Quantifying and modelling the ENSO phenomenon and extreme discharge events relation in the La Plata Basin

Melanie Meis\textsuperscript{a,b}, María Paula Llano\textsuperscript{a,b} and Daniela Rodriguez\textsuperscript{b,c}

\textsuperscript{a}Departamento de Ciencias de la Atmósfera y los Océanos (FCEN-UBA), Buenos Aires, Argentina; \textsuperscript{b}Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina; \textsuperscript{c}Instituto del Calculo (IC-UBA-CONICET), Buenos Aires, Argentina

**ABSTRACT**
Understanding and monitoring extreme events is essential, particularly in river discharges from the La Plata Basin, where a large percentage of the economic resources and population of the region are concentrated. In this article, we seek to quantify the relationship between extreme events in discharge and the seasonal climatic index Niño 3.4. We start by estimating the phase shift between the index and mean seasonal (trimester) discharge values. Based on this result, we align the series and use the copula method to fit a joint distribution. We end up with a model that is particularly useful for quantifying the probability of occurrence of extreme events and monitoring their return periods. As a final step, we generate predictions and validate the model by splitting the series into training and test datasets. We develop a simple effective model for monitoring discharges using the El Niño Southern Oscillation (ENSO) index.

**Introduction**
Studying extreme precipitation and streamflow events in the La Plata Basin (LPB) is essential. These events have adverse effects ranging from economic losses for different social groups to possible disease outbreaks and navigation problems, among others. Flood risk is a crucial hydrological variable in terms of social and economic importance (Kiem et al. 2003). With this in mind, Re and Barros (2009) studied extreme precipitation events in the southeast region of South America during the latter half of the 20th century and found a positive trend in the annual maximum precipitation values, as well as an increased frequency of extreme events. The increased incidence of extreme events in the south region of the LPB was also pointed out by Cavalcanti et al. (2015).

Other studies have established a relation between extreme streamflow events and the El Niño Southern Oscillation (ENSO) phenomenon in the LPB. Camilloni and Barros (2000) analysed this relation from convection in the higher and mid Paraná River, and the sea surface temperature (SST) in the Niño-1 + 2 and Niño-3 regions. Additionally, Camilloni and Barros (2003) noticed a clear relation between the ENSO phases and the discharge anomalies in the higher Paraná. Berri et al. (2002) studied the connection between mean discharge and ENSO events for the upper Paraná River, showing evidence of increased mean streamflow during El Niño events. Furthermore, most great floods that occur in the LPB region appear to be related to the El Niño events. In this sense, Antico et al. (2015) analysed the contribution of different oscillations to the four most significant floods registered in the Paraná River.

The streamflow–ENSO correlation can be studied using stochastic hydrological models. Many authors have used multivariate stochastic models to study the relationships among several meteorological and hydrological variables in different regions of the world. AghaKouchak et al. (2014) quantified the joint risk of abnormal temperature and precipitation in certain gauging stations in California, USA, intending to help decision makers. Khendun et al. (2014) modelled the relationship among the Pacific Decadal Oscillation (PDO), ENSO, and precipitation anomalies in the state of Texas and defined a prediction model for extreme events. Cong and Brandy (2012) estimated the relation between precipitation and temperature in a region of Sweden using a copula model. Their final objective was informing agricultural policy decision makers in the context of climate change.

Wei Fang et al. (2018) used copula models to analyse the relation between the standardized precipitation index of adjacent dry and wet seasons. Additionally, Ganguli and Reddy (2013) analysed the relation between ENSO and drought risk in a region in India. The results suggest that including ENSO in the climatic variability copula model is useful for drought risk studies and the management of water resources in the area. In this sense, Ward et al. (2014) have shown, on a global scale, that ENSO exerts strong and widespread influences in both flood hazard and flood risk. Moreover, the authors highlighted that true anomalies of flood risk exist during El Niño or La Niña years, or both.

As demonstrated, there is a strong relation between the ENSO phenomenon and flood hazard in many parts of the world. Therefore, any efforts that can be made towards quantifying this relationship would be of interest. Thus, this work aims to study extreme discharge events in the Paraná and Uruguay Rivers regarding the ENSO phenomenon, using stochastic methodologies, such as the copula method. This methodology has many advantages over the classical methods considered in the study of
climate and hydrology. Most existing stochastic approaches to this problem assume a linear relationship between the variables, while copula methodology is capable of capturing nonlinear structures in the dependent variables (AghaKouchak et al. 2010, Zhang and Singh 2012). Moreover, we obtain an estimate of the bivariate density, which is a core metric for the relationship between the random variables. Having an estimate of the density allows us to derive a wide variety of results, ranging from very simple ones such as correlations or return periods, to more complex ones, such as conditional expectations given observed values. We estimate the occurrence of extreme streamflow events and their relationship with ENSO events using a probabilistic model, considering the last variable shifted. Moreover, we find ways to quantify this relation and generate forecasts. This approach might help inform policymakers in the LPB region, particularly when managing water resources in the area, since the LPB is a critical water and energy source for the countries in the basin (Popescu et al. 2014). The importance of quantifying and understanding hydrological variability is discussed in the literature (e.g. Kiem et al. 2013).

**Data and methodology**

In this study, we used data on mean monthly streamflow (m$^3$/s) from two different gauging stations: Túnel Subfluvial (Paraná River) and Paso de los Libres (Uruguay River) (Fig. 1, Table 1). We obtained these data from the Subsecretaría de Recursos Hídricos (Argentine Undersecretariat for Water Resource, SRH). We analysed 44 years’ worth of data for both gauging stations: December 1974 to November 2016 for Túnel Subfluvial, and December 1974 to December 2015 for Paso de los Libres. Also, we considered the ENSO climatic index for the same period as the Túnel Subfluvial data. Due to evidence of significant changes in the circulation regime in the mid seventies (García and Vargas 1998), we decided to analyse a period starting in 1974.

![Figure 1. Túnel Subfluvial (Paraná River) and Paso de los Libres (Uruguay River) gauging stations located in the La Plata Basin between 27° and 33° S.](https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/)

1Rayner, N. A., et. al, 2003. https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/ [Accessed 23 Mar 2017].
We divided the time series by season, and employed seasonal discharge means (Dec–Jan–Feb, Mar–Apr–May, Jun–Jul–Aug, and Sep–Oct–Nov). Figure 2 shows the evolution of the seasonal (trimester) time series in the study period for the Túnel Subfluvial and Paso de los Libres streamflow, and the NÍÑO 3.4 index. We used the augmented Dickey-Fuller test (1996) and Phillips-Perron test (1988) to analyse the stationarity condition in the time series at a 5% significance level. Both tests have the same null hypothesis: that the time series presents a unit root in its characteristic equation and therefore is not stationary. The alternative hypothesis depends on the version of the test; in this case, it is that the series is stationary. The main difference between the two methods lies in how the distribution of the statistic under the null hypothesis is obtained. Finally, it is

| Data                              | Period         | Latitude/longitude |
|-----------------------------------|----------------|-------------------|
| Túnel Subfluvial (Paraná River)   | 12/1974–11/2016| 31°43’11”60”31”03”|
| Paso de los Libres (Uruguay River)| 12/1974–12/2015| 29°43’17”57”04”57”|
| NÍÑO 3.4                         | 12/1974–12/2016| -                 |

Figure 2. Seasonal streamflow time series (m³/s) for Túnel Subfluvial (top panel), Paso de los Libres (mid panel) and NÍÑO 3.4 seasonal index (bottom panel).
essential to say that testing for stationarity was necessary as most of the statistical methods we employed rely on it.

For both gauging stations, we analysed the same period of the time series of streamflow and NIÑO 3.4. We started our analysis by exploring the relation between Túnel Subfluvial and NIÑO 3.4, and Paso de los Libres and NIÑO 3.4, using the cross-correlation function of each pair. In both cases, we computed the Spearman correlation (Siegel 1985) between streamflow and index series, applying the lag that presented the highest correlation value.

Afterwards, intending to assess the presence of significant waves in the time series considered, we employed the cross-wavelet and wavelet coherence methodologies (Grinsted et al. 2004, Marau and Kurths 2004).

The cross-wavelet transform spectrum can be defined as:

\[
W_n^{xy}(s) = W_n^x(s)W_n^y(s) \tag{1}
\]

where \(x\) and \(y\) are the two time series, \(W_n^x(s)\) and \(W_n^y(s)\) are the wavelet transforms, and \(W_n^{xy}(s)\) indicates a complex conjugate (Torrence and Compo 1998). As the cross-wavelet spectrum is complex, the cross-wavelet power is defined as:

\[
|W_n^{xy}(s)| \tag{2}
\]

Additionally, from this transform, we can compute cross-wavelet coherence for two given series, which is a powerful tool for detecting common waves in two series that occur in a particular time frame, given a pre-established significance level. Another useful measure for this analysis is the cross-wavelet phase, which is useful for estimating phase shifts in the above-mentioned common waves. For this study, we followed the procedure in Grinsted et al. (2004). We applied this method to estimate the phase shift between the series studied in this work.

Since quantifying multivariate relations in extreme events was one of the aims in this work, we estimated multivariate probabilistic functions using the copula methodology. Following Sklar’s theorem (1959), if \(F_{XY}\) is the joint distribution function of a pair of random variables \((x,y)\) with \(F_X\) and \(F_Y\) marginal distributions, there is a unique 2-copula function such that

\[
F(x, y)_{X,Y} = C(F_X(x), F_Y(y)) \tag{3}
\]

For all \(x, y \in R\). If \(F_X\) and \(F_Y\) are continuous functions, there is a unique \(C\); otherwise, \(C\) is defined only in the \(F_X \times F_Y\) range.

Alternatively, if \(C\) is a 2-copula, and \(F_X\) and \(F_Y\) are distribution functions, the \(F_{XY}\) (the function given by Equation 3) is a joint distribution function with \(F_X\) and \(F_Y\) marginal distributions.

To estimate joint probability distributions through the copula theory, we needed to obtain the marginal distribution fit for each time series first. Thus, we employed the maximum likelihood method for identifying the parameters and distribution of each gauging station and index, among different classical distribution functions (normal, log normal, gamma, Gumbel, Weibull). Additionally, we analysed the corresponding goodness of fit using the Kolmogorov-Smirnov (KS) test through bootstrap resampling with 95% confidence (Meis and Llano 2019). Note that whenever more than one theoretical distribution did not reject the null hypothesis, we chose the one with the lower Akaike information criterion (AIC).

Using the theoretical marginal distributions that provided the best fit, we proceeded to select the copula function that best described the bivariate relation between each streamflow and index pair, using the algorithm proposed by Schepsmeier et al. (2018). In this sense, we ended up choosing a copula among different families (according to the AIC), obtaining an estimated joint probability distribution function in each case. We performed a goodness-of-fit test for the copula family using bootstrap resampling with 95% confidence (Schepsmeier et al. 2018), considering the Cramer-Von Mises (CVM) and KS tests.

To evaluate the models, we split the time series into 70/30 training and test sets. We fitted the copula model in the training set and then assessed the goodness of fit of the calibrated model in both training and test sets. As described above, we always used CVM and KS tests for goodness of fit. We repeated this procedure using two different methods for the 70/30 split. In the first method, following Khendun et al. (2014), we split the series randomly. In the second method, we split the series temporally to replicate a forecast situation, taking the first 70% of the series as training data and keeping the remaining 30% for testing.

Once we fitted a copula to our data, and after performing model validation as well as obtaining the corresponding joint distribution of streamflow and index for each station, we used the model for forecasting. The approach we propose is sampling predicted values for the response variable from the conditional distribution obtained after observing a value for the explanatory one, and then computing metrics of interest, such as conditional expectation or quantiles, which we used for forecasting.

We first explain our methodology for obtaining a sample from the dependent variable \(Y\) (discharge) given an observation \(X = x\) (index) and the copula model estimate of the joint distribution. It consists in the following steps:

1. Transform \(X\) and \(Y\) to uniform variables by applying the inverse of their cumulative distribution function. Obtain \(U_1 = F_X^{-1}(X)\) and \(U_2 = F_Y^{-1}(Y)\) uniform variables and name \(u_1 = F_X^{-1}(x)\) the observed value from \(U_1\) corresponding to \(x\), the observed value from \(X\).
2. Obtain a conditional sample \(u_2\) from \(U_2\) given \(U_1 = u_1\) using the conditional function distribution from the copula (Equation 4).
3. Transform the uniform sample \(u_2\) to the space of the original variable \(Y\) applying \(F_Y\), thus obtaining a sampled value \(y = F_Y(u_2)\).

We carried out this procedure using the implementation proposed by Schepsmeier et al. (2018).

\[
C_2(u_2|u_1) = P(U_2 \leq u_2|U_1 = u_1) = \frac{\partial C(u_1, u_2)}{\partial u_1} \tag{4}
\]

Once equipped with the ability to sample values of \(Y\) from observations of \(X\), we took two different approaches for forecasting. Initially, we estimated the expected value of \(Y\) given an
observation of \( X \) and compared it to the corresponding observed value of \( Y \). For this, we split the series again, using the first 70\% (training data) for calibrating the model and reserving the remaining 30\% (test data). Given the observed index values and the fitted copula model, we generated samples for \( Y \) following the process described above and computed an estimate of the expected streamflow value by averaging them out. We compared the estimated expected values of streamflow with the observed values in both datasets (training and test) by computing Pearson and Spearman correlation between them.

In a second approach, instead of computing the expected value and using it as a forecast, as previously explained, we estimated quantiles of the conditional distribution and used them as boundaries of confidence intervals for streamflow and predictors for extreme seasonal events. Given an observation of the index, we estimated the 5\textsuperscript{th} and 95\textsuperscript{th} quantiles of the streamflow given the index observation, by sampling 1000 values of \( Y \) and computing the corresponding quantiles of the sample. For this second approach we always fitted our model to the whole dataset. We are aware that by fitting the whole dataset, we are only estimating training error, which may not fully agree with forecast error. However, as we count fewer than 170 data points, and the aim of the second approach is measuring a 5\% frequency event, splitting into training and test sets would leave us with very few positive events with which to estimate the test error correctly. In any case, we are moderately optimistic with respect to how well our results could account for data not previously observed. This optimism is based on the initial results we obtained regarding the quality of the copula fits in a test set, which is always inferior but still reasonable.

Using the quantile estimations, we first computed the number of times an observed value of \( Y \) lies outside the confidence interval formed by the estimated 5\textsuperscript{th} and 95\textsuperscript{th} quantiles. We expect this value to be around 10\% of the total samples. Second, we kept the 95\textsuperscript{th} quantile only and used it as a way of predicting evacuations. For this purpose, we built a binary variable indicating evacuation, using river height data (m) for the period December 1974 to November 2016 in the Túnel Subfluvial station (Paraná River). We also obtained this data from the SRH. We computed the following variable: for each season of the period, we checked whether the evacuation level stipulated by Prefectura Naval Argentina,\(^2\) which is set at \( 5 \) m for the Túnel Subfluvial station, was exceeded at least once. In this case, we tagged the season as evacuation; if not, we kept it as non-evacuation. In this way, we turned height data into a binary vector with trimester-period resolution, assigning a 0 to the periods tagged as non-evacuations and a 1 to the periods tagged as evacuations.

To evaluate the performance of the estimated 95\textsuperscript{th} quantile as a predictor of evacuations, we built the receiver operating characteristic (ROC) curve (Robin et al. 2011) using the expected 95\textsuperscript{th} percentile as the explanatory variable and the evacuation variable as the binary response. This curve measures the sensitivity versus the specificity of a binary classification model for different cut points in the explanatory variable. Each point in the ROC curve represents the sensitivity and one minus the specificity we obtain if we put a threshold on the explanatory variable, and predicts a value of one for the response variable whenever the value is above the threshold and zero if it is not.

Sensitivity and specificity are terms that were introduced by Jacob Yershalsky (1947). The first, sensitivity, refers to the capability to correctly predict a positive case, and is defined as the ratio between the number of times the model correctly predicted the evacuation and the number of times evacuation actually occurred. The second, specificity, is related to the ability to measure the proportion of negative cases that are correctly identified as such, and is defined as the ratio between the number of times the model correctly predicted that there will be no evacuation and the number of times that evacuation did not occur. The ROC curve uses one minus the specificity (also called the false positive rate), which is the proportion of the number of times the model wrongly predicted an evacuation to the number of times there was no evacuation.

Furthermore, a derived metric for evaluation of the performance of the prediction is the area under the ROC curve, presenting values between zero and one; the farther from 0.5 this value is, the better the performance of the predictor.

Finally, we picked two values for the expected 95\textsuperscript{th} quantile of discharge that we would use as thresholds to predict an evacuation alarm, i.e. each time the expected 95\textsuperscript{th} quantile is above the threshold we predict there will be an evacuation. We used the values of specificity, sensitivity and a third metric called precision to pick those optimal thresholds. The precision of a prediction is the ratio between the number of times the model correctly predicted an evacuation and the number of times the evacuation was predicted, which means it is a metric conditioned in the forecast (when the model forecasted the evacuation).

The first value was selected as the one that optimizes a metric called the F-score. The F-score is a useful metric to evaluate the performance of the model, which takes into consideration both precision and sensitivity. It is computed as the harmonic mean between these two quantities. The second value was picked visually from the ROC curve. We observed that in our case the maximum value of the F-score corresponds to a threshold where the ROC curve presents a noticeable change in slope, meaning that sensitivity and one minus the specificity start to change at different rates. Inspecting the curves, we noticed that there is another threshold where they present a similar change in slope. We also chose that value, which is usually a more conservative option, meaning that using it will yield a higher number of events predicted as evacuations, thus triggering more alarms with a higher proportion of false positives but covering more cases in which an evacuation actually happens.

Finally, for both thresholds we computed contingency tables and performed a chi-squared test to test the null hypothesis of independence between predicted and observed evacuations, with a 95\% level of confidence.

**Results**

**Exploring the discharge-climatic index relation**

Based on the augmented Dickey-Fuller and Phillips-Perron tests, we rejected the non-stationary null hypothesis with a \( p \)

\(^2\text{https://contenidosweb.prefecturanaval.gob.ar/alturas/mapa.php} \) [Accessed 11 Nov 2019].
value below 0.05 for all the series under consideration. This is an essential assumption for most of the methods used in this work. We studied the relation between the variable pairs Tünel Subfluvial-NIÑO 3.4 and Paso de los Libres-NIÑO 3.4 using, first, the cross-correlation function, which we present in Fig. 3. For the first pair it is easy to see that the maximum correlation between lagged series was obtained for a lag of one and two seasons, while for the second pair it was obtained for a lag of one season.

The presence of significant waves in the paired time series was analysed using the cross and coherence wavelet methodology. This joint study of seasonal discharges and NIÑO 3.4 showed significant waves in different periods, with some phase shift between the variables. Figure 4 shows the relation between Tünel Subfluvial and NIÑO 3.4 using the wavelet coherence methodology. A significant period can be observed from the beginning of the recorded period until the 1980s, with seasonal periods between 8 $\Delta t$ (8 seasons – 2 years) and 16 $\Delta t$ (4 years). Given the inclination phase (black arrows in the plot), we determined that, on average, the phase shift was nearly $2\Delta t$.

Moreover, significant waves were observed in both seasonal time series after the end of the 20th century, with wavelengths of 4–8 seasons, and 8–16 seasons, and around 16 $\Delta t$. In these three examples, the average phase shift is $2\Delta t$. We already analysed this relation in 100-year periods for both measurement stations in Meis and Llano (2019), where a relationship between the climatic index and its dominance regarding streamflow was observed. For the Paso de los Libres and NIÑO 3.4 pairs, we observed the same phase shift (data not shown).

Based on these findings, we established a $2\Delta t$ phase shift (with $\Delta t$ equal to 1 season) between the seasonal discharge variable and the seasonal climatic index. In the rest of this work, whenever we considered pairs of index–streamflow series, the climatic index series was always shifted $2\Delta t$ in time to make significant waves of both series match temporally. Additionally, the Spearman coefficient was estimated considering the phase shift and in both pairs of time series, finding a coefficient of around 0.30 ($p = 0.00010$) for Tünel Subfluvial-NIÑO 3.4 and 0.20 ($p = 0.017$) for Paso de los Libres- NIÑO 3.4. We represent these relations in a dispersion plot for Tünel Subfluvial-NIÑO 3.4 (Fig. 5). A positive relation between the two variables can be seen, meaning that when the index values are higher, the discharge values are also higher. This result matches the above-mentioned results for the relation between ENSO and streamflow in LPB stations.

**A model of the streamflow–climatic index relation**

The first step of the copula method for fitting the joint distribution of each pair of series was estimating the marginal distributions. As described in the methodology section, we

---

**Figure 3.** Cross-correlation function between seasonal NIÑO 3.4 and Tünel Subfluvial discharge (left), and NIÑO 3.4 and Paso de los Libres discharge (right). Blue lines represent the confidence intervals.

**Figure 4.** Wavelet coherence and significant phase difference between the seasonal NIÑO 3.4 index and the Tünel Subfluvial seasonal streamflow, with a 5% significance level.

**Figure 5.** Relation between seasonal streamflow time series for the Tünel Subfluvial gauging station $(m^3/s)$ and the climatic seasonal index NIÑO 3.4, with a phase shift equal to two trimesters.
functions can fit that behaviour. However, we picked Gumbel as a marginal distribution as it is a better fit according to the AIC. We also see difference in favour of the Gumbel distribution fit in the Q-Q plot (Fig. 6).

Regarding the Paso de los Libres station, the Weibull and Gumbel theoretical distributions can fit the seasonal streamflow data with a 5% significance, but according to the AIC, the Gumbel function should be employed (Table 3); this is shown in Fig. 7, which plots the theoretical Gumbel fit with the observed data.

As expected, the normal distribution has the best fit for the seasonal NINO 3.4 data with 5% of significance. Figure 8 shows how the theoretical model – the normal distribution – represents the observed data, although specific extreme values can be seen that slightly differ from a normal distribution. The fit that we show here for seasonal NINO 3.4 is based on the same analysis period as for Tunel Subfluvial. Using the theoretical marginal distributions with the better fit, we selected the copula function that better describes the bivariate relation between each data pair.

As described in the methodology section, we selected the best copula function among several families, using the algorithm proposed by Schepsmeier et al. (2018) (which considers more than 30 families of copulas), and according to the AIC, the Joe copula (Equation 5) was most suitable for both data pairs (NINO 3.4 – Tunel Subfluvial and NINO 3.4 – Paso de los Libres). Note that we were interested in studying the joint probability of extreme events in greater depth, particularly events where both streamflow and climatic indices were above the 95th percentile, but not all copula families can accurately model extreme values. The copula we chose (Joe) is adequate to determine upper tail dependence (Van-Nam et al. 2013).

\[ C_\theta(u, v) = 1 - [(1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta]^{1/\theta} \]

with \( \theta \geq 1 \)

(5)

Where \( u \) and \( v \) are the uniform variables, and \( \theta \) is the estimated parameter. The upper tail dependence coefficient for this copula is \( 2 - 2^\theta \), while for the lower tail it is 0.

Table 4 shows the parameter \( \theta \) for each fit and the goodness-of-fit results for the copula functions. It contains \( p \) values for the CVM and KS tests obtained through bootstrap

---

### Table 2. Seasonal streamflow fit and goodness of fit for the Tunel Subfluvial gauging station, according to different theoretical distributions, with a 5% significance level.

| Theoretical distributions | Parameters | Bootstrap p value |
|---------------------------|------------|-------------------|
| Normal                    | \( \mu = 15885.82 \) | 0.006             |
|                           | \( \sigma = 3347.81 \) |                  |
| Lognormal                 | \( \mu = 9.65 \) | 0.30              |
|                           | \( \sigma = 0.20 \) |                  |
| Weibull                   | \( k = 4.57 \) | 0                 |
|                           | \( \lambda = 1.709532e+04 \) |                  |
| Gamma                     | \( a = 2.431058e+01 \) | 0.18              |
|                           | \( \beta = 1.533227e-03 \) |                  |
| Gumbel                    | \( \mu = 14339.17 \) | 0.62              |
|                           | \( \beta = 2604.48 \) |                  |

### Table 3. Seasonal streamflow fit and goodness of fit for the Paso de los Libres gauging station according to different theoretical distributions, with a 5% significance level.

| Theoretical distribution | Parameters | Bootstrap p value |
|--------------------------|------------|-------------------|
| Normal                   | \( \mu = 4776.1143 \) | 0.007             |
|                           | \( \sigma = 2842.095 \) |                  |
| Log normal               | \( \mu = 8.28477217 \) | 0.012             |
|                           | \( \sigma = 0.64 \) |                  |
| Weibull                  | \( k = 1.7795657 \) | 0.905             |
|                           | \( \lambda = 5388.56 \) |                  |
| Gamma                    | \( a = 1684.6814030 \) | 0                 |
|                           | \( \beta = 2.835 \) |                  |
| Gumbel                   | \( \mu = 3505.55 \) | 0.52              |
|                           | \( \beta = 2129.72 \) |                  |

Considered different classical distribution functions (normal, log normal, gamma, Gumbel, Weibull) for each gauging station, and we estimated each parameter distribution using the maximum likelihood method. Additionally, we performed the corresponding goodness-of-fit tests. The fitted distributions and parameters can be found in Tables 2 and 3. We also plotted the theoretical distribution quantiles against the empirical quantiles’ distribution, a procedure known as a Q-Q plot (Fig. 6), with a 5% significance (Fox and Weisberg 2019). Also, a histogram with the observed data and the proposed theoretical density function is shown. In both figures it can be verified that the fit is coherent.

In Table 2, it can be observed that the log normal distribution and the Gumbel distribution can fit the measured data for seasonal discharges in the Tunel Subfluvial station. We noticed that the variable has a right-tail asymmetry, and both analytical functions can fit that behaviour. However, we picked Gumbel as a marginal distribution as it is a better fit according to the AIC. We also see difference in favour of the Gumbel distribution fit in the Q-Q plot (Fig. 6).

Regarding the Paso de los Libres station, the Weibull and Gumbel theoretical distributions can fit the seasonal streamflow data with a 5% significance, but according to the AIC, the Gumbel function should be employed (Table 3); this is shown in Fig. 7, which plots the theoretical Gumbel fit with the observed data.

As expected, the normal distribution has the best fit for the seasonal NINO 3.4 data with 5% of significance. Figure 8 shows how the theoretical model – the normal distribution – represents the observed data, although specific extreme values can be seen that slightly differ from a normal distribution. The fit that we show here for seasonal NINO 3.4 is based on the same analysis period as for Tunel Subfluvial. Using the theoretical marginal distributions with the better fit, we selected the copula function that better describes the bivariate relation between each data pair.

As described in the methodology section, we selected the best copula function among several families, using the algorithm proposed by Schepsmeier et al. (2018) (which considers more than 30 families of copulas), and according to the AIC, the Joe copula (Equation 5) was most suitable for both data pairs (NINO 3.4 – Tunel Subfluvial and NINO 3.4 – Paso de los Libres). Note that we were interested in studying the joint probability of extreme events in greater depth, particularly events where both streamflow and climatic indices were above the 95th percentile, but not all copula families can accurately model extreme values. The copula we chose (Joe) is adequate to determine upper tail dependence (Van-Nam et al. 2013).

\[ C_\theta(u, v) = 1 - [(1 - u)^\theta + (1 - v)^\theta - (1 - u)^\theta(1 - v)^\theta]^{1/\theta} \]

with \( \theta \geq 1 \)

(5)

Where \( u \) and \( v \) are the uniform variables, and \( \theta \) is the estimated parameter. The upper tail dependence coefficient for this copula is \( 2 - 2^\theta \), while for the lower tail it is 0.

Table 4 shows the parameter \( \theta \) for each fit and the goodness-of-fit results for the copula functions. It contains \( p \) values for the CVM and KS tests obtained through bootstrap

---

![Figure 6. Seasonal streamflow Q-Q plot (5% significance) and Gumbel theoretical quantiles for the Tunel Subfluvial gauging station (left). Seasonal streamflow probabilistic histogram and Gumbel probability density distribution for the Tunel Subfluvial gauging station (right).](image-url)
resampling. In both cases, and according to the \( p \) values (0.21 and 0.39), the null hypothesis is not rejected, and the Joe copula family with 95% confidence fits the pair NIÑO 3.4 – Túnel Subfluvial. We used an analogous procedure for the pair NIÑO 3.4 – Paso de los Libres, with similar results.

### Model validation

We calibrated the copula to the training data that were selected randomly, which turned out to be of the Joe family again. We ran goodness-of-fit tests for the obtained model with both training and test data. In all cases, the null hypothesis of our tests is not rejected, indicating that the proposed model fitted both series. The \( p \) values obtained were 0.19 and 0.38 for the CVM test, and 0.20 and 0.44 for the KS test (training and test sets, respectively).

Also, we repeated the procedure by splitting the series temporally. Again, we obtained a Joe family copula in the training series, and the null hypothesis of the goodness-of-fit tests was not rejected in any case, yielding \( p \) values 0.19 and 0.37 for the CVM test and 0.25 and 0.36 for the KS test (training and test data, respectively). Note that this second approach provided out-of-period validation for the model.

We replicated the above procedure for Paso de los Libres station. In this case, the validation of the model could only be done when the data were split randomly. This could be related to a lesser signal of ENSO in the Uruguay basin than in the Parana basin, combined with an insufficient amount of data. In the case of temporal splitting, the proposed algorithm could not find a suitable copula function that fitted the training set.

### Forecasting from the joint distribution

While we fitted a bivariate distribution to our pairs of series, we always considered seasonal discharge to be the response variable and index values to be the explanatory one. Therefore, it made sense to use the model as a tool for forecasting in the following way: once we observed an index value, we conditioned the joint distribution to that value and used the
obtained streamflow distribution to predict quantities of interest. Note that as the index series has a phase shift of $2 \Delta t$ (half a year), we were forecasting within that time horizon.

To validate these predictions for Túnel Subfluvial discharge, we split the series again, using the first 70% (training data) for calibrating the model. We generated predicted streamflow values, following the above procedure, and computing the conditional expectation for the whole series and compared with the observed values in both datasets (training and test). In both cases, we computed the correlation between observed and forecasted values. The Pearson and Spearman correlation between observed and predicted streamflow values were 0.54 and 0.32 respectively ($p = 3.789e-10$ and 0.00050) for the training series, and 0.52 and 0.37 ($p = 0.00024$ and 0.011) for the test series.

We present scatter plots for observed and predicted values in Fig. 9 (a – train, b – test and c – whole dataset). Furthermore, we performed the same correlation analysis for data points corresponding to values of Niño 3.4, above and under its median separately. The results show a healthy relationship for the positive phase of the index and almost null for the negative phase. Finally, we repeated the same procedure using random train/test splits (instead of using the initial 70% of the series as the training set and reserving the final 30% as the testing set) and obtained Pearson and Spearman correlations of 0.51 and 0.27, respectively, on the training set, and 0.49 and 0.21 on the testing set, with $p$ values of 8.202e-09 and 0.0045, and 0.00033 and 0.02, respectively.

We explored a second approach for forecasting, which is more focused on the prediction of extreme events of streamflow. We followed the same idea as above: once we observed an index value, we could use the model to estimate a conditional distribution for the corresponding streamflow data point. In this case, instead of computing the expected value and using it as a forecast, the idea was to estimate quantiles of the conditional distribution and use them as boundaries of confidence intervals for streamflow. For example, we would expect that streamflow does not surpass the 95% quantile of the conditional distribution with 95% confidence.
Figure 10. Series of 5% and 95% quantiles of the estimated conditional distribution obtained from the observed index value for Túnel Subfluvial.

Figure 11. Proportion of observations that lay below the estimated quantile.

In this case, as we explained in the methodology section, we fitted the model to the whole series; thus we only report one value for each metric and not two (training and test). Following the same procedure as in the previous case, we generated samples for the conditional distribution for each observation of the index. We estimated 5% and 95% quantiles of the conditional distribution from these samples, generating confidence intervals for the Túnel Subfluvial series (Fig. 10). We ended up with the top and bottom series for streamflow, inferred from the index series. The theoretical result that should hold is that for any time point, the streamflow should be under the top series value with 95% confidence, and over the bottom series value with 95% confidence also. To assess the accuracy of the calibration of these series, we computed – in steps of 5% for each value from 5% to 95% – the proportion of streamflow observations that lay below the estimated quantile. The results are shown in Fig. 11 for Túnel Subfluvial, together with the expected theoretical proportion.

Forecasting analysis was not performed for Paso de los Libres-NIÑO 3.4, mainly because we found that model validation does not hold temporally for that series. This indicates that although the model could be calibrated to the whole dataset and this calibration was validated with random splits, we do not have enough data to perform out-of-period validation, and therefore it might not be wise to generate forecasts for this series.

Evaluation of the 95th quantile as a measure of early evacuation

We added a second use for the previously computed quantiles. We proposed to evaluate and quantify a possible evacuation alert system using the expected 95th percentile series for the distribution of Túnel Subfluvial discharge conditional to the previously observed NIÑO 3.4 index.

As described in the methodology section, an ROC curve was built (Fig. 12) between the binary variable that discriminates when the height of the river in Túnel Subfluvial station was over 5 m at least once in the trimester (response variable), and the conditional 95th percentile discharge estimated given the NIÑO 3.4 index (explanatory variable). We computed the area under the curve, which is a metric about the predictive power of the explanatory variable, and obtained a value of 0.83, which is indicative of a strong relation.

From this ROC curve, we picked two thresholds on the expected 95th percentile that can be used for predicting an evacuation. For each of those values, we obtained a new binary variable of predicted evacuation, which is equal to one when the expected 95th percentile is higher than the
threshold, and zero otherwise, obtaining a total of 166 vectors. The first of the thresholds was selected as the one that maximizes F-score. We used the Performance Estimation library for R (Torgo 2014) to compute F-scores for different thresholds and obtained a maximum of 0.6 at the threshold 23 147.76 m$^3$/s. We show the location of this value in the ROC curve in Fig. 13, where it can be seen that at that point, there is a change in the slope of the curve. For this threshold value, the corresponding contingency table, Table 5, was obtained.

| Explanatory | Response 0 | Response 1 |
|-------------|------------|------------|
| Exploratory | 0          | 125        |
|             | 1          | 13         |

Taking into account Table 5, a chi-squared test was done to evaluate the association between the variables with a 95% level of confidence. Since the p value obtained was 0.0004998 the null hypothesis was rejected, and we concluded that there is enough evidence to indicate an association between the variables.

The second threshold was selected in a visual manner, looking for a second change in the slope of the ROC curve. This threshold has an associated value of 20708.24 m$^3$/s. Figure 14 shows the ROC curve, with vertical and horizontal lines associated with one minus the specificity and the sensitivity value, respectively, for that fixed threshold. This second threshold is a more conservative alert system, meaning that more alerts are emitted, improving the sensitivity (near 90%) at the cost of generating a much higher quantity of false positives (near 45%). From this new threshold the contingency table was computed (Table 6). Next, we considered the chi-squared test to evaluate the association among the variables with 95% confidence and obtained a p value of 2.188e-07, enough to infer an association between the variables.

| Explanatory | Response 0 | Response 1 |
|-------------|------------|------------|
| Exploratory | 0          | 76         |
|             | 1          | 62         |

Table 5. Contingency between the response variable (binary river height) and the binary discharge from the threshold obtained from the receiver operating characteristic (ROC) curve that maximizes the F-score.

Table 6. Contingency between the response variable (binary river height) and the binary discharge from the threshold obtained from the receiver operating characteristic (ROC) curve where a change in the slope of the curve was seen.
In summary, two thresholds were proposed as alert systems for possible evacuation. In this way, not only did we study the general relation between the expected discharge variable and the alert system for evacuation, but also the performance of these alert systems was assessed.

**Comparison between the theoretical model and the observed data**

We performed a simulation using the copula model to compare the theoretical function with the measured data. The results are shown for the Túnel Subfluvial and NIÑO 3.4 pairs (Fig. 15). Again, it is possible to observe a positive relationship for each variable in the theoretical behaviour (black dots) as well as in the real data (red dots). This relation follows from the close positive correlation between NIÑO 3.4 and precipitation in the LPB region, as reported in the literature. Additionally, it is essential to mention that the data distribution for extreme values shows adequate theoretical modelling in the upper tail. For Paso de los Libres (data not shown), we also obtained an adequate simulation for the observed data. In this sense, the modelling was satisfactory for the two study cases.

**Probability estimates for extreme events**

To determine the probability of joint occurrence in the class of situations that are of interest in this article, we estimated the cumulative probability for the bivariate relation. This analysis is helpful in estimating the probability of extreme discharge events (in this study, those above the 95th percentile) and relating them to climatic phenomena. In this sense, given that we are studying a bivariate relation, it is also necessary to fix a climatic index threshold. Due to the occurrence of particularly extreme events in seasonal streamflow, and taking into account the phase shift of two seasons from the climatic index, we observed that the seasonal index values from NIÑO 3.4 associated with those events were above the 95th percentile. Figure 16 shows the bivariate probability for Túnel Subfluvial and NIÑO 3.4. The red lines highlight the quadrant that is associated with extreme events, fixed by the 95th percentile thresholds for both variables. According to this theoretical model, the joint probability for extreme events was 2.2%. The relation between Paso de los Libres and NIÑO 3.4 showed similar results, and for that reason, the figure is not shown.

**Bivariate return periods in extreme events**

A useful metric that can be estimated from the joint distribution of the pair of series is the return period (Salvadori and De Michele 2004). In this sense, the distribution data for Túnel Subfluvial-NIÑO 3.4 (black dots), and the return period isolines were obtained with the probability function we adopted (Fig. 17). If we analyse extreme events of seasonal streamflow (above the 95th percentile, red line) associated with NIÑO 3.4 values higher than the 95th percentile (1.54), we observe a return period of approximately 15 years according to empirical data, and of 11.5 years according to simulated data. The events that are above those thresholds are autumn/winter 1983, winter 1992, and autumn 1998 for seasonal streamflow, as well as summer 1998 and summer 2016. Also, it is interesting to point out that the return period of each variable for the 95th percentile would have been 5 years. In this way, if both variables were independent, the result would have been considerably larger, achieving a 100-year return period for the previous joint threshold. In other words, the last result reveals a strong dependence between the two variables.

Moreover, our analysis shows that extreme events for both series seem to be strongly correlated, so it is essential and necessary to model the index-streamflow relation. Additionally, it should be noted that although there were certain events whose values were above the 95th percentile (21 916.61 m$^3$/s), the index value was below that percentile; yet these events count as El Niño events (summer 1983, autumn 1995). Indeed, events in the upper right tail belong to the warm ENSO phase,
not to normal streamflow conditions (they are substantially above the mean). Finally, in Fig. 17, it is possible to observe the most extreme seasonal event, winter 1983. The combination of extreme seasonal discharge and extreme seasonal positive ENSO 3.4 index for this event has a close to 200-year return period. Such a return period is not surprising as this ENSO event is considered exceptional (Camilloni and Barros 2000).

Regarding the relation between Paso de los Libres and Niño 3.4 (Fig. 18), we observed extreme events with streamflow values above the 95th percentile, and above the 95th percentile in the seasonal climatic index (1.39). By computing the fitted return period, we observe that events of this magnitude occur, on average, every 14.5 years. Among the observations, we have five such events in the study period (autumn and winter 1983, winter 1992, summer and autumn 1998). Moreover, there were events above the 95th percentile for streamflow (10 050.38 m³/s), but with climatic index values below that percentile (spring 1982, 1997, and 2002). Additionally, winter 1984 was close to the 95th percentile threshold for seasonal streamflow. This was an isolated event, as it is not associated with a warm ENSO phase. The existing literature mentions the substantial floods that occurred in the LPB region for that season (Righi and de Souza Robaina 2010), where the authors emphasize that precipitation showed extreme values, possibly due to other physical causes.

Both analyses show not only the 95th-percentile thresholds but also the 90th-percentile thresholds, for both the seasonal streamflow and the seasonal climatic index (Figs 17 and 18). In the Tünel Subfluvial–Niño 3.4 analysis, the return period for this type of event was around 5 years, while in Paso de los Libres, the period was slightly lower.

**Conclusion**

We proposed a method for estimating the joint probability density function of the combined phase shift for climatic index and seasonal discharge. The main aim was to provide a useful tool for anticipating extreme events in a runoff, as we are studying the interaction of the observed index with the streamflow occurring two seasons later. Estimating the joint distribution of a pair of response/explanatory variables can be useful in many aspects. In this work, we derived metrics that are particularly useful for quantifying some known relationships, and proposed some simple ideas for forecasting. However, the picture of possible applications far exceeds what is presented here, as the joint distribution of a pair of variables contains all the available probabilistic information in them. For example, the information about the joint distribution can surely be used as an input for more complex models.

In our case, this tool enabled us to quantify probabilities (and their associated return periods) for certain joint extreme events which have already occurred in recent years. A metric like this one can be useful for monitoring the occurrence of extreme events in the streamflow variable and in the climatic index which are related to extreme ENSO events (Takahashi and Dewitte 2016). Moreover, we provided two possible approaches for seasonal discharge forecasting. The first one, which follows the line of Khendun et al. (2014), is about computing the expected value of discharge given the observed index value and using it as a forecast. The second one is about computing confidence intervals for streamflow, given climatic index values. In the first case, we find that there is a signal in the prediction by computing a significant correlation between observed and predicted values. In the second case, the forecasted quantiles are especially useful as they provide a simple way of obtaining a value that the streamflow will not surpass, with an absolute confidence. The 95% quantile of the conditional distribution seems to be a reasonable metric to monitor, as it is tight enough to be meaningful (not raising the alarm on every period) but at the same time will be fulfilled with reasonable confidence.
Additionally, our model was validated by different approaches, showing that it could be considered for monitoring the relationship between streamflow ENSO and the gauging stations from the LPB. We validated our model temporally for Túnel Subfluvial, giving us the certainty that both metrics we monitored and the forecasts we produced should be meaningful moving forward. Moreover, we also checked the consistency of the thresholds we computed for streamflow values.

The performance of the model considered for Túnel Subfluvial station was evaluated using multiple metrics. We computed the area under the ROC curve built using a dichotomous variable obtained from discharge height as the objective variable and the 95th percentile expected discharge as the predictor. Using the information provided by this curve, we explored the possibility of using 95th percentile expected discharge to build a binary variable that anticipates evacuations. We set two different thresholds on this variable and examined the quality of the prediction obtained if we could predict an evacuation each time the 95th percentile expected discharge is above those thresholds. For both cases, we computed contingency tables and carried out a chi-square test in which we rejected the null hypothesis of lack of association between predicted and actual evacuations with 95% confidence. We also computed sensitivity and specificity in all the cases.

In general, our work provides an approach for more deeply studying the relationship between climatic indices and discharge, in a way that is not currently found in the literature about these two variables in the region. Moreover, we propose a new yet simple way of employing the results obtained by this methodology in the case of meteorological variables.

To conclude, our procedure shows that given a certain threshold for the climatic index, it is possible to quantify the probability of extreme discharge events with confidence. In this sense, we were able to provide a simple and validated climatic probabilistic way of monitoring the occurrence of extreme events in two rivers of great socio-economic importance in the southern region of the LPB, from a well-studied variable. In this sense, in the Túnel Subfluvial case, the methodology foresees the occurrence of events higher than the 95th percentile on both variables once every 11.5 years, indicating that we expect roughly half of the extreme streamflow events to correspond to extreme index events. Early identification of these types of events might help to mitigate their negative impact.

Acknowledgements

We thank the anonymous referees for their valuable comments and critical reading of the article. This research was supported by CONICET under Grant 11220130100806, 20021070100330BA from the Universidad de Buenos Aires and PICT-201-0377 from ANPYCT, Argentina.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas [11220130100806]; ANPYCT [PICT-201-0377]; Universidad de Buenos Aires [20020170100330BA].

References

AghaKouchak, A., et al., 2014. Global warming and changes in risk of concurrent climate extremes: insights from the 2014 California drought. Geophysical Research Letters, 41, 8847–8852. doi:10.1002/2014GL062308
AghaKouchak, A., Bárdossy, A., and Habib, E., 2010. Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula. Advances in Water Resources, 33 (6), 624–634. doi:10.1016/j.advwatres.2010.02.010
Antico, A., Torres, M.E., and Diaz, H.F., 2015. Contributions of different time scales to extreme Paraná floods. Climate Dynamics, 46, 3785–3792. doi:10.1007/s00382-015-2804-x
Berri, G., Ghietto, M., and Garcia, N.O., 2002. The influence of ENSO in the flows of the Upper Paraná River of South America over the past 100 years. American Meteorological Society, 3, 57–65.
Camilloni, I. and Barros, V.R., 2003. Extreme discharge events in the Paraná River and their climate forcing. Journal of Hydrology, 278, 94–106. doi:10.1016/S0022-1694(03)00133-1
Camilloni, I.A. and Barros, V.R., 2000. The Paraná river response to El Niño 1982–83 and 1997–98 events. Journal of Hydro meteorology, 1, 412–430. doi:10.1175/1525-7541(2000)001<0412:TPRRT>2.0.CO;2
Cavalcanti, I.F.A., et al., 2015. Precipitation extremes over La Plata Basin - Review and new results from observations and climate simulations. Journal of Hydrology, 523, 211–230. doi:10.1016/j.jhydrol.2015.01.028
Cong, R.G. and Brandy, M., 2012. The interdependence between rainfall and temperature: copula analyses. The Scientific World Journal, 2012, 1–11. doi:10.11002/2012/405675
Fox, J. and Weisberg, S., 2019. An R companion to applied regression. 3rd ed. Sage.
Fuller, W.A., 1996. Introduction to statistical time series. 2nd ed. New York: John Wiley and Sons.
Ganguli, P. and Reddy, M.J., 2013. Analysis of ENSO-based climate variability in modulating drought risks over western Rajasthan in India. Journal of Earth System Science, 122 (1), 253–269. doi:10.1007/s12049-012-0247-x
García, N.O. and Vargas, W.M., 1998. The temporal climatic variability in the “Rio de la Plata” basin displayed by the rivers discharges. Climatic Change, 38, 359–379. doi:10.1023/A:1005386530866
Grinsted, A., Moore, J., and Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics, 11, 561–566. doi:10.5194/npg-11-561-2004
Khendun, C.P., et al., 2014. A copula-based precipitation forecasting model: investigating the interdecadal modulation of ENSO’s impacts on monthly precipitation. Water Resources Research, 50, 580–600. doi:10.1002/2013WR013763
Kiem, A., S. Verdon-Kidd, D. 2013. The importance of understanding drivers of hydroclimatic variability for robust flood risk planning in the coastal zone. Australasian Journal of Water Resources, 17 (2), 126–134.
Kiem, A., Franks, S., and Kuczera, G., 2003. Multi-decadal variability of flood risk. Geophysical Research Letters, 30 (2), 1035.
Marau, D. and Kurths, J., 2004. Cross wavelet analysis: significance testing and pitfalls. Nonlinear Processes in Geophysics, 11, 505–514. doi:10.5194/npg-11-505-2004
Meis, M. and Llano, M.P., 2019. Hydrostatistical study of the Paraná and Uruguay Rivers. International Journal of River Basin Management, 17, 1–12. doi:10.1080/17456530.2018.1446962
Phillips, P.C.B. and Perron, P., 1988. Testing for a unit root in time series regression. Biometrika, 75, 335–346. doi:10.1093/biomet/75.2.335
Popescu, I., Brandimarte, L. and Paviani, M., 2014. Effects of climate change over energy production in La Plata Basin. International Journal of River Basin Management, 12, 1–9. doi:10.1080/17456530.2014.917317
Rayner, N. A., et al., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal Geophysical Research, 108, 4407. doi:10.1029/2002JD002670
Re, M. and Barros, V.R., 2009. Extreme rainfalls in SE South America. Climatic Change, 96, 1–2, 119–136. doi:10.1007/s10584-009-9619-x
Righi, E. and de Souza Robaina, L.E., 2010. Enchentes do rio Uruguaí no rio Grande Do Sul entre 1980 e 2005: uma análise geográfica. Sociedade & Natureza, 22, 35–54. doi:10.1590/S1982-45132010000100004

Robin, X., Turck, N., and Hainard, A., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77. doi:10.1186/1471-2105-12-77

Salvadori, G. and De Michele, C., 2004. Frequency analysis via copulas: theoretical aspects and applications to hydrological events. Water Resources Research, 40, 1–17. doi:10.1029/2004WR003133

Siegel, S., 1985. Nonparametric statistical: for the behavioral sciences. 1st ed. Tokyo: McGraw-Hill Book Company.

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publications De l’Institut De Statistique De l’Université De Paris, 8, 229–231.

Torgo, L., 2014. An Infra-Structure for Performance Estimation and Experimental Comparison of Predictive Models in R, CoRR abs/1412.0436 [cs.MS]. http://arxiv.org/abs/1412.0436

Torrence, C. and Compo, G., 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society, 79, 61–78. doi:10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Van-Nam, H., et al., 2013. Uncertainty analysis in econometrics with applications. Berlin: Springer.

Ward, P., et al., 2014. Strong influence of El Niño Southern Oscillation on flood risk around the world. Proceedings of the National Academy of Sciences, 111 (44), 15659–15664. doi:10.1073/pnas.1409822111

Zhang, L. and Singh, V., 2012. Bivariate rainfall and runoff analysis using entropy and copula theories. Entropy, 14, 1784–1812.