Regional methods for trend detection: Assessing field significance and regional consistency

Benjamin Renard, M. Lang, P. Bois, A. Dupeyrat, O. Mestre, Hélène Niel,
Eric Sauquet, C. Prudhomme, S. Parey, E. Paquet, et al.

Regional methods for trend detection: Assessing field significance and regional consistency

B. Renard,[1] M. Lang,[2] P. Bois,[3] A. Dupeyrat,[4] O. Mestre,[5] H. Niel,[6] E. Sauquet,[2] C. Prudhomme,[7] S. Parey,[4] E. Paquet,[8] L. Neppel,[6] and J. Gailhard[8]

[1]   This paper describes regional methods for assessing field significance and regional consistency for trend detection in hydrological extremes. Four procedures for assessing field significance are compared on the basis of Monte Carlo simulations. Then three regional tests, based on a regional variable, on the regional average Mann-Kendall test, and a new semiparametric approach, are tested. The latter was found to be the most adequate to detect consistent changes within homogeneous hydro-climatic regions. Finally, these procedures are applied to France, using daily discharge data arising from 195 gauging stations. No generalized change was found at the national scale on the basis of the field significance assessment of at-site results. Hydro-climatic regions were then defined, and the semiparametric procedure applied. Most of the regions showed no consistent change, but three exceptions were found: in the northeast flood peaks were found to increase, in the Pyrenees high and low flows showed decreasing trends, and in the Alps, earlier snowmelt-related floods were detected, along with less severe drought and increasing runoff due to glacier melting. The trend affecting floods in the northeast was compared to changes in rainfall, using rainfall-runoff simulation. The results showed flood trends consistent with the observed rainfall.

1.   Introduction

[2]   The impact of climate change on the hydrological regime of rivers is still a subject of active research, especially regarding extreme hydrological events such as floods or droughts.

[3]   Numerous studies have attempted to detect trends in hydrological data in various parts of the world (*Douglas et al.* [2000] in USA, *Hisdal et al.* [2001] in Europe, *Black and Burns* [2002] in Scotland, *Birsan et al.* [2005] in Switzerland, *Cigizoglu et al.* [2005] in Turkey). Significant trends have been reported in some regions but could in most cases be related to natural climatic fluctuations (e.g., because of the influence of the North Atlantic Oscillation) rather than climate change (*Robson* [2002] and *Hannaford and Marsh* [2006] in the UK). Moreover, when considered at the global scale, results appear inconclusive [*Kundzewicz et al.*, 2005; *Svensson et al.*, 2005].

[4]   The lack of clear conclusions from local analyses may reflect the need for regional-scale analysis, which could be more relevant than at-site studies for detecting the impacts of global phenomena such as climate change. Several distinct issues can be investigated at the regional scale. The most thoroughly studied is field significance [*Livezey and Chen*, 1983; *Lettenmaier et al.*, 1994; *Douglas et al.*, 2000; *Ventura et al.*, 2004; *Renard and Lang*, 2007]: the aim is to derive a statistical test for the hypothesis "H_0: data from all sites are stationary," when a test is repeated with a given significance level on several locations. The second aspect involves the consistency of changes detected within a given region. This is usually studied by representing results of at-site tests on a map, but more rigorous approaches exist, through test statistics defined over the entire region [*Douglas et al.*, 2000; *Yue and Wang*, 2002]. Exploiting the concept of regionalization may be a step forward [*Katz et al.*, 2002], by assuming that at-site samples are identically distributed (including an identical trend), up to some scale factor. Unfortunately, both field significance and regional consistency determination are difficult tasks, because specific problems arise at the regional scale. The most challenging difficulty of regional studies stems from the spatial dependence between data from different locations, which can significantly bias conclusions when ignored.

[5]   This paper aims to test existing techniques and develop new methods for detecting changes at the regional scale, and to illustrate their use on extreme discharge in France. It is organized as follows. Section 2 presents methods for assessing field significance (2.1.) and regional consistency (2.2.). These methods are then evaluated and

_____
[1]School of Engineering, University of Newcastle, Callaghan, New South Wales, Australia.
[2]Cemagref, U.R Hydrologie-Hydraulique, Lyon, France.
[3]LTHE, BP 53, Grenoble, France.
[4]EDF /R&D, Chatou, France.
[5]Météo-France, Ecole Nationale de la Météorologie, Toulouse, France.
[6]Maison des Sciences de l'eau Laboratoire Hydrosciences, Montpellier, France.
[7]Centre for Ecology and Hydrology, Wallingford, U. K.
[8]EDF/DTG, Grenoble, France.

compared using synthetic data in section 3. A case study was conducted on runoff series arising from a set of French rivers, at the national (3.1.) and regional (3.2.) scales. Moreover, potential links between changes in river flow and rainfall are investigated for a single hydro-climatic region, using a conceptual rainfall-runoff model (4.3). The main conclusions of the paper are summarized in section 5, together with broader perspectives and future developments.

2. Methods

2.1. Field Significance

[6] Field significance is assessed when a statistical test is repeated on several distinct data series (e.g., from several locations in a given region) and has been studied, for example, by *Livezey and Chen* [1983], *Lettenmaier et al.* [1994], *Douglas et al.* [2000], *Yue and Wang* [2002], *Ventura et al.* [2004], and *Renard and Lang* [2007]. In a change detection context, it aims at testing the hypothesis "H_0: data from all sites are stationary." In order to illustrate this concept, let us assume that 100 stations are tested with a significance level α equal to 0.05. Under the H_0 hypothesis, stationarity should be rejected for approximately five stations, corresponding to the 5% significance level. The question is how to conclude if, for instance, six stations showed significant trends. In other words, how can we estimate the minimum number of locally significant trends to conclude, with a regional significance level α', that the changes are not all due to chance? This requires knowing the distribution of N, the number of locally significant tests under the hypothesis that all series are stationary. If the observed number of locally significant results is larger than the $1 - \alpha'$ quantile of the distribution of N, the changes will be considered field significant for regional significance level α'.

[7] When data from all sites are independent, it can easily be shown that N follows a binomial distribution with parameters p (number of sites) and α (at-site significance level). However, in practice data series are rarely independent, and the distribution of N departs from the binomial distribution when dependence increases in the data set. Specific methods accounting for dependence therefore have to be applied.

2.1.1. Effective Number of Stations

[8] In order to account for the effect of intersite dependence, one approach consists of using an equivalent (or effective) number of stations (ENS) [*Matalas and Langbein*, 1962]. This concept has often been used to quantify the redundancy of information induced by dependence. Given a statistic S and a data set with p stations, the ENS p^* is defined as:

$$p^* = p \frac{Var_{ind}(S)}{Var(S)} \tag{1}$$

where $Var_{ind}(S)$ is the variance of S under the independence hypothesis. This means that, in terms of information content, p-dependent stations are equivalent to p^*-independent stations.

[9] An intuitive approach to assess the field significance of a set of statistical tests is to apply the binomial approximation to the ENS p^*. This results in a critical value N^*,

which can be expressed as a critical percentage $r = 100.N^*/p^*$. The field significance is finally assessed by comparing the observed percentage of significant trends detected in the original data set to r. Note that this approach is purely empirical because the ENS is not uniquely defined, as it depends on the S statistic considered in equation (1). Moreover, although the binomial distribution is the exact distribution of N in the independence case, this does not ensure it can be used to approximate the critical value N^* when using the ENS p^* in lieu of the actual number of stations p.

[10] In this paper, the ENS was derived from equation (1) by using the regional average Mann-Kendall (RAMK) statistic proposed by *Douglas et al.* [2000] and *Yue and Wang* [2002]. Let $X_i^{(j)}$ be the variable of interest computed at site j ($j = 1,\ldots,p$) on year i ($i = 1,\ldots,n$) and $S^{(j)}$ the Mann-Kendall statistic [*Mann*, 1945; *Kendall*, 1975] computed at site j:

$$S^{(j)} = \sum_{k=1}^{n-1} \sum_{i=k+1}^{n} sign\left(X_i^{(j)} - X_k^{(j)}\right) \tag{2}$$

The RAMK statistic S_R is the mean of the at-site Mann-Kendall statistics:

$$S_R = \frac{1}{p} \sum_{j=1}^{p} S^{(j)} \tag{3}$$

Under the null hypothesis H_0 that at-site series are stationary, the distribution of $S^{(j)}$ is asymptotically Gaussian with variance:

$$Var\left(S^{(j)}\right) = n(n-1)(2n+5)/18 \tag{4}$$

If the p sites are independent, it is easy to show that the variance of S_R is equal to $Var(S^{(j)})/p$. However, this result is not true when data series are spatially dependent. In this case, *Douglas et al.* [2000] and *Yue and Wang* [2002] showed that the variance of the RAMK statistic could be estimated by:

$$Var(S_R) = \frac{n(n-1)(2n+5)}{18p}(1 + (p-1)\overline{\rho}),$$

$$\text{where } \overline{\rho} = \frac{2\sum_{i=2}^{p}\sum_{j=1}^{i-1}\rho_{i,j}}{p(p-1)} \tag{5}$$

where $\rho_{i,j}$ is the correlation coefficient between data recorded at sites i and j. Consequently, the impact of intersite dependence on the RAMK statistic can be estimated by:

$$\frac{Var(S_R)}{Var_{ind}(S_R)} = 1 + (p-1)\overline{\rho} \tag{6}$$

Using equation (1), the ENS p^* is:

$$p^* = \frac{p}{1 + (p-1)\overline{\rho}} \tag{7}$$

Although the RAMK statistic is used to derive the ENS, we stress that the assessment of field significance using the ENS is fundamentally different from the RAMK test, which will be presented in section 2.2.2.

2.1.2. Bootstrap Procedure

[11] *Douglas et al.* [2000] suggested the use of a bootstrap procedure to assess field significance. This procedure can be used to estimate the distribution of N, the number of locally significant results under the hypothesis that all series are stationary at a given iteration ($m = 1,\ldots,n_{sim}$), the bootstrap procedure consists of the following steps. (1) A sample of years is created by bootstrapping the years used in the original data set (i.e., sample with replacement the same number of years as the original data set). (2) The new data set is formed by the records of the p stations corresponding to the bootstrapped years. (3) A statistical test is applied at each station of the bootstrapped data set. (4) $N^{(m)}$ is the number of stations with significant results obtained at iteration m. In each bootstrapped data set, the at-site data are free from any trend, because the bootstrap procedure randomly permutes the years. Moreover, the procedure maintains the existing intersite correlations. Consequently, the sample $(N^{(m)})_{m=1,\ldots,n_{sim}}$ can be used as an approximation of the distribution of N under the H_0 hypothesis that all series are stationary. The critical value related to the regional significance level α' can be estimated as the empirical $1 - \alpha'$ quantile of the simulated values $(N^{(m)})_{m=1,\ldots,n_{sim}}$.

2.1.3. Gaussian Copula

[12] *Renard and Lang* [2007] suggested using a Gaussian copula to evaluate the field significance of repeated statistical tests. In short, the Gaussian copula is a convenient tool to model both marginal distributions and the dependence structure of a multivariate data set. The model is parameterized by a positive-definite matrix describing the dependence between pairs of sites, and by marginal parameters (for instance, scale and location parameters of Gumbel distributions). The procedure can be described as follows: (1) estimation of the parameters of the Gaussian copula (dependence matrix + marginal parameters; see *Renard and Lang* [2007] for more details), (2) simulation of n_{sim} new multivariate data sets from the estimated Gaussian copula, and (3) computation of $N^{(m)}$, the number of significant trends for the mth simulated data set, $m = 1,\ldots,n_{sim}$. The critical value with the regional significance level α' is derived from the simulated values $(N^{(m)})_{m=1,\ldots,n_{sim}}$ as the $(1 - \alpha')$ quantile.

[13] There are limitations when using a Gaussian copula. First, since the model is parametric, the adequacy between model and data must be thoroughly checked. Second, given that the dependence matrix is required, the procedure may be unusable when the number of observed data is insufficient compared to the number of sites being considered, because of overparameterization. The Gaussian copula is therefore restricted to field significance assessment using a limited number of sites.

2.1.4. False Discovery Rate

[14] All three methods described in previous sections are based on the counting of at-site significant results, which implies that these approaches ignore the confidence with which at-site null hypotheses have been rejected (i.e., the p values of at-site tests). This could limit their statistical power. An alternative approach, based on the false discovery rate (FDR) concept, has emerged in the statistical literature [*Benjamini and Hochberg*, 1995] and has been applied in a climate context by *Ventura et al.* [2004]. We refer to these papers for a detailed description of this approach, and present only the application of FDR testing to field significance assessment, after *Wilks* [2006].

[15] The purpose of a FDR procedure is to identify at-site significant tests by controlling the FDR, which is defined as the expected proportion of falsely rejected null hypotheses among all rejections. At first sight, a FDR procedure might seem irrelevant for field significance determination, since it aims at making at-site decisions. However, it can also be used for this purpose by setting the nominal FDR to the desired regional significance level α' and declaring field significance if at least one at-site test is FDR significant.

[16] The standard FDR procedure used in this paper can be described as follows: let q_i be the p value related to the test performed at site i ($i = 1,\ldots,p$), and $q_{(i)}$ denotes the ith smallest of these p values. A FDR probability p_{FDR} is defined as follows:

$$p_{FDR} = \max_{i=1,\ldots,p} \left\{ p_{(i)} : p_{(i)} \leq \alpha'(i/p) \right\} \qquad (8)$$

Assuming all local tests are independent, declaring a significant change for sites with p values smaller than p_{FDR} ensures that the FDR remains smaller than α' [*Benjamini and Hochberg*, 1995]. Field significance will therefore be declared if at least one local test has a p value smaller than p_{FDR}.

[17] The assumption of independence might appear surprising, since we noted earlier that in most cases data are affected by spatial dependence and that accounting for this spatial dependence is the most challenging task in regional testing methods. However, contrary to other methods, the FDR procedure has been reported to be very robust when dependence exists between sites [*Ventura et al.*, 2004; *Wilks*, 2006], suggesting that it could still be applied for dependent data sets. This assertion will be evaluated in section 3.2.

2.2. Regional Consistency

[18] Climate is a global phenomenon and its change is likely to have an impact over an extended spatial area. Consequently, river flow in nearby catchments located within the same homogenous climatic area would be expected to be impacted by a similar climate change. In addition to climate, the streamflow regime also depends on the catchment topography and on the underlying geology, which could lessen or exacerbate variations in climate. It is therefore interesting to investigate the consistency of changes for catchments expected to respond similarly to the same climate driver. If all catchments indicate consistent changes, it is likely that these changes would be attributable to a similar cause, such as a significant change in the climate regime.

[19] Homogenous hydro-climatic regions could be defined by cross-referencing two distinct classifications, one characterizing hydrological mechanisms generating specific events (e.g., snowmelt- or rainfall-related floods) and one highlighting atmospheric circulation patterns characterizing the climate. One may argue that in a context of climate change, such a classification could evolve with time

[*Krasovskaia et al.*, 2003], as some rivers could be impacted by gradually moving from one regime to another. We acknowledge this problem, but herein hydro-climatic regions are considered a minimum framework to ensure homogeneity of hydro-climatic river properties rather than a descriptor of runoff behavior whose stationarity has to be assessed. Should these regions evolve with time, this should not be a limitation to the methods presented subsequently, as long as rivers within each region evolve in a similar way.

2.2.1. Regional Variable

[20] A first technique reduces the regional analysis to a univariate analysis by considering regional indices, i.e., variables defined over the entire region. For instance, if the date of occurrence of the annual maximum flood is computed on all sites, the mean date can be used as an indicator of the flood seasonality in the region. However, in most cases, the order of magnitude of the variable of interest greatly varies within the region. For instance, if the annual maximum flood is the variable of interest, the mean of the at-site values for each year has no physical meaning if the range of catchment areas is large. A preliminary data transformation is necessary to avoid such artifacts. In this study, a rank transformation was used, so that the regional variable studied is the mean annual rank of the at-site variables of interest. A Mann-Kendall test was used to assess the stationarity of this regional variable.

2.2.2. Regional Statistic: RAMK

[21] A complementary approach uses a regional statistic rather than a regional variable. *Douglas et al.* [2000] and *Yue and Wang* [2002] therefore suggest the regional average Mann-Kendall test (RAMK), based on the S_R statistic described in equation (3). The distribution of the following Z statistic under H_0 is asymptotically a standard normal distribution:

$$Z = \begin{cases} \dfrac{S_R - 1}{\sqrt{Var(S_R)}} & \text{if } S_R > 0 \\ 0 & \text{if } S_R = 0 \\ \dfrac{S_R + 1}{\sqrt{Var(S_R)}} & \text{if } S_R < 0 \end{cases}, \quad (9)$$

where $Var(S_R)$ can be estimated using equation (5). The RAMK test can then be applied by comparing the observed value of Z to the quantiles of the standard normal distribution.

2.2.3. Issue of Dependence: A Semiparametric Approach

[22] Regional variable analysis and regional statistics can both be used for testing stationarity at a regional scale. However, depending on the dependence structure of the observations, both methods could also detect nonconsistent changes. For example, consider ten strongly dependent sites impacted by an identical upward trend, while an eleventh site, independent of the others, is affected by an opposite downward trend. This could represent ten data series from the same river and one from a different river in a distinct area of the hydro-climatic region. Because of their large contribution to the regional variable (or the regional statistic), the first ten sites will mask the inconsistent behavior of the eleventh site. Nevertheless, in terms of information

content, ten highly dependent sites can be roughly equivalent to a single site.

[23] The approach developed for overcoming this limitation is based on data transformation and the multivariate Gaussian distribution. Let X denote the $n*p$ matrix of data. First, the data series from all sites are transformed by a normal score transformation as follows:

$$\tilde{X}_i^{(j)} = \phi^{-1}\left(\hat{F}_j\left(X_i^{(j)}\right)\right) \quad (10)$$

where ϕ is the cumulative distribution function (cdf) of the standard normal distribution and $\hat{F}_j$ is the empirical cdf of data $(X_i^{(j)})_{i=1,\dots,n}$ from site j.

[24] Second, the transformed data $\tilde{X}$ are assumed to follow a multivariate Gaussian distribution with variance matrix Σ. This is a strong hypothesis that must be verified on the observed series, because the construction creates Gaussian marginal distributions but not necessarily a Gaussian joint distribution.

[25] A first model is constructed under the H_0 hypothesis that all series are stationary. In that case, the mean of the multivariate Gaussian distribution is equal to zero, and the maximum likelihood (ML) estimator of the variance matrix is:

$$\hat{\Sigma} = \frac{1}{n}\tilde{X}^T\tilde{X} \quad (11)$$

The alternative model assumes the trend is consistent for all stations. The mean of the transformed data is:

$$E\left(\tilde{X}_i^{(j)}\right) = \beta\tilde{y}_i, \quad (12)$$

where $\tilde{y}_i = y_i - \bar{y}$ and y_i denotes the year corresponding to the ith data.

[26] The trend β does not depend on the site j; that is, the model assumes all sites to be affected by the same trend. Centering the years by $\tilde{y}_i = y_i - \bar{y}$ only aims to derive easier computations. The trend is estimated by maximizing the likelihood. Assuming the variance matrix Σ to be known, with $(\gamma_{i,j})_{i,j=1,\dots,p}$ the terms of the matrix Σ^{-1}, the ML estimate of the trend is (see Appendix):

$$\hat{\beta} = \frac{\sum\limits_{k=1}^{n}\left\{\tilde{y}_k \sum\limits_{i,j=1}^{p} \gamma_{i,j}\tilde{x}_k^{(j)}\right\}}{\left(\sum\limits_{k=1}^{n} \tilde{y}_k^2\right)\left(\sum\limits_{i,j=1}^{p} \gamma_{i,j}\right)} \quad (13)$$

[27] Unfortunately, in practice the variance matrix is usually unknown. Estimating jointly all parameters by the ML method may quickly become intractable when the number of sites increases, even with powerful maximization methods. Alternatively, the variance matrix can be approximated by equation (11). This does not necessarily reach the maximum of the likelihood, but provides an explicit formula and equation (13) becomes:

$$\hat{\beta} = \frac{I_p^T\left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}^T\tilde{Y}}{\tilde{Y}^T\tilde{Y}I_p^T\left(\tilde{X}^T\tilde{X}\right)^{-1}I_p} \quad (14)$$

where $\tilde{Y}$ is the column vector of centered years, and $\boldsymbol{I_p}$ is a column vector in which all elements are equal to one.

[28] The regional likelihood ratio (LR-reg) test is constructed by computing the deviance statistic:

$$
\begin{aligned}
D &= -2\left(L_0\left(\tilde{\boldsymbol{X}};\hat{\boldsymbol{\theta}}_{\boldsymbol{\theta}}\right) - L_1\left(\tilde{\boldsymbol{X}};\hat{\boldsymbol{\theta}}_{\boldsymbol{I}}\right)\right) \\
&= -2\left(\sum_{k=1}^{n}\log\left(N\left(\tilde{\boldsymbol{x}}_{\boldsymbol{k}};\boldsymbol{0},\hat{\boldsymbol{\Sigma}}\right)\right) - \sum_{k=1}^{n}\log\left(N\left(\tilde{\boldsymbol{x}}_{\boldsymbol{k}};\hat{\beta}\tilde{y}_k\boldsymbol{I_p},\hat{\boldsymbol{\Sigma}}\right)\right)\right)
\end{aligned}
\tag{15}
$$

This deviance should follow asymptotically the χ^2 distribution with one degree of freedom. The two-step procedure used to estimate β and $\boldsymbol{\Sigma}$ does not necessarily lead to the maximum of the likelihood, possibly resulting in a smaller deviance and therefore in a conservative test. Although this is preferable to a liberal test, the ability of the LR-reg test to detect regional trends has to be studied. This will be done in section 3.3.

3. Comparison of Methods Using Synthetic Data

[29] The adequacy of the methods described in section 2. was explored using synthetic data with known dependence properties. More accurately, we first aimed to verify that the methods were unbiased (i.e., the actual test level was in accordance with the nominal regional significance level α'). Moreover, the methods were compared on the basis of their power to detect regional changes. Generating spatially consistent data representative of river flow series is difficult, in particular because their spatial variability depends to a large extent on the hydrographical network structure. The synthetic data used in this section were generated using three techniques, simple enough for easy implementation, and mimicking various spatial dependence structures, at-site distributions, numbers of stations and record lengths. However, they only illustrate part of the spatial variability that could be encountered in real case studies.

3.1. Methodology: Simulation Procedures
3.1.1. Method 1
[30] Method 1 uses a multivariate Gaussian distribution, with mean zero and variance matrix $\boldsymbol{\Sigma}_\rho$:

$$
\boldsymbol{\Sigma}_\rho = \begin{pmatrix}
1 & \rho & \rho^2 & \cdots & \rho^{p-1} \\
\rho & 1 & \rho & \ddots & \vdots \\
\rho^2 & \rho & \ddots & \ddots & \rho^2 \\
\vdots & \ddots & \ddots & 1 & \rho \\
\rho^{p-1} & \cdots & \rho^2 & \rho & 1
\end{pmatrix}
\tag{16}
$$

An 80*50 data set can be generated from that distribution, mimicking 80 years of data recorded on 50 sites. This simulation method is straightforward, and the simple dependence structure allows an easy assessment of the impact of spatial dependence on field significance and regional consistency, by using increasing values of ρ. However, the resulting synthetic data are unlikely to be representative of real hydrologic data, as at-site distributions are rarely Gaussian, and the spatial dependence structure introduced is not realistic.

3.1.2. Method 2
[31] Method 2 is based on a Gaussian copula with Gumbel marginal distributions and is therefore more appropriate for reproducing extreme distributions. The parameters of the Gaussian copula used to simulate data arise from a real-world case study, using a set of 13 hydrometric stations in the northeast of France [Renard and Lang, 2007]. A data set of size 50 years * 13 stations can then be generated. These data could be considered more realistic than using the multivariate Gaussian hypothesis, as Gumbel marginal distributions and the dependence matrix were derived from observed discharge data. However, the Gaussian copula is far from being a universal model for multivariate extremes, and more complex dependence structures may be encountered in practice.

3.1.3. Method 3
[32] Method 3 is based on the rainfall generator TBM [Ramos, 2002] which simulates rainfall fields with given characteristics (at-site distribution, spatial dependence given by type of variogram, range, intermittency). A squared 20 × 20-km area is generated, with rainfall depth r simulated on 200 × 200-m pixels. The chosen at-site distribution is a lognormal distribution with parameters (1;1) (i.e., $\log(r) \sim N(1;1)$). The spatial characteristics of the rainfall field are generated assuming spherical variograms with range 5 km and 50 km. Figure 1 describes examples of such fields. The 20 white squares show the pixels data are extracted from, and are considered equivalent to rain gauges. Contours delimit areas where rainfall totals are accumulated. They are considered as equivalent to catchments, so that totals mimic areal precipitations. Because some areas are nested, the spatial dependence structure of these synthetic data is representative of dependence observed with streamflow data. Twenty areal rainfall series are computed for each simulated field. In Figure 1a, a rainfall field with a 5-km range is shown. Because dependence quickly decreases with distance, simulated data show weak spatial dependence. Conversely, with a range of 50 km (Figure 1b), dependence between data is greater. A total of 80 independent fields are generated, leading to a multivariate data set of size 80 years * 20 sites (or areas).

3.2. Field Significance
3.2.1. Adequacy With the Nominal Significance Level
[33] The actual level of the tests presented in section 2.1. was evaluated with the three simulation methods previously described. To this aim, 1000 data sets were simulated with each simulation method, and the level of each test was estimated by computing the rejection rate of H_0 among these 1000 simulated data sets. For unbiased tests, these rejection rates should remain close to the regional significance level α', set at 10% in this study.

[34] Table 1 shows the result of this analysis. The first line relates to the binomial approximation, which ignores spatial dependence. When the dependence in data is low (multivariate Gaussian with $\rho < 0.6$, or TBM data with a 5-km variogram range), the test is conservative, with the actual test level smaller than α'. Conversely, the test becomes liberal with highly dependent data (multivariate Gaussian with $\rho > 0.6$, or TBM data with a 50-km
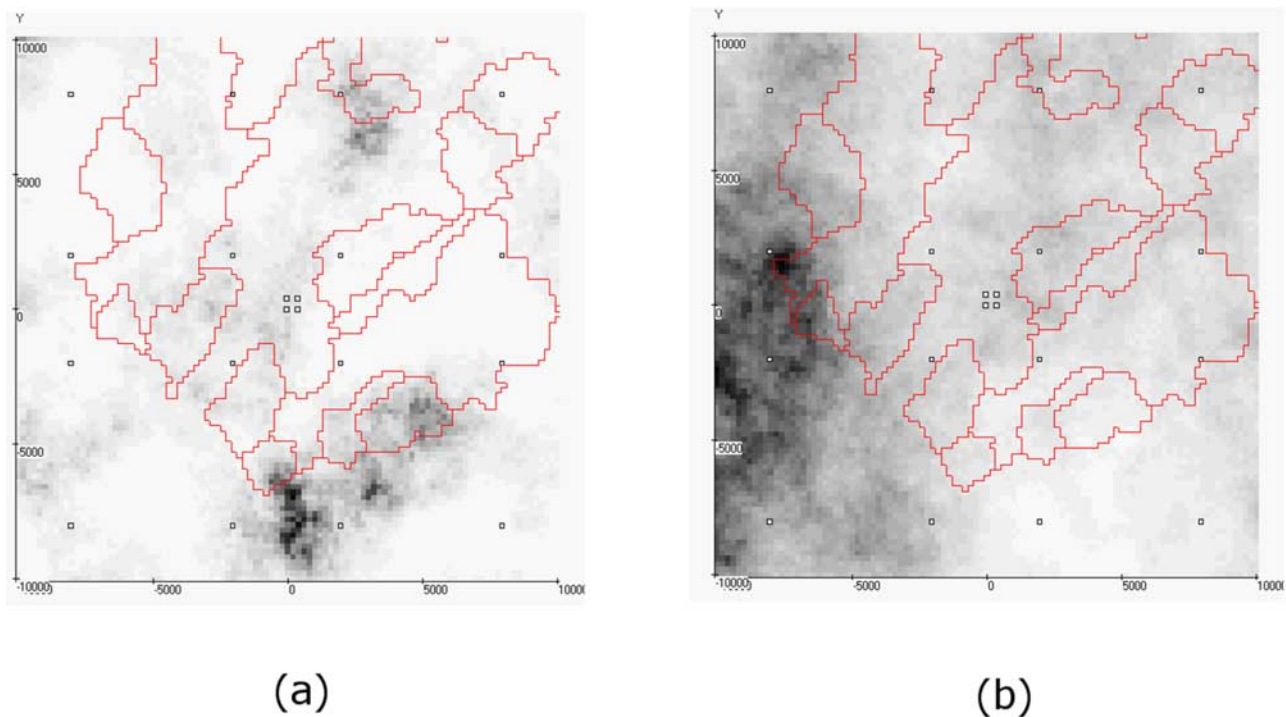
Figure 1. Rainfall fields generated by TBM with (a) low 5-km and (b) high 50-km variogram range.

variogram range). The results related to multivariate Gaussian data indicate that the actual test level monotonically increased as a function of ρ, reaching approximately 18% for $\rho = 0.9$. It can therefore be concluded that dependence significantly affects the behavior of this procedure, leading to a liberal test for high levels of dependence. The binomial approximation should therefore not be used for assessing field significance with spatially dependent data. The second line shows the results related to the ENS procedure: all actual levels are significantly smaller than the nominal level $\alpha' = 10\%$, leading to a highly conservative test. Although this is preferable to a liberal test, it could result in a very low power for detecting changes; this will be evaluated in the next section. The bootstrap and Gaussian copula procedures showed a similar behavior, with actual levels remaining acceptably close to α' (although the value of 13.5% for the Gaussian copula procedure with TBM data might appear suspicious). Last, the FDR procedure also leads to acceptable rejection rates,

although a tendency to become conservative for highly dependent data can be observed (multivariate Gaussian with $\rho = 0.9$ or TBM data with a 50-km variogram range).

3.2.2. Power

[35] The power of field significance procedures was estimated by computing the rejection rate of H_0 on 1000 nonstationary data sets. A trend was introduced in the simulated data as follows: let $\boldsymbol{X}^{(j)} = (X_i^{(j)})_{i=1,\ldots,n}$ denote the data at site j simulated by any of the methods presented in section 3.1. The trend δ is introduced at site j according to $Y_i^{(j)} = (1 + \delta \frac{i}{N}) X_i^{(j)}$, $i = 1,\ldots,n$. Contrasted types of changes can be considered at the regional scale: for instance, all sites could be affected by a weak trend, or conversely only a small number of sites could be affected by a large trend. The estimated power of the field significance procedures (except the binomial procedure, which was shown to be biased in the previous section) for detecting these different types of regional changes is shown in Figure 2 for synthetic data sets arising from TBM with high spatial dependence. Figure 2a

Table 1. Rejection Rates of the Null Hypothesis With Field Significance Procedures Applied on Stationary Data With Significance Level $\alpha' = 10\%$[a]

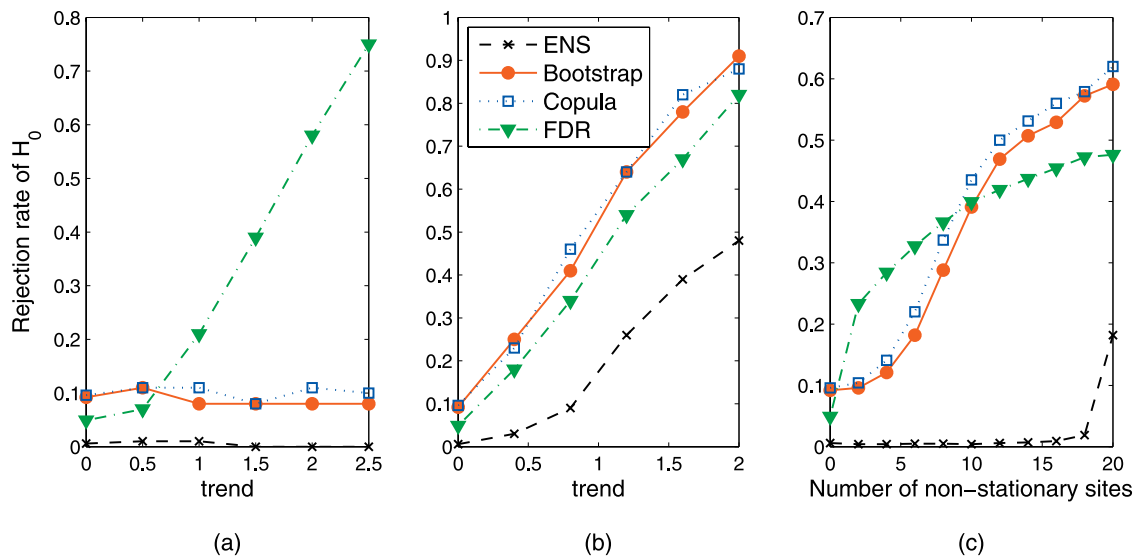| Field Significance Procedure	Simulation Procedure										
	Multivariate Gaussian, $\rho =$							TBM			
	0	0.2	0.4	0.6	0.8	0.9	Gaussian Copula	Point, 5 km	Areal, 5 km	Point, 50 km	Areal, 50 km
Binomial	5.5	6.0	8.1	10.8	13.6	18.2	10.8	6.5	6.2	14.2	13.7
ENS	5.2	2.9	1.7	0.7	0	0.4	0.6	1.8	0.4	0.2	0.6
Bootstrap	7.5	7.7	8.5	9.3	8.1	9.5	11.2	12.2	10.6	9.8	9.2
Gaussian copula	9.7	9.7	11.4	10.6	8.7	10.4	11.9	13.5	11.8	10.2	9.6
FDR	9.4	7.3	8.6	8.0	7.6	5.6	9.3	9.8	10.5	4.9	4.9

[a]Rejection rates are in percent.

Figure 2. Rejection rates of the null hypothesis computed on nonstationary TBM data with four field significance procedures ($\alpha' = 10\%$). (a) Only two sites are affected by the trend. (b) All sites are affected by the trend. (c) N_S sites are affected by a trend $\delta = 1$.

was obtained with only two sites affected by a trend. Only the FDR procedure was able to detect this change, while the power of all three other procedures did not increase with the trend's magnitude. This can easily be explained by the fact that these procedures are based on counting at-site significant tests and are not sensitive to how significant the at-site tests are. Figure 2b is related to a trend affecting all 20 sites. In that case, the power of all methods increased with the trend's magnitude, but contrasts appeared between them: the ENS procedure had a significantly lower power, because it is very conservative, as emphasized in the previous section. Bootstrap and copula methods had a similar behavior and were more powerful than the FDR procedure for such a generalized change, with a difference of approximately 10 points. Last, Figure 2c is related to a trend of given magnitude ($\delta = 1$), affecting an increasing number of sites. It summarizes the previous observations: when the regional change only affects a limited number of sites, the FDR procedure is the most powerful. However, as the number of sites affected by the trend increases, the bootstrap and copula methods catch up with the FDR procedure, and become more powerful for detecting a generalized change. On the other hand, the ENS procedure was unable to detect a trend with this magnitude.

3.2.3. Discussion

[36] Guidelines can be deduced from the results described in the previous sections. The binomial method is not a reliable tool for assessing field significance for spatially dependent data, because it becomes progressively more liberal when dependence increases. On the other hand, the ENS method is too conservative to be useful, as its power for detecting changes is very low. This might be due to one of following reasons: (1) the binomial approximation does not hold for the ENS or (2) the ENS is not uniquely defined since it is related to a given statistic, and the use of the RAMK statistic might be a poor choice for deriving equation (1). Note that the latter does not imply that the RAMK test itself performs poorly, since this approach is fundamentally different from the ENS method.

[37] The bootstrap, copula and FDR procedures are adequate tools for assessing field significance, since the simulations we performed did not highlight any significant bias, although the FDR procedure seems to become slightly conservative for highly dependent data. On the one hand, the bootstrap and copula methods behave in a very similar way; however, the bootstrap procedure is easier to apply and requires no parametric assumption about marginal and joint distributions of data. Consequently, this method might be favored in most applications. On the other hand, the FDR procedure is significantly more powerful for detecting changes affecting only a limited part of the sites, but is less powerful for detecting weaker generalized change. In conclusion, the choice between the bootstrap and the FDR procedures depends on the expected type of change. When no prior information about the regional change is available, a pragmatic approach would simply consist in applying both tests to the data.

3.3. Regional Consistency

3.3.1. Adequacy With the Nominal Significance Level

[38] The performance of the three regional tests (regional variable, RAMK and LR-reg) was evaluated using the three simulation procedures of section 3.1. The methods were first tested for bias, i.e., verification of the significance level α. The rejection rate of H_0 was computed using 1000 synthetic stationary data, with $\alpha = 10\%$. For all simulation methods, rejection rates remained close to α for the three test procedures (Table 2; lines 1, 3, 5). The RAMK and LR-reg tests were also applied with an independence hypothesis. In this case, rejection rates quickly increased with dependence, leading to very liberal tests. This shows the need to account for spatial dependence at a regional scale to avoid erroneously detecting a regional change.

3.3.2. Ability to Detect a Consistent Regional Trend

[39] The type of regional trend that can be detected by the three regional tests is explored here by generating 11 data series, ten of which are highly correlated, and the eleventh series independent of the others. Highly correlated streamflow

Table 2. Rejection Rates of the Null Hypothesis With Regional Testing Procedures Applied on Stationary Data With Significance Level $\alpha = 10\%$[a]

	Simulation Procedure										
	Multivariate Gaussian, $\rho =$							TBM			
Regional Test	0	0.2	0.4	0.6	0.8	0.9	Gaussian Copula	Point, 5 km	Areal, 5 km	Point, 50 km	Areal, 50 km
Regional variable	10.7	9.1	9.7	10.0	10.2	9.3	10.7	8.4	9.4	9.3	8.7
RAMK, independence hypothesis	9.2	17.8	25.5	40.4	55.7	66.5	52.0	22.2	31.1	66.4	68.3
RAMK	9.4	10.7	7.4	9.9	8.4	9.2	11.0	9.7	10.1	10.7	9.8
LR-reg, independence hypothesis	4.2	7.6	16.4	24.6	37.2	44.8	38.8	15.4	14.1	54.5	47.7
LR-reg	9.1	10.1	11.2	9.9	10.8	10.6	10.7	9.8	10.5	9.0	11.6

[a]Rejection rates are in percent.

series could represent data from the same river but different gauging stations. Forty years of data were thus generated from a multivariate Gaussian distribution with dimension 11, mean vector $\boldsymbol{0}$ and variance matrix $\boldsymbol{\Sigma}$:

$$
\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.99 & \cdots & 0.99^9 & 0 \\ 0.99 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0.99 & \vdots \\ 0.99^9 & \cdots & 0.99 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \quad (17)
$$

[40] An additive trend was added to these data. In the case of a consistent trend, this trend was the same for the eleven stations. In the case of a nonconsistent change, the trend applied to the eleventh station was the opposite of the trend of the ten other stations. One thousand synthetic data sets were generated and rejection rates were computed to estimate the power of the different tests for detecting such regional changes. All tests were able to detect a consistent regional trend (Figure 3a). The regional variable and the RAMK tests were nearly equivalent, whereas the LR-reg test was more powerful. In the case of a nonconsistent trend (Figure 3b), the difference between the methods is clear:

while the regional variable and RAMK tests still detected a trend, the rejection rate of the LR-reg procedure remained close to the prescribed level (10%), thus indicating LR-reg will generally not detect a trend which is not consistent at the regional scale.

3.3.3. Discussion

[41] These results can be used as a basis to choose the most appropriate method at a regional scale. Since all three tests are unbiased (at least on the basis of the simulations we performed), they are all candidates for regional stationarity analysis. In terms of power, no method was found to perform better but instead they showed an ability to detect different changes. The regional variable and the RAMK tests can detect dominant changes within a region. Although this may be an interesting task, this assessment is very sensitive to the choice of the stations analyzed. Conversely, the LR-reg test is more restrictive, as it forces the regional trend to be consistent to be detectable. In the following case study, since climate-based changes are investigated, the LR-reg will be favored, as it is felt that to be significant, trends need to be consistent over the homogeneous region of interest.

4. Regional Trend Detection in France

4.1. Data Sets and Variables

[42] The stationarity of hydrological extremes in France was studied at a regional scale, using the methodologies described above. Daily discharge series arising from 195
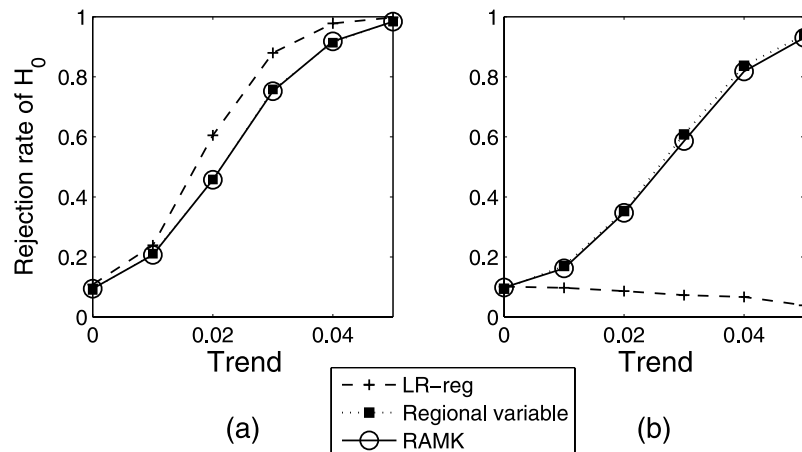
Figure 3. Rejection rates of the null hypothesis with three regional testing procedures, applied with $\alpha = 10\%$: (a) consistent trend and (b) nonconsistent trend (see text for further explanation).
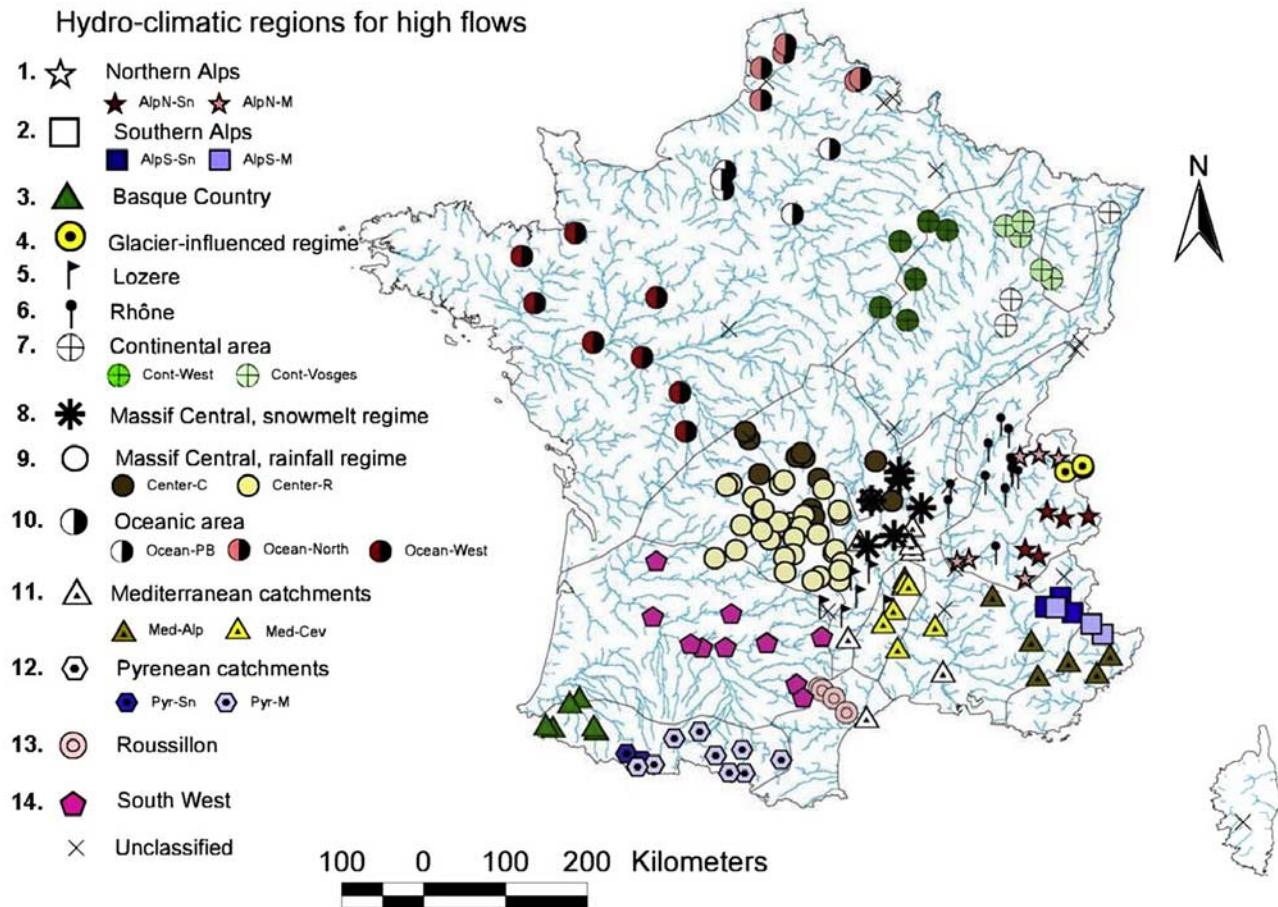
Figure 4. Hydro-climatic regions for high flows. Contours in thin lines show rainfall-based climate regions [after *Champeaux and Tamburini*, 1995].

gauging stations with a minimum record length of 40 years are available. Among these stations, 179 were used for high-flow analysis and 128 for low-flow analysis, the selection being based on data quality requirements. These stations were grouped in homogeneous hydro-climatic regions by combining a climatic rainfall-based classification defined by *Champeaux and Tamburini* [1995] and hydrological classifications based on the seasonality of extreme events. Regions were defined prior to (and thus independently from) the stationarity analysis in order to ensure a fair assessment of the trend's consistency within each region. Figure 4 and Figure 5 show these hydro-climatic regions for high flows and low flows, respectively. Additional details on the definition and the description of these regions can be found in the online material.[1]

[43] Extensive preliminary analyses were performed before attempting to detect any regional change. More accurately, at-site tests for trends and step changes were applied to numerous variables for each station [*Lang et al.*, 2006; *Renard et al.*, 2006], and the detected changes were reviewed by analyzing the gauging station archives. Focusing on the years suggested by the step change tests, we found that nonclimatic factors (notably measurement problems such as modification of the rating curve, displace-

ment of the station, equipment changes) explained a large part of the changes detected. As we wished to focus on climate-related changes, such stations were removed from the data set, reducing the number of data series available for the regional trend analysis. More accurately, 99 stations were available for high-flow analysis, 25 for snowmelt-related floods, 90 for low flows and 130 for mean annual flow.

[44] The regional trend analysis was based on a set of hydrological variables mainly derived for describing flood and drought phenomena. Two variables were used for high flows: (1) annual maximum (*Max*), extracted from daily discharge series and (2) date of the annual maximum (*Date*$_{max}$), converted to an angular value (*Date*$_{max}$ = *Julian date* x $(2\pi/365)$).

[45] Floods generated by both snowmelt and intense rainfall can affect some of the stations studied in this paper. Because these two phenomena result from different climatic and hydrological processes, they cannot be studied in the same way. Snowmelt-related events were therefore treated specifically. Analysis was restricted to the season of snowmelt occurrence in France (March–September), except for three stations mainly influenced by glaciers (June–October). Three variables were considered: (1) mean seasonal flow γ, which provides an indication on the snowmelt-generated volume, (2) center of mass χ, a seasonality measure suggested by *Stewart et al.* [2005] defined as the day when half of the total seasonal flow has run off, and (3) seasonal
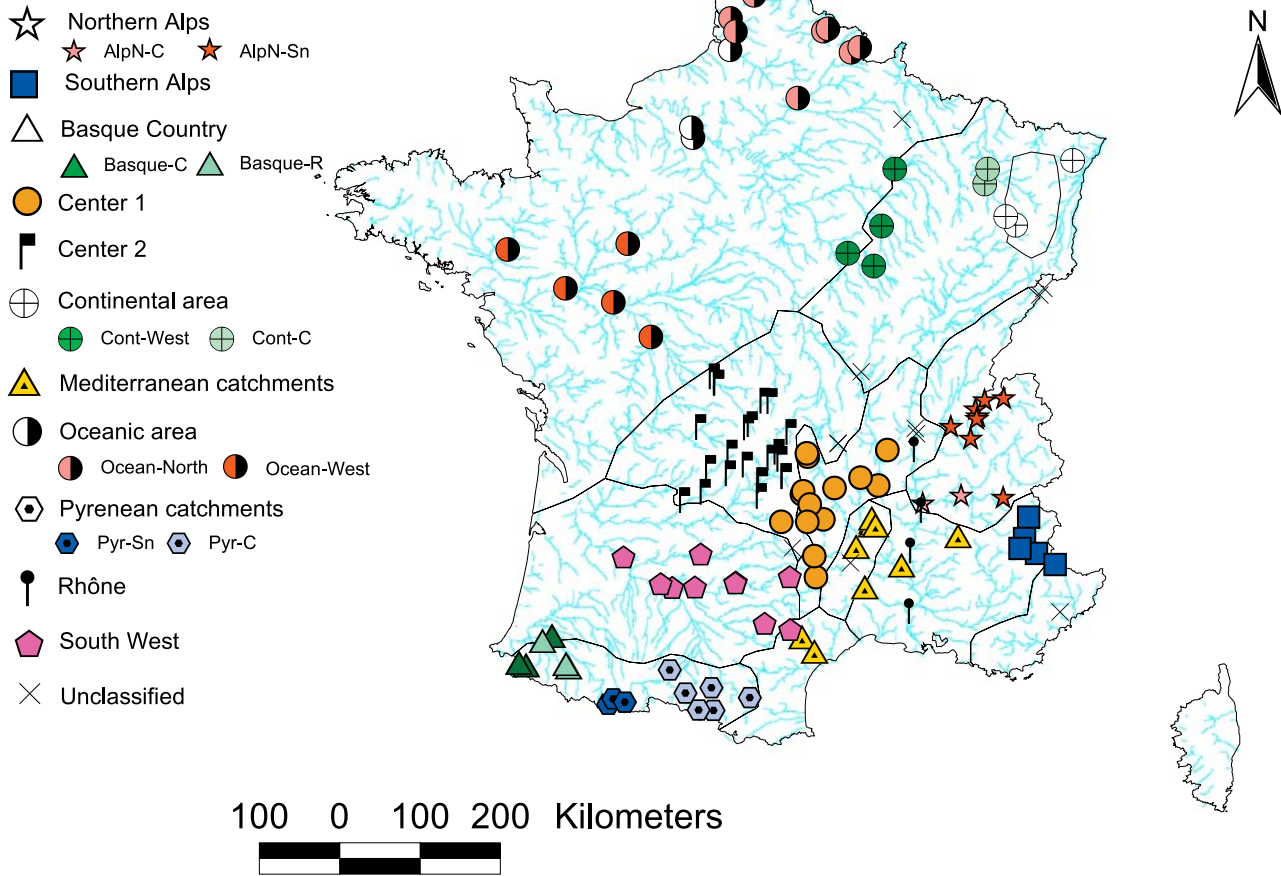
Figure 5. Same as Figure 4 for low flows.

maximum value of the base flow (β), describing snowmelt peak. Base flow is considered as a way of removing the influence of moderate rainfalls that can be superimposed on high discharge generated by snowmelt. β is estimated by applying the base flow separation algorithm suggested by *Tallaksen and Van Lanen* [2004].

[46] Hydrological droughts can be defined in a number of ways, summarized by *Tallaksen and Van Lanen* [2004]. Four variables were used to describe low flows: (1) annual minimum (*Min*) of 7-day mean discharge, (2) angular date of annual minimum of 7-day mean discharge (*Date*$_{min}$), (3) annual drought duration (*d*), and (4) the annual volume deficit (*v*). The derivation of variables *d* and *v* is based on a low-flow threshold, which was defined in this study as the 15th percentile of the flow duration curve. This choice is a tradeoff between minimizing the number of years with no draught event (*d* = 0 and *v* = 0) and ensuring the selection of extreme events; *d* is computed as the annual number of days with discharge lower than the threshold, and *v* is the corresponding volume deficit. Finally, although this paper mainly focuses on extremes, mean annual flow (*Mean*) was also studied as a descriptor of the average flow regime.

4.2. National Scale: Field Significance

[47] Field significance of at-site trends was assessed at the scale of France, using the bootstrap and the FDR procedures, found to be the most efficient methods in

section 3.2. The main difficulty relates to the handling of missing data, which is not obvious in the case of the bootstrap procedure. In the data sets considered, the number of years shared by all stations was equal to zero. Consequently, the bootstrap procedure was only applied to a subset of stations whose years between 1965 and 2000 had no missing data (extended to 1960–2000 for snowmelt flood variables). The second column of Table 3 shows the number of stations used for field significance assessment. Conversely, the FDR procedure can analyze series with nonconcomitant years, because it only relies on the p values of at-site tests. Consequently, it is also applied to the whole data sets.

[48] Table 3 shows the result of this analysis. A Mann-Kendall [*Mann*, 1945; *Kendall*, 1975] test was used as at-site test for high-flow variables, and a modified Mann-Kendall test [*Hamed and Rao*, 1998] accounting for potential autocorrelation was used in the case of low and mean flows. Local and regional significance levels are set to $\alpha = \alpha' = 5\%$. For the subset of stations, the critical percentage of at-site significant results as estimated by the bootstrap procedure is shown in fifth column of Table 3. This percentage was always larger than the observed percentage of at-site significant trends (fourth column), which means that none of the variables studied reached field significance at the national scale. The FDR procedure leads to the same conclusion, given that the number of FDR significant results was always equal to zero. Consequently,

Table 3. Summary of Field Significance Assessment Results[a]

	Reduced Data Set					Whole Data Set		
Variable	Number of Stations	Number of Common Years	Percent of At-Site Significant Results	Critical Percentage	Number of FDR-Significant Results	Number of Stations	Percent of At-Site Significant Results	Number of FDR-Significant Results
			High Flows					
Max	83	36	7.2	18.1	0	99	10.1	2
$Date_{max}$	83	36	3.6	19.3	0	99	2	0
			Snowmelt-Related Floods					
χ	15	41	6.7	46.7	0	25	12	0
β	15	41	0	20	0	25	12	1
γ	15	41	6.7	26.7	0	25	12	2
			Low Flows					
Min	58	36	6.9	19	0	90	6.7	0
$Date_{min}$	58	36	0	22.4	0	90	6.7	0
d	58	36	1.7	29.3	0	90	3.3	0
v	58	36	3.4	31	0	90	1.1	0
			Regime					
Mean	86	36	9.3	25.6	0	130	4.6	0

[a]Local and national levels are equal to 5%.

the assessment of field significance using the subset of stations leads to the conclusion that the H_0 hypothesis (all stations are stationary) cannot be rejected.

[49] Applying the FDR procedure to the whole data set provided slightly different results, as the trends detected on respectively two, one and two stations for *Max*, β and γ variables were FDR-significant. To the best of our knowledge, these trends cannot be linked to any particular measurement problem or direct influence. It can be noted that one of the stations showing a FDR-significant result for variable *Max* was studied by *Andréassian* [2002], who detected a trend in the rainfall-runoff relationship. However, as the number of FDR-significant results remains very small, they are likely to reflect particular behaviors of the related watersheds rather than some generalized evolution at the scale of France. Consequently, the search for a generalized change in extreme hydrological events at the national scale through field significance assessment remains largely inconclusive. However, the study undertaken was limited by the power of at-site tests to detect a change. A search for consistent changes at a smaller regional scale could increase the power to detect trends, because it could be anticipated that several weak but consistent changes in a given region may reflect a significant regional change.

4.3. Regional Consistency in Hydro-climatic Regions

[50] Because of the existence of varied climate and river flow regimes, it was felt important to analyze potential trends at a smaller spatial scale than the whole of France if one is interested in consistency. This section describes the application of the regional LR-reg procedure to assess the existence and consistency of potential trends in the hydro-climatic regions defined in section 4.1.

[51] The hypothesis of joint normality of the normal-score-transformed data underlying the LR-reg test was verified using paired scatterplots of transformed data and qq plots based on a quadratic function of the data (see *Renard and Lang* [2007] for examples of such graphics and *Mardia* [1980] for additional methods for verification of normality). In most cases, the data verified the LR-reg test

assumptions. Exceptions were detected for annual maxima series in the Center, Mediterranean and Oceanic regions, which could result from the limited number of years compared with the number of stations, implying over-parameterization problems. Consequently, only subregions were subsequently investigated. The transformed date of annual maximum was found not to be normal in most regions. Consequently, the regional variable approach was used for the date of the annual maximum by performing a Mann-Kendall test on the regional mean annual date. For snowmelt-related floods, all transformed variables verified the hypothesis of joint normality. For low-flow variables, the mean date of the annual minima was also studied using a regional variable approach. Finally, the LR-reg test was used to study annual mean flows, but only large regions were investigated, because subregions reflected differences in the drought regime that were not relevant for mean annual flow regime. Because autocorrelation is not taken into account by the LR-reg test, the results should be considered with caution for regions where groundwater control prevails (e.g., the northern part of the Oceanic region and the Paris Basin).

[52] The LR-reg test was applied with 1, 5 and 10% significance levels. The results for high-flow variables are shown in Table 4, the percentages indicating the minimum level leading to a significant result (e.g., +5% denotes a positive trend significant with $\alpha = 5\%$ but not with $\alpha = 1\%$). A first comment can be made regarding the limited number of stations available in some regions or subregions. Because stations where changes have been related to nonclimatic factors were discarded, some regions only include two or three stations. In such cases, the conclusions should be considered with caution. A significant result obtained on two stations will indeed indicate consistency, but this consistency should not be extrapolated to all existing rivers in the region. A solution would have been to merge some regions or subregions, but it was felt that this did not produce a fair analysis, as the hydro-climatic regions must be defined independently from the stationarity analysis.

Table 4. Results of Regional Testing Procedures Applied on High-Flow Variables[a]

Region[b]	Number of Stations	Number of Common Years	Significance Max	Significance $Date_{max}$
AlpN-M	3	40	x	x
AlpS-M	2	31	x	x
Basque	3	35	x	x
Centre	26	18	/	x
Centre-C	7	37	x	x
Centre-R	12	37	x	x
Cont	10	31	x	x
Cont-west	4	40	x	x
Cont-Vosges	5	37	+10%	x
Lozère	4	38	x	x
Med	20	22	/	x
Med (without Alps and Cevennes)	8	41	x	x
Med-Alp	3	28	x	x
Med-Cev	9	31	x	x
MassCent	4	41	x	x
Ocean	10	21	/	x
Ocean-PB	2	29	+5%	x
Ocean-north	4	36	+5%	x
Ocean-west	4	63	x	x
Pyr-M	4	32	−5%	x
Rhône	6	43	x	x
Roussillon	3	36	x	x
Southwest	9	24	+5%	x

[a]Symbols are as follows: x, nonsignificant result; /, test not performed on the region (see text for further explanations); and %, significance of the regional trends, positive (negative) for upward (downward) trend.
[b]Regions are bold; subregions are not.

[53] A significant increase in annual maxima was found in the Vosges subregion (continental region), in two subregions of the Oceanic area and in the southwest. Conversely, annual maxima of rainfall-related floods significantly decreased in the Pyrenees. For snowmelt-related floods (Table 5), a trend toward earlier floods was detected in the northern Alps, and the three glacier-controlled rivers showed a significant increase in mean seasonal flow. Although only based on 17 common years and three stations, this is consistent with the ablation of these glaciers observed by *Vincent* [2002] and *Vincent et al.* [2004]. For low-flow variables (Table 6), only the Oceanic region showed a significant trend (increase) in annual minimum flow, but more changes were detected for variables duration (d) and volume deficit (v). Less severe droughts were detected in the Alps (both northern and southern), while more severe droughts were found in the Basque region, the Pyrenees and the Oceanic region. In the latter case, the increase in drought duration is in apparent contradiction with the increase in annual minima. Finally, mean annual flow was found to decrease in the Pyrenees (last column of Table 6).

[54] In sum, three geographical regions were found to have experienced consistent regional change.

[55] 1. The northeast experienced a slight increase in flood peaks. Although detected only in the Vosges subregion (five stations, Table 4), additional significant results related to other flood-based variables were also detected in the west subregion and in the whole Continental region (e.g., annual maxima of d days mean discharges, where d is a characteristic flood duration [*Renard*, 2006]).

[56] 2. The Alps experienced less severe low flows for three stations in the north and four stations in the south,

Table 5. Same as Table 4, for Snowmelt-Related Floods

Region	Number of Stations	Number of Common Years	Significance χ	Significance β	Significance γ
AlpN	5	40	−5%	x	x
AlpS	5	30	x	x	x
Pyr	6	30	x	x	x
MassCent	4	55	x	x	x
Gla	3	17	x	x	+5%

along with earlier snowmelt in the northern part (five stations) and increased volume related to glacier melting in three stations.

[57] 3. The Pyrenees (including the Basque region) experienced a decreasing trend in high (Pyrenees, five stations), low (three stations in Pyrenees and five in Basque region) and mean (Pyrenees, six stations) flows.

[58] In addition, significant changes were also detected in the north and Paris Basin subregions of the Oceanic area. However, there is little confidence in these results, due to the prevailing groundwater control in these regions (chalk area). This particularity induces long-term persistence in flow (including high flows for the north subregion), which can bias the conclusions of the statistical tests. Conversely, data in the three geographical regions described above do not present significant autocorrelation.

[59] Given that the results found at the regional level are based on thoroughly quality-checked data, they are likely to be representative of genuine changes in the regime rather than a result of measurement problems or direct influences. However, these findings are not sufficient to demonstrate the existence of an impact of climate change on hydrological extremes in France. Other factors, such as long-term modification of the rainfall-runoff relationship and catchment evolution, could also result in changes in the hydrological regime. The next section describes a first attempt at exploring this possibility.

4.4. Investigation of the Causes of River Flow Changes

[60] As noted by *Andréassian et al.* [2003], the changes in runoff depend both on climate drivers (e.g., rainfall,

Table 6. Same as Table 4, for Low- and Mean-Flow Variables

Region	Number of Stations	Number of Common Years	Min	$Date_{min}$	d	v	Mean
AlpN	3	41	x	x	x	−5%	x
AlpS	4	32	x	x	−10%	x	x
Basque	5	34	x	x	+5%	+10%	x
Basque-C	3	35	x	x	+5%	+10%	/
Basque-R	2	90	x	x	x	x	/
Center1	8	33	x	x	x	x	x
Center2	13	35	x	x	x	x	x
Cont	8	31	x	x	x	x	x
Cont-C	2	38	x	x	x	x	/
Cont-west	4	40	x	x	x	x	/
Med	6	30	x	x	x	x	x
Ocean	6	26	+5%	x	x	x	x
Ocean-north	5	34	x	x	+10%	x	/
Pyr	3	30	x	x	+5%	x	−5%
Pyr-C	2	33	x	x	x	x	/
Rhone	3	81	x	x	x	x	x
Southwest	5	27	x	x	x	x	x

Table 7. Hydrological Properties and Goodness-of-Fit for the Four Modeled Catchments

River @ Station	Catchment Area (km^2)	Station Elevation (m)	Mean Annual Flow (m^3 s^{-1})	10-Year Daily Flow (m^3 s^{-1})	Nash (%) Calibration (1968–1979)	Nash (%) Validation (1980–1998)	R^2 Between Observed and Simulated MAXAN (1968–1998)
Moselle @ Saint-Nabord	621	371	24	360	85.3	87.4	0.70
Moselle @ Epinal	1220	324	39	470	85.5	87.0	0.75
Madon @ Pulligny	940	225	10	220	78.6	83.0	0.63
Moselle @ Toul	3350	201	63	790	87.9	89.1	0.78

evapotranspiration) and on the stationarity of the watershed hydrological behavior. A change in either (or possibly both) of these phenomena could result in a change in runoff. Changing rainfall or evapotranspiration could occur as a result of climate change, while changing hydrological behavior could occur as a result of changes in land use (e.g., deforestation). In this section, the problem of assessing the potential cause of a change in streamflow is addressed.

[61] The evolution of temperature and rainfall over the last few decades has been extensively studied in France [*Mestre*, 2000; *Moisselin et al.*, 2002; *Spagnoli et al.*, 2002; *Dubuisson and Moisselin*, 2006; *Parey et al.*, 2007]. Significant and consistent changes have been reported for temperatures, including extreme indices. Annual or seasonal rainfall has also shown some evidence of change, although less pronounced and more regionally contrasted than those observed for temperature. At present, no consistent pattern has been highlighted for extreme rainfalls. However, simply comparing trends detected on observed rainfall or temperatures with trends detected on observed discharges makes little sense, because of the strong nonlinearity of the rainfall-runoff relationship. For instance, there is no reason to believe that a trend in annual maxima of rainfall will result in an identical trend in annual maxima of discharge. Similarly, if a change in the hydrological behavior is investigated, comparing a trend detected on a descriptor of the watershed (e.g., percentage of forest cover) with the trend detected on runoff is groundless. This nonlinearity has to be tackled by using a rainfall-runoff model. A preliminary attempt was made to evaluate whether the trend detected in runoff was consistent with the potential trends in rainfall and whether a possible change in the rainfall-runoff relationship could be brought out.

4.4.1. Input Data and Model

[62] The case study considered here involves four hydrometric stations located in northeast France, which were chosen because good quality rainfall data were available. The general hydrological properties of the related catchments are shown in Table 7. An increase in annual maximum flow was detected in this region (see section 4.3.). Consequently, our main interest was to evaluate whether this change was related to evolving rainfall patterns or modifications in the watershed behavior. We will therefore only focus on high flows.

[63] Daily input variables of the hydrological model (Potential Evapotranspiration PE and rainfall) are available from 1968 to 1998. PE is derived from temperature and relative humidity data series recorded in Nancy, the closest meteorological station. The derived PE values are interannual mean values with a time step of 10 days, obtained with

the Penman equation [*Penman*, 1948]. A polynomial interpolation was used to derive daily PE values. Because we were only interested in high flows, whose sensitivity to PE is limited, using interannual PE data was deemed acceptable. Daily areal rainfall for each catchment was derived from the dense rain gauge network used by *Javelle* [2001].

[64] Areal precipitation and PE series were used as input of the lumped rainfall-runoff model GR4J [*Perrin et al.*, 2003] to derive daily flow series. This four-parameter model was calibrated to maximize the Nash criterion [*Nash and Sutcliffe*, 1970] on daily river flow, using the 1968–1979 period for calibration. The first year was used for the initialization of the stores in the model. The Nash criteria showed good agreement between observed and simulated streamflow for each station, both in calibration and validation (Table 7). The last column of Table 7 also shows R^2 coefficients between observed and simulated annual maximum runoff (*Max*), which is of primary interest because stationarity analysis will be performed on this variable. Although acceptable, these values clearly indicate that the fit between observed and simulated *Max* values is far from perfect.

4.4.2. Comparison of Trends on Observed and Simulated Streamflow Series

[65] The LR-reg test was applied to the annual maximum daily discharge *Max* derived from observed and simulated flow series over the entire 1968–1998 period (Figure 6). A 5% significant positive trend was detected on observed *Max*, consistent with the trend detected on the stations of the Continental area (see section 4.3.). A positive trend was also detected on *Max* derived from the simulated series, differing significantly from zero at 10% level. The important point is that neither model parameters nor interannual evapotranspiration evolved from year to year in the hydrological model. Consequently, the trend detected in simulated *Max* can only be explained by the evolving rainfall pattern. Because a comparable trend was detected on observed *Max*, this indicates that this change could be at least partly related to changes in rainfall.

[66] It is then interesting to assess whether changes can be detected in observed rainfall. To this end, variables describing high rainfall were derived, such as annual maximum 1-day ($RX_{1\ day}$) and 3-day ($RX_{3\ day}$) depth and annual number of days with rainfall depth greater than 10 mm ($R_{10\ mm}$). The LR-reg test was then applied to each rainfall variable. Only weak changes were detected in the rainfall characteristics (Figure 7), the only significant change being a slight increase in $R_{10\ mm}$. In particular, maximum annual rainfall did not exhibit any significant trend for 1-day depth or 3-day depth, which highlights the strong nonlinearity of the rainfall-runoff relationship: the
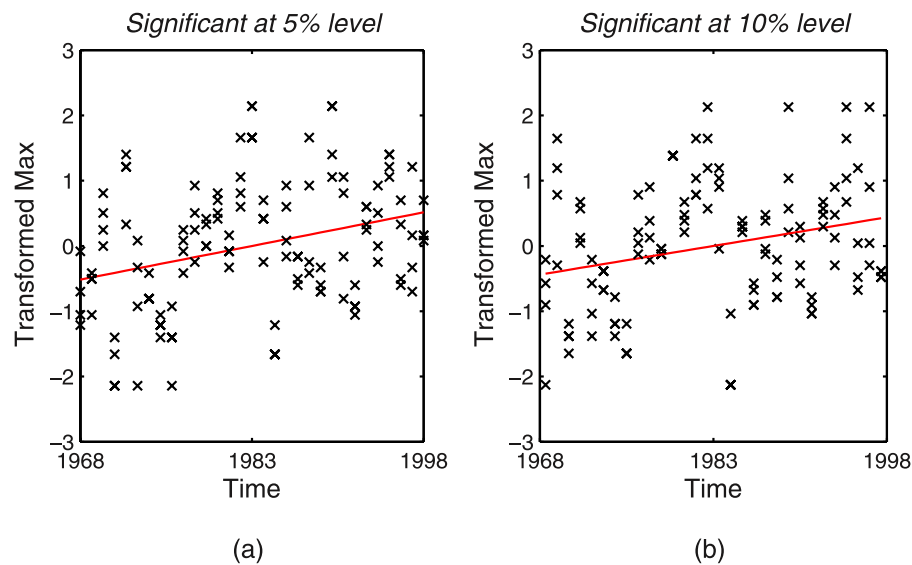
Figure 6. Regional trends on annual maximum flow for four modeled catchments for (a) observed values and (b) simulated values. Variables are transformed by normal score.

trend in extreme runoff cannot be explained by an identical behavior of extreme rainfall, but might be related to an increased frequency of wet (but not extreme, as a 10-mm depth can hardly be viewed as extreme) conditions.

[67] This analysis must be considered as a preliminary attempt to link changes in rainfall and runoff. The results are encouraging, given that the hydrological model was able to recreate a trend detected on observed runoff, even though the ability to correctly model extreme discharges was far from perfect (Table 7). However, this analysis does not provide quantified evidence of a relation between rainfall and detected trend in annual maximum discharges. A more rigorous approach would require using a statistical test, in order to assess whether the consistency between rainfall and runoff changes is significant. This test could consist in investigating the equality of the trend detected in observed and simulated discharge. This would imply accounting for sampling uncertainty (because the trend is estimated on only 30 years of data) and modeling uncertainty, related to the

calibration process. Although the former uncertainty might be quantified by using asymptotic statistical results, the latter one is more challenging. Methods such as BATEA [*Kavetski et al.*, 2006a, 2006b] might be useful for this purpose. Although such a development would be of great interest, it is beyond the scope of this paper, and is left for future development.

[68] Finally, a similar study was undertaken for the snowmelt-related floods in the Alps (not detailed in this paper), showing earlier melting consistent with a significant increase in temperature in this area [*Renard*, 2006].

4.4.3. Investigation of Potential Changes in the Rainfall/Runoff Relationship

[69] The analysis on four catchments in the northeast showed that the positive trend detected in maximum annual flood flows might be related to a climatic mechanism because it is consistent with observed rainfall. However, the hydrological behavior of studied watersheds may also have changed at the same time, potentially attenuating or
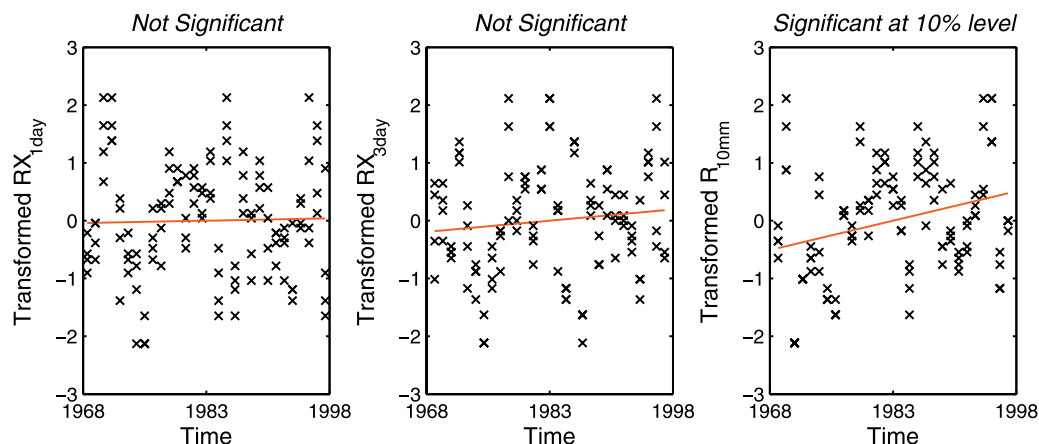

Figure 7. Regional trends on three rainfall variables for the four catchments studied. Variables are transformed by normal score.

enhancing the trend due to rainfall. This possibility is addressed in this section, using the statistical test developed by *Andréassian et al.* [2003] (hereinafter referred to as APM test). The APM test aims at identifying watershed behavior over distinct time periods by successive calibration of the model using the data of each period. It then uses a resampling approach to quantify the significance of trends. *Andréassian et al.*'s [2003] can be consulted for a complete description of this test.

[70] The APM test was applied to the four catchments studied in the previous section. Following the guidelines of *Andréassian et al.* [2003], the 1968–1998 period was divided into six subperiods of 5 years. Because we are mainly interested in floods in this case study, an adequate variable has to be derived for describing high flows of each subperiod. Unfortunately, using annual maximum discharge is impossible, because it would require using 1-year subperiods, which is too short to allow for a proper calibration of the model. Consequently, high flows are described by the volume exceeding a high threshold (chosen as the 95% percentile of the observed daily discharges), as suggested by *Andréassian et al.* [2003].

[71] No significant change was detected for three of the four stations. Conversely, the Madon at Pulligny station showed a significant change at 10% level, which means that the watershed behavior significantly evolved during the study period. More accurately, this change depicts a less responsive catchment (i.e., for identical rainfall, the generated high-flow volume decreased). This could have attenuated the increasing trend detected on observed annual maximum runoff. However, we are unable to provide a physical explanation for this finding, since no particular change has been reported for this watershed over the last 30 years. This example therefore emphasizes the difficulty in studying the stationarity of discharge data, because several forcings (climatic or watershed hydrologic behavior) may have contrasted impacts on runoff.

5. Conclusion

[72] The impact of climate change on hydrological regimes is still an open question, as illustrated by the lack of a clear signal emerging from large-scale studies [*Kundzewicz et al.*, 2005; *Svensson et al.*, 2005]. Some authors offered explanations regarding this inability to detect any robust trend [*Svensson et al.*, 2006; *Wilby*, 2006]. A possible cause (among many others) might be the weak power of at-site statistical methods to detect a trend in hydrological data affected by a very high level of natural variability, especially if extreme regimes are investigated. This paper therefore explored methods for regional trend detection and provided an example of application using French river data.

[73] First, methods for regional-scale trend analysis were described and evaluated. Methods for assessing the field significance of a set of at-site tests were compared, using different synthetic data sets. The bootstrap and the FDR methods were found to be adequate and robust tools. Consistency of trends within a homogeneous hydro-climatic region was also assessed. A semiparametric test (LR-reg) was developed and compared with methods based on regional variables or regional statistics. Although none of the methods were biased, they detected different changes: the regional variable and the RAMK tests can be used to

detect dominant changes within a region, but they will also detect nonconsistent changes. Conversely, the LR-reg test is more restrictive, as it forces the regional trend to be consistent to be detectable.

[74] Regional methods were then applied to ten variables representative of the high-, low- and mean-flow regimes on a set of gauging stations in France. At the scale of the entire country, the search for a generalized change in extreme hydrological events through field significance assessment remained largely inconclusive. At the smaller scale of hydro-climatic regions, encompassing catchments with comparable hydrological behavior subjected to similar climate forcings, the LR-reg procedure showed no significant result on most regions. However, consistent changes were detected in three geographical areas. In the northeast, a slight increase in annual maximum flow was found. In the Alps, less severe droughts were detected, along with earlier snowmelting in the northern Alps, and increasing volume related to glacier melting in glacier regime rivers. Finally, in the Pyrenees, a decreasing trend was detected in low, mean and high flows.

[75] The relationship between rainfall and flow changes was also explored, on the basis of a case study of four stations in the northeast. The positive trend detected on annual maximum flow was found to be consistent with observed rainfall, which may indicate climate-driven changes. Possible changes in the hydrological behavior of the watershed were also explored, using the test proposed by *Andréassian et al.* [2003]. A significant trend toward a less responsive catchment in high flows was detected for one of the four stations. These results show that the trend in climatic variables may interact with changes in the watershed hydrological behavior, leading to a complex response of runoffs.

[76] From a broader perspective, the results obtained in this paper highlighted some of the challenges related to the detection of runoff changes in a context of nonstationary climate.

[77] Although not developed in detail in this paper, preliminary analyses showed many stations from the initial data set were affected by significant changes, but most of these changes could be explained by nonclimatic factors, principally measurement problems. Such biases are unlikely to be specific to France and might be encountered in any river flow series. Consequently, an intensive quality check is recommended as a necessary step when undertaking trend detection analysis, as such biases might hide a smoother and more consistent signal due to climatic factors. No generalized change was then found after this quality check at the scale of France, based on a field significance assessment. This result is not really surprising, considering the large natural variability in extreme discharge [*Svensson et al.*, 2006]. However, it is important to note that a nonsignificant result does not mean that the null hypothesis can be accepted, but only that it cannot be rejected. In other words, our results do not imply that climate change has no impact on hydrological extremes, but should this impact exist, it cannot be detected today using at-site methods. The same analysis on the same data (but longer series) should be conducted in the future and could provide different conclusions [*Wilby*, 2006].

[78] Hydrologists routinely use the concept of regionalization to overcome the difficulty in characterizing natural variability with short record lengths. This concept can also be a very useful tool in detecting consistent trends in hydroclimatic regions. Indeed, the LR-reg test can be viewed as a trend regionalization procedure, as it assumes all stations in a homogeneous hydro-climatic region are affected by an identical trend. The results showed that for three regions, a regional trend could be detected, while this did not appear from at-site analyses.

[79] Finally, providing a physical explanation for the streamflow trends remains difficult. Here hydrological modeling was considered a preliminary approach to link the changes in rainfall and runoff. A more rigorous study would require quantifying the uncertainty in model calibration. Had this been done, it would not be sufficient to establish causality from climate change, as the changes in the rainfall regime might merely reflect cycles in the natural variability. Searching for a formal link between changes in hydrological extremes and climatic change would require using fingerprint methods [*Hasselmann*, 1993, 1998]. Unfortunately, this would involve using GCM outputs, and therefore dealing with new uncertainties, for example, from downscaling or parameterization of the GCM. Methodological improvements are still needed, both for quantifying these uncertainties and combining them into the analysis.

Appendix A

[80] Let X be a $n*p$ data matrix, and $\tilde{X}$ the matrix of data transformed by normal score, with variance matrix Σ. $\tilde{X}_i^{(j)}$ denotes the transformed data recorded at site j and year y_i. The data are assumed to be affected by a regional trend, so that $E(\tilde{X}_i^{(j)}) = \beta \tilde{y}_i$, with $\tilde{y}_i = y_i - \bar{y}$.

[81] The log likelihood of the data $\tilde{X}$ can be written as follows:

$$
\begin{aligned}
L(\tilde{X};\beta) = & -\frac{n}{2}\log(|\Sigma|) - \frac{np}{2}\log(2\pi) - \frac{1}{2}\sum_{k=1}^{n} \\
& \cdot \left(\tilde{x}_k^{(1)} - \beta\tilde{y}_k, \ldots, \tilde{x}_k^{(p)} - \beta\tilde{y}_k\right)\Sigma^{-1} \\
& \cdot \left(\tilde{x}_k^{(1)} - \beta\tilde{y}_k, \ldots, \tilde{x}_k^{(p)} - \beta\tilde{y}_k\right)^T
\end{aligned}
\tag{A1}
$$

Deriving this expression and developing the matrix product gives:

$$
\begin{aligned}
\frac{\partial}{\partial\beta}L(\tilde{X};\beta) = & -\frac{1}{2}\sum_{k=1}^{n}\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j} \\
& \cdot \left[-\tilde{y}_k(\tilde{x}_k^{(j)} - \beta\tilde{y}_k) - \tilde{y}_k(\tilde{x}_k^{(i)} - \beta\tilde{y}_k)\right] \\
= & \frac{1}{2}\sum_{k=1}^{n}\tilde{y}_k\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j}\left[\tilde{x}_k^{(j)} + \tilde{x}_k^{(i)} - 2\beta\tilde{y}_k\right] \\
= & \frac{1}{2}\sum_{k=1}^{n}\tilde{y}_k\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j}\tilde{x}_k^{(j)} + \frac{1}{2}\sum_{k=1}^{n}\tilde{y}_k\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j}\tilde{x}_k^{(i)} \\
& - \beta\sum_{k=1}^{n}\tilde{y}_k^2\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j}
\end{aligned}
\tag{A2}
$$

Consequently,

$$
\frac{\partial}{\partial\beta}L(\tilde{X};\beta) = 0 \Leftrightarrow \beta = \frac{\sum_{k=1}^{n}\tilde{y}_k\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j}\tilde{x}_k^{(j)}}{\sum_{k=1}^{n}\tilde{y}_k^2\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j}}
\tag{A3}
$$

and

$$
\frac{\partial^2}{\partial\beta^2}L(\tilde{X};\beta) = -\sum_{k=1}^{n}\tilde{y}_k^2\sum_{i=1}^{p}\sum_{j=1}^{p}\gamma_{i,j} = -\left(\sum_{k=1}^{n}\tilde{y}_k^2\right)I_p^T\Sigma^{-1}I_p,
\tag{A4}
$$

This last expression is strictly negative, because Σ^{-1} is definite positive. With equation (A3), this proves that $\hat{\beta}$ maximizes the likelihood.

References

Andréassian, V. (2002), Impact de l'évolution du couvert forestier sur le comportement hydrologique des bassins versants, Ph.D thesis, 781 pp., Univ. Pierre et Marie Curie, Paris, France.

Andréassian, V., E. Parent, and C. Michel (2003), A distribution-free test to detect gradual changes in watershed behavior, *Water Resour. Res.*, 39(9), 1252, doi:10.1029/2003WR002081.

Benjamini, Y., and Y. Hochberg (1995), Controlling the false discovery rate: A practical and powerful approach to multiple testing, *J. R. Stat. Soc., Ser. B.*, 57, 289–300.

Birsan, M. V., P. Molnar, P. Burlando, and M. Pfaundler (2005), Streamflow trends in Switzerland, *J. Hydrol. Amsterdam*, 314, 312–329, doi:10.1016/j.jhydrol.2005.06.008.

Black, A. R., and J. C. Burns (2002), Re-assessing the flood risk in Scotland, *Sci. Total Environ.*, 294, 169–184, doi:10.1016/S0048-9697(02)00062-1.

Champeaux, J. L., and A. Tamburini (1995), Zonage climatique de la France à partir des séries de précipitations [1971–1990] du réseau climatologique d'État, *Meteorologie*, 14, 44–54.

Cigizoglu, H. K., M. Bayazit, and B. Onoz (2005), Trends in the maximum, mean and low flows of Turkish rivers, *J. Hydrometeorol.*, 6, 280–290, doi:10.1175/JHM412.1.

Douglas, E. M., R. M. Vogel, and C. N. Kroll (2000), Trends in floods and low flows in the United States: Impact of spatial correlation, *J. Hydrol. Amsterdam*, 240, 90–105, doi:10.1016/S0022-1694(00)00336-X.

Dubuisson, B., and J. M. Moisselin (2006), Observed changes in climate extremes in France, *Houille Blanche*, 6, 42–47, doi:10.1051/lhb:2006099.

Hamed, K. H., and A. R. Rao (1998), A modified Mann-Kendall trend test for autocorrelated data, *J. Hydrol. Amsterdam*, 204, 182–196, doi:10.1016/S0022-1694(97)00125-X.

Hannaford, J., and T. Marsh (2006), An assessment of trends in UK runoff and low flows in a network of undisturbed catchments, *Int. J. Climatol.*, 26, 1237–1253, doi:10.1002/joc.1303.

Hasselmann, K. (1993), Optimal fingerprints for the detection of time-dependent climate change, *J. Clim.*, 6, 1957–1971, doi:10.1175/1520-0442(1993)006<1957:OFFTDO>2.0.CO;2.

Hasselmann, K. (1998), Conventional and Bayesian approach to climate-change detection and attribution, *Q. J. R. Meteorol. Soc.*, 124, 2541–2565, doi:10.1002/qj.49712455202.

Hisdal, H., K. Stahl, L. M. Tallaksen, and S. Demuth (2001), Have streamflow droughts in Europe become more severe or frequent?, *Int. J. Climatol.*, 21, 317–333, doi:10.1002/joc.619.

Javelle, P. (2001), Caractérisation du régime des crues: Le modèle débit-durée-fréquence convergent. Approche locale et régionale, Ph.D. thesis, 268 pp., Inst. Natl. Polytech. de Grenoble/Cemagref, Lyon, France.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*, 1287–1304.

Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, *42*, W03407, doi:10.1029/2005WR004368.

Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, *42*, W03408, doi:10.1029/2005WR004376.

Kendall, M. G. (1975), *Rank Correlation Methods*, 202 pp., Griffin, London.

Krasovskaia, I., L. Gottschalk, E. Leblois, and E. Sauquet (2003), Dynamics of river flow regimes viewed through attractors, *Nord. Hydrol.*, *34*, 461–476.

Kundzewicz, Z. W., D. Graczyk, T. Maurer, I. Pinskwar, M. Radziejewski, C. Svensson, and M. Szwed (2005), Trend detection in river flow series: 1. Annual maximum flow, *Hydrol. Sci. J.*, *50*, 797–810, doi:10.1623/hysj.2005.50.5.797.

Lang, M., B. Renard, P. Bois, A. Dupeyrat, O. Mestre, H. Niel, J. Gailhard, C. Laurent, L. Neppel, and E. Sauquet (2006), A study on national trend and variation in French floods and droughts, in *Water Resource Variability: Processes. Analyses and Impacts, IAHS Publ.*, *308*, 514–518.

Lettenmaier, D. P., E. F. Wood, and J. R. Wallis (1994), Hydro-climatological trends in the continental United-States, 1948–88, *J. Clim.*, *7*, 586–607, doi:10.1175/1520-0442(1994)007 < 0586:HCTITC > 2.0.CO;2.

Livezey, R. E., and W. Y. Chen (1983), Statistical field significance and its determination by Monte Carlo techniques, *Mon. Weather Rev.*, *111*, 46–59, doi:10.1175/1520-0493(1983)111 < 0046:SFSAID > 2.0.CO;2.

Mann, H. B. (1945), Nonparametric tests against trend, *Econometrica*, *13*, 245–259, doi:10.2307/1907187.

Mardia, K. V. (1980), Tests of univariate and multivariate normality, in *Handbook of Statistics 1: Analysis of Variance*, edited by P. R. Krishnaiah, pp. 279–320, North-Holland, Amsterdam, Netherlands.

Matalas, N. C., and W. B. Langbein (1962), Information content on the mean, *J. Geophys. Res.*, *67*, 3441–3448, doi:10.1029/JZ067i009p03441.

Mestre, O. (2000), Méthodes statistiques pour l'homogénéisation de longues séries climatiques, Ph.D. thesis, 229 pp., Univ. Paul Sabatier, Toulouse, France.

Moisselin, J. M., M. Schneider, C. Canellas, and O. Mestre (2002), Les changements climatiques en France au vingtième siècle, *Meteorologie*, *38*, 45–56.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models: 1. A discussion of principles, *J. Hydrol. Amsterdam*, *10*, 257–274.

Parey, S., F. Malek, C. Laurent, and D. Dacunha-Castelle (2007), Trends and climate evolution: Statistical approach for very high temperatures in France, *Clim. Change*, *81*, 331–352, doi:10.1007/s10584-006-9116-4.

Penman, H. L. (1948), Natural evaporation for open water, bare soil and grass, *Proc. R. Soc. London*, *193*, 120–145.

Perrin, C., C. Michel, and V. Andreassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol. Amsterdam*, *279*, 275–289, doi:10.1016/S0022-1694(03)00225-7.

Ramos, M. H. (2002), Analyse de la pluviométrie sous des systèmes nuageux convectifs. Etude de cas sur des données de la ville de Marseille et de la méthode ISIS de Météo-France, Ph.D. thesis, 165 pp., Univ. Joseph Fourier, Grenoble, France.

Renard, B. (2006), Détection et prise en compte d'éventuels impacts du changement climatique sur les extrêmes hydrologiques en France, Ph.D. thesis, 364 pp., Inst. Natl. Polytech. de Grenoble, Lyon, France.

Renard, B., and M. Lang (2007), Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology, *Adv. Water Resour.*, *30*, 897–912, doi:10.1016/j.advwatres.2006.08.001.

Renard, B., M. Lang, P. Bois, A. Dupeyrat, O. Mestre, H. Niel, J. Gailhard, C. Laurent, L. Neppel, and E. Sauquet (2006), Observed changes in hydrological extremes in France, *Houille Blanche*, *6*, 48–54, doi:10.1051/lhb:2006100.

Robson, A. J. (2002), Evidence for trends in UK flooding, *Philos. Trans. R. Soc. London, Ser. A*, *360*, 1327–1343, doi:10.1098/rsta.2002.1003.

Spagnoli, B., S. Planton, M. Deque, O. Mestre, and J. M. Moisselin (2002), Detecting climate change at a regional scale: The case of France, *Geophys. Res. Lett.*, *29*(10), 1450, doi:10.1029/2001GL014619.

Stewart, I. T., D. R. Cayan, and M. D. Dettinger (2005), Changes toward earlier streamflow timing across western North America, *J. Clim.*, *18*, 1136–1155, doi:10.1175/JCLI3321.1.

Svensson, C., Z. W. Kundzewicz, and T. Maurer (2005), Trend detection in river flow series: 2. Flood and low-flow index series, *Hydrol. Sci. J.*, *50*, 811–824, doi:10.1623/hysj.2005.50.5.811.

Svensson, C., J. Hannaford, Z. W. Kundzewicz, and T. Marsh (2006), Trends in river floods: Why is there no clear signal in observations?, in *Frontiers in Flood Research, IAHS Publ.*, *305*, 1–18.

Tallaksen, L. M., and H. A. J. Van Lanen (2004), *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*, 580 pp., Elsevier, New York.

Ventura, V., C. J. Paciorek, and J. S. Risbey (2004), Controlling the proportion of falsely rejected hypotheses when conducting multiple tests with climatological data, *J. Clim.*, *17*, 4343–4356, doi:10.1175/3199.1.

Vincent, C. (2002), Influence of climate change on French glaciers mass balance over the 20th century, *Houille Blanche*, *8*, 20–24.

Vincent, C., G. Kappenberger, F. Valla, A. Bauder, M. Funk, and E. Le Meur (2004), Ice ablation as evidence of climate change in the Alps over the 20th century, *J. Geophys. Res.*, *109*, D10104, doi:10.1029/2003JD003857.

Wilby, R. L. (2006), When and where might climate change be detectable in UK river flows?, *Geophys. Res. Lett.*, *33*, L19407, doi:10.1029/2006GL027552.

Wilks, D. S. (2006), On "field significance" and the false discovery rate, *J. Appl. Meteorol. Climatol.*, *45*, 1181–1189, doi:10.1175/JAM2404.1.

Yue, S., and C. Y. Wang (2002), Regional streamflow trend detection with consideration of both temporal and spatial correlation, *Int. J. Climatol.*, *22*, 933–946, doi:10.1002/joc.781.

————————————

P. Bois, LTHE, BP 53, F-38041 Grenoble CEDEX 09, France.

A. Dupeyrat and S. Parey, EDF /R&D, 6 quai Watier, BP49, F-78401 Chatou CEDEX, France.

J. Gailhard and E. Paquet, EDF/DTG, 21 av. de l'Europe, F-38040 Grenoble CEDEX 9, France.

M. Lang and E. Sauquet, Cemagref, U.R Hydrologie-Hydraulique, 3 bis quai Chauveau, F-69336 Lyon Cedex 03, France.

O. Mestre, Météo-France, Ecole Nationale de la Météorologie, 42, Av. G. Coriolis, F-31057 Toulouse CEDEX, France.

L. Neppel and H. Niel, Maison des Sciences de l'eau Laboratoire Hydrosciences, 300 avenue Emile Jeanbrau, F-34095 Montpellier, France.

C. Prudhomme, Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK.

B. Renard, School of Engineering, University of Newcastle, Callaghan, NSW 2308, Australia. (benjamin.renard@newcastle.edu.au)