1. **INTRODUCTION**

The effects of a change in the frequency and intensity of hydrologic events (floods and droughts) have motivated trend detection studies to support water resources management planning. Several factors influence the ability to detect true trends, such as the presence of temporal autocorrelation, the question of multiplicity of tests in regional analyses, and the fact that the utility of the test used in the analysis and the degree of bias in the estimation of the magnitude of the trend depend on characteristics of the historical series (size, degrees of variability and asymmetry, chosen distribution, and actual magnitude of the trend).

The lack of knowledge about how these factors affect the detection analysis leads to errors in the identification of stationarity and have been a motivation for studies that aimed to contribute to fill this gap. (Yue et al., 2002a; Yue et al., 2004; Pittock, 1980; Chiew and McMahon, 1993; Radziejewski e Kundzewicz, 2004 and Gelman and Carlin, 2014)

Based upon Monte Carlo simulations results, Yue et al. (2002a) investigated the power of Mann Kendall’s test and Spearman’s Rho when applied for trend detection in independent and non-correlated time series. The authors generated 2,000 samples distributed as Normal, Log-normal, Pearson type III, EV1, EV2 and EV3 with different variances, samples sizes, trend’s magnitude and significance level of the test. As result, Yue et al. (2002a) found that the power of both tests raises as trend’s magnitude and sample size increase. On the other hand, increase on variance causes reduction in capacity of detect a trend when it is statistically true. Authors also pointed that the type of statistical distribution and skewness influence power of the test when a time series is declared stationary. In other words, EV3 and Log-Normal are the distributions, respectively, with greater and lower associated power when results are compared under the same conditions.

Also using Monte Carlo simulations, Yue et al. (2004) proposed to evaluate the power of MK, t-test, *BS-slope* and *BS-*MK when applied in linears and non-linear trend’s detection process. The authors used normally and non-normally (P3, Gumbel, EV2 and Weibull) independent time series with mean (, coefficient of variation (, linear trend’s magnitude (, significance level and sample size. Specifically for non-normal samples, authors also adopted: EV2: ; Weibull: ; Pearson III (P3): . The simulations results showed that (a) o *t-test* and *BS-slope* have the same power when samples are normally distributed. The same was seen for MK and *BS-MK*; (b) if the sample is Normal, the power of the tests based in slope coefficient analysis, *t-test* and *BS-slope*, are slightly greater than others evaluated. (c) Considering non-Normal samples, MK and *BS-MK* are the ones pointed as better tests to be used by Yue et al. (2004).In addition, this study reveals that the ability of detect truly trends is slightly sensible to the type of non-stationarity, i.e. if the trend is linear or not.

It is important to keep in mind that sometimes tests will not be able to declare a sample as non-stationary even when a true trend is present. Pittock (1980) and Chiew and McMahon (1993) evaluate this capability for hydrologic time series. The authors, in general, concluded that detect a trend is unlike if it does not appear consistently for a specific time interval. Considering this statement and difficulties in detection process when the signal of change is week and sample’s variability is high, Radziejewski e Kundzewicz (2004) proposed to generate synthetic series by the association of two distinct parts: (a) *base*: obtained by a normally distributed and independent random process ( and (b) *change*: came from a two types of deterministic process: *gradual* (linear) and *jump*. The parcels’ superimpose was controlled by the authors and objected create time series with two distinct portions and limited by a point called *changing point.* This point established the beginning of time series’ portion that was affected by the change (*run up time*) or the end of interval that trend does not exist.(*reference period*). Radziejewski e Kundzewicz (2004) defined 30 years as fixed value for *reference period* and 10 to 70 years for the *run up time*. The authors also defined the variance of synthetic times series as multiple of the base period’s standard deviation (SD) In other words, the variance for *gradual* and *jump* time series was defined, respectively, as 0.05SD e 1SD. After applied MK, Spearman’s Rho, linear regression for monotonic series and CUSUM, cumulative deviation (Buishand, 1982) for the ones with *jumps*, Radziejewski e Kundzewicz (2004) concluded that if a test did not detect a trend, does not mean necessarily that serie is stationary. The authors noticed that this fact is more likely to happen when trend’s magnitude is weak or appear in a short period of time, as previously informed by Pittock (1980) and Chiew and McMahon (1993).

Gelman and Carlin (2014) proposed a different approach for a trend detection analysis. The authors suggested that the evaluation of stationarity can be complemented by the analysis of Type S (*signal*) and Type M (*magnitude*) associated with B\_SEN. Instead using exclusively the conventional approach (errors type I and II), these authors pointed that is possible to estimate erroneously the signal and trend’s magnitude, especially when the samples tested have high noise and small sample size.

The absence of studies that focus on trend’s detection analysis in hydrological extremes time series associated with a poorly evaluation seen in literature about bias on trend’s estimates provided by B\_SEN related to hydrology motivated the development of this paper. In other words, the purpose of this paper is realize an evaluation of trends based on Monte Carlo simulation studies, which vary the size of the historical series (n); the magnitude of the trend (b); the coefficient of variance (Cv) and the coefficient of asymmetry of the historical series; the form of the function (k) of the distribution of Generalized Extreme Values (GEV); local influence of the power of the Mann-Kendall (MK) test; the probability of obtaining an estimate of the magnitude of the trend with an opposite sign of the actual (error type S) employing the estimator suggested by Sen (1968); and the degree of bias of this estimate (type M error).

The next three sections are fairly distinct. In section 2, we describe the methods used for a trend detection analysis, Mann-Kendall, and to estimate the magnitude and signal of a time series trend, B\_SEN. The third section presents the methodology used to generate GEV’s time series with trend and expressions defined to evaluate Monte Carlo simulations results using MK test and B\_Sen. Section 4 presents results and their analysis. The last one, Section 5, reveals the authors’ conclusions based upon content of previously sections.

1. **METHODS**

Given a dataset X consisting of *x* values with sample size *n*, the MK calculation starts by estimating the S statistic.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

where

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

As indicated in Mann (1945) and Kendall (1945); Kendall (1975) when the distribution of *S* approaches the Gaussian form with mean E(S) = 0 and variance V(S) given by:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

where is the number of ties of length *m*.

The statistic S is then standardized resulting in the MK final value. The significance of the MK statistic can be estimated from the normal cumulative distribution function. Positive (negative) MK values indicate the presence of increasing (decreasing) trends.

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Considering a two-tailed test with a significance level equals to a%, the null hypothesis is rejected if or , i.e. if the estimated value of *Z* falls into a critical region *R.*

Sen (1968) proposed an estimator of trend’s magnitude due the fact that the least squares estimator of a regression coefficient (*b*) is vulnerable to gross errors and the associated confidence interval is, in addition, sensitive to non-normality of the parent distribution. This estimator can be defined as the median of the set of slopes joining pairs of points with

1. **SIMULATION STUDY**

Monte Carlo simulation was conducted to evaluate the power of the MK test when applied to series that try to mimic observed extreme hydrological time series. The experiment generates 10.000 independent GEV’s time series with mean of 1.0, considering different values of coefficient of variation ( and shape parameter (. Then a trend with slope is superimposed onto each onto of generated series, as given by:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

As (0,, , , …, ), correspondingly, the mean (=1.0) over 100 years will increase ( = , i.e., increases/decrease 0 (10) 500%. The rate of rejecting the null hypothesis of no trend can be given by:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

where *N* is the total number of simulation experiments and is the number of experiments that fall in the critical region *R*, estimated considering MK test’s significance level equals to 0.05 (5%).

The evaluation of B\_SEN, in terms of type S and M errors when ,, *b* and *n* vary, was conducted considering the generated times series that were evaluated as stationary by MK test. The type error S was estimated by:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

where the number of times that estimated trend’s signal by B\_SEN is different from the assumed one during the time series generation process (*b*) and

Type M error was evaluated using an indirect metric that tries to capture the range of trend’s magnitude estimates by B\_SEN for stationary time series associated with a specific,, *b* and *n.*  Named as type M error relative width (, this metric can be given by:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

**4. RESULTS**

The relationship among power, sample size (*n*), shape parameter (κ), and ratio signal-noise is summarized in Figure 1 for a fixed significance level (α = 0.05). The power of MK test is an increasing function of both the sample size and ratio signal-noise. In other words, as the sample size and ratio signal-noise increase, the power of the test increases, leading to an increased ability to discern the existence of a trend. In Figure 1, considering fixed ratio and κ values, it is possible to notice that the power associated with a specific curve increases as values of *n* become higher (i.e. moving in the positive direction of x-axis). As an example, if and κ=-0.30, then the powers of MK test are 0.45 and 0.70, respectively, when n =30 and 40. This same behavior can be observed for a ratio signal-noise. As values of increase, positive direction of y-axis, the power is higher. Figure 1 also shows that distances between power’s curves for κ = -0.30 are smaller than ones seen for κ = +0.30, considering same value of *n*. This fact indicates that the relationship among the power and shape parameters is negative. An example, when n=30, = 0.01 and shape parameters are equal to -0.30 and +0.30, powers are, respectively, 0.20 and less than 0.1.

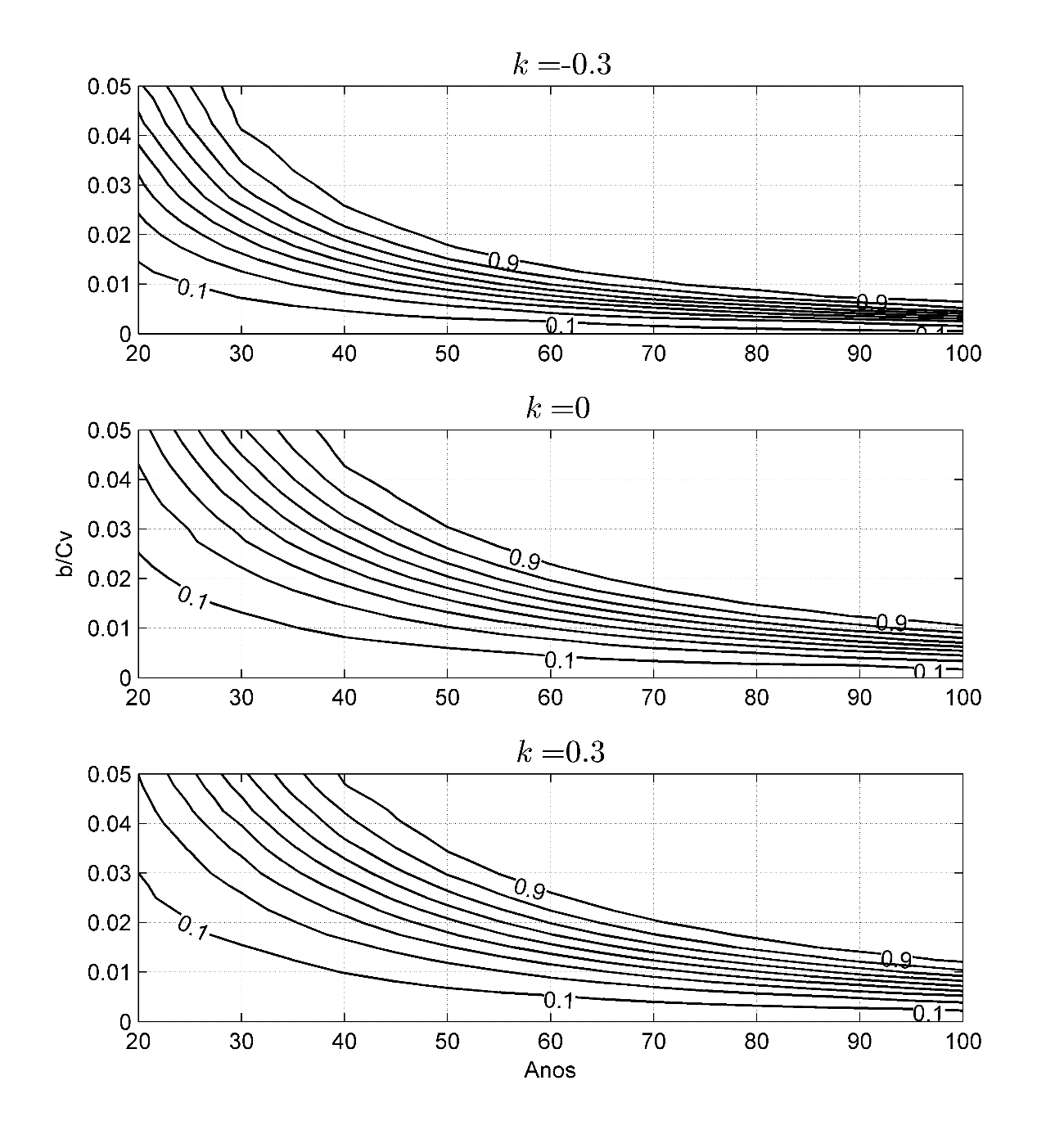


Figure 1

Sometimes the variation of MK test’s power is evaluated in function of two independent variables, the trend’s magnitude (b) and coefficient of variation (. The coefficient of variation is a decreasing function of power, i.e. decrease the ability of MK test in detecting true trends as increases. The opposite behavior is seen for the trend’s magnitude (b). The simulation’s results confirmed that effect of b(signal) and (noise) can be expressed as a single ratio (, which a high trend magnitude can compensate the drawback caused by a high coefficient variation in terms of power and vice versa. Thus, the use of is more appropriated to represent the effect of those variables in a trend detection analysis.

Figure 2 tries to express how confident MK test trend results are, in terms of power, and based upon different values of ratio signal-noise. This figure defines three distinct regions: n = 30 (yellow), n = 50 (green) and n = 80 (purple), where it is possible to observe their upper (UL) and lower limits (LL). As expected, the yellow region, associated with the smallest sample size (n=30), is the widest one for a specific value of signal-noise ratio, i.e. the difference between UP and LL is greater than is seen for the other two regions.

As previously mentioned, an increase in signal-noise ratio results in higher MK test’s power. This aspect also is showed in Figure 2, where moving in positive direction of x-axis results in more power and a wider region. An example is the yellow region which varying from 0.01 to 0.02 results in 100% greater UL and LL. Another important aspect still related with yellow region is the fact that for high signal-noise ratio values, the possibility to detect a true trend in an extreme time series is relatively low, for example, 0.50 is the power, on average, associated with = 0.03 and n = 30. Thus, recommend 30 as the minimum sample size to realize a trend analysis seems to be a questionable assumption, especially seen the differences in power when you use n = 50 instead 30. Considering the same ratio signal-noise, for example 0.02, the upper limit of the green region is more than twice as large as the yellow one. In other words, the value of yellow UL (n=30) is 0.40 and increase for approximately 0.97 when is considered a twenty years longer time series (n =50).

In addition, the importance of having longer time series is showed in Figure 2, especially in cases where the ratio signal-noise is lower than 0.01. In this case, only to illustrate, Figure 2 indicates that the maximum power associated with a sample size equals to 50 is 0.50, i.e. the MK test has 50% chance to detect a true trend. Then, the use of longer time series should be preferred in situations where the availability of gages is representative of the study area.

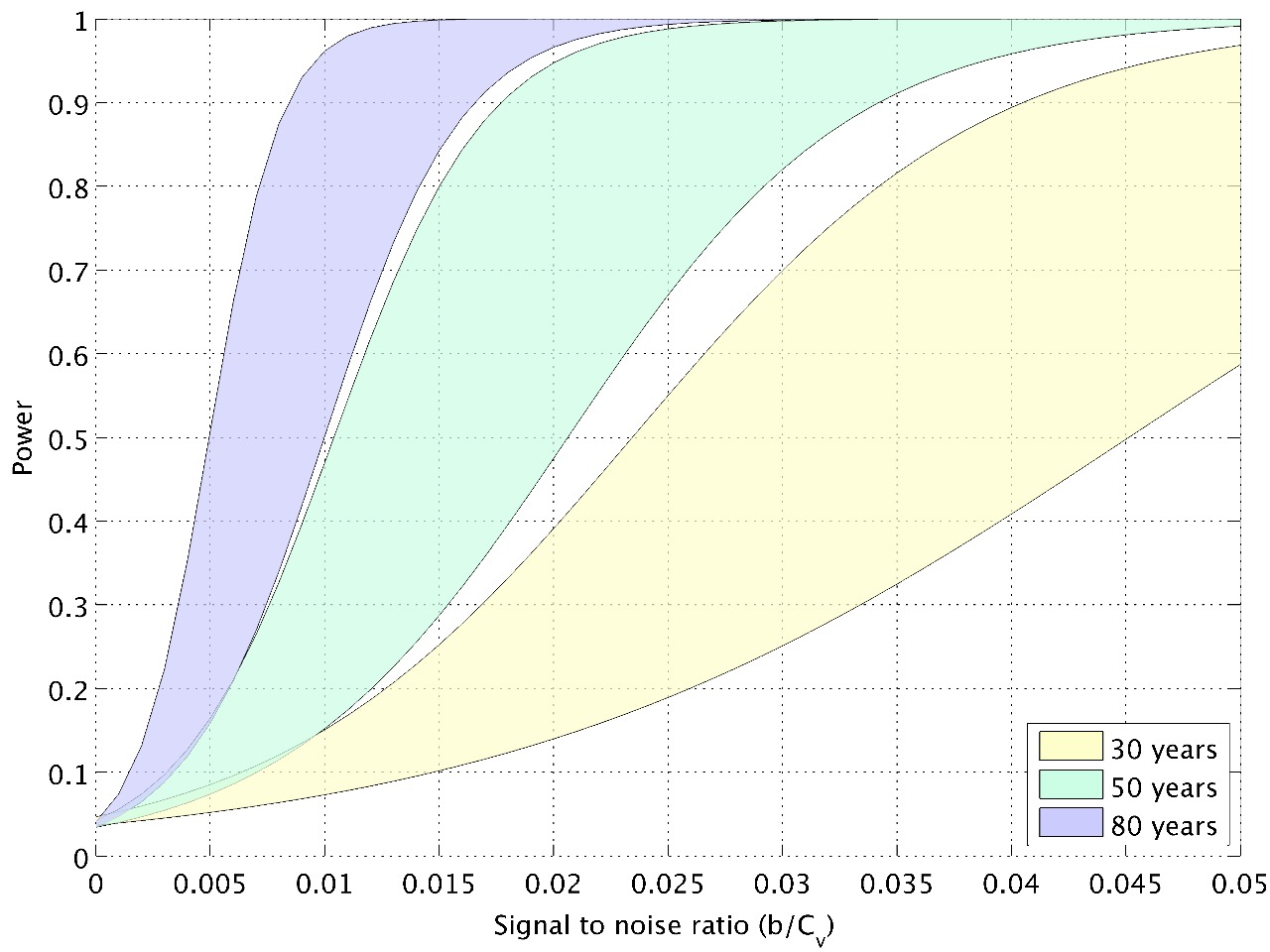


Figure 2

Figure 3 shows the relationship among probability of occur a type S error using B\_SEN and GEV sample sizes, coefficient of variation and trends magnitude when κ = -0.30. Regarding sample size, this Figure indicates that trend’s signal estimated by B\_SEN, i.e. decreasing or increasing, has greater probability of being correct as sample size increase. Thus, type S error is a decreasing function of sample size. As an example, with = 1.00 and =2.00%, more extreme case considered in this paper, type S error occurrence probability is 36.4% for n = 20, 18% for n = 40, 8% for n = 60, 2.5% for n = 80 and less than 1% for time series with n = 100.

Similar behavior seen for sample size occur with an increase of trend’s magnitude. The decreasing relationship is expressed by a reduction in a size of Figures 3’s bars as move from left to right, positive direction of x-axis. Considering n =40, = 1.0 e b = 2.00%, type S error occurrence probability is equal to 18.4% varying to 6.3% when trend’s magnitude increases to 4.00%.

Also an increase is noted in Figure 3 in the type S error occurrence probability as become greater. This fact can be illustrated for time series with n = 20, =0.20 and b =2.00% that have the coefficient of variation raised to 0.60 and 1.00. As a result, the probability of erroneously estimates the signal of trend increases from 7.6% (0.20) to 29.5% (0.60) and to 36.4% (1.00). Even for different values of shape parameter, the positive relationship among type S error occurrence probability and coefficient of variation persists when other parameters evaluated here were kept constants, as showed in Figure 4.

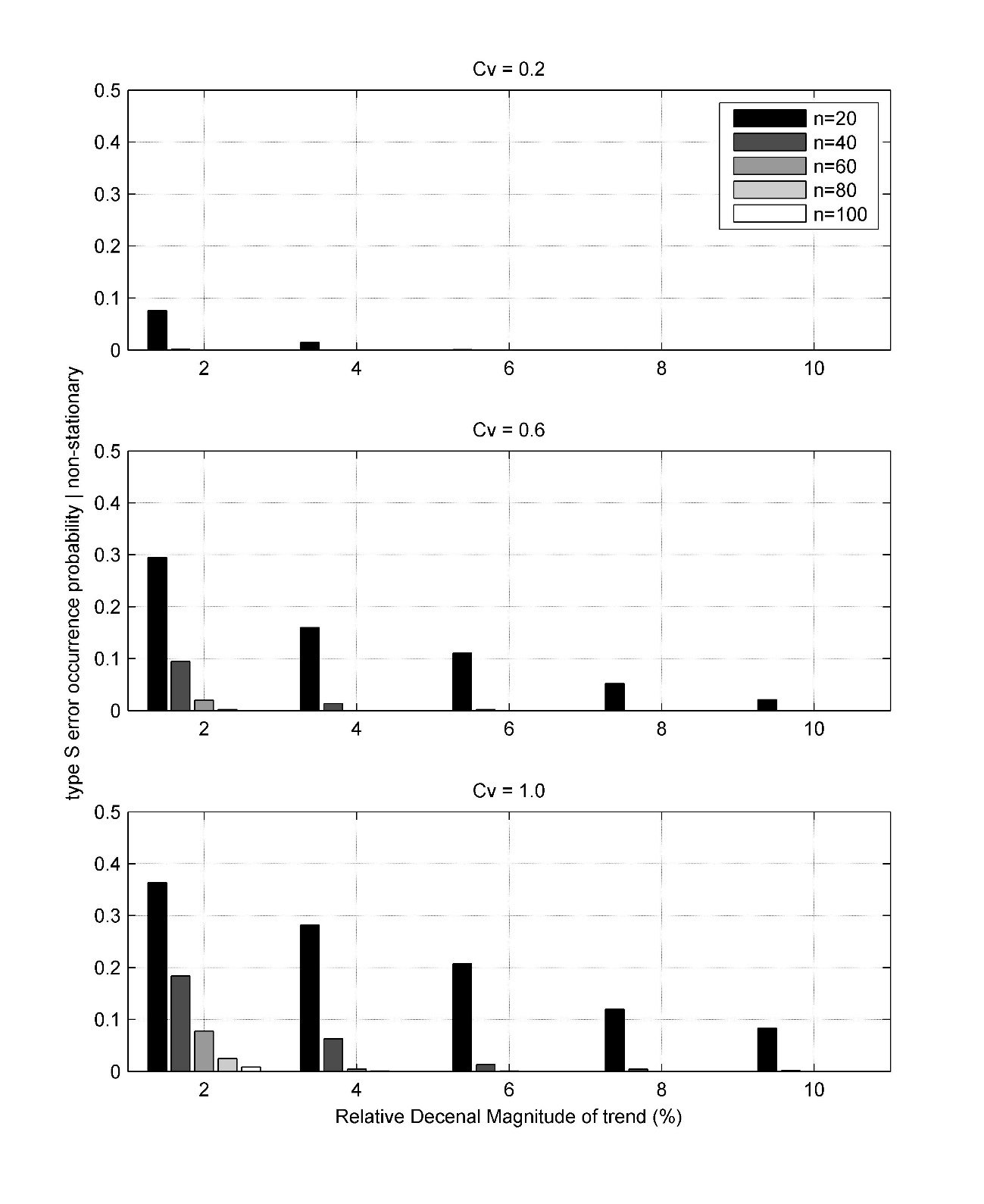


Figure 3

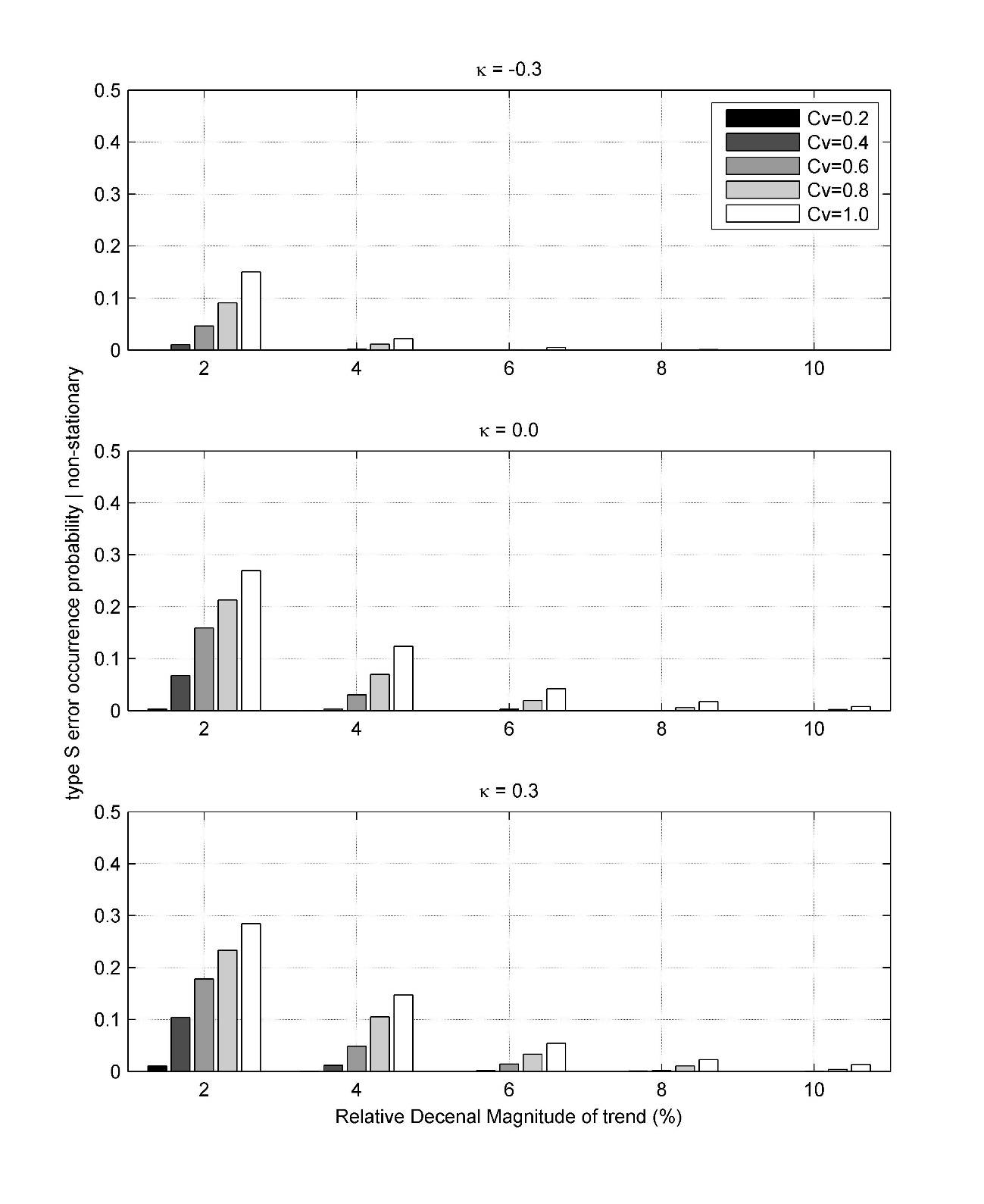


Figure 4

Figure 4 also express the positive dependency between GEV’s shape parameter and the capability of B\_SEN estimates correctly signal of change, i.e. lower and greater values of type S error occurrence probability are associated with, respectively, lower (-0.3) and greater (+0.3) κ values when n, and *b* are fixed. As an example, the difference between type S error occurrence probability for κ=-0.30 e κ=0.30 is equal 13.2%, considering n = 50, = 0.60 and b = 2.00% constants.

Until now, the results reveal that the probability of erroneously identify the signal of change is relatively high in time series with high annual variability and low trend’s magnitude (also κ’s values dependent), especially in samples with size lower than 60 years of record.

Figure 5 shows that type S error occurrence probability can be significant when the power of the test is lower than 0.15. This fact can complicate trend detection studies, resulting, for an example, in a situation that neighbor’s gages are identified as nonstationary but with opposite signal of change. This problem can be aggravated when the power of test is low, because B\_SEN tends to overestimate the real trend’s magnitude.

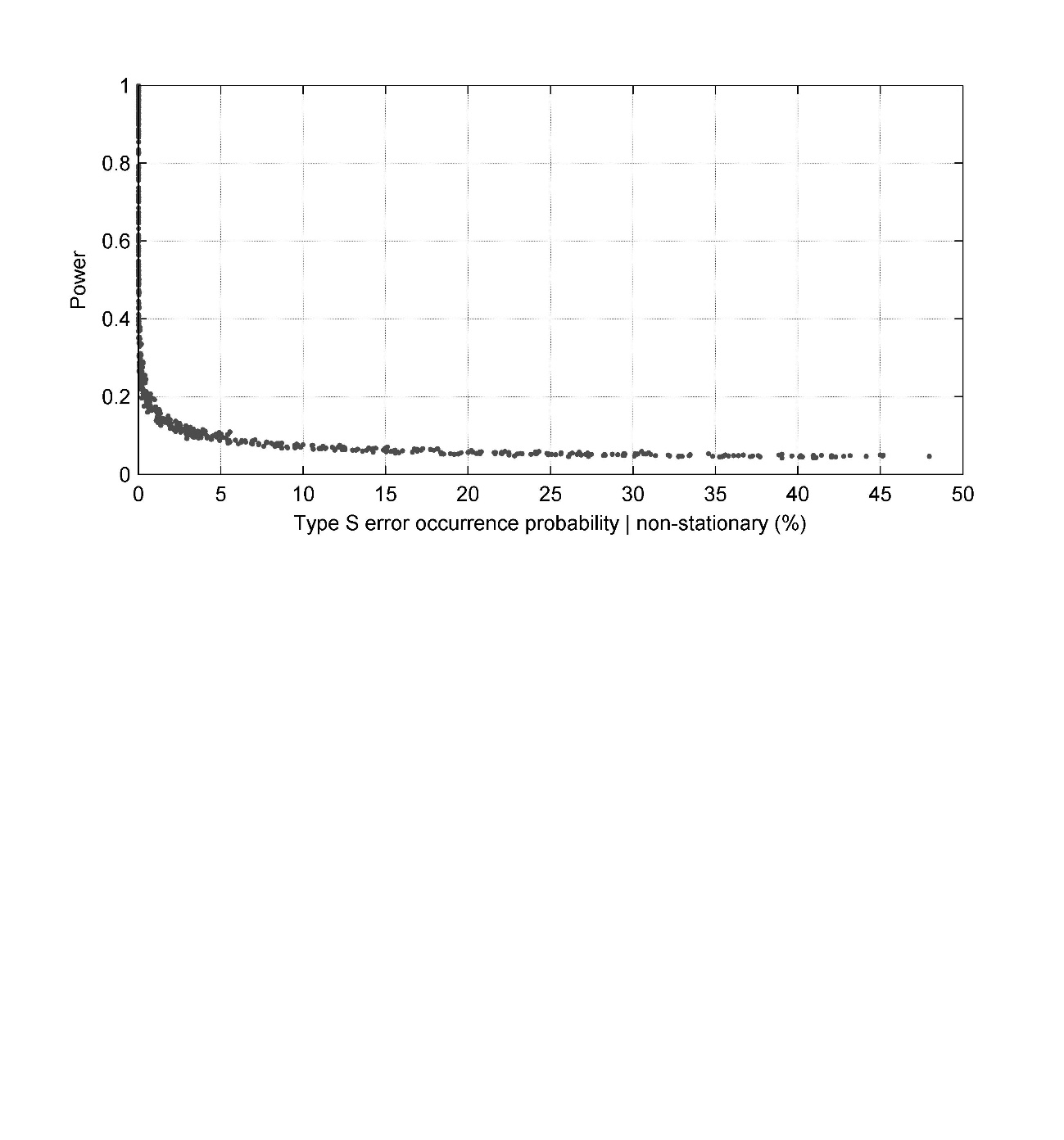


Figure 5

Monte Carlo’s results also allowed evaluate the bias and variance of B\_SEN when applied for a different extreme time series, i.e. GEV’s with different sample size (n’s), coefficient of variation (), trend’s magnitude (b’s) and shape parameter (κ’s). Figure 6 express the relationship among statistical test’s power and relative bias of B\_SEN when a sample is declared as nonstationary by MK test. In Figure 6 can be seen that B\_SEN tends to overestimate the real trend’s magnitude, especially when the power of test is low. As an example, in situations where the power of MK test is approximately 0.50, the estimated value of change by B\_SEN is, on average, 40% greater than the real one. In addition, if the power’s test is equal to 0.20, B\_SEN can predict magnitudes of trends 100% greater, i.e. double. This distortion can affect interpretation of results and, consequently in water resource management and planning or in other decisions that are taken based upon a possible wrong estimated trend.

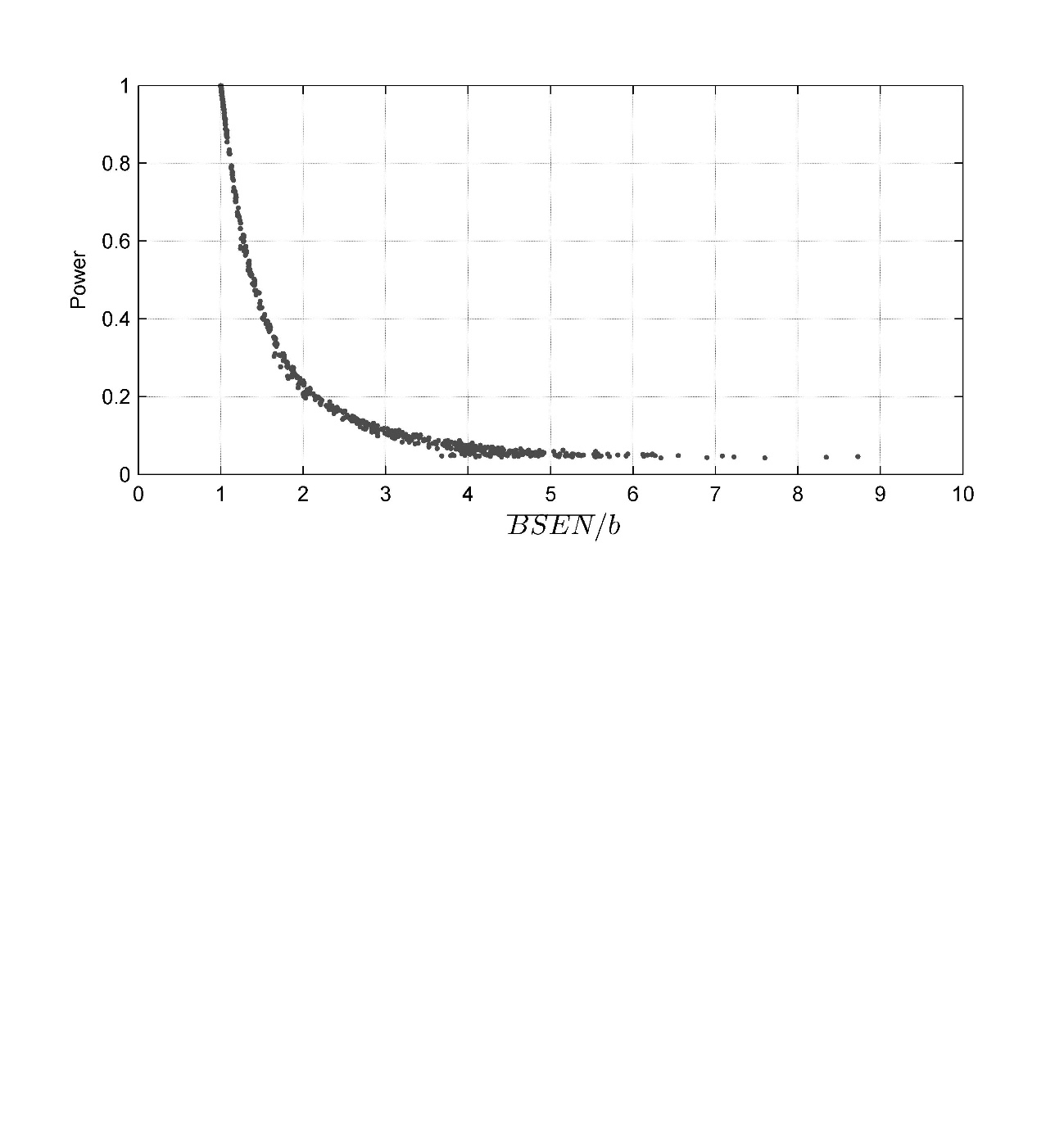


Figure 6

The variance of B\_SEN was analyzed considering the difference between the 2.5% and 97.5% percentiles of estimated trend’s magnitude divided by the real value of change (*b*) used during the generation process of 10,000 GEV series with different sample sizes, coefficient of variation and shape parameter. Named as *Relative Width of Type M Error (L)*, this ratio was estimated only taking account of series that were previously considered as nonstationary by MK test. In general, the results presented on Figures 7 and 8 showed that *Relative Width of Type M Error* can be equal to 20 in situations where the power of the test is low.

The Figure 7 represents the dependency between sample size and *Relative Width of Type M Error,* considering a fixed value of κ = -0.30 and b = 0.006 and differents coefficient of variations’ magnitudes (). This figure shows that as the number of years available increases the confidence interval (IC) become narrower. Opposite behavior is seen for coefficient of variation: an increasing in Cv results in wider IC, i.e. increase in variability of trends’ values estimated by B\_SEN. The coefficient of variation effect over the IC is more pronounced in reduced samples (n< 40 years), an example is the difference in *Relative Width of Type M Error* estimate with Cv = 0.20 and 1.00, respectively, 1.1 and 25 times b’s values.

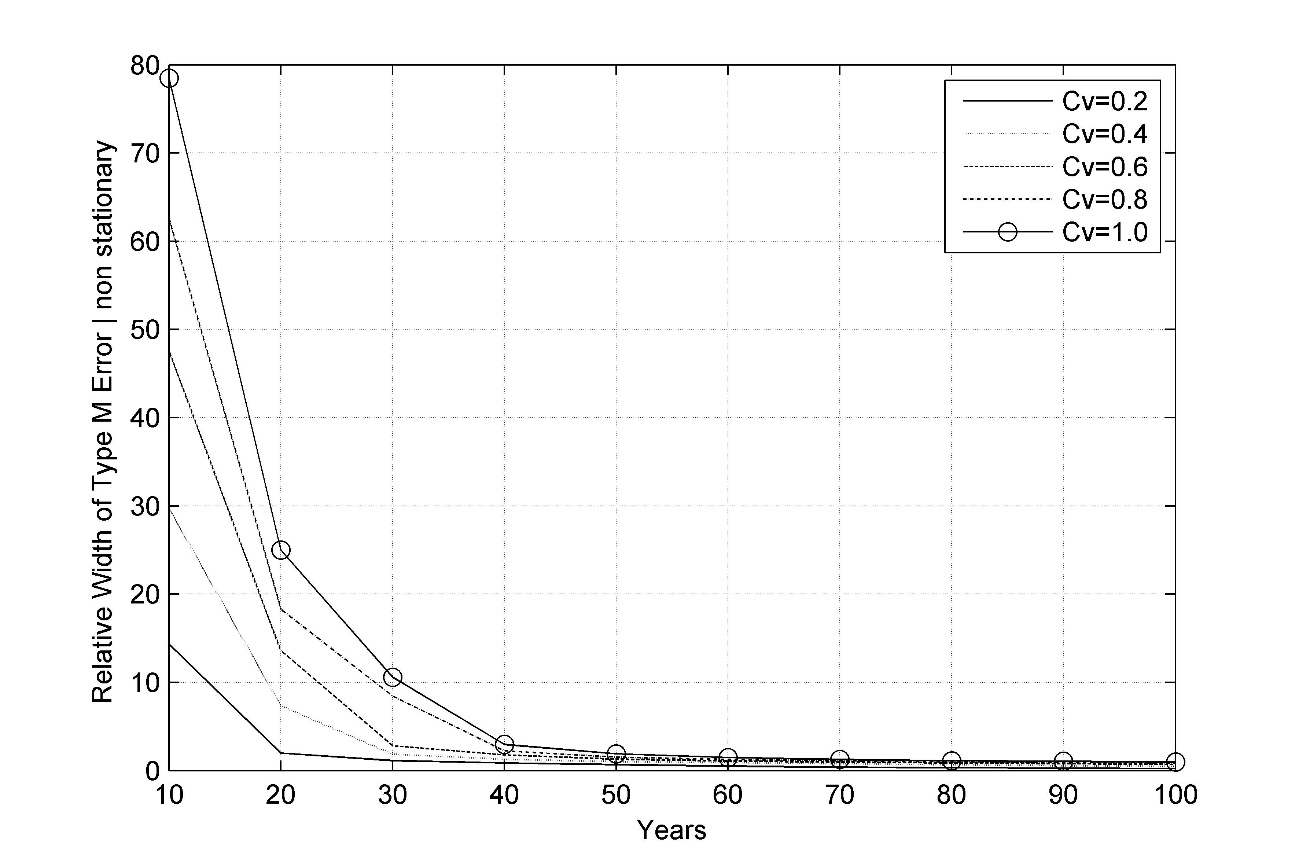


Figure 7

The relationship among GEV’s shape parameter and *Relative Width of Type M Error* is showed in Figure 8. Adopting a level of confidence equals 5%, b = 0.006, Cv = 0.6 and κ = [-0.3 -0.1 0 +0.1 +0.3], curves express similar behavior as seen for Cv: IC become narrower as sample size increases, reaching the expected value *b* for time series with sample sizes greater or equal to 60 years. The shape parameter effect is more severe for reduced samples (n < 40 for κ < - 0.1 and n < 60 for κ > 0) and act to scatter B\_SEN’s results. For example, the *Relative Width of Type M Error* is equal to 25 times the value of b when n = 20 and κ = +0.30.

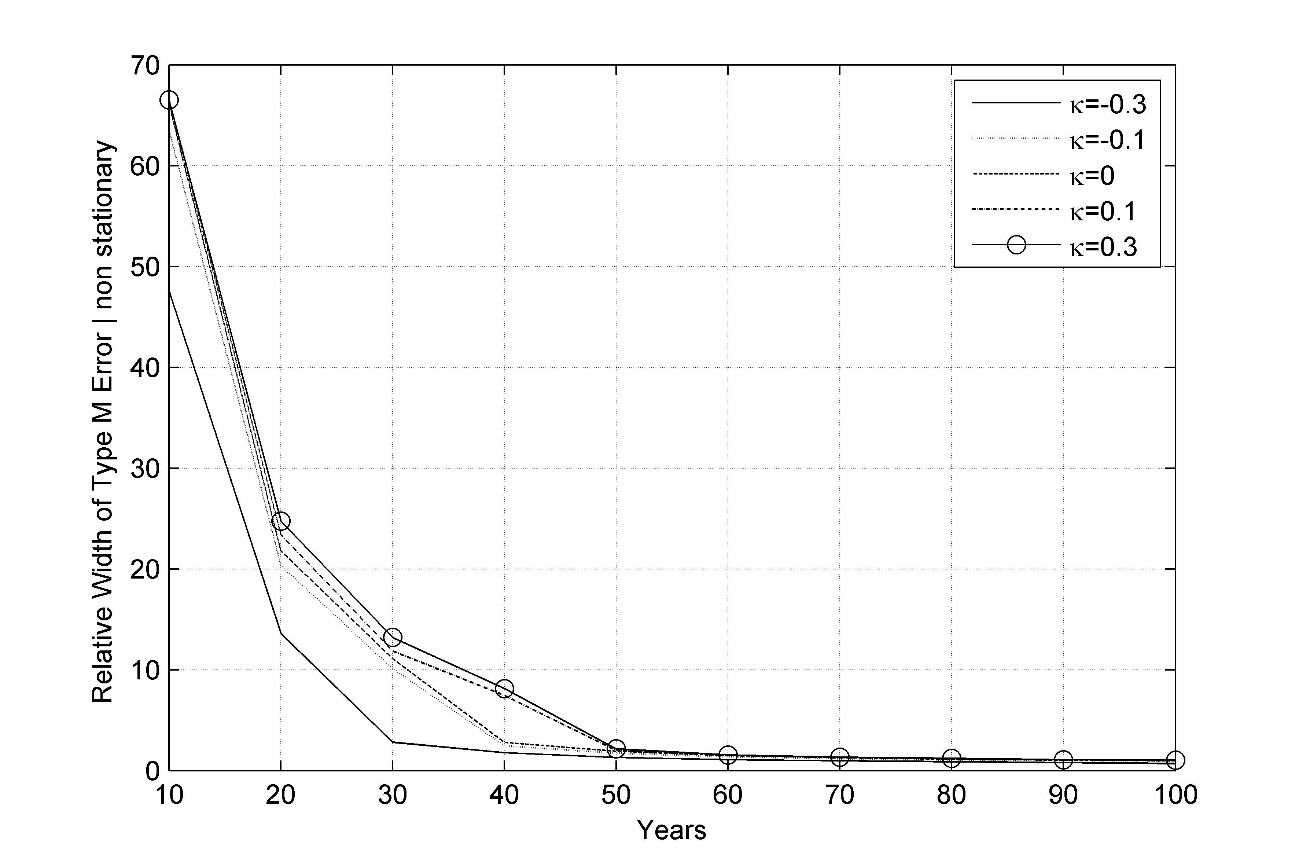


Figure 8

**CONCLUSIONS**

This study realized an evaluation of trends based on Monte Carlo simulation, which vary the size of the historical series (n); the magnitude of the trend (b); the coefficient of variance (Cv) and the coefficient of asymmetry of the historical series; the form of the function (k) of the distribution of Generalized Extreme Values (GEV); local influence of the power of the Mann-Kendall (MK) test; the probability of obtaining an estimate of the magnitude of the trend with an opposite sign of the actual (error type S) employing the estimator suggested by Sen (1968); and the degree of bias of this estimate (type M error).

The simulation experiments have allowed mapping the power of the MK test to different combinations of n, Cv, κ and b for extreme series of GEV, enabling a better understanding of the likelihood of identifying a season with a tendency when a true trend it is present. As a result and in agreement of previously studies developed for normally distributed time series, an increase in a sample’s size or a reduction in signal-noise ratio (b/Cv) result in an increase of MK’s test power for GEV times series. On the other hand, as asymmetry coefficient k reduces (become more negative) the power of MK test decreases. For example, a series with 30 years, Cv = 1, k = -0.3 and an annual relative trend of 2%, the likelihood of the MK test identifying this trend is only 0.45. If this same series had a length of 50 years, the likelihood of detection would increase to 0.90. This aspect suggests that a use of a time series with a length of 30 years, value usually accepted in the literature as adequate for detection studies, is controversial and need to be more discussed. In addition, another example that confirm the previously statement is the low percentage of MK’s correct detection, 45%, 20% e 17%, when is evaluated a time series with n=30, Cv = 1, b =20% of change in a decade and, respectively, ** = -0.30, ** = 0.0 e ** = +0.30. Considering just an increase in the sample size to 50, for example, the power of MK test increases to 90%, 60% e 50%.

Monte Carlo’s results also suggests that the likelihood of estimating a trend with the opposite sign of the true trend (when the station is determined to be significant) depends on the power of the test and can assume values greater than 5%, and can reach 40% for test powers less than 10%. It was also verified that there is a tendency of overestimation of the magnitude of the trends, and that the estimate value (on average) can be 1.5 times greater than the actual value when the test power is equal to 0.40, and can still be up to 5 times great than the actual value (on average) when the power of the test is equal to 0.05.

Indeed, neglecting the conclusions presented during this study in a water resources framework, for example, can induced the decision-makers to adopt an approach that cannot result in future benefits to society and also can induce the use of engineering measures or a planning process when it’s in fact not necessary.