of annual precipitation. We also fit Normal and Lognormal pdfs to each record, and then compare various statistics (mean, standard deviation, and skewness) and quantiles to assess the performance of the proposed DD versus the "traditional" methodology. This procedure allows us to generate enough subsamples to draw statistically significant conclusions. It is important to note though, that the resampling destroys any long-range dependency that could be present in the original record. In this method, replication (200 x) is achieved through resampling.

(ii) In order to provide a more realistic setting but still allow for statistical comparisons, we also conduct full blind tests of the DD approach, at 52 different locations in Switzerland, each with the same 32 years of high-quality rainfall data. In these, we analyze the ability of the DD to estimate the long-term (32-yr) standard deviation of annual rainfall, as compared to direct computation from the sample of annual totals, when only a shorter, continuous record is available. The tests are done for all possible short record durations N between 3 and 15 years (i.e., N = 3, 4, 5, … , 14, 15 years). For each value of N, at each one of the 52 stations, we randomly choose a N-yr long, continuous record; for example, for N = 5 years at Genève-Cointrin, we randomly selected the record between January 1st, 1996, and December 31st, 2000. We next fit the DD to this short record and compute the standard deviations as obtained both from the DD and from the sample of size 5 years, and then compare them with the long-term (1981-2012) deviation at the station by computing their relative errors. In this second method, replication (52 x) is achieved by considering a large number of stations.

The second focus of our work is the effect of data resolution on the DD. At many locations precipitation is observed only once a day, so that higher resolution, continuously gauged records are not available. Thus, it is interesting to test how applicable the DD approach is when using such low-resolution, daily data. To this end we aggregate the continuously gauged data for Concepción and Lugano every 24 h (between 08:00 LT in a given day and 08:00 LT next day, as is commonly done in meteorological practice). When decreasing the data resolution, the MIT is accordingly changed to 1 day, at both locations, to accommodate the minimum identifiable dry spell under the new scenario. In this case we are interested in the differences within the DD method, between using continuously-gauged rainfall data and daily data, considering the same 200 n-year-long (n = 3, 5, 7, 10, 15 years) shortened records that were previously assembled.

3 Results

3.1 Event properties

Storm events in Concepción are dominated by mid-latitude extra-tropical cyclones, which produce fronts resulting in low to mid-intensity rainfall events that occur throughout the year, but with higher frequency and magnitude during the winter months (Falvey and Garreaud, 2007). When using a 12 h MIT in order to discriminate independent storms, we obtain a total of 1350 rainfall events over the 1975–1999 period. This count neglects storms with a total depth below 1 mm, because of the difficulties involved in extracting the properties for such small events from paper pluviograms. These discarded trace events amount on average for 4.1 mm per year, whereas mean annual rainfall in Concepción is around 1200 mm, so we neglect this source of bias. In Concepción, the Gamma and Exponential distributions fit the data adequately, as visually shown in Fig. 2. In the case of the storm depths, we fit a Gamma with parameters $\lambda = 0.02784$ mm^{-1}, and $\kappa = 0.6157$ (χ^2 = 58.5, df = 10, p = $7 * 10^{-9}$). For the inter-arrival times, we choose an Exponential distribution with parameter $\omega = 0.006261$ h^{-1} (χ^2 = 239.7, df = 10, p = 0). Both marginal distributions are shown in Figure 2. Even though both distributions would be rejected at the 0.05 significance level, we still use them because of the reasons explained in Section 2.3.

In Lugano, storms are dominated by local-scale convective systems in the summer and fall, which produce mid to high intensity showers during a few hours, accentuated by orographic effects (Schiesser et al., 1995; Panziera et al., 2014). Consequently, a 4 h MIT was chosen, resulting in 1794 independent storms over the 1981–2012 period. Storm depths are fitted with a Gamma distribution with parameters $\lambda = 0.0255$ mm^{-1}, and $\kappa = 0.3281$ (χ^2 = 57.2, df = 11, p = $3 * 10^{-8}$), whilst the inter-arrival times are fitted with an exponential distribution with parameter $\omega = 0.0136$ h^{-1} (χ^2 = 541.5, df =11, p = 0), as shown in Figure 3.

3.2 Effect of record length

3.2.1 Resampling tests

At Concepción and Lugano, for the complete records, we obtained the cumulative distribution functions of annual precipitation using the DD, based on the marginal distributions for storm counts and depths, and compared them with fitted Normal and Lognormal distributions (Figure 4). At both locations, we find that there are no significant visual differences between the three distributions, at least in the range where the bulk of the data lies. Goodness-of-fit tests indicate similar performance for the Normal and Lognormal distributions, at both locations. In Concepción (n = 25), the Normal has a K-S statistic of 0.0846 (p = 0.987), and a (χ^2 statistic of 0.465 (p = 0.793, df = 2), while the Lognormal has K-S = 0.0978 (p = 0.952) and (χ^2 = 0.0829 (p = 0.959, df = 2). In Lugano, the Normal has K-S = 0.0961 (p = 0.902) and (χ^2 = 0.194 (p = 0.979, df = 3), while the Lognormal has K-S = 0.0819 (p = 0.971) and (χ^2 = 0.824 (p = 0.975, df = 5). Because of these results, and considering that our focus is on the relative improvement that can be had by using the DD approach, we omit the Lognormal results from the rest of this paper. It should be noted that it is basically impossible to reject any specific model for the small samples we are interested in.

The sampling of shorter records was conducted on an annual basis in order to maintain seasonal coherence within a given year, i.e., years were selected at random and all storms within those years were sampled. As this sampling destroys any correlation in precipitation between years, if it exists, we verified the lack of temporal correlation in annual precipitation using Kendall's τ statistic (Ferguson et al., 2000) for lag-1 autocorrelation. In Concepción, Kendall's $\tau = -0.1522$ (p-value = 0.3128) for $n = 25$ years. For Lugano Kendall's $\tau = 0.0151$ (p-value = 0.9195) for $n = 32$ years. Thus, the null hypothesis that there is no lag-1 autocorrelation is not rejected in either case. This gives some support to the decision of resampling shorter records based on whole years.

The consequences of reducing record lengths to $n = 5$ and $n = 10$ years are shown in Figures 5 and 6 for Concepción, and Figures 7 and 8 for Lugano, respectively. At both locations the results show a clear increase in dispersion of the pdfs of annual precipitation with shorter records, but the variability is significantly lower when using the DD. The results for other subsample sizes ($n = 3, 7, 15$) show the same tendency. The reproduction of the standard deviation of annual precipitation in Figure 9 clearly shows that the DD approach reduces the uncertainty in this estimate of inter-annual variability for shorter record lengths.

In hydrological practice, we commonly fit distributions in order to estimate quantiles. These results are presented in Tables A1-A4, where we show the values of annual precipitation for different return periods, as computed with both methods for all record lengths, using continuously-recorded data, both in Concepción and Lugano. The tables show the mean, standard deviation and skewness, as well as 10 selected quantiles, obtained both with the DD and by fitting a Normal pdf. These statistics are computed for both the complete records and the resampled, shorter records, in which case the mean and standard deviation of 200 samples are presented.

The uncertainty in estimating the variability of annual precipitation from short records is very large if only annual totals are used. Based on the resampling tests, it would seem that the DD for the same short records significantly reduces this uncertainty. Still, one should be careful with the results obtained from these resampled records, as they hinge on the applicability of the model's assumptions.

3.2.2 Full blind tests

These were conducted in order to test the ability of the DD approach to estimate the long-term inter-annual variability of precipitation based on short records, without the need for any assumptions. For each record length (3 to 15 years), at each one of the 52 locations, we computed the percentual relative errors between the 'true' long-term (32 yr) standard deviation of annual rainfall, and two different estimations from the short record: (i) using the DD, and (ii) direct computation from the (small-size) sample of annual totals. Figure 10 shows the mean (± standard deviation) and median (N=52) decrease in percentual relative error that is achieved when using the DD (positive values indicate that the DD performs better). Smoothing our results, we conclude that for these 52 Swiss stations, both in the mean and the median, the DD improves estimation of the long-term variability of annual rainfall when record length is shorter than about $6 \sim 8$ years, with a marked gain when $N \leq 4$ yrs. When records longer than 8 yrs are available, there is little difference between both estimation methods.

3.3 Effect of data resolution

Most weather stations world-wide are not equipped with continuously recording rain gauges; in such cases precipitation is usually measured only once per day. In order to test the suitability of daily rainfall data for the DD method, we totalized our high resolution data over 24 h long periods and then applied the DD approach to these daily data. Figure 11 shows that when we use daily instead of continuously recorded precipitation data we obtain similar values for the central tendency indicators, but the extremes of the distribution show important biases.

This occurs as well, at both locations, when combining the use of aggregated data and shorter record lengths, as shown in Figures 12 and 13. The spread of annual precipitation increases both when reducing the record length as well as when data are aggregated, which may lead to an important bias, especially in wet years. This result points to the benefits of installing high-resolution gauges to derive accurate storm statistics as well as to the limited value of short daily records for deriving reliable precipitation statistics at the annual scale, using the DD approach.

4 Discussion and conclusions

The distributions of annual precipitation obtained with Eagleson's (1978) derived distribution approach are very similar to the fitting of a Normal or Lognormal distribution when using the complete, continuously gauged record (Figure 4). Thus, whenever a long and homogeneous (stationary) rainfall record is available, the traditional approach of fitting a Normal or Lognormal distribution should be adequate. On the other hand, the amount of information used in the DD approach is much greater, because it explicitly includes statistics from the many storm events that occur within each year with data, instead of considering only annual sums.

More importantly, the DD method still yields good results when attempting to estimate the annual rainfall distribution with shorter records, as long as these consist of continuously gauged data. Shorter records yield larger variability in annual precipitation, but using the DD, measures of both the central tendency and dispersion are still more consistent with those estimated using all of the information available. This is because even if, say only 3 years of data are available, there is still a sufficiently large number of rainfall events (e.g., an average of 54 yearly storms for Concepción and 119 for Lugano) to allow for an adequate probabilistic description of their external characteristics.

On the other hand, when one attempts to fit a distribution (Figures 5-8), or estimate sample moments (Figures 9 and 10) with only a few annual rainfall totals, there is a large uncertainty in the estimates. Furthermore, years with incomplete records may still be used with the DD method in order to extract storm properties and estimate model parameters, while fitting a distribution to annual totals requires years with complete data. However, our results also show that a bias is introduced when only daily rainfall records are available (Figures 11-13), so that the use of low-resolution rainfall data cannot be recommended for the DD method.

Overall, our results show that Eagleson's derived distribution approach is a better way of estimating the probability distribution of annual precipitation, when only a short, high-resolution record is available, because the uncertainty in estimates is reduced. The importance of these results lies not only in the possibility of estimating annual rainfall and its variability when only short records are available. When there is suspicion of non-stationarity in a rainfall record, the DD method should be useful for describing the long-term behaviour of annual precipitation (even if a long series is available) by breaking the longer record into shorter series over which it is more tenable to assume stationarity. In turn, one could also think of using the DD approach as part of a test for homogeneity of rainfall records, under the basic assumption that if annual rainfall is showing trends, these should be reflected in event frequency and in the distribution of storm depths.

An important conclusion of this work is that installing high-resolution (hourly or less) precipitation gauges in previously ungauged locations next to the customary, manual rain-measuring instrument, even for short periods, has tangible benefits in the estimation of long-term precipitation statistics, such as inter-annual variability and quantiles of annual precipitation with high return periods. This is important because accurate gauge-level precipitation estimates remain vital for the correction of remotely sensed data and in merging different precipitation data types, e.g., weather radar, satellite, etc. (e.g., Xie and Arkin, 1996), as well as for the spatial interpolation of precipitation, especially in areas with complex topography (e.g., Masson and Frei, 2014).

Acknowledgements. The ETH Global R4D Seed Money Grants scheme is acknowledged for providing financial support for this research. The publication of this manuscript has been funded through ETH Zürich membership in the open-access Copernicus Publishers. Conicyt (Chile) funded C. Meier and J. Moraga through project Fondef CA13I10190. We thank the Swiss Federal Office of Meteorology and Climatology, MeteoSwiss, for providing the rainfall data for Switzerland, and Freddy Echeverría and Alberto Foppiano, Departamento de Física de la Atmósfera y el Océano, Universidad de Concepción, for those for Concepción. Comments from Günter Blöschl and an anonymous reviewer much improved our original manuscript.

References

Barlow, M., Nigam, S., and Berbery, E.H.: ENSO, Pacific decadal variability, and US summertime precipitation, drought, and streamflow, J. Climate, 14(9), 2105–2128, 2001.

Benjamin, J.R., and Cornell, C.A.: Probability, Statistics, and Decision for Civil Engineers, McGraw-Hill, New York, 1970.

Dai, A. G.: Drought under global warming: a review, Wiley Interdisc. Rev. Climate Change, 2(1), 45–65, 2011.

Dai, A. G., Trenberth, K.E., and Qian, T.T.,: A global dataset of Palmer Drought Severity Index for 1870-2002: Relationship with soil moisture and effects of surface warming, J. Hydrometeorol., 5(6), 1117–1130, doi:10.1175/JHM-386.1, 2004.

Díaz-Granados, M.A., Valdés, J.B., Bras, R.L.: A physically based flood frequency distribution. Water Resources Research, 20 (7), pp. 995-1002, 1984.

Driscoll, E. D., Palhegyi, G. E., Strecker, E. W., and Schelley, P. E.: Analysis of storm event characteristics for selected rainfall gages throughout the United States, U.S. Environmental Protection Agency, Washington D.C., 1989.

Dufrêne, E., Davi, H., François, C., Le Maire, G., Le Dantec, V., Granier, A.: Modelling carbon and water cycles in a beech forest. Part I: Model description and uncertainty analysis on modelled NEE . Ecological Modelling, 185 (2-4), pp. 407-436, 2005.

Dunkerley, D.: Identifying individual rain events from pluviograph records: a review with analysis of data from an Australian dryland site, Hydrol. Process., 22(26), 5024–5036, 2008.

Eagleson, P.S.: Climate, soil, and vegetation. 2. The distribution of annual precipitation derived from observed storm sequences, Water. Resour. Res., 14(5), 713–721, 1978.

Falvey, M., and Garreaud, R.: Wintertime precipitation episodes in central Chile: Associated meteorological conditions and orographic influences, J. Hydrometeorol., 8, 171–193, doi:10.1175/jhm562.1, 2007.

Fatichi, S., and Ivanov, V.: Interannual variability of evapotranspiration and vegetation productivity, Water Resour. Res., 50, 3275–3294, doi:10.1002/2013WR015044, 2014.

Fatichi, S., Ivanov, V. Y., and Caporali, E.: Investigating interannual variability of precipitation at the global scale: Is there a connection with seasonality? J. Climate, 25, 5512–5523, 2012.

Ferguson, T. S., Genest, C. and Hallin, M.: Kendall's tau for serial dependence, Can. J. Stat., 28, 587-–604, 2000.

Freeze, R.A.: A stochastic-conceptual analysis of rainfall-runoff processes on a hillslope. Water Resources Research, 16 (2), pp. 391-408, 1980.

Gaál, L., Molnar, P., and Szolgay, J.: Selection of intense rainfall events based on intensity thresholds and lightning data in Switzerland, Hydrol. Earth Syst. Sci., 18, 1561—1573, doi:10.5194/hess-18-1561-2014, 2014.

Groisman, P. Y., Knight, R. W., Easterling, D. R., Karl, T. R., Hegerl, G. C., and Razuvavev, V. A. N.: Trends in intense precipitation in the climate record, J. Climate, 18(9), 1326–1350, 2005.

Higgins, R. W., Chen, Y., and Douglas, A. V.: Interannual variability of the North American warm season precipitation regime, J. Climate, 12(3), 653–680, 1999.

Ivanov, V.Y., Bras, R.L., Vivoni, E.R.: Vegetation-hydrology dynamics in complex terrain of semiarid areas: 1. A mechanistic approach to modeling dynamic feedbacks. Water Resources Research, 44 (3), art. no. W03429, 2008.

Knapp, A. K., and Smith, M. D.: Variation among biomes in temporal dynamics of aboveground primary production, Science, 291(5503), 481–484, 2001.

Linsley, R.K., Kohler, M.A and Paulhus, J.L.H.: Hydrology for Engineers, 3rd Edition, McGraw-Hill, New York, 1982.

Markovic, R. D.: Probability functions of best fit to distributions of annual precipitation and runoff, Hydrology Paper No 8, Colorado State University, Fort Collins, Colorado, 1965.

Masson, D and C. Frei: Spatial analysis of precipitation in a high-mountain region: exploring methods with multi-scale topographic predictors and circulation types, Hydrol. Earth Syst. Sci., 18(11), 4543–4563, doi:10.5194/hess-18-4543-2014, 2014.

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier and D. P., Stouffer, R. J.: Stationarity is dead: Whither water management?, Science, 319, 573—574, 2008.

Molnar, P., Fatichi, S., Gaál, L., Szolgay, J., and Burlando, P.: Storm type effects on super Clausius–Clapeyron scaling of intense rainstorm properties with air temperature, Hydrol. Earth Syst. Sci., 19, 1753–1766, doi:10.5194/hess-19-1753-2015, 2015.

Onof, C., Mackay, N.G., Oh, L., Wheater, H.S.: An improved rainfall disaggregation technique for GCMs. Journal of Geophysical Research Atmospheres, 103 (D16), art. no. 98JD01147, pp. 19577-19586, 1998.

Panziera, L., James, C. N. and Germann U.: Mesoscale organization and structure of orographic precipitation producing flash floods in the Lago Maggiore region, Q. J. Roy. Meteor. Soc., 141, 224-–248, doi: 10.1002/qj.2351, 2014.

Pranzini, G.: Annual rainfall distribution in Concepción as derived from the characteristics of the storm sequence, Thesis presented for the profesional degree of Civil Engineer, Depto. de Ingeniería Civil, Universidad de Concepción, Chile, 2000 (in Spanish).

Pranzini, G. and C. I. Meier: Annual rainfall distribution in Concepción as derived from the characteristics of the storm sequence, in Proceedings of the XV Chilean Conference of Hydraulic Engineering, Concepción, 7-9 November 2001, 27–38, 2001 (in Spanish).

Restrepo-Posada, P. J. and Eagleson, P. S.: Identification of independent rainstorms, J. Hydrol., 55, 303-–319, doi:10.1016/0022-1694(82)90136-6, 1982.

Reyer, C.P.O., Leuzinger, S., Rammig, A., Wolf, A. Bartholomeus, R.P., Bonfante, A., de Lorenzi, F., Dury, M., Gloning, P., Abou Jaoude, R., Klein, T., Kuster, T.M., Martins, M., Niedrist, G., Riccardi, M., Wohlfahrt, G., de Angelis, P., de Dato, G., Francois, L., Menzel, A., and Pereira, M.: A plant's perspective of extremes: terrestrial plant responses to changing climatic variability, Global Change Biology, 19(1) 75–89, doi:10.1111/gcb.12023, 2013.

Schiesser, H. H., Houze, R. A. and Huntrieser, H.: The mesoscale structure of severe precipitation systems in Switzerland, Mon. Weather Rev., 123(7), 2070–2097, 1995.

Tucker, G.E., Bras, R.L.: A stochastic approach to modeling the role of rainfall variability in drainage basin evolution. Water Resources Research, 36 (7), pp. 1953-1964, 2000.

Willems, P.: A spatial rainfall generator for small spatial scales. Journal of Hydrology, 252 (1-4), pp. 126-144, 2001.

Xie, P. P., and Arkin, P. A.: Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, J. Climate, 9(4), 840–858, 1996.

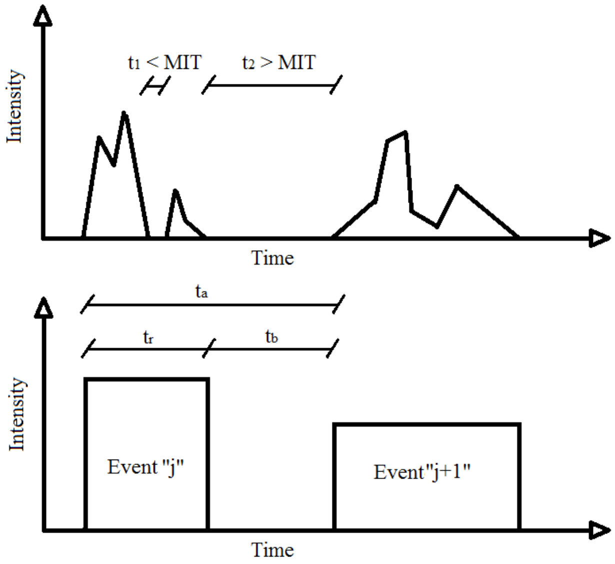

Fig. 1. External independent-storm properties obtained from the rain gauge records.

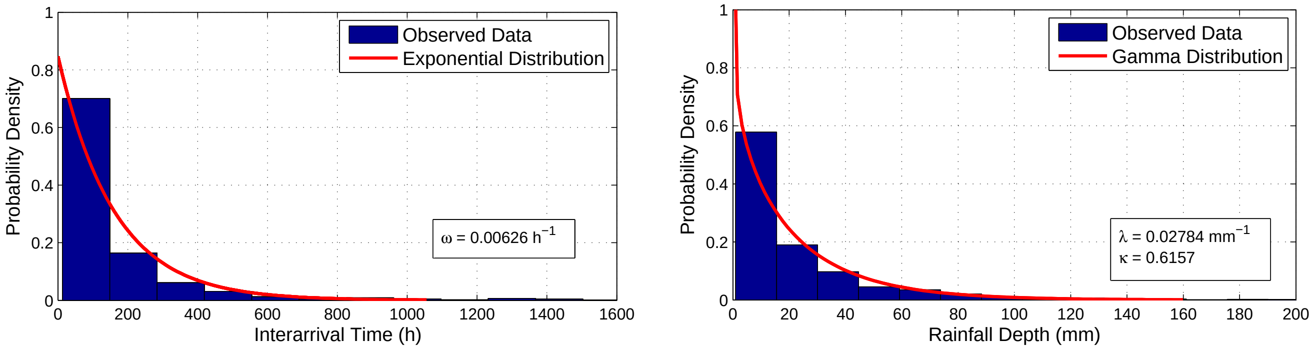

Fig. 2. Fitted distributions to interarrival times (left) and rainfall depths (right) for 25 years of data in Concepción.

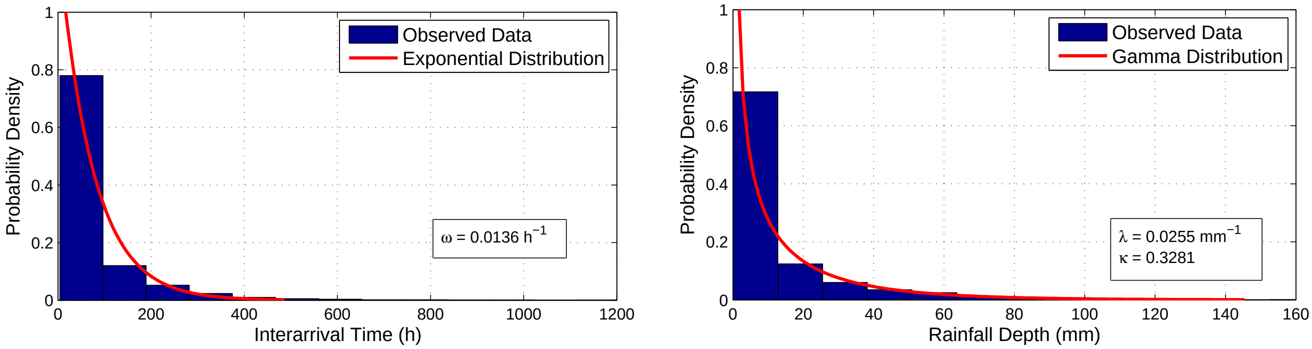

Fig. 3. Fitted distributions to interarrival times (left) and rainfall depths (right) for 32 years of data in Lugano.

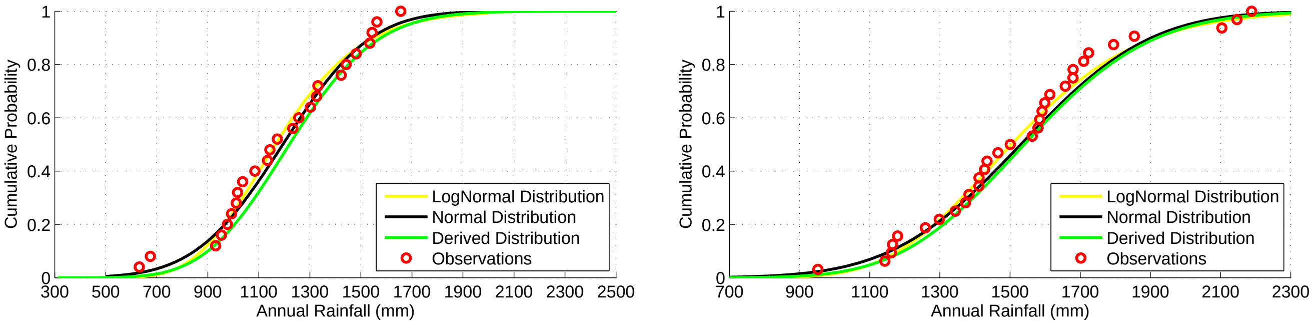

Fig. 4. Cumulative distributions derived with the DDA and fitted as a Normal and Lognormal distribution to annual precipitation totals for Concepción, Chile (left) and Lugano, Switzerland (right).

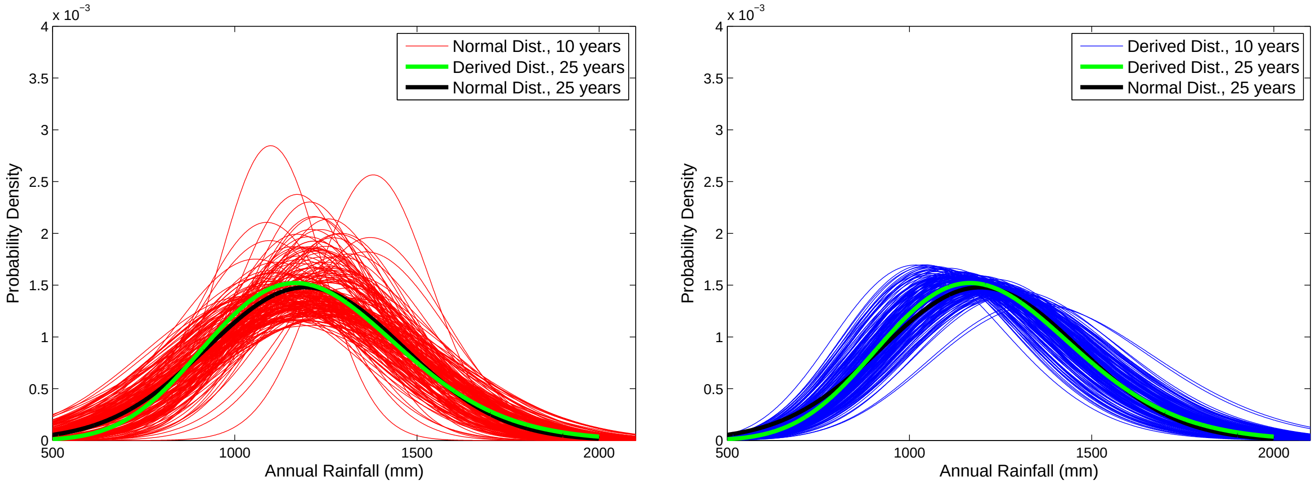

Fig. 5. Estimated PDFs for 10-year subsamples using Normal Distributions (red, left) and Derived Distributions (blue, right) in Concepción.

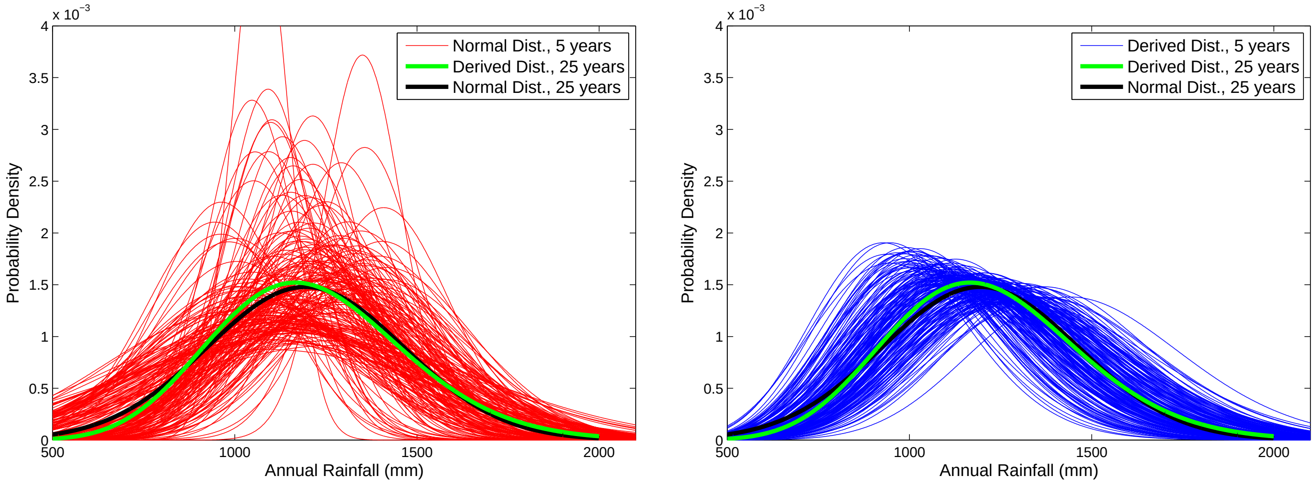

Fig. 6. Estimated PDFs for 5-year subsamples using Normal Distributions (red, left) and Derived Distributions (blue, right) in Concepción.

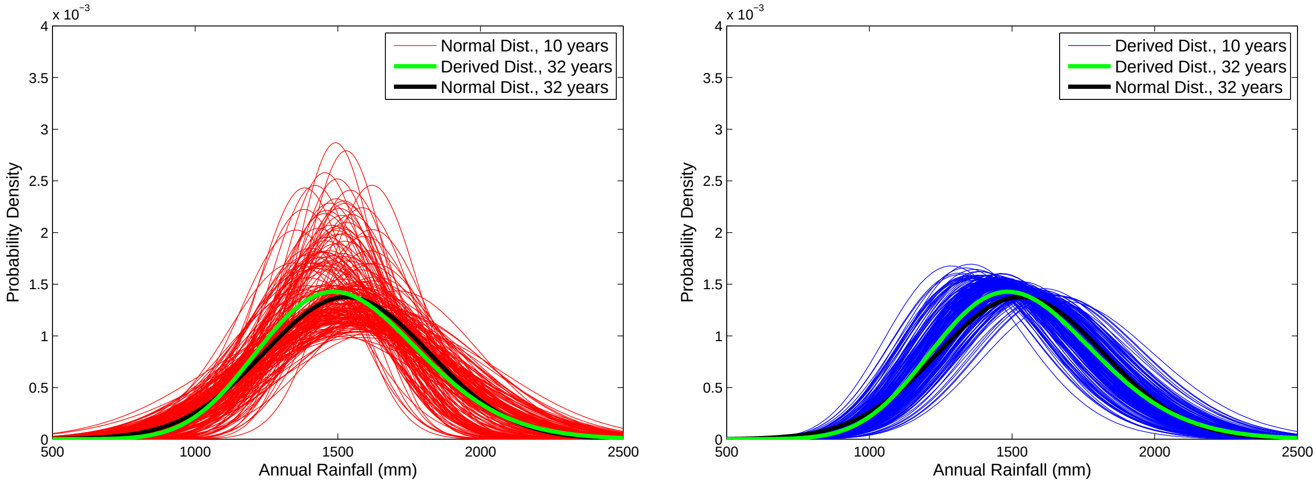

Fig. 7. Estimated PDFs for 10-year subsamples using Normal Distributions (red, left) and Derived Distributions (blue, right) in Lugano.

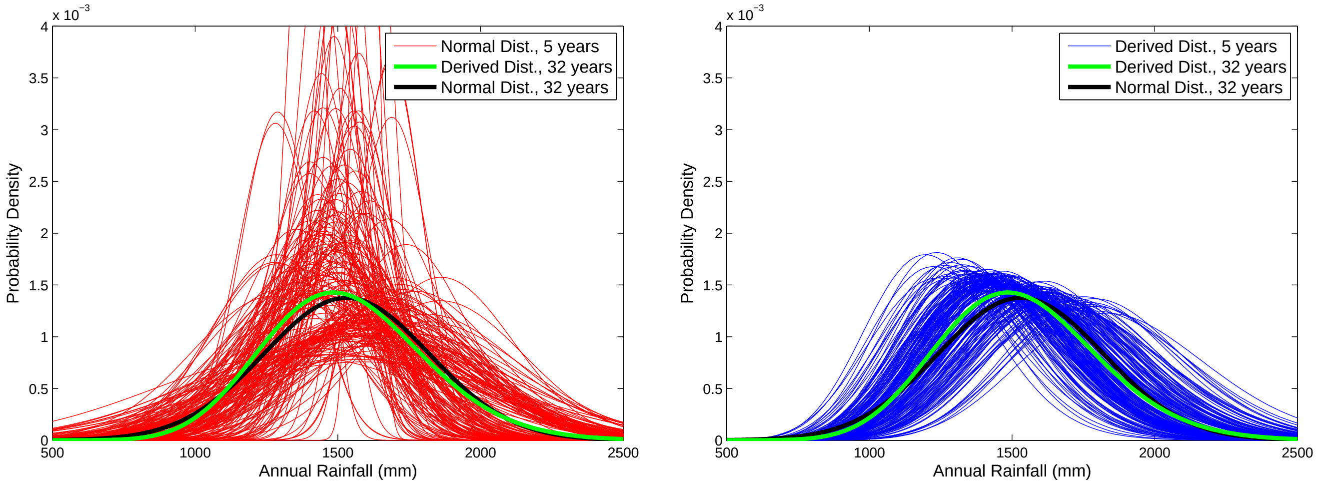

Fig. 8. Estimated PDFs for 5-year subsamples using Normal Distributions (red, left) and Derived Distributions (blue, right) in Lugano.

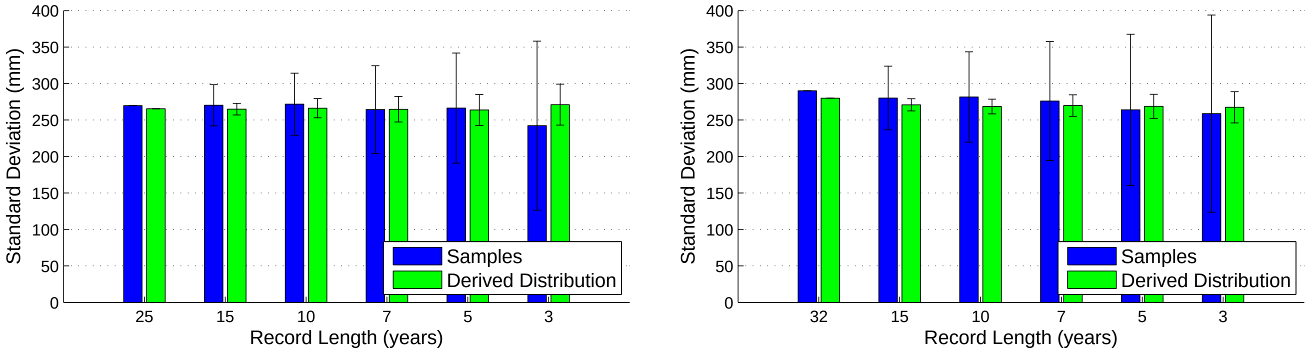

Fig. 9. Sample standard deviations of annual precipitation computed from yearly totals (in blue), compared to the corresponding population standard deviations estimated with the DDA (in green). For record lengths ⩽ 15 yr, the whiskers show the range ± 1 std from resampling (n=200). Concepción is on the left and Lugano on the right.

Table A1. Quantiles of the distribution of annual rainfall in Concepción, as obtained with different record lengths resampled 200 times and continuous data using derived distributions (in mm).

Sample Size		Mean	STD	Skew.	QUANTILES									
					Q .01	Q .02	Q .05	Q .10	Q .20	Q .80	Q .90	Q .95	Q .98	Q .99
25 Years		1213.6	265.4	0.35	670	727	817	900	1006	1451	1579	1688	1814	1900
15 Years	m	1209.8	264.9	0.35	667	725	814	897	1002	1447	1575	1683	1809	1895
	s	39.5	8.0	0.01	27.1	28.4	30.3	32.1	34.5	45.3	48.5	51.4	54.7	57.0
10 Years	m	1217.3	266.2	0.35	672	730	820	903	1009	1456	1584	1693	1820	1906
	s	60.0	13.2	0.01	38.8	40.9	44.2	47.3	51.5	70.2	75.9	80.8	86.6	90.5
7 Years	m	1214.0	264.8	0.35	672	729	818	902	1007	1451	1579	1687	1813	1899
	s	82.1	17.6	0.02	53.9	56.6	61.0	65.2	70.7	95.5	103.0	109.5	117.1	122.4
5 Years	m	1205.0	263.8	0.35	665	722	811	894	998	1442	1569	1677	1802	1888
	s	99.1	21.2	0.02	66.4	69.6	74.6	79.5	85.9	114.9	123.8	131.4	140.6	146.9
3 Years	m	1242.2	271.1	0.35	687	746	837	922	1030	1485	1616	1727	1856	1944
	s	140.7	28.1	0.04	98.4	102.5	109.2	115.4	123.6	160.6	171.8	181.6	193.2	201.2

Table A2. Quantiles of the distribution of annual rainfall in Concepción, as obtained with different record lengths resampled 200 times fitting Normal distributions (in mm).

Sample Size		Mean	STD	QUANTILES									
				Q .01	Q .02	Q .05	Q .10	Q .20	Q .80	Q .90	Q .95	Q .98	Q .99
25 Years (DDA)		1213.6	265.4	670	727	817	900	1006	1451	1579	1688	1814	1900
25 Years (Normal)		1194.1	269.8	566	640	750	848	967	1421	1540	1638	1748	1822
15 Years	m	1189.8	270.2	561	635	745	843	962	1417	1536	1634	1745	1818
	s	42.3	28.2	80.9	74.5	65.3	57.8	50.1	46.8	53.4	60.3	69.0	75.2
10 Years	m	1197.9	271.7	566	640	751	850	969	1427	1546	1645	1756	1830
	s	63.8	42.6	123.7	113.9	99.9	88.4	76.6	69.7	79.3	89.5	102.5	111.9
7 Years	m	1194.9	264.4	580	652	760	856	972	1417	1534	1630	1738	1810
	s	87.6	60.1	172.8	158.8	138.9	122.7	105.8	96.4	110.3	124.9	143.6	156.9
5 Years	m	1186.7	266.4	567	640	749	845	963	1411	1528	1625	1734	1807
	s	104.1	75.5	215.6	197.8	172.2	151.2	128.9	114.5	132.4	151.1	174.9	191.9
3 Years	m	1224.2	242.2	661	727	826	914	1020	1428	1535	1623	1722	1788
	s	152.0	115.9	325.2	297.6	257.8	225.0	190.3	170.4	199.3	229.1	266.6	293.2

Table A3. Quantiles of the distribution of annual rainfall in Lugano, as obtained with different record lengths resampled 200 times and continuous data using derived distributions (in mm).

Sample Size		Mean	STD	Skew.	QUANTILES									
					Q .01	Q .02	Q .05	Q .10	Q .20	Q .80	Q .90	Q .95	Q .98	Q .99
32 Years		1529.8	280.0	0.27	942	1006	1105	1197	1311	1786	1919	2032	2163	2252
15 Years	m	1531.2	270.8	0.27	960	1023	1120	1208	1319	1778	1907	2016	2142	2227
	s	60.4	8.3	0.03	44.1	45.9	48.8	51.4	54.7	68.9	72.9	76.3	80.3	83.1
10 Years	m	1520.5	268.6	0.27	955	1017	1112	1200	1310	1765	1893	2001	2126	2210
	s	71.5	10.2	0.03	52.7	54.8	58.1	61.1	64.9	81.4	86.2	90.3	95.0	98.3
7 Years	m	1531.6	269.8	0.26	962	1024	1121	1210	1321	1779	1908	2017	2143	2228
	s	101.8	14.7	0.05	73.8	76.9	81.9	86.5	92.5	117.9	125.3	131.6	139.0	144.0
5 Years	m	1532.0	268.8	0.25	964	1026	1123	1211	1322	1780	1908	2017	2142	2228
	s	121.1	16.5	0.07	89.0	92.7	98.4	103.9	110.7	140.1	148.6	155.9	164.4	170.1
3 Years	m	1529.5	267.4	0.24	963	1025	1121	1210	1320	1778	1906	2015	2140	2225
	s	150.9	21.4	0.09	110.2	114.9	122.2	129.2	137.9	175.8	186.8	196.3	207.2	214.8

Table A4. Quantiles of the distribution of annual rainfall in Lugano, as obtained with different record lengths resampled 200 times fitting Normal distributions (in mm).

Sample Size		Mean	STD	QUANTILES									
				Q .01	Q .02	Q .05	Q .10	Q .20	Q .80	Q .90	Q .95	Q .98	Q .99
32 Years (DDA)		1529.8	280.0	942	1006	1105	1197	1311	1786	1919	2032	2163	2252
32 Years (Normal)		1530.5	290.1	856	935	1053	1159	1286	1775	1902	2008	2126	2205
15 Years	m	1530.5	280.2	879	955	1070	1171	1295	1766	1890	1991	2106	2182
	s	56.8	43.7	101.9	92.1	78.4	68.0	58.6	75.6	90.0	103.1	118.6	129.3
10 Years	m	1521.7	281.6	867	943	1058	1161	1285	1759	1883	1985	2100	2177
	s	69.0	62.0	137.7	123.1	102.4	86.2	71.3	99.5	121.4	141.0	163.9	179.5
7 Years	m	1530.0	276.0	888	963	1076	1176	1298	1762	1884	1984	2097	2172
	s	98.8	81.6	177.4	158.8	132.9	113.2	96.4	140.1	169.0	194.6	224.6	245.1
5 Years	m	1532.6	264.0	918	990	1098	1194	1310	1755	1871	1967	2075	2147
	s	116.4	103.6	239.7	215.2	180.5	152.9	126.6	162.0	197.3	229.2	266.9	292.7
3 Years	m	1530.0	258.8	928	999	1104	1198	1312	1748	1862	1956	2062	2132
	s	141.9	135.2	322.9	290.1	243.2	205.0	166.6	196.1	241.5	282.9	332.1	365.8

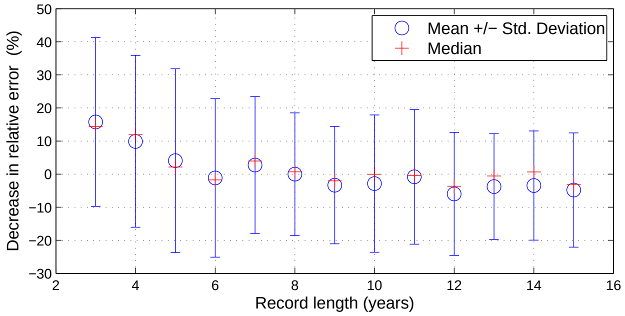

Fig. 10. Mean (± std. dev.) and median decrease in relative percentual error when using short records (3 to 15 yrs long) and the derived distribution instead of direct computation from the sample annual totals, to estimate the long-term (32 yr) standard deviation of annual rainfall at 52 locations in Switzerland.

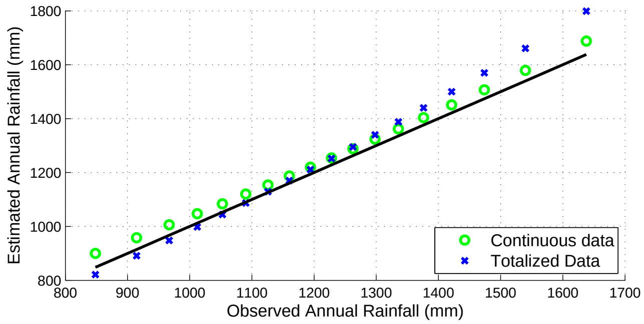

Fig. 11. Effects of data resolution on the distributions obtained with Derived Distributions for Concepción (left) and Lugano (right). The diagonal black line represents a perfect agreement.

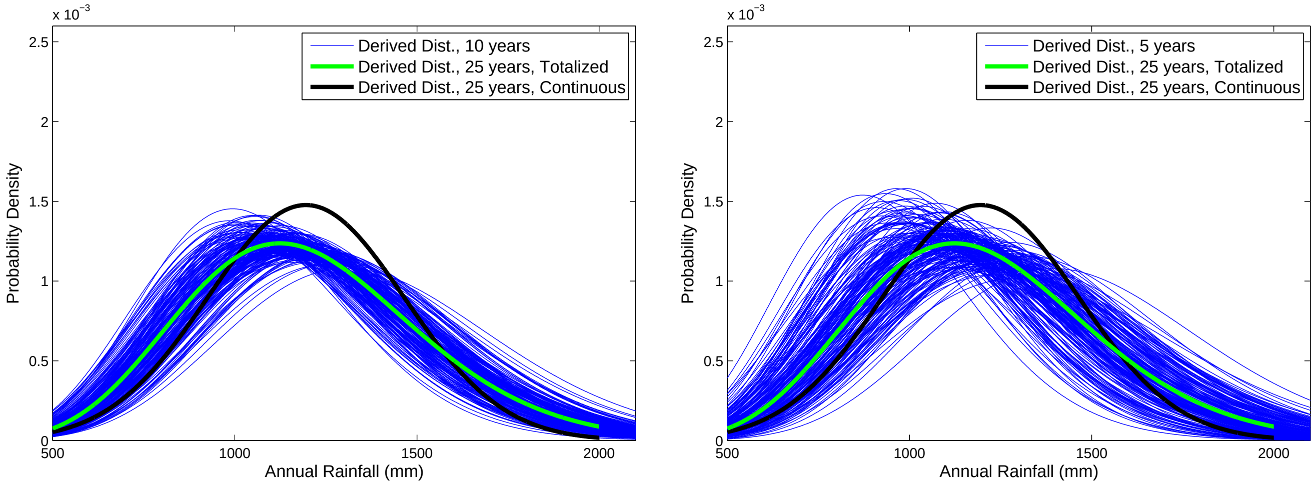

Fig. 12. Estimated PDFs for 10-year (left) and 5-year (right) subsamples using a 24-h totalized record in Concepción.

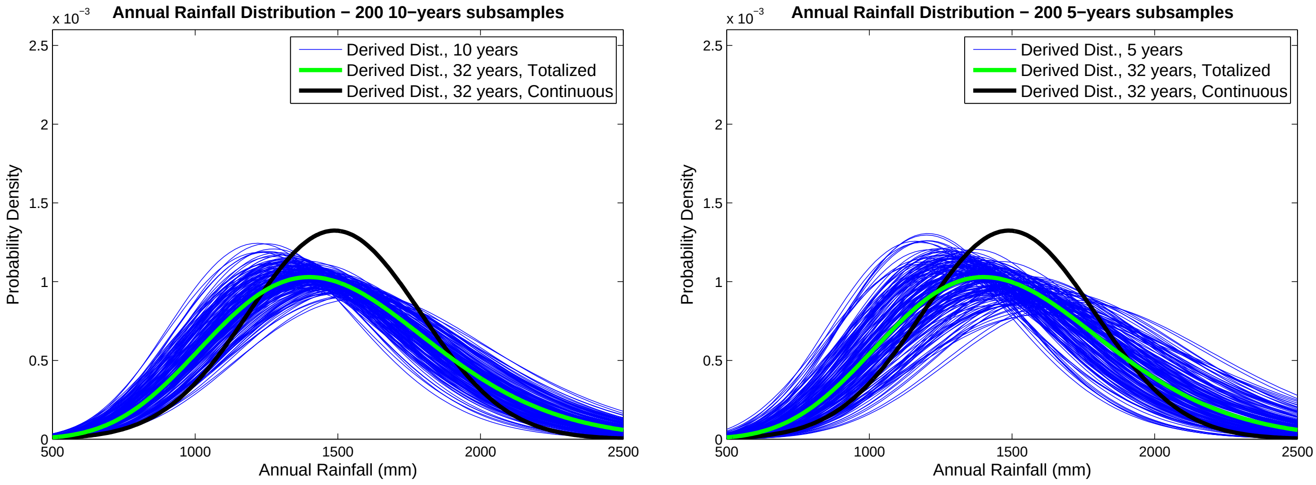

Fig. 13. Estimated PDFs for 10-year (left) and 5-year (right) subsamples using a 24-h totalized record in Lugano.