environmental sciences proceedings

MDPI

Proceeding Paper

Stationary/Non-Stationary Modelling for Extreme Value Distribution: Analysis of Rainfall Annual Maxima in Italy in a Climate Change Context †

Davide Luciano De Luca [1,*], Benedetta Moccia [2], Fabio Russo [2] and Francesco Napolitano [2]

[1] Department of Informatics, Modelling, Electronics and System Engineering, University of Calabria, 87036 Rende, Italy
[2] Department of Civil, Constructional and Environmental Engineering (DICEA), Sapienza University of Rome, 00184 Rome, Italy
* Correspondence: davide.deluca@unical.it
† Presented at the International Conference EWaS5, Naples, Italy, 12–15 July 2022.

Abstract: In this paper, EXTRASTAR software (EXTRemes Abacus for STAtistical Regionalization) is illustrated, which provides a general framework for extreme value analysis, specifically for Annual Maxima (AM) time series. The proposed methodology represents a useful and quick approach to carry out onsite and regional statistical investigations of extreme events, which is also suitable for climate change scenarios. EXTRASTAR was tested with the AM time series at daily resolutions for the entire Italian rain gauge network by implementing EV1 (Extreme Value Type I), GEV (Generalized Extreme Value) and TCEV (Two-Components Extreme Value) probability distributions.

Keywords: annual rainfall maxima; statistical regionalization; stationary and non-stationary models

check for updates

Citation: De Luca, D.L.; Moccia, B.; Russo, F.; Napolitano, F. Stationary/Non-Stationary Modelling for Extreme Value Distribution: Analysis of Rainfall Annual Maxima in Italy in a Climate Change Context. *Environ. Sci. Proc.* **2022**, *21*, 65. https://doi.org/10.3390/environsciproc2022021065

Academic Editors: Vasilis Kanakoudis, Maurizio Giugni, Evangelos Keramaris and Francesco De Paola

Published: 2 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

From the literature, statistical analysis of hydrological extremes can be diversified according to: (1) the adopted probability function; (2) the type of approach: stationary (i.e., temporal invariance of model parameters) or non-stationary (i.e., model parameters are time-varying to account for climate change scenarios); (3) the method for parametric estimation (onsite or regional), based on the sample size; (4) the specific technique for parametric estimation (method of moments, maximum likelihood, least squares, and so on). Therefore, a technician has many combinations for statistical analysis of extreme values, even in possible climate change scenarios.

In this framework, the paper describes a quick methodology implemented within the EXTRASTAR software (EXTRemes Abacus for STAtistical Regionalization), a user-friendly Excel file with macros in VBA (Visual Basic for Application).

EXTRASTAR was tested for the Italian network of annual maxima (AM) of daily rainfall time series, related to the SCIA database of the Italian Institute for Environmental Protection and Research-ISPRA (www.scia.isprambiente.it, accessed on 31 January 2022).

The obtained results have shown that EXTRASTAR can be considered a useful tool for obtaining quick indications about possible clusters of different time series in statistically homogeneous areas, as well as a reduction in the computational cost for onsite parametric estimates.

2. Methods and Material

2.1. Theoretical Background on Adopted Probability Functions

Firstly, assuming a stationary context, the following probability models were implemented in EXSTRAR for the statistical analysis of the AM series in terms of Cumulative Density Functions (CDF):

EV1 [1]

$$F_X(x) = P[X \leq x] = e^{-e^{-(\frac{x}{\theta} - ln\Lambda)}} \quad (1)$$

where Λ represents the mean annual number of occurrences with a magnitude greater than a prefixed threshold (under the hypothesis of a homogeneous Poisson process for the occurrences), while θ corresponds to the annual mean value for the associated intensities (by assuming an exponential distribution for this process).

GEV [2]

$$F_X(x) = P[X \leq x] = e^{-[1-b(\frac{x}{\theta} - ln\Lambda)]^{1/b}} \quad (2)$$

where, in addition to Λ and θ, there is the shape parameter b. In this case, preserving the hypothesis of a homogeneous Poisson process for the occurrences, it is assumed that the process of intensities above a threshold follows the Generalized Pareto distribution. It is well-known that GEV coincides with the EV1 for $b = 0$, while EV2 and EV3 distributions are obtained when $b < 0$ and $b > 0$, respectively.

TCEV [3]

$$F_X(x) = P[X \leq x] = e^{-\Lambda_1 e^{-\frac{x}{\theta_1}} - \Lambda_2 e^{-\frac{x}{\theta_2}}} \quad (3)$$

in which Λ_1 and Λ_2 (with $\Lambda_1 > \Lambda_2$) are the mean annual number for ordinary and outlier events, respectively, while θ_1 and θ_2 (with $\theta_1 < \theta_2$) are the correspondent mean values for intensities. A well-known TCEV formulation, used in contexts of statistical regionalization, is obtained by introducing two dimensionless parameters $\theta_* = \theta_2/\theta_1$ and $\Lambda_* = \Lambda_2/\Lambda_1^{1/\theta_*}$:

$$F_X(x) = e^{-\Lambda_1 e^{-\frac{x}{\theta_1}} - \Lambda_* \Lambda_1^{1/\theta_*} e^{-\frac{x}{\theta_* \theta_1}}} \quad (4)$$

Obviously, an ad hoc procedure for statistical regionalization can be developed for any probability distribution [4]. For the implemented models in EXTRASTAR, the reduced EV1 variable [3] is considered:

$$Y = \frac{X - \theta ln\Lambda}{\theta} = \frac{X - \theta_1 ln\Lambda_1}{\theta_1} \quad (5)$$

As Y is a linear transformation of X, these two random variables clearly present the same value for skewness.

Then, it is possible to rewrite Equations (1)–(4) as:

$$\text{EV1} : F_Y(y) = e^{-e^{-y}} \quad (6)$$

$$\text{GEV} : F_Y(y) = e^{-(1-by)^{1/b}} \quad (7)$$

$$\text{TCEV} : F_Y(y) = e^{-e^{-y} - \Lambda_* e^{-y/\theta_*}} \quad (8)$$

and to identify Homogeneous Regions (HRs), where the theoretical skewness, and therefore the parameters Λ_* and θ_* for the TCEV, or the parameter b (with $b = 0$ in the case of EV1) for the GEV, can be assumed as constant [5] or they follow laws based on geomorphological covariates (altitude, hydrographic basin area, etc.).

The developed methodology in EXTRASTAR (useful for either onsite or regional approaches) is based on the following steps:

1. Sample sizes N between 20 and 200 were considered. For each N and for each distribution, 5000 series of the variable Y were generated using the Monte Carlo methodology [6]. In detail: (i) for GEV, the following six values of parameter b were considered: 0 (EV1), −0.05, −0.1, −0.15, −0.2, −0.25; (ii) 32 combinations of (Λ_*, θ_*) were used for TCEV, with $0.1 \leq \Lambda_* \leq 0.5$ (step 0.1) and $1.5 \leq \theta_* \leq 5$ (step 0.5). Overall, 38 × 5000 samples of the standardized variable Y were generated for each value of N;

2. for each set of 5000 series, the 90% confidence band was evaluated for the sample skewness g; therefore, 38 confidence bands for each N can be represented within an abacus, implemented within the EXTRASTAR software.

By inserting, into the abacus, the information from AM time series, in terms of N and g, it is possible to quickly assess which distributions (and with reference values of the shape parameters) can model the sample skewness (and then the possible presence of outliers), and therefore to cluster some/many series in HRs.

It should be highlighted that this methodology can also be useful in climate change contexts, mainly for specific time resolutions (e.g., daily) for which the hypothesis of stationarity could not be rejected in many cases. A detailed, state-of-the-art trend analysis for the Italian AM rainfall series is reported in [7].

Moreover, it should be stated that the concept of "change" is not mutually exclusive with the term "stationarity" [8]. In fact, according to the examples with Newton's laws: (1) without an external force, the position of a body in motion changes in time, but the velocity is unchanged; (2) a constant force implies a constant acceleration and a changing velocity. Consequently, change is a general notion applicable to the real world, while stationarity and non-stationarity only regard the adopted models to explain the observed data. In some/many cases, changes are more evident for specific spatial and temporal resolutions, with respect to others (for which a stationary approach can be sufficient to model the sample series).

In this context, results from simple numerical experiments are below illustrated. Specifically, starting from an EV1 distribution in terms of the Y variable (Equation (6)):

1. A total of 5000 sets of 100 Y data were generated with the Monte Carlo methodology;
2. From Equation (5), the authors calculated the correspondent values for the X variable as:

$$X(t) = \theta(t) \cdot (Y + ln\Lambda) = \theta_0 \cdot (1 + \alpha \cdot t) \cdot (Y + ln\Lambda) \quad (9)$$

   i.e., by considering Λ as invariant, while an increasing linear trend for θ is assumed, where θ_0 and α are prefixed initial value and trend rate, respectively, and $1 \leq t \leq N$;
3. For a considered trend rate α, and varying the sample size N from 20 to 100, the Mann–Kendall test [9,10] was applied for each synthetic sample of X, for which it is easy to demonstrate that the statistic Z_{MK} does not depend on the initial value θ_0, because of these specific assumptions of EV1 population and invariance for Λ;
4. The percentages of synthetic samples with $|Z_{MK}| > 1.96$ (i.e., the null hypothesis of no trend is rejected at 5% significance level) are represented in Figure 1 for three values of trend rate α (10%, 20% and 50% in 100 years) and different invariant values for Λ.

The obtained results highlight that there is a significant percentage of synthetic samples for which the assumed change for θ does not imply $|Z_{MK}| > 1.96$, and then a stationary modelling could be adopted for statistical analysis in these cases.

2.2. Data Set

The main input for EXTRASTAR is a two-column .txt file (without headers), in which the ith row contains sample size N (first column) and sample skewness g (second column) of the ith AM series. In this work, the used input file was obtained from the SCIA database of the Italian Institute for Environmental Protection and Research-ISPRA (www.scia.isprambiente.it, accessed on 31 January 2022) regarding the network of daily AM rainfall series. Overall, 3351 samples with $N > 20$ years in the time interval 1860–2020 were analyzed. The mean value of N is approximately 50 years, with a maximum of 146 years (Milan rain gauge). The values of g varied from -0.71 to 5.10, with an average of 1.34. The software allows for the display of the pairs (N, g) on the abacus, together with the 90% confidence bands relating to EV1 law (inserted by default) and to GEV and TCEV distributions with values of parameters b (GEV) and Λ_*, θ_* (TCEV), which are selected by the user through a dedicated form. The abacus is structured in such a way as to represent

only one option at a time for both GEV and TCEV to simplify the visualization of the results (Figure 2).

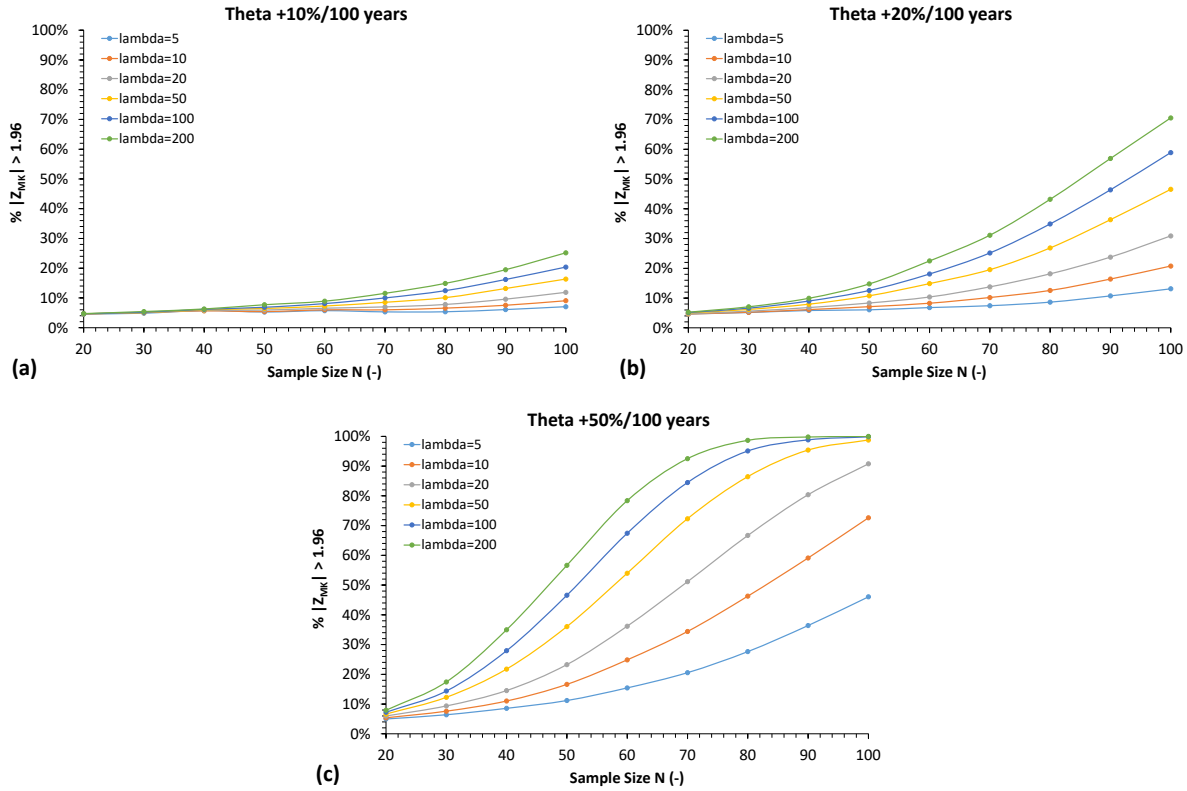


Figure 1. Percentages of synthetic samples (from a transient EV1 distribution) with $|Z_{MK}| > 1.96$ for a different trend rate related to θ: (**a**) +10% in 100 years; (**b**) +20% in 100 years; (**c**) +50% in 100 years.

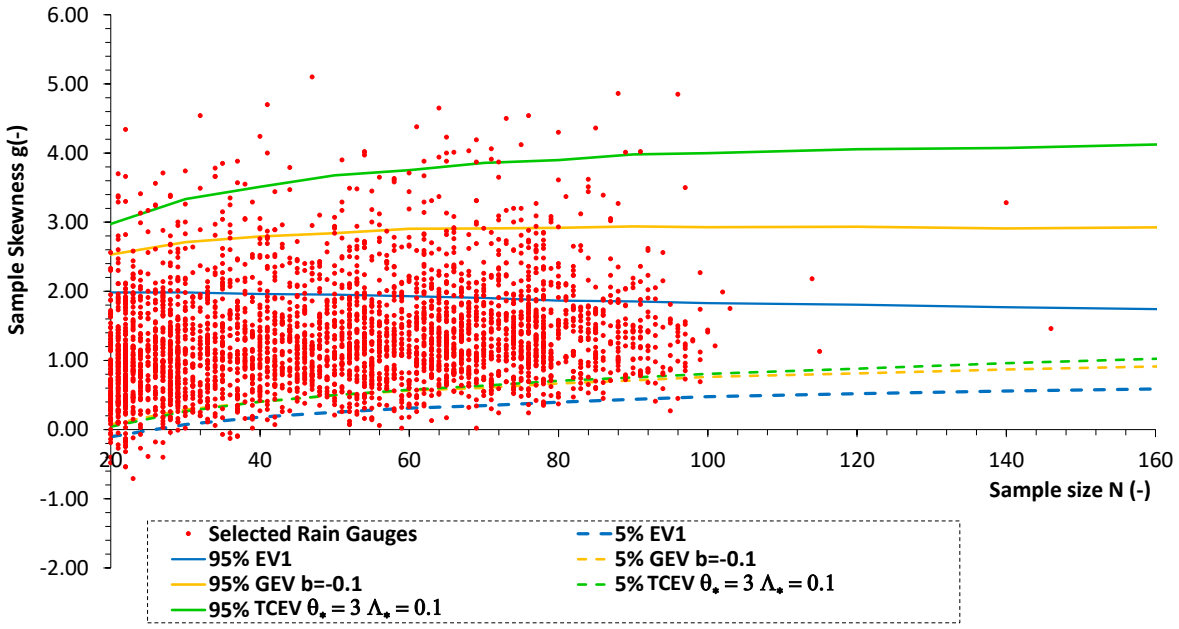


Figure 2. Plot of EXTRASAR abacus, with information from SCIA database of ISPRA (www.scia.isprambiente.it, 31 January 2022) concerning the Italian network of daily AM rainfall series.

3. Results and Discussion

From the analysis of the abacus, it is possible to assess that: (1) the EV1 model allows for the reconstruction of sample skewness for almost 80% of cases (i.e., 2679 samples); (2) the use of the shape parameter $b = -0.1$ (among all the considered values) provides the best modeling with the GEV law: 88.7% of cases, equal to 2971 sample series. Moreover, 2526 series from this subset (i.e., the 75.4% of the whole data set) can be modeled with EV1; (3) the sample skewness of 3064 time series (91.4% of the entire data set) can be reconstructed through the TCEV law with $\Lambda_* = 0.1$ and $\theta_* = 3$. From this subset, the 2520 series (75.2 % of the whole data set) can be modelled with the EV1 and 2961 series (88.4% of the entire data set) with GEV distribution with $b = -0.1$.

As an example of an application for a single daily AM time series, the authors focused on the Subiaco rain gauge (close to Rome), characterized by $N = 103$, g = 1.53, and an average of 68.8 mm with a standard deviation of 21.1 mm. The Mann–Kendall test provided a value of the statistic Z_{MK} equal to 1.39 over the entire time series and to 1.91 if we consider the data from 2000. In both cases, the absolute value of Z_{MK} is less than 1.96, then it is possible to not reject the hypothesis of a stationary process with a significance level of 5%. By inserting the information ($N = 103$, g = 1.53) on the EXSTRASTAR abacus, it emerges that Subiaco AM series can be modeled with EV1 and with GEV$-$b = -0.1 and TCEV$-\Lambda_* = 0.1$ and $\theta_* = 3$.

The Maximum Likelihood (ML) estimates for EV1 distribution provided $\theta = 15.2$ mm and $\Lambda = 50.7$. The ML technique usually allows for an excellent fitting of the data with respect to ordinary values. Assuming these estimates also for the corresponding parameters of GEV and the ordinary component of TCEV, authors obtained the comparison on the EV1 probabilistic plot (Figure 3a), from which the EV1 and TCEV laws could be considered inadequate. On the contrary, if 5000 synthetic Monte Carlo generations (each one with $N = 103$) are carried out for each considered probabilistic laws, and the corresponding 90% confidence bands are estimated, it is clear that each distribution may not be rejected for the modeling of the investigated sample series (Figure 3b), thus confirming what is hypothesized with the abacus analysis. Obviously, in the context of parsimony, a two-parameter distribution would be preferred.

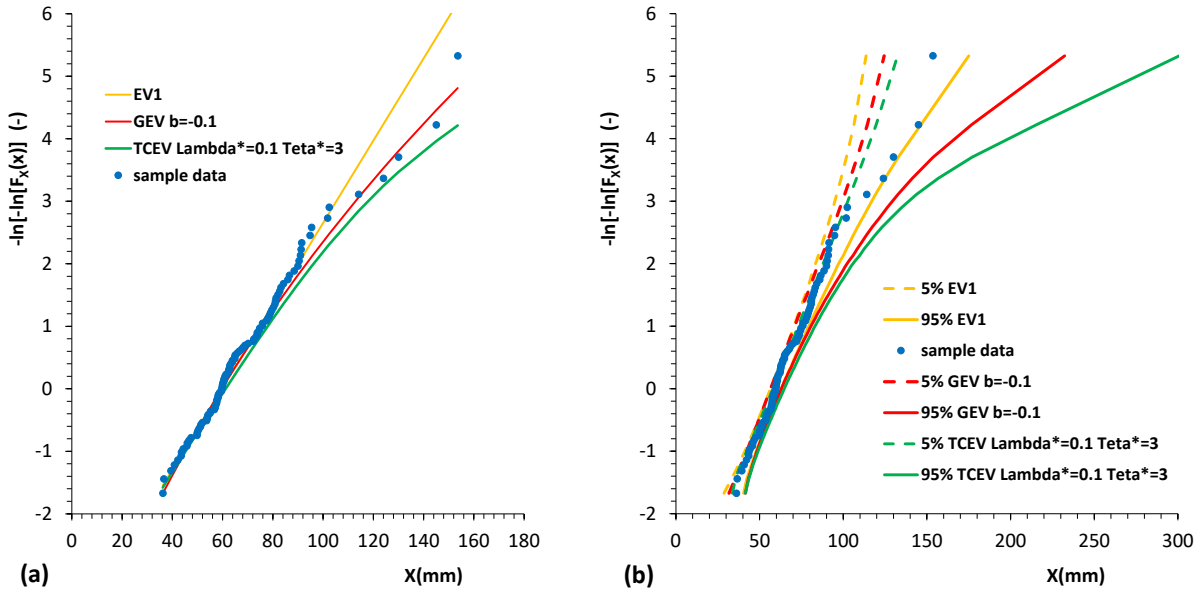


Figure 3. Daily AM series of Subiaco rain gauge: (**a**) comparison with EV1, GEV ($b = -0.1$) and TCEV ($\Lambda_* = 0.1$ e $\theta_* = 3$) distributions; (**b**) comparison with 90% confidence bands for the investigated distributions.

4. Conclusions

The EXTRASTAR software can undoubtedly represent a user-friendly tool for the statistical analysis of annual maxima, even in the context of a climate change scenario. For an observed time series of interest, it is possible to quickly test the modelling capability for EV1 and GEV or TCEV, also with reference values for the shape parameters (b for GEV, and Λ_* and θ_* for TCEV), on the basis on the sample size and skewness.

Furthermore, by simultaneously analyzing the sample skewness of several time series, a user can obtain a quick indication of potential clusters in homogeneous regions.

Furthermore, the use of EV1 Maximum Likelihood estimates also for GEV and the ordinary component of TCEV allows for a reduction of the computational costs with respect to carry out a specific calibration procedure for each probability function.

Future developments will concern the implementation, within EXTRASTAR, of specific modules for modeling with non-stationary approaches to account for any effects of climate change.

Author Contributions: Conceptualization, D.L.D.L., B.M., F.R. and F.N.; methodology, D.L.D.L., B.M., F.R. and F.N.; software, D.L.D.L., B.M., F.R. and F.N.; validation, D.L.D.L., B.M., F.R. and F.N.; formal analysis, D.L.D.L., B.M., F.R. and F.N.; investigation, D.L.D.L., B.M., F.R. and F.N.; data curation, D.L.D.L., B.M., F.R. and F.N.; writing—original draft preparation, D.L.D.L., B.M., F.R. and F.N.; writing—review and editing, D.L.D.L., B.M., F.R. and F.N.; visualization, D.L.D.L., B.M., F.R. and F.N.; supervision, D.L.D.L., B.M., F.R. and F.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gumbel, E.J. *Statistics of Extremes*; Columbia University Press: New York, NY, USA, 1958.
2. Jenkinson, A.F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.* **1955**, *81*, 158–171. [CrossRef]
3. Rossi, F.; Fiorentino, M.; Versace, P. Two-component extreme value distribution for flood frequency analysis. *Water Resour. Res.* **1984**, *20*, 847–856. [CrossRef]
4. Ferrari, E.; Versace, P. La Valutazione delle Piene in Italia. Rapporto di Sintesi GNDCI. 1994. Available online: http://www.idrologia.polito.it/gndci/rapportiPdf/Vapi_Nazionale.pdf (accessed on 31 January 2022). (In Italian).
5. Gabriele, S.; Arnell, N. A hierarchical approach to regional flood frequency analysis. *Water Resour. Res.* **1991**, *27*, 1281–1289. [CrossRef]
6. Kottegoda, N.T.; Rosso, R. *Applied Statistics for Civil and Environmental Engineers*; Blackwell: Oxford, UK, 2008.
7. Caporali, E.; Lompi, M.; Pacetti, T.; Chiarello, V.; Fatichi, S. A review of studies on observed precipitation trends in Italy. *Int. J. Climatol.* **2021**, *41*, E1–E25. [CrossRef]
8. Koutsoyiannis, D.; Montanari, A. Negligent killing of scientific concepts: The stationarity case. *Hydrol. Sci. J.* **2014**, *60*, 1174–1183. [CrossRef]
9. Mann, H.B. Non-parametric tests against trend. *Econometrica* **1945**, *13*, 245–259. [CrossRef]
10. Kendall, M.G. *Rank Correlation Measures*; Charles Griffin: London, UK, 1975.