**Hydrology and
Earth System
Sciences**

Flood trends and variability in the Mekong river

J. M. Delgado, H. Apel, and B. Merz

German Research Centre for Geosciences, Section 5.4, Hydrology, Potsdam, Germany

Abstract. Annual maximum discharge is analyzed in the Mekong river in Southeast Asia with regard to trends in average flood and trends in variability during the 20th century. Data from four gauging stations downstream of Vientiane, Laos, were used, covering two distinct hydrological regions within the Mekong basin. These time series span through over 70 years and are the longest daily discharge time series available in the region. The methods used, Mann Kendal test (MK), ordinary least squares with resampling (OLS) and non-stationary generalized extreme value function (NS-GEV), are first tested in a Monte Carlo experiment, in order to evaluate their detection power in presence of changing variance in the time series. The time series are generated using the generalized extreme value function with varying scale and location parameter. NSGEV outperforms MK and OLS, both because it resulted in less type II errors, but also because it allows for a more complete description of the trends, allowing to separate trends in average and in variability.

Results from MK, OLS and NSGEV agreed on trends in average flood behaviour. However, the introduction of a time-varying scale parameter in the NSGEV allowed to isolate flood variability from the trend in average flood and to have a more complete view of the changes. Overall, results showed an increasing likelihood of extreme floods during the last half of the century, although the probability of an average flood decreased during the same period. A period of enhanced variance in the last quarter of the 20th century, estimated with the wavelet power spectrum as a function of time, was identified, which confirmed the results of the NSGEV.

We conclude that the absence of detected positive trends in the hydrological time series was a methodological misconception due to over-simplistic models.

1 Introduction

Detecting trends in hydrological variables has been given emphasis in recent years, due to an increasing scientific consensus on anthropogenic climate change. Indeed, climatic mechanisms are being triggered that increase the potential for intense precipitation around the world (Kundzewicz and Schellnhuber, 2004) and particularly in Asia (Cruz et al., 2007). However, this change is considered not to be spatially or temporally uniform: different studies show significant increases in extreme precipitation and discharge in many countries (Petrow and Merz, 2009; Robson, 2002; Kunkel et al., 1999), whereas many others do not find evidence on this regard (Robson et al., 1998; Svensson et al., 2006; Kundzewicz et al., 2005; Mudelsee et al., 2003). Nevertheless, global climate models claim that climate change would drive up extreme precipitation and river discharge (Nijssen et al., 2001; Palmer and Räisänen, 2002; IDAG, 2005).

Although Katz and Brown (1992) prove the importance of change in variability (also referred to as the scale parameter of certain statistical distributions), and despite the existence of several frequency models in the literature that take non-stationarity of the scale parameter into account (see Strupczewski et al. (2001), Hundecha et al. (2008) and Villarini et al. (2009), or Khaliq et al. (2006) for a review), many studies in meteorology and hydrology still do not attempt to detect a trend in this parameter. In fact, the effect that change in variability produces in the detection of usual trends in average flood is only poorly understood.

After a first approach to studying the variability of the flood regime of the Mekong river, our case study, we were motivated into a deeper investigation on how trend detection methods are affected by a time-dependent change in variance. The methods, some of them not explicitly taking into account change in variance, were chosen mainly because of their simple underlying concepts, widespread use and
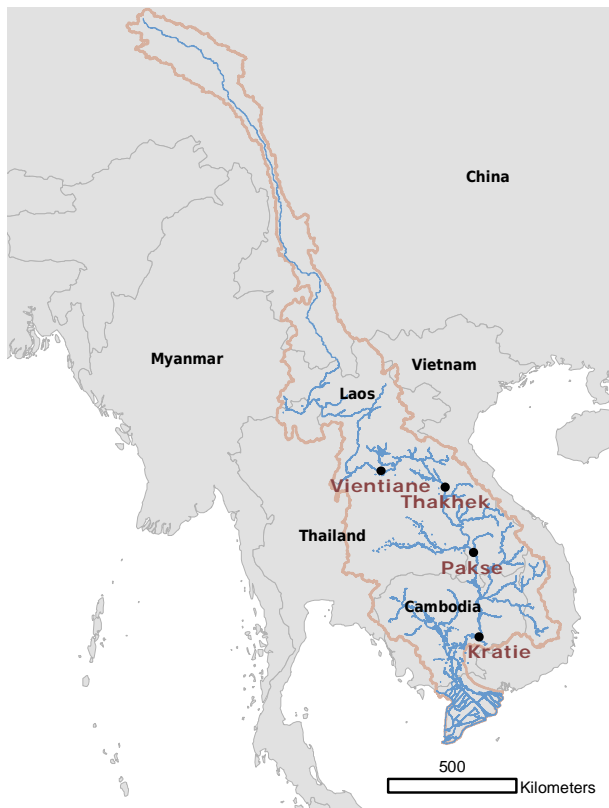
Fig. 1. Map showing part of Southeast Asia and main waterways (blue line) within the Mekong basin (delimited by the red line). Discharge gauging stations used in this study are marked with dark dots.

for being fundamentally different: the ordinary least squares with statistical significance obtained from resampling (OLS), the Mann-Kendall test (MK) and the non-stationary generalized extreme value function (NSGEV) with location and scale parameter as a function of time. The performance of trend detection tests in the presence of time-varying variability is investigated in a Monte Carlo experiment. We generate many synthetic time series with a priori knowledge of their variance based on the generalized extreme value function and try to detect a trend with the aforementioned methods. These methods are conceptually different and each of them focus on different definitions of what a trend is, yielding different responses to a change in variance.

Another important aspect of variability is its link to vulnerability on the societal level. One of the drivers of vulnerability is variability and change in the environmental conditions (Turner et al., 2003), and the probability of exposure to stress or perturbations of the system is a part of the vulnerability equation (Luers et al., 2003; Adger, 2006). Therefore, methods for identifying periods of enhanced variability are crucial to contextualize and provide a quantitative background to vulnerability assessments in the field. Additionally, a framework that assumes a non-stationary approach to frequency analysis is necessary to quantify the change in the probability

of an extreme event. That is accomplished in this work by using the wavelet power spectrum and the NSGEV model, respectively.

Further motivating our work is a general public consensus on an increase in the flood damage and risk during the last century in the Mekong basin (Campbell, 2007; Käkönen, 2008), although the scarce published studies that attempt to identify trends in river discharge or precipitation point to a negative trend (Campbell, 2007; Lu and Siew, 2006). Model outputs also point to a future increase in the intensity of flood events in the region due to climate change (Milly et al., 2002; Hoanh et al., 2003; Kiem et al., 2008). Even disregarding anthropogenic climate change, trends are expected, as an effect of an interannual to decadal organization in climate (Black, 2002) as well as changes in monsoon intensity over centennial to millenial timescales (Zhang et al., 2008).

The purpose of this work is to evaluate whether there is a trend in average flood and in flood variability on four stations along the Mekong river and evaluate how such a change in variability might affect the power of usual trend detection tests.

2 Data and geographical extent

The present study analyzes the only available long daily discharge time series in the river Mekong. These are available for Vientiane (1913–2000), Thakhek (1924–2000), Pakse (1923–2000) and Kratie (1924–2007) and shown in Fig. 2. The time series were used in their full length. The data was provided by the Southern Institute of Water Resources Research in Ho Chi Minh City, Vietnam, and are not publicly available. The daily discharge is estimated by the use of a rating curve and daily water level readings. Discharge measurements do not exist for the years before 1960 and therefore the values here presented were estimated using the rating curves from 1960 on.

The data was checked for quality. The extent to which errors in the time series would influence the results of the NSGEV had to be tested. In order to do that, we removed noise by applying a wavelet filter to the data. This noise was then shuffled on annual blocks, amplified by 10% and added to the denoised time series. Then, the NSGEV was applied. This procedure was repeated 1000 times and its results recorded. Over 90% of the trials yielded a trend with the same slope as the one obtained for the original series.

A second quality assessment could be performed only for data from Vientiane, Thakhek and Pakse, for which rating curves were available. Discharge was measured directly on different campaigns, which yielded different rating curves. We used rating curves from 1960 and from 2002 to transform the available water level time series (1960–2000) in discharge time series. On both cases, the choice of the rating curve did not affect the significant trends detected by the NSGEV.
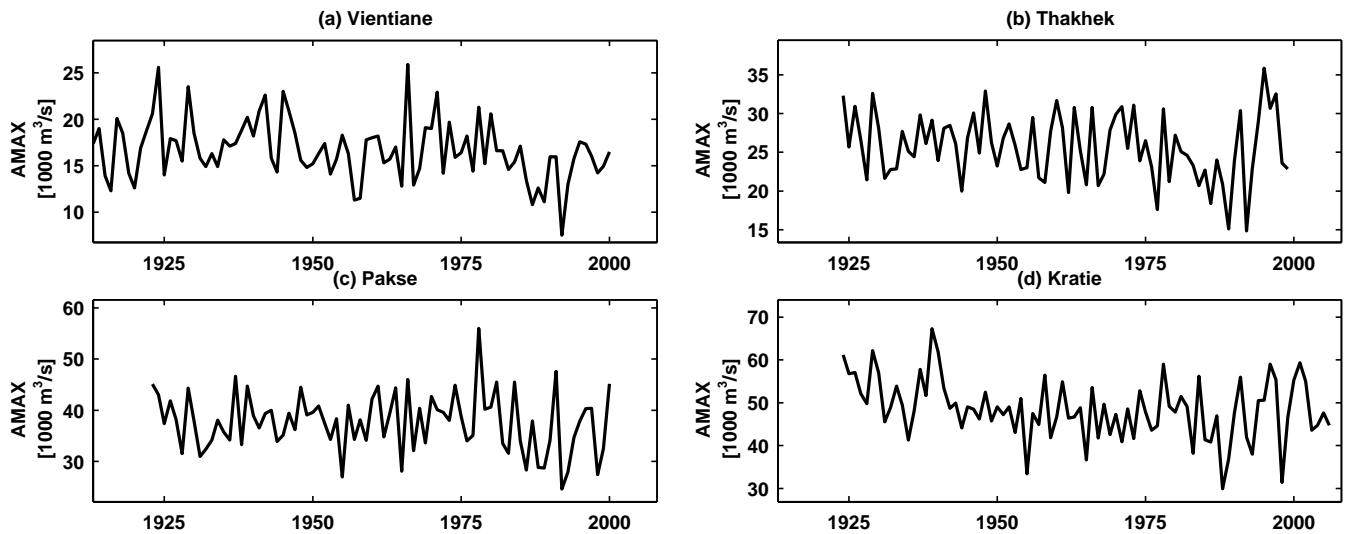
Fig. 2. Time series of annual maximum discharge used in this study.

Special care should be payed to Kratie, where rating curves only exist for the 1960s and after 2000 and for which we did not have access to water level records. However, the data was allegedly corrected and gaps were filled based on the station of Stung Treng, about 100 km upstream (MRC, 2004).

The flood index used was the annual maximum discharge series (AMAX), obtained from daily discharge. This describes well the flood hydrograph, which depends on the same forcing mechanisms and arrives roughly at the same time for every year.

The Mekong river lies in Southeast Asia and its 800 000 km² catchment is shared by China, Myanmar, Thailand, Laos, Cambodia and Vietnam (Fig. 1). In China, the river flows on the Tibetan plateau through the Yunnan province, mainly fed by snow melting in Spring and receiving a small proportion of monsoon precipitation. The Yunnan component makes up for 16% of the whole annual runoff (MRC, 2005). In the lower basin, the Mekong may be divided into three main reaches: from the Chinese border to the beginning of the eastern highlands on the Laos-Vietnam border (more or less near Vientiane), from there on to Kratie and from Kratie to the delta. The main differences concern the flood generation during the monsoon season, the first reach being mainly fed by moisture from the bay of Bengal (thus related with the Indian monsoon – IM); the second being fed by strong orographic precipitation from westerly air masses that cross Southeast Asia until they meet the eastern highlands, although they are forced by the monsoon system over the Western North-Pacific, east of Southeast Asia; and finally the third sharing the same source of moisture as the second, but in a relatively flatter terrain. These two moisture sources have different forcing large scale atmospheric circulation patterns and onset times.

3 Methods

We start by a methodological definition of the different types of trends we are aiming at. As we use different methods that detect trends in different aspects of the data, we distinguish two groups of trends: a trend in average flood, which is a change in a statistic related to or describing the expected value of the time series, may it be the mean, the location parameter of the underlying distribution or another related parameter; and a trend in flood variability, which may be detected by an estimation of average variance on a given year or of the changes in the scale parameter of the underlying distribution. In the case of the NSGEV, where both a trend in average flood and a trend in flood variability may be detected at the same time, caution is needed when interpreting the results. From a negative trend in the location parameter and a positive on the scale parameter may emerge an almost zero trend in the mean of the distribution (Zhang et al., 2004), although the statistical properties of the sample are being affected by a change. To avoid that, when dealing with NS-GEV, we never refer to a trend in the mean, but explicitly to a trend in the location parameter or in the scale parameter.

Three methods were used to estimate trends in the average of the time series (as in Zhang et al., 2004): linear regression in a least square sense, an inappropriate but straightforward and often used method for detecting trends in extreme values; the Mann-Kendall test (Kendall, 1938), a powerful non-parametric trend test for every kind of time series; and the non-stationary generalized extreme value model (Coles, 2001), a parametric statistical test that accounts for the skewness of the data.

Assessing the significance of a linear regression as an estimate for a trend was done following the resampling methodology given by Kundzewicz and Robson (2004). According
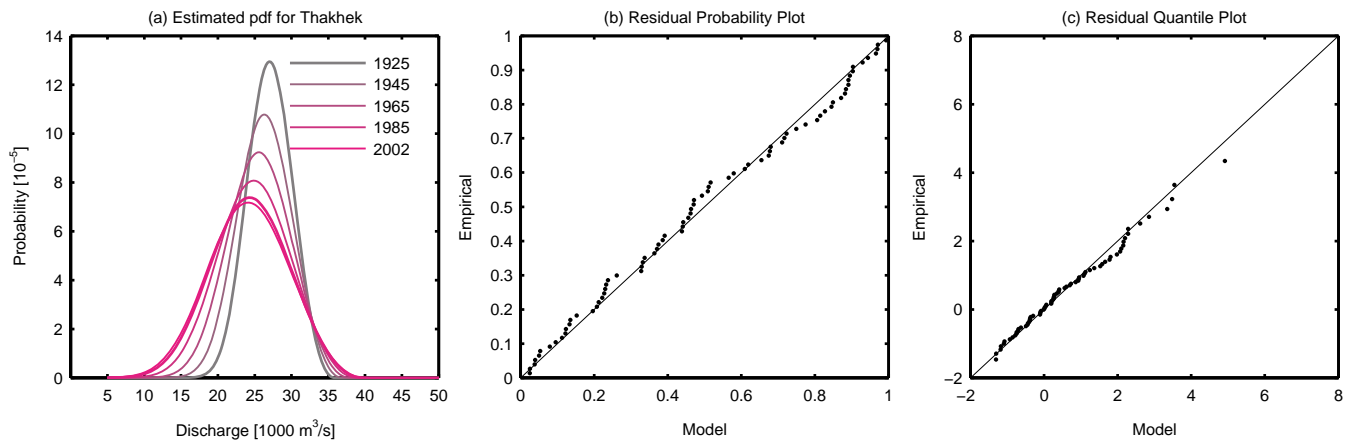
Fig. 3. (a) the estimated probability density function of AMAX for different years according to NSGEV with a negative trend in the location parameter and a positive trend in the scale parameter. **(b)** the residual probability plot, presented as a diagnostic of the NSGEV model application. **(c)** same as (b) but for the residual quantile plot.

to this method, a time series with a trend is represented by

$$x(t) = b_1 + b_2 t + \epsilon_t \qquad (1)$$

where b_1 and b_2 are the terms of the linear trend and are estimated by the method of the least squares and ϵ_t is the deviation of the trend line to the time series. If ϵ_t is normally distributed, then its expected value is 0. ϵ_t is often not normally distributed and does not even have a symmetric distribution in the case of climate variables. We use this method nevertheless as a reference, because it is easy to use and often adopted in trend assessments.

The Mann-Kendall test (Kendall, 1938) is a non-parametric statistical test that evaluates whether there is a trend in a time series x_i of size n. Each element of x_i is compared with its successors x_{i+j}, with $0 < j < n$. z_{ij} is defined as equal to 1 if $x_{i+j} > x_i$, to -1 if $x_{i+j} < x_i$ and to 0 if $x_{i+j} = x_i$. We then compute

$$Z = \frac{\sum_{i=1}^{n-1} \sum_{j=1}^{n-i} z_{ij}}{\sqrt{n(n-1)(2n+5)/18}} \qquad (2)$$

Z follows a normal distribution with standard deviation 1 and expected value 0, when there is no trend in x_i. By computing Z, it is possible to test the statistical significance of rejecting the null hypothesis "no trend in x_i".

The NSGEV function was used following the methodology in Coles (2001). This model is an extended parametrization of the generalized extreme value function (GEV), a combination of three families of extreme value distribution, Gumbel, Fréchet and Weibull (Jenkinson, 1955). The cumulative GEV distribution function is written as:

$$F(x) = \begin{cases} \exp\left[-\left(1 - \frac{\xi}{\sigma}(x-\mu)\right)^{\frac{1}{\xi}}\right] & \text{if } \xi \neq 0 \\ \exp\left[-\exp\left(-\frac{(x-\mu)}{\sigma}\right)\right] & \text{if } \xi = 0 \end{cases} \qquad (3)$$

where x is the random variable and the rest are the distribution parameters, which are fit to the sample with the maximum likelihood estimator. According to this parametrization, the location (μ, which defines the position of the function with regard to the origin) and scale parameter (σ, which defines the spread of the distribution) are made time-dependent following a desired function. The shape parameter (ξ, defining additional shape characteristics of the function) is left constant. Starting with this distribution, we fitted different combinations of linear and second degree time dependent parameters and evaluated the contribution of each combination by estimating the deviance statistic (see below). The combinations start from a stationary model, to which we add terms one by one, as in Hundecha et al. (2008). This explicitly accounts for changes in average and variance over time, as seen in the example shown in Fig. 3a for Thakhek, yielding a different probability distribution each year (tails grow fatter with time). A first approach for the time dependent parameterization of μ and σ is a linear dependency:

$$\beta(t) = \beta_0 + \beta_1 t \qquad (4)$$

The parameter β represents generically the location and the scale parameter. The linear model was extended with a second order term in a consecutive step. However, the significance of a second degree time dependence was investigated only on one of the parameters, in order to guarantee the convergence of the numerical procedure:

$$\beta(t) = \beta_0 + \beta_1 t + \beta_2 t^2 \qquad (5)$$

The second degree extension may become important, because it accounts for the change in the sign of the trend, which may occur at the time scale analyzed. Accepting more than one maximum or minimum would mean that we would be considering a cycle and not a trend (Wu et al., 2007), so higher degree co-variation was not considered.

To find the best fit of the parameter set to the sample, the maximum likelihood criterium was used. Instead of the location and scale parameter, the whole expressions of $\mu(t)$ and $\sigma(t)$ were inserted in the likelihood function:

$$L = \prod_{t=1}^{n} \sigma(t)^{-1} \exp\left[-\left(1 - \xi \frac{x(t) - \mu(t)}{\sigma(t)}\right)\right] \quad (6)$$

where $x(t)$ is the element of the time series corresponding to time t.

Both the linear and second degree model are a generalization of the stationary GEV. They necessarily yield a set of parameters that are at least as good as the particular case of $\beta_0, \beta_1 = 0$. However, if the results are very similar to the stationary case, it can be argued that the differences were obtained by chance and not due to an improvement in the description of non-stationarity. Therefore, a likelihood deviance statistic was used to raise confidence in the model. Let M_0 be a submodel of model M_1, stationary and non-stationary, respectively, whose log-likelihood is l_0 and l_1. The deviance statistic is given by:

$$T = 2(l_1 - l_0) \quad (7)$$

where T is χ_q^2 distributed, and q is the difference between the number of free parameters in M_1 and M_0. We reject M_0 at $1 - \alpha$ significance level, if the integral of the χ_q^2 distribution from T to infinity is smaller than α.

The addition of parameters to the model was done on a forward selection: we started with the stationary GEV and a higher term was added only when a significant improvement in the fit was indicated by the deviance statistic. The final parametrization was tested by deriving its covariance matrix and testing each parameter with t tests. This was done using a distribution-based bootstrap method with stratified design.

After obtaining a statistically significant model, either stationary, linear, or second order, baseline values are established for the stationary case with the GEV distribution. As a measure of an average flood, we used the 2 year return period flood according to the stationary GEV fitted to the sample. As a measure of extreme floods we used the threshold of the 20 year return period discharge of the sample, when estimated with a stationary GEV model. Note that the return period is given by

$$R = \frac{1}{1 - F(x)} \quad (8)$$

where $F(x)$ is the cumulative probability from Eq. 3.

The goodness of fit of the NSGEV may be also visually inspected by plotting a diagnostic. Two types are given in Coles (2001): a residual probability plot and a residual quantile plot. Both diagnostic plots represent standardized variables: first the modeled probability against the plotting positions and second the observed discharges against the modeled discharge corresponding to the respective plotting position. These are presented in Fig. 3b and c as an example.

The estimation of variance against time was done with the wavelet power spectrum (WPS) (Torrence and Compo, 1998), which is the squared absolute value of the wavelet transform. The wavelet transform may be described as a correlation coefficient between the time series and a given and well known function that slides over the time domain and is scaled to account for different frequencies. A coefficient is therefore given for every scale and time step, building a two dimensional plot. The present application used a *Morlet* wavelet, which is a complex valued, nonorthogonal function.

The average variance over the time domain was also obtained by the wavelet. If we integrate the power spectrum with respect to the scales, we obtain for each year the localized variance over a chosen scale range. This is an useful tool for validating the NSGEV in terms of variability, as it explicitly shows the changes in variance over time.

In order to test the ability of the different models to detect trends under presence of a trend in variance, we performed a Monte Carlo experiment using synthetic AMAX time series. The synthetic annual maximum discharge time series were generated without simulating the annual cycle or modeling the temporal occurrence of flood peak. The reason is that the annual cycle is very stable, defined by the monsoon precipitation that arrives approximately at the same time of the year (MRC, 2005). The same may be said about the flood season. More than one flood peak per year can occur, but always within the same flood season, close to the maximum, and they are imposed on the annual flood hydrograph, which is unique for any given year but similar in shape between different years. The annual maximum discharge is able to represent the magnitude of the flood.

The chosen distribution for the generation of the synthetic annual maximum discharge time series was the GEV distribution with time varying location and scale parameter (a NSGEV model), which are the analogues to mean and standard deviation of a normal distribution. We used a NSGEV with linearly varying location and scale parameter fitted to the AMAX of Pakse as a baseline, as in Eq. 4. The trend in the location parameter was kept constant and equal to the baseline model. The different trends in the scale parameter tested in the Monte Carlo experiment may be defined as: $\sigma_k(t) = \sigma_0^{\text{Pakse}} + k\sigma_1^{\text{Pakse}}t$, $k = 0, 0.2, 0.4, \ldots, 2$, where $\sigma_0^{\text{Pakse}} = 3.42 \times 10^3 \, m^3/s$ and $\sigma_1^{\text{Pakse}} = 45 \, m^3/s$ (the lower index identifies the term in Eq. 4). For the location parameter $\mu_0^{\text{Pakse}} = 3.63 \times 10^4 \, m^3/s$ and $\mu_1^{\text{Pakse}} = -26 \, m^3/s$; the shape parameter was kept constant: $\xi^{\text{Pakse}} = -0.19$. Note that an analogue experiment was performed by Zhang et al. (2004), where the range of trends in the location parameter used was greater than the one used in the present work. The only restrictions for the synthetic time series are that each data point may not be greater than 1.5 times the maximum historical discharge and not lower than half the minimum recorded AMAX. Next, we tested the occurence of type I errors (detecting a trend when there is none in the data) by running the same Monte Carlo experient this time with a constant scale
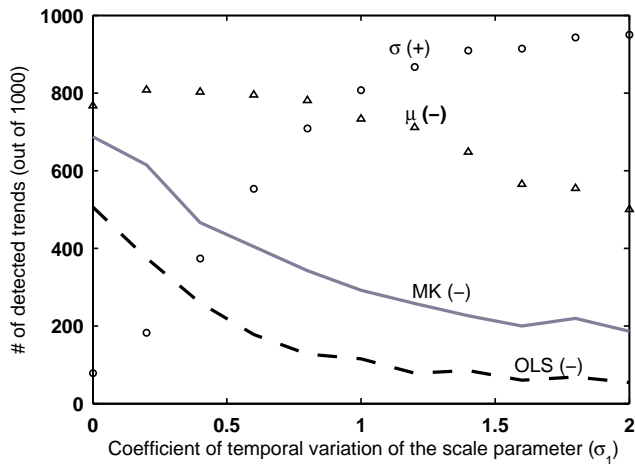
Fig. 4. Number of trends detected at a 90% significance level in 1000 synthetic time series using OLS, MK and NSGEV. The time series were generated using the NSGEV model with a constant trend in the location parameter and a varying one in the scale parameter. Different rates of change for the scale parameter are given in the abscissa. μ refers to the detected trends in the location parameter and σ to the ones in the scale parameter derived by the NSGEV model. Plus and minus signs indicate positive and negative trend.

parameter (no trend in the scale parameter) and a varying rate of change in the location parameter, analogous to the one used previously for the scale parameter. If the results show a rate of detected trends in the scale parameter significantly above or below the nominal significance level (10%), the experience will be not valid, because either it is biased (below the nominal significance level) or produces too many type I errors (above the nominal significance level).

The three trend detection methods, OLS, MK and NSGEV with the forward selection of linear terms described earlier, were applied to the 1000 synthetic time series. Results are presented and discussed in Sect. 4.

4   Results and discussion

4.1   Trend detection with changing variance

Figure 4 shows the number of detected negative trends among 1000 trials for each of the positive scale parameter rate of change (Eq. 4), given a constant negative trend in the location parameter of the synthetic time series. A first observation is that NSGEV is the most powerful method to detect a trend in average flood ($\mu(-)$ in Fig. 4). In second comes MK and finally OLS. Their performance for constant scale parameter was 77%, 69% and 51% for NSGEV, MK and OLS, which is greater than in Zhang et al. (2004), because of different significance levels. However, it is also seen that all the methods loose power in detecting the negative average trend, when the samples are driven with a strong positive scale parameter trend. For example, MK detects 3
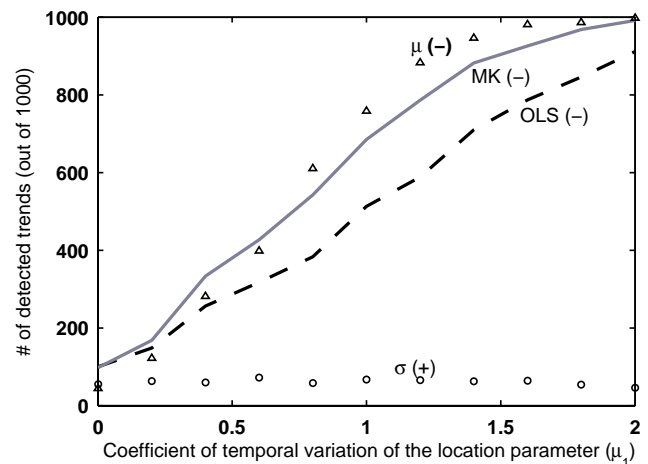

Fig. 5. Number of trends detected at a 90% significance level in 1000 synthetic time series using OLS, MK and NSGEV. The time series were generated using the NSGEV model with a varying trend in the location parameter and a constant scale parameter. Different rates of change for the location parameter are given in the abscissa. μ refers to the detected trends in the location parameter and σ to the ones in the scale parameter derived by the NSGEV model. Plus and minus signs indicate positive and negative trend. The sum of the power of detection of positive and negative trends on the scale parameter is within the 90% confidence intervals around the nominal significance level chosen for this work (10%).

times less negative trends in the presence of a strong trend in the scale parameter than it would with a constant scale parameter, whereas OLS more than 9 times less. This means that regarding the detection of trends in average floods, an error of type II (failure to detect an existing trend) is more likely to occur in the presence of a strong temporal change in variability.

Secondly, we observed that the rate of trend detection in the scale parameter was 12% (presented are only the rate of detection of positive trends, which is 8%), which is within acceptable ranges of the 10% nominal significance level. This was thoroughly tested in the second Monte Carlo experience, where the confidence intervals were also estimated, as described below. Although not shown in the figures, the NSGEV was the method with less errors of type I.

Thirdly, detecting a trend in the scale parameter with NSGEV appears to be free of problems. The power of the NSGEV increases with a steeper trend in the scale parameter. The stepwise forward selection of parameters yielded better results than when we estimated each parameter individually. By testing the model improvement by adding one additional parameter at a time, instead of constructing a complex model and then testing it, we were able to validate each of the parameters involved.

Regarding our test for false positives, the results in Fig. 5 show an almost constant rate of detection of trends in scale parameter, around 5%. However, if we add the detected

negative trends, we get an average of 10% type I errors, which is the nominal significance level. As for the detection of trends in the location parameter when $\mu_1 = 0$, MK and OLS show 9 and 11% of false positives. NSGEV presents again a total of 10%, when we sum the positive and negative detected trends (shown are only the number of detected negative trends).

Another observation in Fig. 5 is that the number of detected trends in the scale parameter remains constant. This means that the number of false positives is not dependent on the intensity of the trend in the location parameter.

Although we used the 90% significance level for all the methods, this does not mean that we can trust the results equally, due to the fact that statistical significance was computed following three different methods. On the same line, the results must be interpreted according to the method used, because each of them is conceptually different. For example, it is expected that OLS places a greater weight on greater magnitudes than NSGEV: the method is based on gaussian assumptions, whereas the sample that it is applied to has more frequent high peaks than it had if it was driven from a normal distribution, given GEV's heavy tail. Regarding MK, we cannot expect to cover the change in the frequency of extreme high floods, which itself may induce a significant perception of a trend, because it places the same weight on an upper percentile value as on a median value. This affects its ability to incorporate the more frequent occurence of extremes, which is well described by the NSGEV, for example. In summary, the different methods focus on different aspects of the time series, which means that the user should be aware of each method's limitations. As it will be seen in the application to the case study, the use of NSGEV allows the study of different sets of magnitudes, for example greater magnitudes of the time series or average values: we can focus on which percentile of the time series we want to analyze and estimate its change over time. Moreover, it allows to perform both a trend detection test and a frequency analysis.

We learn from this exercise that different methods are affected differently by a change in variance in the time series. Namely, the power of detection of an average trend decreases greatly for MK and OLS, to a level where they resulted in a type II error in most of the test trials. OLS even detects more positive trends than negative when the trend in scale parameter is greater than $\sigma_1^{\mathrm{Pakse}}$, probably due to being based on a normal distribution, when the data is clearly non-normal. When suspecting changes in variance, NSGEV should be used, as it explicitly accounts for change in the scale and location parameter. Even when only considering a trend in the location parameter, it was by far the best method tested.

Results from trend detection should be considered with caution and always validated against other methods. Further, and equally important, a possible change in variance should be considered, as it can affect the trend detection results even with high significance levels, as shown in this section. Simple methods are available that give an idea of the change in

Table 1. Summary of the trend analysis of AMAX in the lower Mekong river. "–" stands for negative trend and "+" for positive trend. Bold lettering indicates 90% statistical significance.

		Vientiane	Thakhek	Pakse	Kratie
OLS		**–**	**–**	–	**–**
Mann-Kendall		**–**	**–**	–	**–**
NSGEV	μ trend	**–**	**–**	–	**–**
	σ trend	**–**	**+**	**+**	**–**

variance of the time series over time. Computing a moving window variance or the average variance obtained from the wavelet power spectrum (Torrence and Compo, 1998) are straightforward choices, although in the case of skewed datasets, as normally meteorological and hydrological data are, the NSGEV or the Generalized Additive Models for Location, Scale and Shape (GAMLSS) (Villarini et al., 2009) could be a better option.

4.2 Flood trends in the Mekong river

A summary of the trend analysis of four stations on the Mekong river is presented in Table 1, where the results of MK, OLS and NSGEV with linearly varying parameters are shown. A first inspection reveals apparent consensual results: a negative trend is affecting all four stations. Only in Pakse there is some uncertainty regarding the trend, because it is not statistically significant. This may be due to the fact that it has the strongest scale parameter trend, identified by NSGEV, which, according to the results of the previous section, leads to a relatively large type II error. However, when we distinguish the trend in the average flood and the trend in variability, we obtain different conclusions regarding how we see the flood regime of the Mekong during the 20th century. This is analyzed with respect to trends in flood variability in Sect. 4.3.

Table 1 shows overall agreement between methods in detecting average flood trends: MK, OLS and NSGEV all detect negative trends in average flood in all stations. This contrasts with public and local managers' perceptions as stated in Campbell (2007) and with the hypothesis of a strengthening large-scale monsoon system (Anderson et al., 2002). We know, however, that average flood trend detection methods like OLS and MK do not capture what may be the most interesting aspect of change in the flood regime: variability (Katz and Brown, 1992; Kundzewicz and Schellnhuber, 2004). Indeed, the trends in the greater flood magnitudes at Thakhek and Pakse become ascendant if the AMAX are modeled by NSGEV with a linear trend in both parameters. This means that the flood regime became more variable during the 20th century. Extremely high flood events were experienced more often than before, although intercalated with years of below-average flooding. Therefore, in present and according to the NSGEV model, the probability of experiencing a greater

Table 2. Estimated parameters for the nonstationary and stationary GEV. Zeros are displayed for terms that did not satisfy the criterium of the forward selection described in Sect. 3.

Station	Location parameter			Scale parameter				Shape parameter	
	Stationary	Nonstationary		Stationary	Nonstationary			Stationary	Nonstationary
	μ	μ_0	μ_1	σ	σ_0	σ_1	σ_2	ξ	ξ_0
Vientiane	15 457.3	16 930.1	−32.6	3085.9	2947.3	0	0	−0.20	−0.18
Thakhek	24 250.4	26 047.1	−46.4	4376.0	2960.0	30.1	0	−0.32	−0.30
Pakse	35 460.9	35 590.8	0	5530.8	3831.3	39.0	0	−0.20	−0.30
Kratie	46 100.2	49 749.0	−96.7	7074.1	9202.3	−196.6	2.5	−0.28	−0.29

than average flood in Thakhek and Pakse is greater than before. This is an interesting result, not only because it matches projections from regional and global climate models (Milly et al., 2002; Kiem et al., 2008), but also because it adds on the discussion of trend detection: within certain hydrological systems, MK, OLS or NSGEV with only varying location parameter may not fully describe change in the flood regime.

But why does not Vientiane present the same behaviour? The answer lies probably in the regions of influence of the two components of the monsoon (Sect. 4.3): the Indian monsoon (IM) and the Western North-Pacific monsoon (WNPM). These two components had different periods of enhancement during the 20th century. They also influence regional patterns of precipitation (Wang et al., 2001). Vientiane receives its flood waters from moisture entering the continent through the bay of Bengal and from melting of snow in the Tibetan plateau. Downstream of Vientiane, the contribution from the highlands on the border between Laos and Vietnam is dominant (MRC, 2005); there, the flood generation is linked with a combination of WNPM and IM, whereas in the south it is linked predominantly with WNPM (Delgado et al., 2010).

The variability trend in Kratie does not match Pakse and Thakhek further upstream. Indeed, a light negative linear change in the scale parameter was found to be significant, although this was the only station where a second degree trend in the scale parameter of the NSGEV proved to be significant (Table 2), when compared to the linear model. The analysis of this trend is done in Sect. 4.3, where it is also compared with other measures of variability.

When focusing on trends in average flood, the three methods seem to agree that floods decreased on average over the 20th century. However, the scale parameter obtained by the NSGEV model presents a significant trend, revealing that the underlying distribution may be changing in a way that may affect extremes differently than it affects average floods. This is discussed in the next section.

4.3 Trends in flood variability

Variability was assessed in two different ways. First, the number of parameters of the NSGEV was increased one by one, and the improvements evaluated by a deviance statistic.

The maximum parametrization allowed was a second degree variation of each of the parameters. This procedure is explained in Sect. 3 and in Coles (2001). The diagnostic plots for Thakhek are presented, showing a fair fit of the linear NSGEV (Fig. 3b and c). Secondly, a more adaptive method is used, the wavelet power spectrum, that is able to outline both the dominant modes of variability and how they vary with time (Torrence and Compo, 1998). The power spectrum was computed for the whole scale domain showing periods of short term variability.

The result of the parameter forward selection is presented in Table 2. The values presented are all above the 90% significance level. Note that for Pakse, only the scale parameter has a trend and that Kratie has a second degree positive term in the scale parameter. The covariance matrix of the models obtained were verified with t tests, as explained previously. As in Table 1, trends in the location parameter are negative, except for Pakse, where the trend in the scale parameter is the strongest of all stations. The main change introduced by the forward selection of model parametrization is the second degree in the scale parameter for Kratie. The fact that this second degree is significant means that there is an inflection in the scale parameter during the 20th century.

The results of the NSGEV regarding baseline values (probability of exceedence of the 20 year return period and of the 2 year return period of the distribution according to the stationary GEV) are given against time in Figs. 6 and 7. The second degree variation of the scale parameter for Kratie is evident in the figure. According to this model, the probability of exceeding $Q_{T=20}^{\mathrm{GEV}}$ decreased in Vientine during the 20th century by 0.08, whereas it increased by 0.03 and 0.10 in Thakhek and Pakse. During the same period, the probability of an average flood was decreasing in all stations. This difference between Vientiane and the two downstream stations may be explained by the different hydrological regimes within the Mekong river: downstream of Vientiane, the contribution of the flow generated in the highlands on the Laos-Vietnam border, whose variability is modulated by the WNPM, starts to affect the flood hydrograph, whereas upstream it is still mainly affected by the Yunnan component. According to the model, the probability of exceeding $Q_{T=20}^{\mathrm{GEV}}$ in Kratie presents an inflection point around the 1970s. This
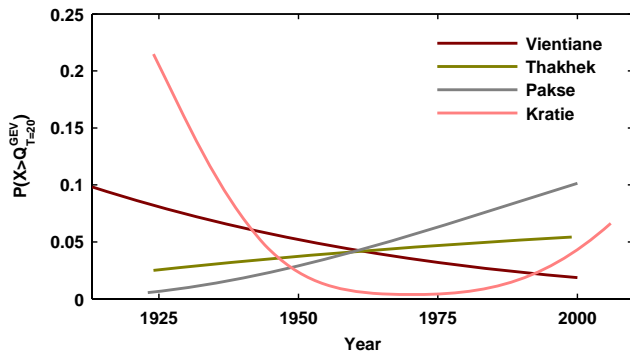
Fig. 6. Probability (computed with NSGEV) of exceeding the 20-year flood estimated by the stationary GEV in Vientiane, Thakhek, Pakse and Kratie.

Fig. 7. Same as in Fig. 6, but for the probability of exceeding the 2-year flood.

decade might represent the beginning of an enhancement in flood variability, as detected by Wang et al. (2001) for the activity of the WNPM. By the end of the century, the estimated increase is of 0.06. Although this inflection is also seen in the average variance plots (Fig. 8), it is not present if one uses a linear model for describing the scale parameter. This possible change of behaviour between Kratie and upstream stations may be explained with the important contribution of tributaries with their mouth between Pakse and Kratie, like the Se San and Se Kong (Tonle San and Tonle Kong in Cambodia) and the more southward landfall of typhoons in the past decades (Ho et al., 2004).

The fact that in the beginning of the time series, the probability of an extreme flood is very high for Kratie may be due to errors in rating curves, or filling of gaps in the record using an upstream station, as discussed previously. The NS-GEV model fit was here driven by the high flows recorded in the early decades of the time series, which are difficult to validate, due to lack of other sources of data (for this region, reliable reanalysis climatic datasets are available only after 1950 and earlier tributary discharge records do not exist). For the later 20th century, the discharge could be compared and validated with precipitation data, which suggests climatic causes for the increase in variability reported. Indeed Wang et al. (2001) and Ho et al. (2004) show an enhancement of the WNPM index variance and typhoon activity in the 1980s, respectively. One can find, additionally to atmospheric, other plausible forcing mechanisms for change in the hydrological system. Some of these factors were approached by Lu and Siew (2006) and Haddeland et al. (2006). Regarding water use, the latter modeled the irrigation demand for river Mekong and, as far as model results are to be trusted, the impact of irrigation on the monthly average discharge is relatively little and only evident on the dry season, i.e. irrigation does not affect flood discharge because of insufficient volumes compared to flood volume and also because during the monsoon season irrigation requirement is at its minimum. Regarding changes caused by dam building,
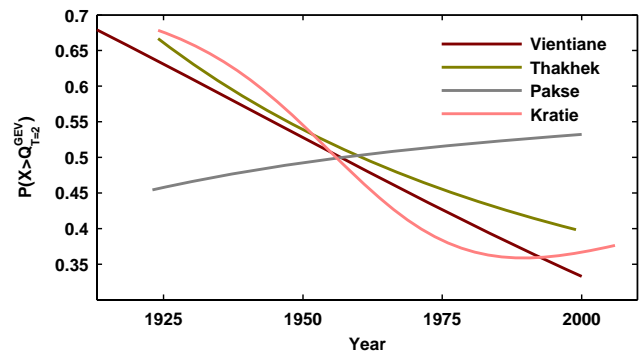
Lu and Siew (2006) argues that this effect is limited to the upper reaches of the Mekong. Furthermore, the first of the Chinese dams was commissioned in 1993, while the reported enhancement in variance starts earlier, in the early eighties. This is enough to dismiss dams as the cause for different behavior of the flood variability during the last 20 years of the 20th century in Vientiane and downstream stations, although their impact is not fully assessed, especially since other dams have been commissioned in recent years both on the main stem and on tributaries. Regarding land use and land use change, it is difficult, if not impossible at present, to evaluate its effect on the Mekong floods due to the lack of long term data on land use. Furthermore, the "effects of land use change on the magnitude of flood peaks in large rivers are difficult to evaluate because such changes are rarely fast and consistent (except perhaps where population pressure is very high) and often compounded by climatic variability" (Bruijnzeel, 2004). This is however not enough to dismiss a possible contribution of land use change to change in extreme floods in a region under dramatic social and economic changes in the last decades (MRC, 2005).

The average variance obtained by integration of the wavelet power spectrum over the scale domain, presented in Fig. 8, confirms the result of NSGEV: a period of enhanced variance is observed in the last quarter of the 20th century for all stations except Vientiane. This enhancement is made evident by the bold contours that represent times and frequencies of significant variability (Torrence and Compo, 1998). The average variance in the middle plot shows increasing variability in the last quarter of the 20th century in the two downstream stations of Pakse and Kratie, this feature being less evident in Thakhek and residual in Vientiane. The descending-ascending behaviour of variance in Kratie is reproduced by the probability of exceeding $Q_{T=20}$ shown in Fig. 6, due to the 2nd degree variation of the scale parameter of the NSGEV. It is also visible in the power spectrum of Kratie (Fig. 8) indicated by the areas of significance.

The separation between trend in average flood and trend in flood variability by NSGEV proves to be more useful than
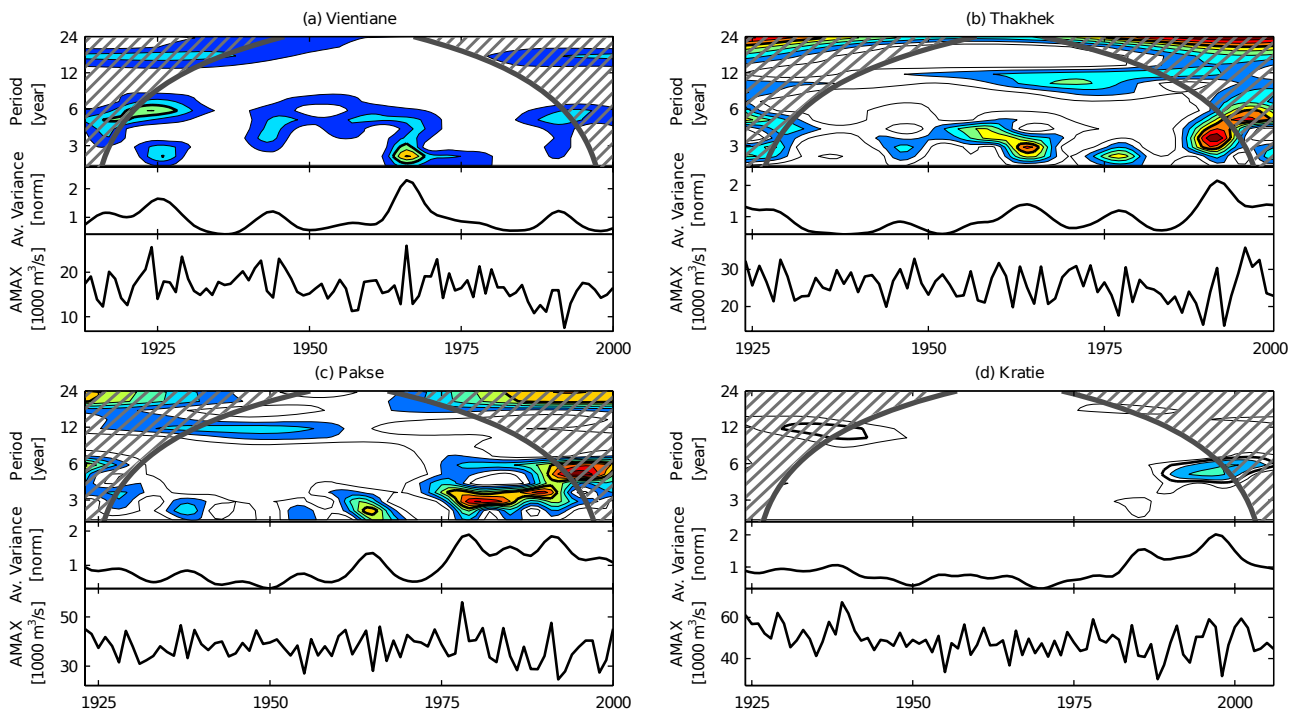
Fig. 8. Top: wavelet power spectrum of AMAX for Vientiane **(a)**, Thakhek **(b)**, Pakse **(c)** and Kratie **(d)**. Colder colors correspond to smaller wavelet coefficients and warmer colors to greater wavelet coefficients. Bold contours enclose significant times and frequencies, whereas the shaded area is outside the cone of influence and should be interpreted with caution. Middle: Average variance (normalized) over the scale domain. Bottom: AMAX time series.

usual trend detection methods like OLS or MK, as it provides a probabilistic interpretation of the trend, including describing the change in probability of occurence of a certain flood. In this sense, although negative average flood trends in all stations are found, the theoretical probability of an extreme event, for example exceeding the 20-year return period, increases over time in the three donwstream stations Thakhek, Pakse and Kratie, at least in the last years of the 20th century (Fig. 6).

5   Conclusions

Usual methods of trend detection like linear regression (OLS) and Mann-Kendall test (MK) proved to loose detection power in presence of changes in variance. In a Monte Carlo experiment it was shown that the introduction of a trend in the scale parameter made the number of detected trends drop to less than half with MK and less than a quarter with OLS. Therefore, the number of type II errors increases with increasing trends in the scale parameter. The use of NS-GEV is advantageous both because of its power of detection in presence of changing variance and because it allows to detect trends in different flood magnitudes with a probabilistic approach.

Regarding the flood regime of the Mekong, it is clear that although average magnitude floods have a negative trend, variability is increasing, both shown by an increase in variance and by a positive trend in the scale parameter of a fitted NSGEV model, for stations downstream Vientiane. According to the fitted distribution, the increase in the theoretical probability of extreme floods is driven by the scale parameter. In this conceptualization, both very large floods and very small floods increase in frequency, with a decrease in frequency of average floods. This motivates further research on the causes and temporal scale of this variability change.

Differences between Vientiane and downstream stations were explained by the influence of regional patterns of precipitation and runoff generation. In the first case the floods mostly originate from rainfall and snowmelt on the upper Mekong basin, and in the second case from intense rainfall over the highlands on the Laos-Vietnam border. These two sources of runoff originate from two distinct atmospheric processes, having therefore different periods of enhancement.

The causes for the detected changes in variance are still unknown and probably range from climate oscillations, climate change and changes in the land and water use. A period of enhanced variance in the WNPM was identified in the literature, that matches the presented results. If these changes are an oscillation in the climate system or a permanent

feature is not known, and will not be understood by only analyzing instrumental records. Analyzing global climate model outputs with regard to variability and links between both monsoon components and precipitation over the Mekong basin would also be useful for understanding this.

Edited by: B. van den Hurk

References

Adger, W. N.: Vulnerability, Global Environ. Chang., 16, 268–281, 2006.

Anderson, D. M., Overpeck, J. T., and Gupta, A. K.: Increase in the Asian Southwest Monsoon during the past four Centuries, Science, 297, 596–599, 2002.

Black, D. E.: The Rains May Be A-Comin', Science, 297, 528–529, 2002.

Bruijnzeel, L. A.: Hydrological functions of the tropical forests: not seeing the soil for the trees, Agr. Ecosyst. Environ., 104, 185–228, 2004.

Campbell, I. C.: Perceptions, data, and river management: Lessons from the Mekong River, Water Resour. Res., 43, W02407, doi:10.1029/2006WR005130, 2007.

Coles, S.: An introduction to Statistical Modeling of Extreme Values, Springer, London, 2001.

Cruz, R. V., Harasawa, H., Lal, M., Wu, S., Anokhin, Y., Punsalmaa, B., Honda, Y., Jafari, M., Li, C., and Ninh, N. H.: Climate Change 2007: Impacts, Adaptations and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, chapter 10, Cambridge University Press, Cambridge, UK, 2007.

Delgado, J. M., Zimmer, J., Apel, H., and Merz, B.: Large scale circulation as a forcing of the variability of floods in the river Mekong, in preparation, 2010.

Haddeland, I., Lettenmaier, D. P., and Skaugen, T.: Effects of irrigation on the water and energy balances of the Colorado and Mekong river basins, J. Hydrol., 324, 210–223, 2006.

Ho, C.-H., Baik, J.-J., Kim, J.-H., and Gong, D.-Y.: Interdecadal Changes in Summertime Typhoon Tracks, J. Climate, 17, 1767–1776, 2004.

Hoanh, C. T., Guttman, H., Droogers, P., and Aerts, J.: Water, Climate, Food And Environment Under Climate Change – The Mekong Basin In Southeast Asia, Tech. rep., International Water Management Institute (Sri Lanka), Mekong River Commission Secretariat (Cambodia), FutureWater (The Netherlands), Institute of Environmental Studies (The Netherlands), 36–45, 2003.

Hundecha, Y., St-Hilaire, A., Ouarda, T. B. M. J., El Adlouni, S., and Gachon, P.: A Nonstationary Extreme Value Analysis for the Assessment of Changes in Extreme Annual Wind Speed over the Gulf of St. Lawrence, Canada, J. Appl.Meteorol. Clim., 47, 2745–2759, 2008.

IDAG: Detecting and Attributing External Influences on the Climate System: A Review of Recent Advances, J. Climate, 18, 1291–1314, 2005.

Jenkinson, A. F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Q. J. Roy. Meteor. Soc., 81, 158–171, 1955.

Käkönen, M.: Mekong Delta at the Crossroads: More Control or Adaptation, Ambio, 37, 205–212, 2008.

Katz, R. W. and Brown, B. G.: Extreme events in a changing climate: variability is more important than averages, Climatic Change, 21, 289–302, 1992.

Kendall, M. G.: A New Measure of Rank Correlation, Biometrika, 30, 81–93, 1938.

Khaliq, M., Ouarda, T., Ondo, J.-C., Gachon, P., and Bobee, B.: Frequency analysis of a sequence of dependent and/or nonstationary hydro-meteorological observations: A review, J. Hydrol., 329, 534–552, 2006.

Kiem, A. S., Ishidaira, H., Hapuarachchi, H. P., Zhou, M. C., Hirabayashi, Y., and Takeuchi, K.: Future hydroclimatology of the Mekong River basin simulated using the high-resolution Japan Meteorological Agency (JMA) AGCM, Hydrol. Process., 22, 1382–1394, 2008.

Kundzewicz, Z. W. and Robson, A. J.: Change detection in hydrological records – a review of the methodology, Hydrolog. Sci. J., 49, 7–19, 2004.

Kundzewicz, Z. W. and Schellnhuber, H.-J.: Floods in the IPCC TAR perspective, Nat. Hazards, 31, 111–128, 2004.

Kundzewicz, Z. W., Graczyk, D., Maurer, T., Pinskwar, I., Radziejewski, M., Svensson, C., and Szwed, M.: Trend detection in river flow series: 1. Annual maximum flow, Hydrolog. Sci. J., 50, 797–810, 2005.

Kunkel Jr., K. E., Roger, A., Pielke Jr., R. A., and Changnon, S. A.: Temporal Fluctuations in Weather and Climate Extremes That Cause Economic and Human Health Impacts: A Review, B. Am. Meteorol. Soc., 80, 1077–1098, 1999.

Lu, X. X. and Siew, R. Y.: Water discharge and sediment flux changes over the past decades in the Lower Mekong River: possible impacts of the Chinese dams, Hydrol. Earth Syst. Sci., 10, 181–195, 2006,
http://www.hydrol-earth-syst-sci.net/10/181/2006/.

Luers, A. L., Lobell, D. B., Sklar, L. S., Addams, C. L., and Matson, P. A.: A method for quantifying vulnerability, applied to the agricultural system of the Yaqui Valley, Mexico, Global Environ. Chang., 13, 255–267, 2003.

Milly, P. C. D., Wetherald, R. T., Dunne, K. A., and Delworth, T. L.: Increasing risk of great floods in a changing climate, Nature, 415, 514–517, 2002.

MRC: Review of hydro-meteorological database of the MRC Secretariat, Progress report (draft), Mekong River Commission, 2004.

MRC: Overview of the Hydrology of the Mekong Basin, Tech. rep., Mekong River Commission, Vientiane, 2005.

Mudelsee, M., Börngen, M., Tetzlaff, G., and Grünewald, U.: No upward trends int he occurrence of extreme floods in central Europe, Nature, 425, 166–169, 2003.

Nijssen, B., O'Donnell, G. M., Hamlet, A. F., and Lettenmaier, D. P.: Hydrologic sensitivity of global rivers to climate change, Climatic Change, 50, 143–175, 2001.

Palmer, T. N. and Räisänen, J.: Quantifying the risk of extreme seasonal precipitation events in a changing climate, Nature, 415, 512–514, 2002.

Petrow, T. and Merz, B.: Trends in flood magnitude, frequency and seasonality in Germany in the period 1951–2002, J. Hydrol., 371, 129–141, 2009.

Robson, A., Jones, T., Reed, D., and Bayliss, A.: A study of national trend and variation in the UK floods, Int. J. Climatol., 18, 165–182, 1998.

Robson, A. J.: Evidence for trends in UK flooding, Philos. T. Roy. Soc., 360, 1327–1343, 2002.

Strupczewski, W. G., Singh, V. P., and Feluch, W.: Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation, J. Hydrol., 248, 123–142, 2001.

Svensson, C., Hannaford, J., Kundzewicz, Z. W., and Marsh, T. J.: Trends in river flows: why is there no clear signal in observations?, edited by: Tchiguirinskaia, I., Ni Ni Thein, K., and Hubert, P., IAHS/IHP-UNESCO, 305, 2006.

Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, B. Am. Meteorol. Soc., 79, 61–78, 1998.

Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., Polsky, C., Pulsipher, A., and Schiller, A.: A framework for vulnerability analysis in sustainability science, Proc. Natl. Acad. Sci. USA, 100, 8074–8079, 2003.

Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., and Krajewski, W. F.: Flood frequency analysis for nonstationary annual peak records in an urban drainage basin, Adv. Water Resour., 32, 1255–1266, 2009.

Wang, B., Wu, R., and Lau, K.-M.: Interannual Variability of the Asian Summer Monsoon: Contrasts between the Indian and the Western North Pacific-East Asian Monsoons, J. Climate, 14, 4073–4090, 2001.

Wu, Z., Huang, N. E., Long, S. R., and Peng, C.-K.: On the trend, detrending, and variability of nonlinear and nonstationary time series, in: Proceedings of the National Academy of Science, National Academy of Sciences, 104, 14889–14894, 2007.

Zhang, P., Cheng, H., Edwards, R. L., Chen, F., Wang, Y., Yang, X., Liu, J., Tan, M., Wang, X., Liu, J., An, C., Dai, Z., Zhou, J., Zhang, D., Jia, J., Jin, L., and Johnson, K. R.: A Test of Climate, Sun, and Culture Relationships from an 1810-Year Chinese Cave Record, Science, 322, 940–942, 2008.

Zhang, X., Zwiers, F. W., and Li, G.: Monte Carlo Experiments on the Detection of Trends in Extreme Values, J. Climate, 17, 1945–1952, 2004.