

# The importance of prewhitening in change point analysis under persistence

Francesco Serinaldi · Chris G. Kilsby

Published online: 19 February 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

**Abstract** The presence of serial correlation in hydro-meteorological time series often makes the detection of deterministic gradual or abrupt changes with tests such as Mann–Kendall (MK) and Pettitt problematic. In this study we investigate the adverse impact of serial correlation on change point analyses performed by the Pettitt test. Building on methods developed for the MK test, different prewhitening procedures devised to remove the serial correlation are examined, and the effects of the sample size and strength of serial dependence on their performance are tested by Monte Carlo experiments involving the first-order autoregressive [AR(1)] process, fractional Gaussian noise (fGn), and fractionally integrated autoregressive [ARFI-MA(1, $d$ ,0)] model. Results show that (1) the serial correlation affects the Pettitt test more than tests for slowly varying monotonic trends such as the MK test both for short-range and long-range persistence; (2) the most efficient prewhitening procedure based on AR(1) involves the simultaneous estimation of step change and lag-1 autocorrelation  $\rho$ , and bias correction of  $\rho$  estimates; (3) as expected, the effectiveness of the prewhitening procedure strongly depends upon the model selected to remove the serial correlation; (4) prewhitening procedures allow for a better control of the type I error resulting in rejection rates reasonably close to the nominal values. As ancillary results, (5) we show the ineffectiveness of the original formulation of the so-called trend-free prewhitening (TFPW) method

and provide analytical results supporting a corrected version called TFPWcu; and (6) we propose an improved two-stage bias correction of  $\rho$  estimates for AR(1) signals.

**Keywords** Pettitt test · Change point analysis · Prewhitening · Autoregressive process · Fractional Gaussian noise · Hurst parameter

## 1 Introduction

Climate fluctuations and human activities can cause statistical shifts in long-term means of hydro-meteorological variables. Recognition and attribution of these changes is fundamental for infrastructure design, water management strategies, and risk mitigation policies. In this respect, appropriate statistical diagnostics and change detection methods can help understand the nature of historic fluctuations in hydrological time series [e.g., Rougé et al. (2013); Guerreiro et al. (2014) and references therein]. Among many available statistical testing procedures devised for assessing the significance of a change [e.g., Kundzewicz and Robson (2004)], the Pettitt test (Pettitt 1979) is one of the widely used rank-based nonparametric tests to check the presence and timing of abrupt changes in the mean or median of hydro-meteorological variables such as rainfall, runoff, and temperature [e.g., Villarini et al. (2009, 2011); Ferguson and Villarini (2012); Rougé et al. (2013); Trambly et al. (2013); Guerreiro et al. (2014); Sagarika et al. (2014) among others].

According to Pettitt (1979), given a set of independent random variables  $\{X_1, X_2, \dots, X_T\}$ , the sequence is said to have a change point at  $\tau$  if  $X_t$  for  $t = 1, \dots, \tau$  have a common distribution  $F_1(x)$  and  $X_t$  for  $t = \tau + 1, \dots, T$  have a common distribution  $F_2(x)$ , and  $F_1(x) \neq F_2(x)$ . Thus, the test tackles the problem of testing the null hypothesis of

---

F. Serinaldi (✉) · C. G. Kilsby  
School of Civil Engineering and Geosciences, Newcastle  
University, Newcastle upon Tyne NE1 7RU, UK  
e-mail: francesco.serinaldi@newcastle.ac.uk

F. Serinaldi · C. G. Kilsby  
Willis Research Network, 51 Lime St., London EC3M 7DQ, UK

“no change”,  $H_0 : \tau = T$ , against the alternative of “change”,  $H_1 : 1 \leq \tau < T$ . The test is based on the statistic

$$K_T = \max_{1 \leq t < T} |U_{t,T}|, \quad (1)$$

where

$$U_{t,T} = \sum_{i=1}^t \sum_{j=i+1}^T \text{sgn}(X_i - X_j), \quad (2)$$

where  $\text{sgn}(x) = 1$  if  $x > 0$ ,  $0$  if  $x = 0$ , and  $-1$  if  $x < 0$ . The statistic  $U_{t,T}$  is equivalent to a Mann–Whitney statistic for testing that two samples  $(x_1, \dots, x_t)$  and  $(x_{t+1}, \dots, x_T)$  come from the same population. This correspondence highlights that the actual alternative of both tests (Mann–Whitney U test and Pettitt test) is that one distribution stochastically dominates the other, meaning that  $F_1(x) < F_2(x)$  for every value of  $x$  or vice versa. Thus, even though this hypothesis is commonly restricted to a shift in the location parameter  $\mu$ ,  $F_1(x) = F_2(x + \mu)$ , these tests are sensitive to all possible conditions resulting in a stochastic ordering. It should be noted that the equivalence mentioned above implies a formal relationship between the Pettitt test and the MK test (Rougé et al. 2013), which is one of the widely used nonparametric approaches for testing slowly varying monotonic trends in hydro-meteorological time series.

Different aspects of such tests (Pettitt and MK) have been widely studied in the literature. However, the MK test has always received much more attention than the Pettitt test despite their common theoretical background and the potential interest of regime shift detection in hydrological and climate studies compared with monotonic trends. For example, the power of the MK test under different conditions (i.e., sample size, magnitude of deterministic trend, type of the parent distribution) was studied by extensive Monte Carlo simulations about one decade ago (Yue et al. 2002a; Önöz and Bayazit 2003; Yue and Pilon 2004), whereas, to the best of our knowledge, an analogous study was performed only recently for the Pettitt test (Xie et al. 2014; Mallakpour and Villarini 2015).

The same holds for the effect of serial correlation (also referred to as autocorrelation or serial dependence) on the outcome of Pettitt and MK tests. It is well known that a basic assumption for a correct application of tests such as Pettitt and MK is that the data should be randomly ordered (i.e. observations should be serially independent), which is a condition seldom fulfilled by real-world hydro-meteorological data (e.g., Hamed 2009). The effect of the autocorrelation on tests

devised for independent data is a general increase of the rejection rate of the null hypothesis (“no change”) of the statistical test, even if no change is present in the data. This over-rejection (compared with the nominal rejection rate) is due to the information redundancy which makes the effective sample size smaller than the observed size, thus implying that the effective variance of the test statistics to be used in the testing procedure under serial dependence is larger than that provided by standard results obtained under the hypothesis of independence (e.g., Bayley and Hammersley 1946; Koutsoyiannis and Montanari 2007). This phenomenon is known as variance inflation. In this respect, there is an extensive literature on the study of the effect of serial correlation on the MK test (see Sect. 2), whereas, to the best of our knowledge, only Busuioac and von Storch (1996) and Rybski and Neumann (2011) (see Sect. 3) tackled the problem for the Pettitt test.

In this study we provide a comprehensive investigation of the effects of serial dependence on the Pettitt test, and propose a set of so-called prewhitening methods (see Sect. 3) in order to make the test procedure suitable for serially correlated data. Such methods involve different autocorrelation structures, and take into account the mutual influence of serial correlation and structural abrupt changes. The capability of controlling the type I error and the sensitivity to model misspecification are tested by extensive Monte Carlo simulations. Since the proposed prewhitening procedures are derived from techniques developed for the MK test, an overview of these methods is given in Sect. 2. Prewhitening approaches for Pettitt are therefore presented in Sect. 3, whilst simulation results are discussed in Sect. 4. Finally, conclusions are drawn in Sect. 5.

## 2 Some aspects of MK analysis of gradual changes under serial correlation

In order to deal with the problem of variance inflation, two approaches have been suggested: the explicit calculation of the inflated variance (e.g., Hamed and Rao 1998; Koutsoyiannis 2003; Yue and Wang 2004c; Hamed 2008b, 2009) and prewhitening procedures (e.g., Katz 1988; Kulkarni and von Storch 1995; von Storch 1999; Yue et al. 2002b; Yue and Wang 2002; Bayazit and Önöz 2007; Hamed 2009). In more detail, Hamed and Rao (1998) showed that the mean and variance of MK  $S$  statistics are (for meta-Gaussian serial dependence structure)

$$\begin{cases} E[S] = 0 \\ \text{Var}[S] = \sum_{i=1}^{T-1} \sum_{j=i+1}^T \sum_{k=1}^{T-1} \sum_{l=k+1}^T \frac{2}{\pi} \arcsin \left( \frac{\rho_{l-j} - \rho_{l-i} - \rho_{k-j} + \rho_{k-i}}{\sqrt{(2 - 2\rho_{j-i})(2 - 2\rho_{l-k})}} \right), \end{cases} \quad (3)$$

where the symbol  $\rho_{j-i}$  denote the value of the empirical autocorrelation function at lag  $(l-j)$  (Hamed and Rao 1998) or the theoretical autocorrelation function corresponding to a selected model which is deemed to correctly represent the serial correlation structure of the process. Referring to Hamed (2009) for a list of candidates and a comparison, possible options are models such as  $AR(p)$ , autoregressive moving average  $ARMA(p, q)$ ,  $fGn(H)$ , or fractionally integrated  $ARMA$  [ARFIMA( $p, d, q$ )], where  $p, q, d$ , and  $H$  denote the AR order, the MA order, the fractional order of differencing, and the Hurst parameter, respectively. As an alternative to using the inflated variance in Eq. 3 or analogous variance inflation factors (Matalas and Sankarasubramanian 2003), one can apply prewhitening procedures, which consist of the removal of the autocorrelation structure by fitting one of the models mentioned above and thus performing the statistical test on the (approximately) uncorrelated residuals (e.g., Katz 1988; Kulkarni and von Storch 1995; von Storch 1999).

Both procedures (inflated variance correction and prewhitening) require the estimation of the autocorrelation terms at different lags (for nonparametric approaches or ARMA models),  $d$  (for ARFIMA models), or  $H$  (for  $fGn$ ). However, the presence of deterministic (gradual or abrupt) changes tends to strengthen the autocorrelation among data, resulting in biased estimates of the models' parameters, and eventually in overestimating the terms of the autocorrelation function. Using such inflated correlation values in computing the variance in Eq. 3 results in an over-inflation of the variance of the test statistic  $S$ , thus making the test too liberal (i.e., the rejection rate of the null hypothesis is smaller than expected). Analogously, the effect of inflated correlation on prewhitening is a removal of a portion of the trend (Yue and Wang 2002), thus increasing the chances of not rejecting the null hypothesis when the original MK test is applied to model residuals. The interaction between deterministic trends and autocorrelation structure prompted a rather heated debate about the suitability of the prewhitening procedure and its effect on the test significance level and power (e.g., Bayazit and Önöz 2004; Yue and Wang 2004a, b; Zhang and Zwiers 2004; Hamed 2008a; Bayazit and Önöz 2008).

In this respect, focusing on prewhitening by AR(1) correlation structure, the preliminary removal of the apparent deterministic trend (e.g., Hamed and Rao 1998; Yue et al. 2002b; Yue and Wang 2004c) was shown to reduce the inflation of the lag-1 autocorrelation  $\rho$  used in prewhitening, thus avoiding the problem of overcorrection (also known as over-whitening). However, Hamed (2009) highlighted that the removal of the apparent trend leads to an underestimation of  $\rho$ , resulting in an insufficient removal of the autocorrelation, and thus in the persistence of

the original problem of over-rejection. He concluded that no prewhitening, prewhitening without trend removal, or prewhitening with trend removal all exhibit a poor performance owing to the presence of the autocorrelation, the overestimation and underestimation of  $\rho$ , respectively. To overcome such problems, Hamed (2009) suggested a procedure allowing for the simultaneous estimation of  $\rho$  and the slope  $\beta$  of a possible deterministic linear trend. This approach was shown to balance between under- and over-correction improving the effectiveness of prewhitening and also correcting the bias in the  $\rho$  estimates.

Since the Hamed's method will be adapted for the Pettitt test, it is worth recalling basic equations and highlighting its relationship with the prewhitening procedures proposed by Zhang et al. (2000) and Yue et al. (2002b). As the AR(1) model and linear trends are the most used options in studies concerning trend analyses, Hamed (2009) assumed the following model:

$$y_t = \rho y_{t-1} + \alpha + \beta t + \varepsilon_t, \tag{4}$$

where  $y_t$  and  $y_{t-1}$  are observed records at time  $t$  and  $t-1$ ,  $\rho$  is the lag-1 autocorrelation coefficient,  $\alpha$  is the intercept of the linear trend,  $\beta$  is the trend slope, and  $\varepsilon_t$  indicates uncorrelated residuals. The corresponding prewhitened time series are written as

$$y_t - \rho y_{t-1} = \alpha + \beta t + \varepsilon_t. \tag{5}$$

Zhang et al. (2000) and Yue et al. (2002b) suggested considering a process as the superposition of an AR(1) process  $X_t$  and a linear trend with slope  $\beta'$

$$\begin{cases} y_t = \rho' x_t + \alpha' + \beta' t \\ x_t = \rho' x_{t-1} + \varepsilon'_t \end{cases}, \tag{6}$$

which yields prewhitened time series (Cochrane and Orcutt 1949; Wang and Swail 2001)

$$y_t - \rho' y_{t-1} = (1 - \rho')\alpha' + \rho' \beta' + (1 - \rho')\beta' t + \varepsilon'_t. \tag{7}$$

From Eqs. 5 and 7, it follows

$$\begin{cases} \rho = \rho' \\ \alpha = (1 - \rho')\alpha' + \rho' \beta' \\ \beta = (1 - \rho')\beta' \\ \varepsilon_t = \varepsilon'_t \end{cases} \iff \begin{cases} \rho' = \rho \\ \alpha' = \frac{(1 - \rho)\alpha - \rho\beta}{(1 - \rho)^2} \\ \beta' = \frac{\beta}{1 - \rho} \\ \varepsilon'_t = \varepsilon_t \end{cases}. \tag{8}$$

Equation 8 helps highlight some aspects that should be accounted for in prewhitening procedures. Under the assumption that the data come from the superposition of an AR(1) signal and a linear trend  $\beta't$ , Hamed's method tests the *equivalent* trend (Hamed 2009, p. 148) with *effective*

slope  $(1 - \rho')\beta'$  corresponding to prewhitened observations  $y_t - \rho y_{t-1}$ . In order to obtain a prewhitened time series with the same trend slope  $\beta'$  of the observed sequences, Wang and Swail (2001) suggested dividing the prewhitened values by  $(1 - \rho')$ , obtaining

$$\begin{aligned} \frac{y_t - \rho' y_{t-1}}{1 - \rho'} &= \alpha' + \frac{\rho' \beta'}{1 - \rho'} + \beta' t + \frac{\varepsilon'_t}{1 - \rho'}, \\ &= \alpha'' + \beta' t + \varepsilon''_t \end{aligned} \quad (9)$$

Equation 9 shows that re-inflating the slope of the prewhitened values from  $(1 - \rho')\beta'$  to  $\beta'$  implies also the inflation of the variance of the white noise residuals from  $\varepsilon'_t$  to  $\varepsilon'_t/(1 - \rho')$ . In other words, prewhitening involves the reduction of the slope to be tested (the variance of the residuals being unchanged) or the increase of the variance of the residuals (being the slope unchanged). The latter approach is coherent with the variance inflation procedures applied to the original signal (Hamed and Rao 1998; Yue and Wang 2004c; Hamed 2008b). In this respect, it is worth highlighting that the TFPW method introduced by Yue et al. (2002b) does not consider the inflation of the variance of  $\varepsilon'_t$ . The steps involved in implementing the TFPW approach are summarized as (Yue et al. 2002b; Khaliq et al. 2009): (1) for a given time series of interest  $\{y_t\}$ , linear trend slope is estimated using the rank-based Sen's method (Sen 1968); (2) the linear trend is removed from the time series and the lag-1 autocorrelation coefficient  $\rho'$  is estimated; (3) if  $\rho'$  is non-significant at the chosen significance level then the trend identification test is applied to the original time series; and otherwise (4) the trend identification test is applied to the detrended prewhitened series recombined with the estimated slope of trend from step 1.

As TFPW implies trend removal, residuals prewhitening, and trend reintroduction, it follows that the MK test is applied to the variable

$$\begin{aligned} \varepsilon'_t + \beta' t &= x_t - \rho' x_{t-1} + \beta' t \\ &= y_t - \beta' t - \rho' (y_{t-1} - \beta' (t-1)) + \beta' t \\ &= y_t - \rho' y_{t-1} + \rho' \beta' (t-1) \\ &= y_t - \rho' x_{t-1} \end{aligned} \quad (10)$$

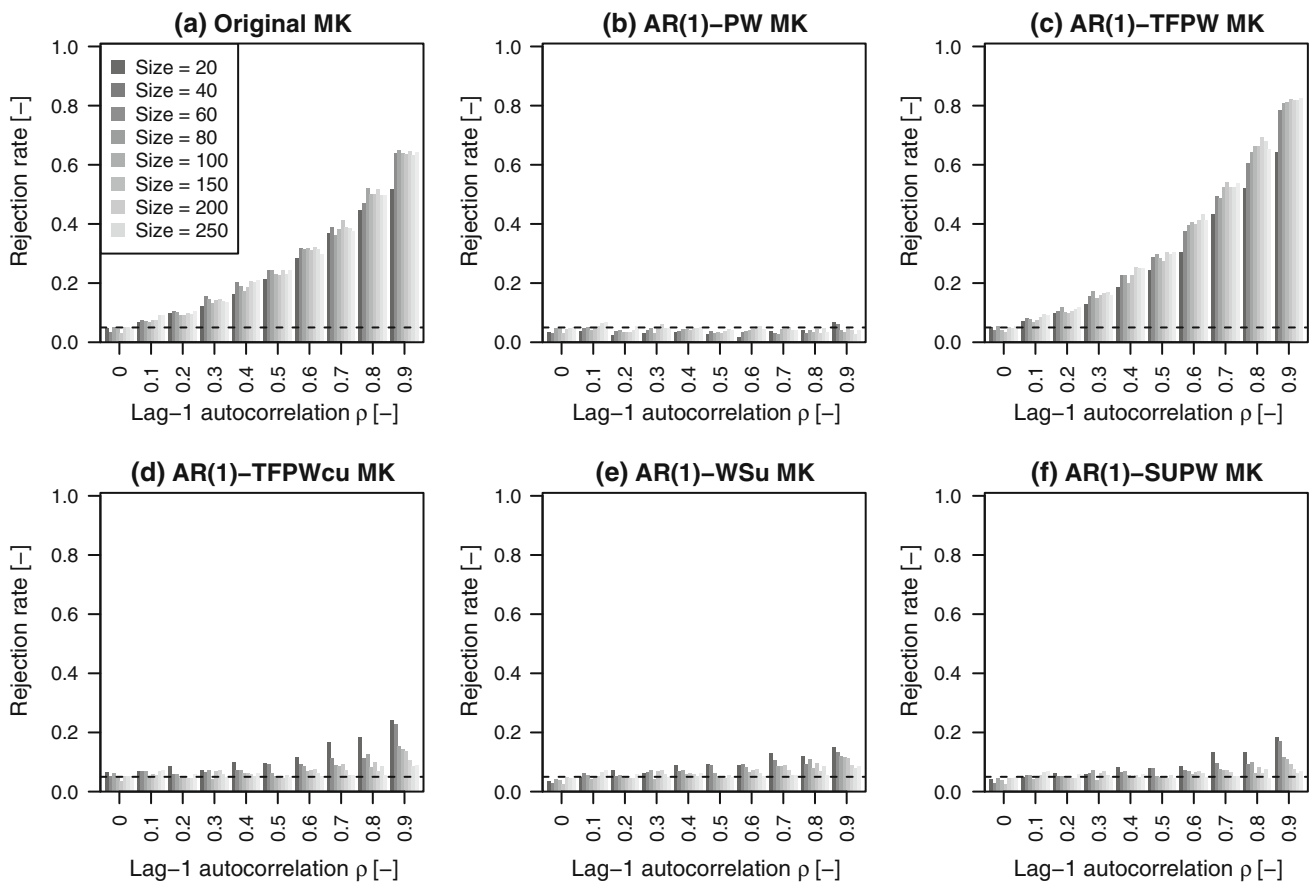
where we omitted the intercept  $\alpha'$  for the sake of simplicity and without loss of generality. Equation 10 clearly shows that the time series tested by MK in the TFPW procedure is not prewhitened at all. Indeed the rationale of TFPW is to make the residuals  $x_t$  around the trend serially independent, whereas MK and Pettitt tests require that the series of data  $y_t$  have to be serially independent or made independent by  $y_t - \rho y_{t-1}$  (under the hypothesis of AR(1) dependence structure). To make TFPW consistent with Wang-Swail's and Hamed's methods,  $\varepsilon'_t$  in Eq. 10 should be replaced with the inflated value  $\varepsilon'_t/(1 - \rho')$ , thus making the tested time

series similar to that in Eq. 9 (the main difference being the efficiency of the procedure used to estimate the model parameters). As this option is actually implemented in R (R Development Core Team 2014) in the package `zyp` (Bronaugh and Werner 2013) based on empirical analyses, our discussion provides the theoretical proof that such an option is actually required to control the type I error.

Monte Carlo simulations confirm the above statements. We simulated 1000 time series from an AR(1) model with  $\rho$  ranging between 0 and 0.9 by 0.1 steps with no trend to check the actual rejection rate of the MK test (conducted at the 5% significance level) using different methods to account for serial correlation. Figure 1a, b show the actual rejection rate obtained applying MK to AR(1) time series and sequences prewhitened without accounting for possible trends, i.e. taking the differences  $y_t - \hat{\rho}^* y_{t-1}$ , where  $\hat{\rho}^*$  is the estimate of  $\rho$  corrected for the bias of the ordinary least square estimator according to the two-stage procedure described in the Appendix. Such results are well-known, and the effectiveness of prewhitening in reproducing the nominal rejection rate (5%) under correct model specification is expected (see e.g., Kulkarni and von Storch 1995, among others). However, Fig. 1a, b can be used to assess the performance of the other prewhitening methods. Indeed, Fig. 1c shows the complete ineffectiveness of TFPW, thus quantifying the consequences of using Eq. 10. Figure 1d, e highlight that the inflation of the variance of the trend residuals  $x_t$  allows the correction of the over-rejection problem (the method is denoted as TFPWcu, where “c” indicates “corrected” and “u” denotes the “unbiased” estimation of  $\rho$ ). This makes the performance of TFPW similar to that of Wang-Swail's method (referred to as WSu in Fig. 1e), which is based on an iterative estimation procedure of the model parameters (see Wang and Swail 2001, for further details). Finally, Hamed's method (referred to as simultaneous unbiased prewhitening (SUPW) in Fig. 1f) performs slightly better than TFPWcu and similarly to WSu, as the estimation method of the model parameters is specifically devised for an AR(1) with linear trend, and provide an efficient treatment and removal of the bias affecting the parameter estimates. Thus, in spite of the presence of the linear trend in the model structure, TFPWcu, SWu, and SUPW yield a rejection rate similar to that of the pure prewhitening shown in Fig. 1b (except for high values of  $\rho$ ). These results are used in the next section to set up prewhitening procedures for the Pettitt test.

### 3 Prewhitening methods for the Pettitt test

As mentioned above, unlike the MK test, the Pettitt test has received less attention in the literature. Dealing with the impact of serial correlation, Busuioac and von Storch (1996)



**Fig. 1** Rejection rate of MK test applied to samples drawn from AR(1) for different values of lag-1 autocorrelation  $\rho$ , several sample sizes, and 5% nominal significance level. Several variants of MK test are considered: **a** original MK test without prewhitening; **b** MK with AR(1) prewhitening; **c** MK with original TFPW; **d** MK with modified

TFPW involving corrected linear trend slope and unbiased  $\rho$  (TFPWcu); **e** MK with Wang-Swail’s prewhitening and unbiased  $\rho$  (WSu); **f** MK with Hamed’s simultaneous unbiased prewhitening (SUPW)

showed the adverse effect of the autocorrelation (namely, AR(1) correlation structure) and the presence of possible gradual (linear) trends on the rejection rate. Busuioc and von Storch (1996) recommend prewhitening before performing the test, and highlight the detrimental effects of the presence of linear trends. Indeed, the preliminary removal of a linear trend corrects for the over-rejection of the Pettitt test if only a linear trend is present. However, when both linear trend and one or more abrupt changes are present, spurious trends can result from the presence of abrupt changes, and trend removal reduces the power of the test making it sometimes useless. Thus they “recommend using the Pettitt test as a mere exploratory tool and calculating Pettitt’s statistic and dealing with change points as unproven hypotheses, which plausibility should be supported by physical arguments”. Similarly, Rybski and Neumann (2011) discussed the over-rejection introduced by a long-range power-law decaying correlation structure, thus confirming the results of Busuioc and von Storch (1996) and suggesting the modification of the expression of the distribution of  $K_T$  under the null hypothesis

accounting for short-range and long-range correlation. However, they do not discuss such procedures. Dealing with a sequential regime shift detection method (Rodionov 2004), which is different to the Pettitt test but is similarly affected by serial correlation, Rodionov (2006) investigated the effect of prewhitening, highlighting the importance of performing a bias correction of the ordinary least squares (OLS) or maximum likelihood estimates of  $\rho$ .

Based on these remarks and the results reported in the previous section concerning the MK test, in this study, we investigate the effect of the autocorrelation on the rejection rate of the Pettitt test and the effectiveness of prewhitening, bearing in mind the concealing effects of the interaction between serial correlation and “true” abrupt changes, and the bias affecting the parameters’ estimates.

### 3.1 TFPWcu adapted for the Pettitt test

Based on results in Sect. 1, under the hypothesis of AR(1) serial dependence, we do not consider the WSu method as



its rationale is similar to TFPWcu but involves an iterative estimation procedure that does not provide significant improvements and can be avoided. TFPWcu was adapted for the Pettitt test replacing the linear trend by a step change. Thus, model in Eq. 6 becomes

$$\begin{cases} y_t = \rho'x_t + \Delta' \cdot \mathbf{1}_{\{t > \tau\}}, \\ x_t = \rho'x_{t-1} + \varepsilon'_t \end{cases}, \tag{11}$$

where  $\mathbf{1}_{\{\bullet\}}$  is the indicator function. The testing procedure is as follows:

- Step 1: The Pettitt test is applied to the original data. If the value of the test statistic  $K_T$  is not significant, it can be concluded that there is no evidence to reject the null hypothesis (“no change”).
- Step 2: If  $K_T$  is significant, the position  $\tau$  of the possible change point is used to split the time series in two sub-series (before and after  $\tau$ ), the difference of the medians or means,  $\hat{\mu}_b$  and  $\hat{\mu}_a$ , of the two sub-series is computed as  $\hat{\Delta}' = \hat{\mu}_b - \hat{\mu}_a$  and used to remove the step change as follows:

$$x_t = y_t - \hat{\Delta}' \cdot \mathbf{1}_{\{t > \tau\}}. \tag{12}$$

- Step 3: The value of the lag-1 autocorrelation  $\rho$  of  $x_t$  is estimated by the OLS estimator and corrected for bias using the two-stage bias correction described in the Appendix; then the AR(1) structure is removed by

$$\varepsilon'_t = x_t - \hat{\rho}^*x_{t-1}, \tag{13}$$

where  $\hat{\rho}^*$  is the bias corrected estimate of  $\rho$  and  $\varepsilon'_t$  should be an uncorrelated series.

- Step 4: The step change and the residuals  $\varepsilon'_t$  are combined by

$$\hat{\Delta}' \cdot \mathbf{1}_{\{t > \tau\}} + \frac{\varepsilon'_t}{1 - \hat{\rho}^*}, \tag{14}$$

and the Pettitt test is applied to these prewhitened series to assess the significance of the abrupt change.

As mentioned in the previous section, dividing the step change residuals  $\varepsilon'_t$  by  $(1 - \hat{\rho}^*)$  allows the appropriate prewhitening of the series to be tested preserving the original step change  $\Delta'$ .

### 3.2 Hamed’s methods adapted for the Pettitt test

#### 3.2.1 AR(1) prewhitening

As mentioned in Sect. 1, it is well known that the OLS estimator of the correlation coefficient is negatively biased

(see e.g., Wallis and O’Connell 1972; Lenton and Schaake 1973; Mudelsee 2001; Koutsoyiannis 2003, and references therein). In the case of linear trend and AR(1) correlation structure, Hamed (2009) proposed the simultaneous estimation of the model parameters in Eq. 4 by the OLS method as follows:

$$[\hat{\rho} \ \hat{\alpha} \ \hat{\beta}]^\top = (\mathbf{z}^\top \mathbf{z})^{-1} \mathbf{z}^\top \mathbf{y}, \tag{15}$$

where  $\mathbf{z}$  is a  $(T - 1) \times 3$  design matrix containing observations from  $y_1$  to  $y_{T-1}$  in the first column, a vector of  $(T - 1)$  ones in the second column, and a sequence of integers from 2 to  $T$  in the third column;  $\mathbf{y}$  is the vector of observation from  $y_2$  to  $y_T$ . The simultaneous estimation allows for the correction of the bias in  $\rho$  related to the estimation of nuisance parameters, i.e. the coefficients of the linear (or polynomial) mean function. In particular, for both OLS and maximum likelihood estimators, and a linear trend, Kang et al. (2003) and van Giersbergen (2005) showed that  $E[\hat{\rho} - \rho] = -(2 + 4\rho)/T$ , yielding the bias-corrected value

$$\hat{\rho}^* = -\frac{T\hat{\rho} + 2}{T - 4}. \tag{16}$$

Using the simultaneous estimation for the Pettitt test and an abrupt change instead of a linear trend is possible because the framework refers to models that are linear in the coefficients, and the bias correction in Eq. 16 is independent of the values of the explanatory variables. Indeed, the sequence  $2, \dots, T$  used by Hamed (2009) can be replaced by a sequence of dates or a standardized series  $2/T, \dots, 1$  (van Giersbergen 2005). Thus, our proposal is to replace the sequence  $2, \dots, T$  with an auxiliary variable described by the indicator function  $\mathbf{1}_{\{t > \tau\}}$ , which is zero for  $t \leq \tau$  and 1 for  $t > \tau$ , obtaining the model

$$y_t = \rho y_{t-1} + \alpha + \Delta \cdot \mathbf{1}_{\{t > \tau\}} + \varepsilon_t. \tag{17}$$

This way, the  $\beta$  parameter in Eqs. 4 and 15 represents the magnitude  $\Delta$  of a step change instead of the slope of a linear trend. Similarly to the case of  $\beta$  and  $\beta'$  in Sect. 2,  $\Delta = (1 - \rho)\Delta'$  is the effective magnitude of the step change. Thus, the testing procedure consists of applying the original Pettitt test to the prewhitened signal

$$y_t - \hat{\rho}^*y_{t-1} = \hat{\alpha} + \hat{\Delta} \cdot \mathbf{1}_{\{t > \tau\}} + \varepsilon_t. \tag{18}$$

#### 3.2.2 Prewhitening with models different from AR(1)

In spite of the widespread use of AR(1) as a prewhitening model, it is well known that the success of prewhitening depends on the correctness of the model selected to describe the autocorrelation structure (Kulkarni and von Storch 1995). Other models should therefore be considered

if the AR(1) does not provide a satisfactory prewhitening. In this respect, Hamed (2009) showed the effect of model misspecification on the variance inflation factor. For such alternative (and generally more complex) models, the simultaneous estimation of the model parameters and gradual or abrupt changes might be no feasible or impractical. Thus, in these cases, we apply a more classical approach which can be summarized by a procedure similar to that suggested by Hamed (2008b) for fGn and linear trends, and adapted for abrupt changes as follows

- Step 1: The Pettitt test is applied to the original data. If the value of the test statistic  $K_T$  is not significant, it can be concluded that there is no evidence to reject the null hypothesis (“no change”).
- Step 2: If  $K_T$  is significant, the abrupt change is removed as for Step 2 of the TFPWcu approach (Sect. 3.1), and the parameters of the selected model are calculated on this detrended time series.
- Step 3: The original data are prewhitened by the model calibrated in the previous step and the Pettitt test is applied. If the value of the test statistic  $K_T$  is not significant, it can be concluded that there is no evidence to reject the null hypothesis (“no change”), otherwise the null hypothesis can be rejected at a given significance level.

The selection of the model used in Step 2 should be based on a preliminary exploratory analysis in order to identify a set of suitable candidates. For fGn, which is parameterized by the Hurst parameter  $H$ , Hamed (2008b) suggested to tests the significance of  $H$  estimated in Step 2 and proceed to the subsequent step only if  $H$  is significantly different from 0.5 (corresponding to white noise). Such a procedure introduces a conditional prewhitening (CPW), whereas prewhitening regardless of the statistical significance of the model parameters is called unconditional (UPW). For MK and linear trends, Kulkarni and von Storch (1995) found that UPW outperforms CPW, and suggested the use of the former method, which is also the approach adopted by Hamed (2009). In this study, we compare both approaches, which are denoted as *model*-UPW and *model*-CPW, where *model* refers to the model used to prewhiten (e.g., AR(1)).

#### 4 Monte Carlo results

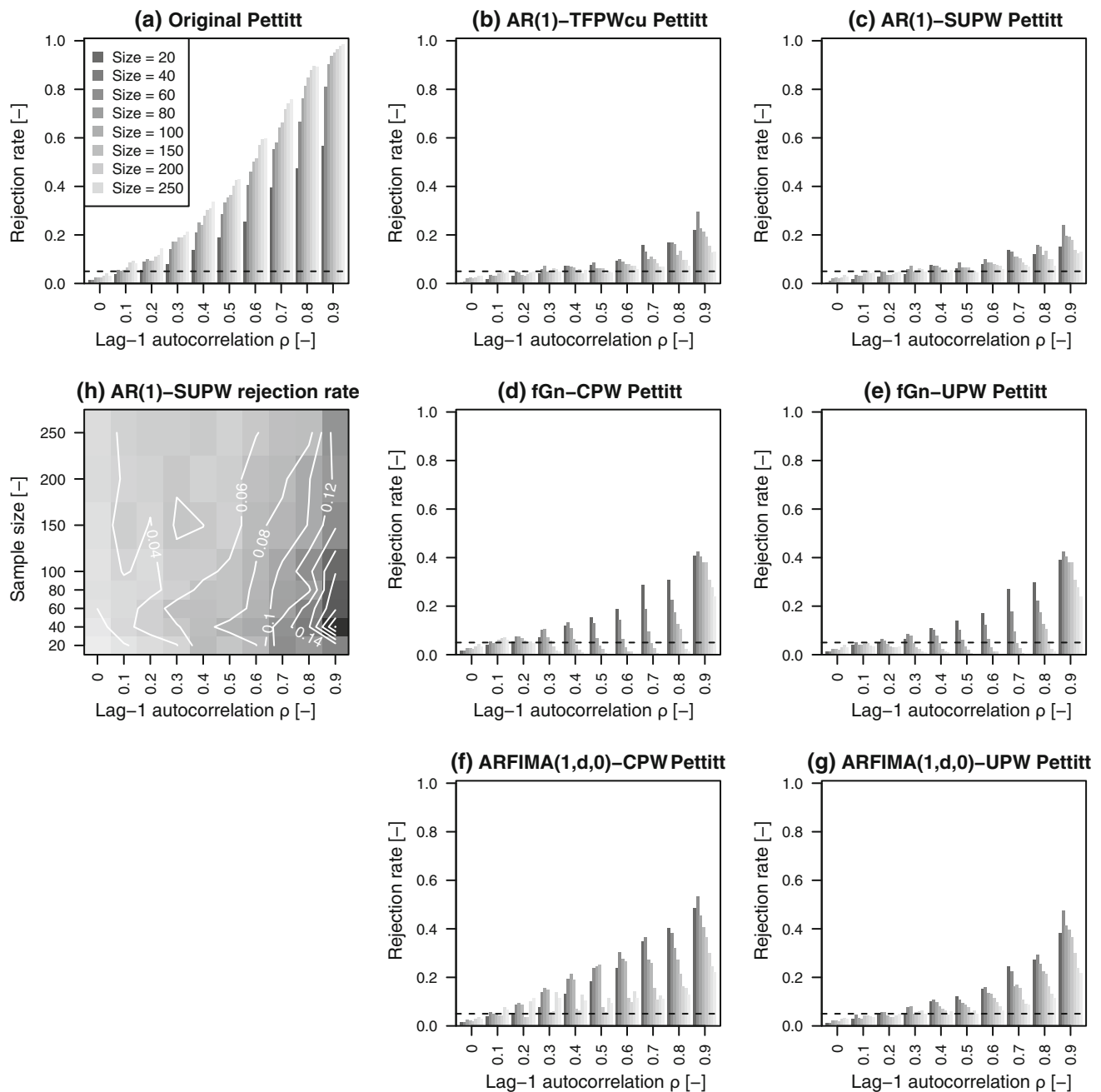
To test the effectiveness of the procedures described in Sect. 3, we used a set of models accounting for both short-range and long-range serial correlation, namely, AR(1), fGn, and ARFIMA(1, $d$ ,0). The analyses are based on

Monte Carlo simulations of samples from AR(1) with  $\rho$  ranging from 0 to 0.9 by 0.1, fGn with Hurst parameter ranging from 0.5 to 0.95 by 0.05, and ARFIMA(1, $d$ ,0) with ten combinations of the parameters  $\rho$  and  $d$  (detailed below), and sample size  $T \in \{20, 40, 60, 80, 100, 150, 200, 250\}$ . For each configuration, 1000 time series were simulated.

Figure 2 shows results corresponding to AR(1) signals. The rejection rate of the original Pettitt test (without prewhitening) quickly increases as  $\rho$  increases, and is larger than that of MK test shown in Fig. 1, thus indicating the greater sensitivity of Pettitt to the influence of the serial correlation. TFPWcu and SUPW provide a rejection rate much closer to the nominal value (5%), with SUPW slightly outperforming TFPWcu. However, both methods are less effective for Pettitt than for MK, further confirming the sensitivity to the effects of serial correlation, especially for  $\rho$  values higher than 0.7.

Figure 2 also shows the effect of model misspecification. In particular, fGn-based methods do not provide a sufficient prewhitening (which is known as under-whitening) for small sample sizes owing to the difficulty of reliably estimating the Hurst parameter in these cases (e.g., Tyralis and Koutsoyiannis 2011). On the other hand, fGn-CPW and fGn-UPW yield over-whitening, and so under-rejection, as the sample size increases and the removed fGn dependence structure is stronger than the actual AR(1). ARFIMA(1, $d$ ,0)-CPW and ARFIMA(1, $d$ ,0)-UPW provide results similar to fGn-UPW and fGn-CPW for small sample sizes, whereas their short-range correlation component prevents over-whitening for larger sample sizes. Finally, there is no significant difference between conditional and unconditional prewhitening. A map of the rejection rate as a function of  $\rho$  and sample size  $T$  is also provided for the “best” performing method to highlight the dependence of the rejection rates on the pairs  $(\rho, T)$ .

Figure 3 shows results concerning the application of the Pettitt test to fGn time series. As expected, AR(1)-based methods (i.e. TFPWcu and SUPW) yield over-rejection owing to the under-whitening of long-range correlated signals. fGn-CPW and fGn-UPW perform better than the other methods; however, both fGn-CPW and fGn-UPW under-whiten the signals even though the model is correctly specified. We argue that this result might be ascribed to two factors: (1) the difficulty of reliably estimating  $H$  for such small sample sizes, and (2) the intrinsic nature of fGn time series, which are characterized by persistent fluctuations that can easily (but erroneously) be recognized as structural change points. In this context, ARFIMA(1, $d$ ,0)-CPW and ARFIMA(1, $d$ ,0)-UPW perform slightly better than TFPWcu and SUPW, but the under-whitening related to the short-range component seems to dominate the



**Fig. 2** Rejection rate of the Pettitt test applied to samples drawn from AR(1) for different values of lag-1 autocorrelation  $\rho$ , several sample sizes, and 5% nominal significance level. Several variants of the Pettitt test are considered: **a** original Pettitt test without prewhitening; **b** TFPWcu adapted for Pettitt; **c** Pettitt with unconditional prewhitening, and simultaneous estimation of  $\rho$  and equivalent step change

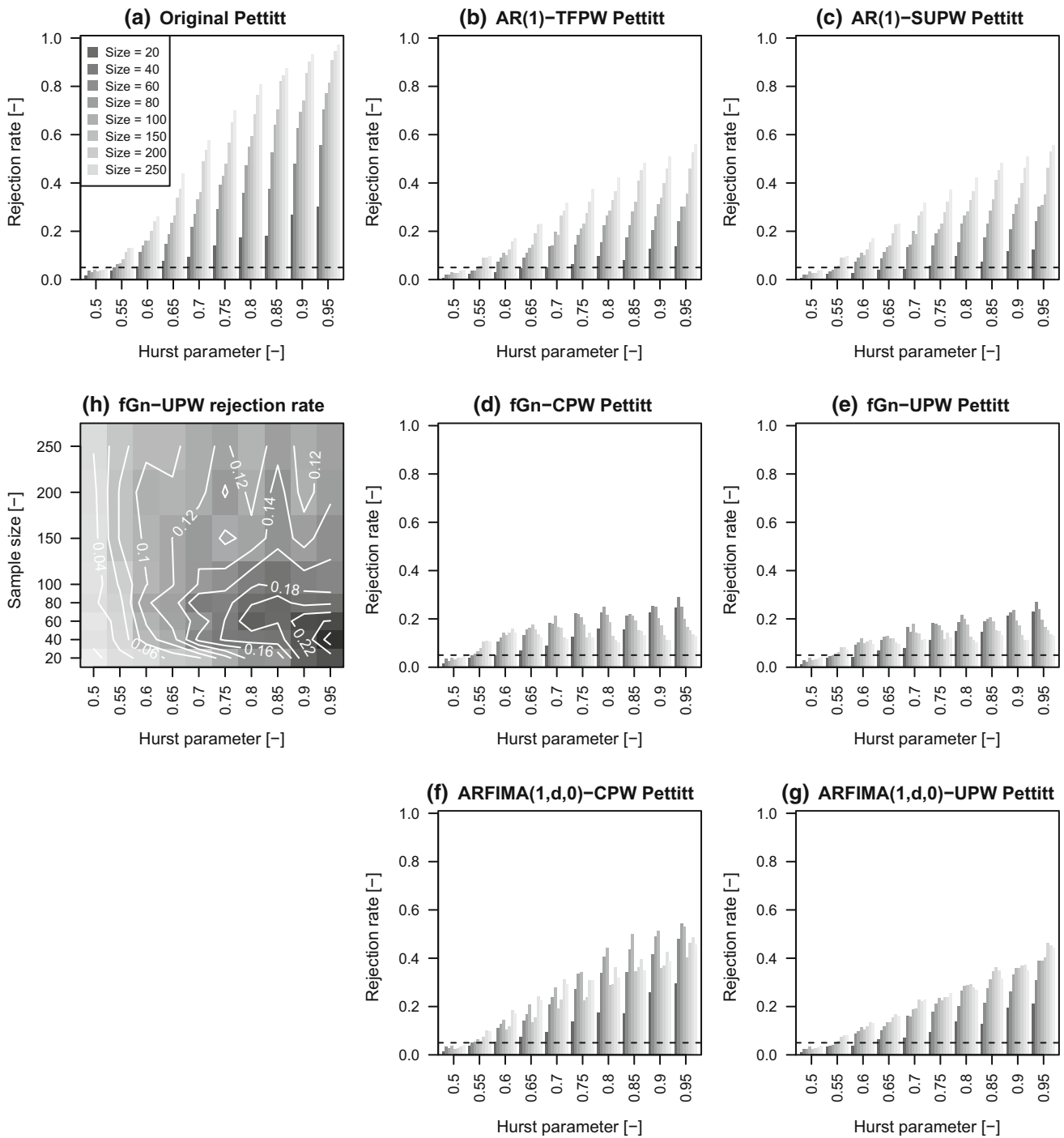
magnitude (SUPW); **d** fGn-based conditional prewhitening (fGn-CPW); **e** fGn-based unconditional prewhitening (fGn-UPW); **f** ARFIMA(1,d,0)-based conditional prewhitening (ARFIMA(1,d,0)-CPW); **g** ARFIMA(1,d,0)-based unconditional prewhitening (ARFIMA(1,d,0)-UPW); **h** map of rejection rates as a function of  $\rho$  and sample size  $T$  for the “best” performing method

outcome of the test, thus yielding rejection rates greater than those of fGn-CPW and fGn-UPW.

For time series simulated from ARFIMA(1,d,0) models, results in Fig. 4 depend on the strength of the long-range and short-range components. However, TFPWcu and SUPW generally yield rejection rates closer to the nominal

values than those provided by ARFIMA(1,d,0)-CPW and ARFIMA(1,d,0)-UPW under correct model specification. Also fGn-CPW and fGn-UPW often outperform ARFIMA-based prewhitening for some combinations of  $\rho$ ,  $d$ , and  $T$ . We argue that these results are partly related to the small sample sizes ( $T \leq 250$ ) that prevent the reliable recognition



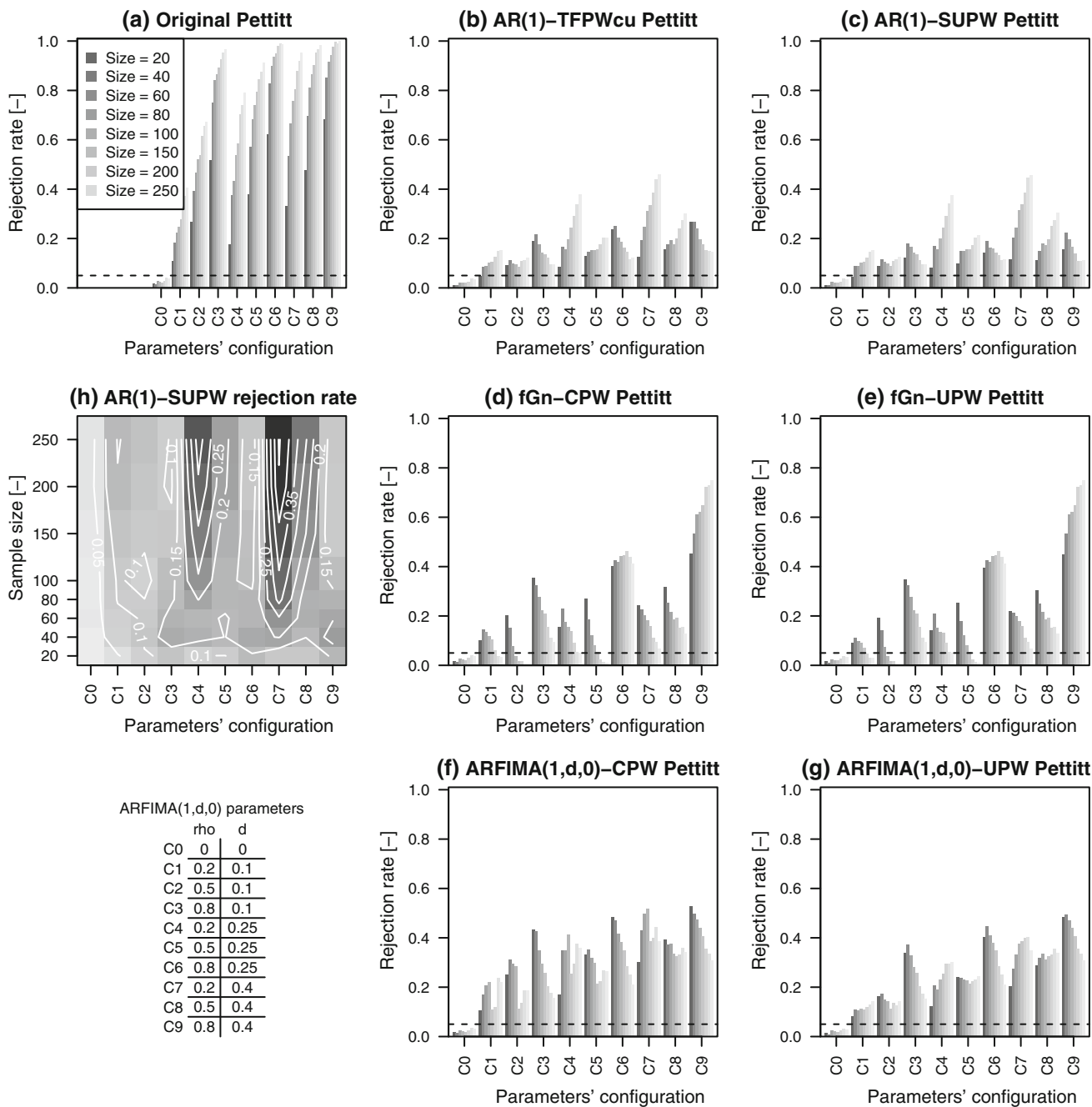


**Fig. 3** As Fig. 2, but for sequences drawn from fGn for different values of Hurst parameter  $H$  ( $H = 0.5$  denotes white noise)

of the long-range component, whereas the short-range component dominates the signal behavior, thus explaining the good performance of the AR(1)-based methods.

Finally, we explored a complementary aspect concerning the location of the change point. Theoretical arguments (Hawkins 1977) and extensive Monte Carlo experiments reported in the literature (Gurevich 2009; Gurevich and Raz

2010; Xie et al. 2014) showed that the Pettitt test can detect change points located in the middle of a time series more easily than those at other positions. However, this property can also be a drawback as it causes a tendency to erroneously detect changes in the middle of the series when no changes exist. Figure 5 confirms this behavior for some of the signals and prewhitening procedures discussed above.



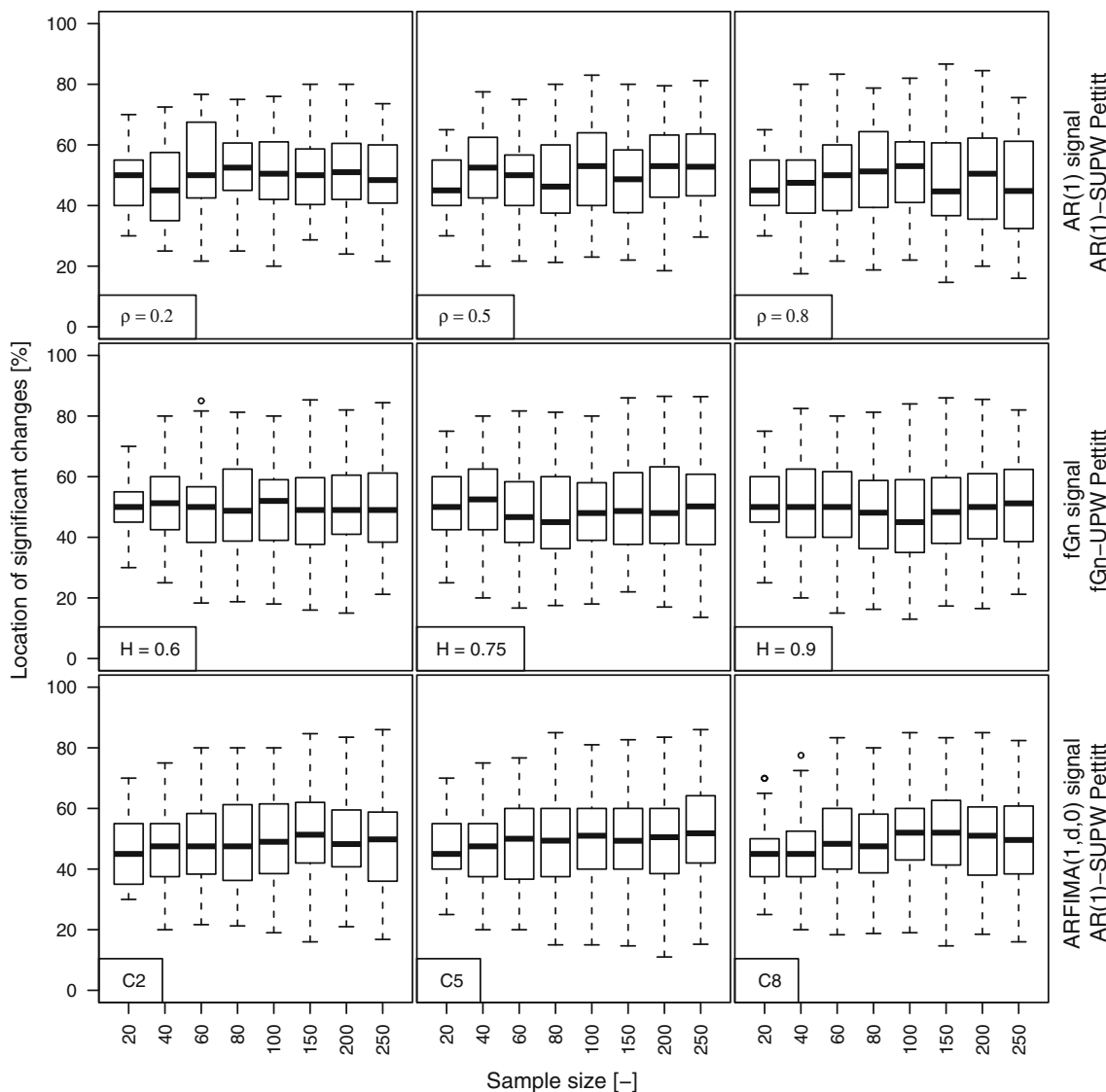
**Fig. 4** As Fig. 2, but for sequences simulated by ARFIMA(1,d,0) for different combinations of the pairs of parameters ( $\rho, d$ ) reported in the bottom left corner

**5 Conclusions**

In this study we have investigated the performance of a range of prewhitening techniques that were developed for the MK test (for gradual monotonic changes) and are suitable to be adapted to the Pettitt test (for abrupt changes). We paid attention to some critical aspects such as the bias affecting the model parameters (especially the autocorrelation terms) owing to the interaction between

deterministic (gradual or abrupt) changes and serial correlation. The analysis was supported by extensive Monte Carlo simulations devised to check the performance of the selected procedures in terms of rejection rate under the null hypothesis in order to assess their capability to control the type I error. Results can be summarized as follows:

1. A preliminary analysis of prewhitening techniques developed for MK showed that the well-known TFPW



**Fig. 5** Distribution of the relative location (percentage of the time series length) of the detected change points (at 5 % significance level) for different signals and testing procedures (see main text and captions of Figs. 2, 3, 4). Location of detected changes is expected to

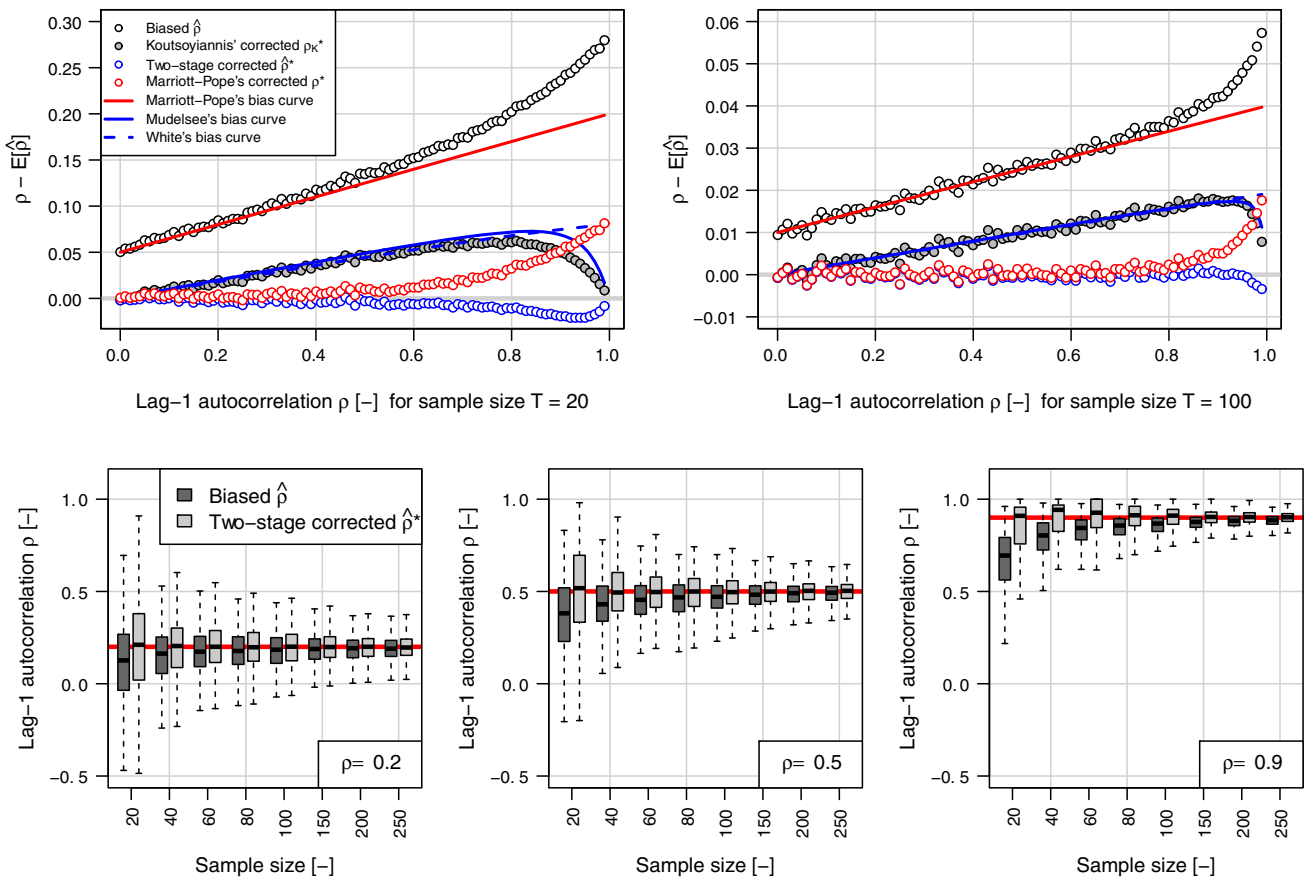
be uniformly distributed along the time series when real change points are not present. Bias toward the centre of the time series confirms previous results reported in the literature

method as introduced by Yue et al. (2002b) can provide an effective prewhitening of the series only if the trend residuals are multiplied by a magnification factor equal to  $1/(1 - \rho)$ . As this correction was introduced for instance in software such as *zyp* (Bronaugh and Werner 2013) based only on empirical results, we provide a theoretical justification showing that it is not an option but a must to guarantee the actual prewhitening of the series and the fulfillment of the basic hypotheses required for a correct application of the MK test.

2. Focusing on AR(1) signals and Pettitt test, we found that the simultaneous estimation of the model parameters ( $\rho$  and  $\Delta$ ) provides the best results, thus

confirming the suitability of this method not only for the MK test but also for the Pettitt test. On the other hand, model misspecification yields systematic over- or under-whitening, and thus under- and over-rejection, respectively. In this respect, it should be noted that we considered a range of sample sizes corresponding with hydro-meteorological series at annual or seasonal time scales, which often makes the estimation of the parameters of long-range dependence components difficult.

3. As far as fGn signals are concerned, the long-range dependence further increases the actual rejection rate confirming the difficulty of distinguishing between deterministic change points and long-range persistence



**Fig. 6** Bias correction factors for the lag-1 autocorrelation estimates of AR(1) time series. *Top panels* show the patterns described by different bias correction formulae, whereas the *bottom panels* show

the performance of the two-stage bias correction algorithm (see the Appendix for further details)

(see e.g., Beran et al. 2013, pp.700–701, and references therein). However, also in this case, prewhitening provides significant reduction of the over-rejection, even though the correction is not as effective as in the case of AR(1). For fGn, model misspecification yields only under-whitening as the alternative models exhibit autocorrelation structures weaker than fGn.

- When short-range and long-range serial dependence structures are mixed via ARFIMA(1,d,0), the performance of the Pettitt test depends on the combination of the model parameters. However, the overall result is that AR(1)-based prewhitening generally yields better results than the correct model specification. Indeed, the small sample size prevents the reliable estimation of the model parameters, especially of the long-range component, which is not easy to recognize in short time series. This partly explains the performance of AR(1)-based methods for ARFIMA(1,d,0) time series.

To summarize, prewhitening procedures do not show significant negative effects on the type I error when the data are not correlated, whereas they always provide rejection rates closer to the nominal when serial dependence is

present, the performance depending on model specification, sample size, and correlation structure and strength. Since the true process underlying real-world observations is unknown and the sample size is usually small (we refer to time series at annual or seasonal time scale commonly analyzed in the literature), AR(1)-based prewhitening is surely useful to obtain more realistic rejection rates in presence of serial correlation. fGn-based prewhitening could lead to under-rejection when long-range dependence is not present, whereas the use of more complex models could be speculative owing to the small sample sizes. Therefore, we suggest the use of AR(1)-based methods together with fGn-based technique in order to compare the results. Of course, results should be complemented with the assessment of the values of  $\rho$  and  $H$  and their significance. For a correct application of the above testing procedures, it should also be mentioned that the serial correlation in the data causes a loss of power that reduces the ability to detect real trends/changes and is independent of the prewhitening procedures. If the power is of major concern, it could be restored by increasing the significance level of the test, providing that the correct significance of the test is known (Hamed 2009).

Finally, it should be mentioned for the sake of completeness that the methods described in this study represent simple approaches (adapted for the Pettitt test) similar to those commonly applied in MK trend analyses of hydro-meteorological data. However, there is quite an extensive literature concerning other tests, especially the so-called CUSUM test, and providing asymptotic results in terms of inflation factors to be used in presence of short-range and long-range serial correlation (see e.g. Basseville and Nikiforov 1993; Beran et al. 2013 (Chap. 7.9), and references therein for an overview].

**Acknowledgments** This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) grant EP/K013513/1 “Flood MEMORY: Multi-Event Modelling Of Risk & recoverY”, and Willis Research Network. The comments of two anonymous reviewers are gratefully acknowledged. The analyses were performed in R (R Development Core Team 2014) by using the contributed packages fArma (Wuertz et al. 2013), FGN (McLeod and Veenstra 2012), fracdiff (Fraley 2012), fractal (Constantine and Percival 2014).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

**Appendix: Technical details**

We report some technical details useful for the practical implementation of the methods described in the main text. Under the hypothesis that a time series  $\{y_t\}$  is a realization of a AR(1) process, denoting the estimators of the mean and standard deviation of  $Y$  as  $\hat{\mu}$  and  $\hat{\sigma}$ , respectively, the OLS estimator of the lag-1 autocorrelation

$$\hat{\rho} = \frac{\frac{1}{(T-1)} \sum_{t=1}^{T-1} (Y_t - \hat{\mu})(Y_{t+1} - \hat{\mu})}{\frac{1}{T} \sum_{t=1}^T (Y_t - \hat{\mu})^2}, \quad \text{where } \hat{\mu} = \frac{1}{T} \sum_{t=1}^T Y_t, \tag{19}$$

is affected by two types of bias: the bias related to the correlation between the deviations of the sample covariance and variance from the population covariance and

variance, and the bias arising when the mean is not known and has to be estimated from the data. This second bias is always negative and is present even if the autocorrelation is zero (Orcutt 1948). For the case of unknown mean, which is the most common in real-world analyses, Marriott and Pope (1954) found that

$$E[\hat{\rho}] = -\frac{1}{T} + \left(1 - \frac{3}{T}\right)\rho, \tag{20}$$

which provides an approximately unbiased estimate of  $\rho$  as

$$\rho^* = \left(\hat{\rho} + \frac{1}{T}\right) \left(\frac{T}{T-3}\right). \tag{21}$$

The performance of this correction was tested by simulating 10000 samples with size  $T$  equal to 20 and 100 and  $\rho$  ranging from 0 to 0.99 by 0.01 increments, and then computing the average value of  $\hat{\rho}$  for each value of  $\rho$ . Figure 6 shows that the Marriott-Pope’s correction factor performs satisfactory up to  $\rho \approx 0.85$ , where discrepancies arise owing to the order of the series expansions used by Marriott and Pope (1954) to obtain Eq. 20.

Even though Marriott and Pope (1954) stated that “the two sources of bias may reinforce each other, or may act in opposite directions; they are not independent and cannot be investigated separately”, we found that actually distinguishing the two effects is possible, at least empirically. To show this aspect we applied a two-stage bias correction involving the correction formula proposed by Koutsoyiannis (2003) for the autocorrelation of the fGn process (also known as Hurst–Kolmogorov process)

$$\rho_K^* = \hat{\rho} \left(1 - \frac{1}{T'}\right) + \frac{1}{T'}, \tag{22}$$

where  $T'$  is the effective sample size for an AR(1) process (Koutsoyiannis and Montanari 2007, Eq. 7)

$$T' = T \frac{(1-\rho)^2}{(1-\rho^2) - 2\rho(1-\rho^T)/T}. \tag{23}$$

The obtained values  $\rho_K^*$  were therefore further corrected using a combination of the White’s and Mudelsee’s correction formulae devised to correct the bias under the hypothesis of AR(1) process with known (zero) mean (Mudelsee 2001)

---


$$\begin{cases} E[\hat{\rho}] \simeq E[\hat{\rho}]_W = \left(1 - \frac{2}{T} + \frac{4}{T^2} - \frac{2}{T^3}\right)\rho + \frac{2}{T^2}\rho^3 + \frac{2}{T^2}\rho^5 & \text{for } \rho < 0.88 \\ E[\hat{\rho}] \simeq E[\hat{\rho}]_M = \rho - \frac{2\rho}{(T-1)} + \frac{2}{(T-1)^2} \frac{(\rho - \rho^{2T-1})}{(1-\rho^2)} & \text{for } \rho \geq 0.88 \end{cases} \tag{24}$$


---



Figure 6 shows that the residual bias, after Koutsoyiannis' correction, follows closely the curve described by Eq. 24, and the further correction by this equation provides an approximately complete bias removal, thus indicating that Eq. 22 mainly accounts for the bias associated to the estimation of the unknown mean. As the two-stage bias correction (described by Eqs. 22, 23, and 24) performs better than the Marriott-Pope's formula, it is used in Step 2 of TFPWcu and SWu methods involving AR(1) prewhitening. It should be noted that such equations can be combined in a unique function representing the total bias correction and solved (numerically) for  $\rho$  in order to obtain a bias corrected estimate  $\hat{\rho}^*$ .

As far as the fGn-based procedures are concerned, prewhitening is performed using the Cholesky decomposition method described by Hamed (2009), whereas, following Hamed (2008b), the Hurst parameter  $H$  is computed by the maximum likelihood estimator (McLeod and Hipel 1978; McLeod et al. 2007) applied to the normal quantile transformed values  $\Phi^{-1}(F_n(y))$ , where  $\Phi^{-1}$  denotes the inverse of the standard Gaussian cumulative distribution function and  $F_n(y) = 1/(T+1) \sum \mathbf{1}_{\{y_i \leq y\}}$  is the Weibull version of the empirical cumulative distribution function. The maximum likelihood estimator of  $H$  has the advantage to be very accurate (Tyrallis and Koutsoyiannis 2011) and not to rely on graphical diagnostic plots unlike other estimators (see e.g., Serinaldi 2010).

ARFIMA(1, $d$ ,0) prewhitening relies on the computation of model residuals, which are calculated recursively by Eq. 5.7 and 5.9 reported in Shumway and Stoffer (2011) and adapted to account for the AR(1) component.

## References

- Basseville M, Nikiforov IV (1993) Detection of abrupt changes: theory and application. Prentice Hall, New Jersey
- Bayazit M, Önöz B (2004) Comment on "Applicability of prewhitening to eliminate the influence of serial correlation on the Mann–Kendall test" by Sheng Yue and Chun Yuan Wang. *Water Resour Res* 40(8):W08801
- Bayazit M, Önöz B (2007) To prewhiten or not to prewhiten in trend analysis? *Hydrol Sci J* 52(4):611–624
- Bayazit M, Önöz B (2008) Reply to discussion of "To prewhiten or not to prewhiten in trend analysis?". *Hydrol Sci J* 53(3):669–669
- Bayley GV, Hammersley JM (1946) The "effective" number of independent observations in an autocorrelated time series. *Suppl J R Stat Soc* 8(2):184–197
- Beran J, Feng Y, Ghosh S, Kulik R (2013) Long-memory processes: probabilistic properties and statistical methods. Springer, New York
- Bronaugh D, Werner A (2013) zyp: Zhang + Yue-Pilon trends package. <http://www.CRAN.R-project.org/package=zyp,rpackageversion0.10-1>
- Busuioc A, von Storch H (1996) Changes in the winter precipitation in Romania and its relation to the large-scale circulation. *Tellus A* 48(4):538–552
- Cochrane D, Orcutt GH (1949) Application of least squares regression to relationships containing auto-correlated error terms. *J Am Stat Assoc* 44(245):32–61
- Constantine W, Percival D (2014) Fractal: fractal time series modeling and analysis. <http://www.CRAN.R-project.org/package=fractal,rpackageversion2.0-0>
- Ferguson CR, Villarini G (2012) Detecting inhomogeneities in the Twentieth Century Reanalysis over the central United States. *J Geophys Res Atmos* 117(D5):D05–123
- Fraley C (2012) fracdiff: fractionally differenced ARIMA aka ARFIMA(p, d, q) models. URL <http://www.CRAN.R-project.org/package=fracdiff,rpackageversion1.4-2> (S original by Chris Fraley and U. Washington and Seattle. R port by Fritz Leisch at TU Wien; since 2003–2012: Martin Maechler; fdGPH and fdSperio and etc by Valderio Reisen and Artur Lemonte.)
- Guerreiro SB, Kilsby CG, Serinaldi F (2014) Analysis of time variation of rainfall in transnational basins in Iberia: abrupt changes or trends? *Int J Climatol* 34(1):114–133
- Gurevich G (2009) Asymptotic distribution of Mann–Whitney type statistics for nonparametric change point problems. *Comput Model New Technol* 13(3):18–26
- Gurevich G, Raz B (2010) Monte Carlo analysis of change point estimators. *J Appl Quant Methods* 5(4):659–669
- Hamed KH (2008a) To prewhiten or not to prewhiten in trend analysis? *Hydrol Sci J* 53(3):667–668
- Hamed KH (2008b) Trend detection in hydrologic data: the Mann–Kendall trend test under the scaling hypothesis. *J Hydrol* 349(3–4):350–363
- Hamed KH (2009) Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *J Hydrol* 368(1–4):143–155
- Hamed KH, Rao AR (1998) A modified Mann–Kendall trend test for autocorrelated data. *J Hydrol* 204(1–4):182–196
- Hawkins DM (1977) Testing a sequence of observations for a shift in location. *J Am Stat Assoc* 72(357):180–186
- Kang W, Shin DW, Lee Y (2003) Biases of the restricted maximum likelihood estimators for ARMA processes with polynomial time trend. *J Stat Plan Inference* 116(1):163–176
- Katz RW (1988) Statistical procedures for making inferences about climate variability. *J Clim* 1(11):1057–1064
- Khalilq MN, Ouarda TBMJ, Gachon P, Sushama L, St-Hilaire A (2009) Identification of hydrological trends in the presence of serial and cross correlations: a review of selected methods and their application to annual flow regimes of Canadian rivers. *J Hydrol* 368(1–4):117–130
- Koutsoyiannis D (2003) Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrol Sci J* 48(1):3–24
- Koutsoyiannis D, Montanari A (2007) Statistical analysis of hydro-climatic time series: uncertainty and insights. *Water Res Res* 43(5):W05–429
- Kulkarni A, von Storch H (1995) Monte Carlo experiments on the effect of serial correlation on the Mann–Kendall test of trend. *Meteorol Z* 4(2):82–85
- Kundzewicz ZW, Robson AJ (2004) Change detection in hydrological records—a review of the methodology. *Hydrol Sci J* 49(1):7–19
- Lenton RL, Schaake JC (1973) Comments on 'Small sample estimation of  $\rho_1'$  by James R. Wallis and P. Enda O'Connell. *Water Resour Res* 9(2):503–504
- Mallakpour I, Villarini G (2015) A simulation study to examine the sensitivity of the Pettitt test to detect abrupt changes in mean. *Hydrol Sci J*. doi:10.1080/02626667.2015.1008482
- Marriott FHC, Pope JA (1954) Bias in the estimation of autocorrelations. *Biometrika* 41(3–4):390–402
- Matalas NC, Sankarasubramanian A (2003) Effect of persistence on trend detection via regression. *Water Resour Res* 39(12):1342
- McLeod AI, Hipel KW (1978) Preservation of the rescaled adjusted range: 1. A reassessment of the Hurst Phenomenon. *Water Resour Res* 14(3):491–508

- McLeod AI, Veenstra J (2012) FGN: fractional Gaussian noise, estimation and simulation. <http://www.CRAN.R-project.org/package=FGN,rpackageversion2.0>
- McLeod AI, Yu H, Krougly ZL (2007) Algorithms for linear time series analysis: with R Package. *J Stat Softw* 23(5):1–26
- Mudelsee M (2001) Note on the bias in the estimation of the serial correlation coefficient of AR(1) processes. *Stat Pap* 42(4):517–527
- Önöz B, Bayazit M (2003) The power of statistical tests for trend detection. *Turk J Eng and Environ Sci* 27(4):247–251
- Orcutt GH (1948) A study of the autoregressive nature of the time series used for Tinbergen's model of the economic system of the United States, 1919–1932. *J R Stat Soc Ser B* 10(1):1–53
- Pettitt AN (1979) A non-parametric approach to the change-point problem. *J R Stat Soc Ser C* 28(2):126–135
- R Development Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>, ISBN3-900051-07-0
- Rodionov SN (2004) A sequential algorithm for testing climate regime shifts. *Geophys Res Lett* 31(9):L09204
- Rodionov SN (2006) Use of prewhitening in climate regime shift detection. *Geophys Res Lett* 33(12):L12707
- Rougé C, Ge Y, Cai X (2013) Detecting gradual and abrupt changes in hydrological records. *Adv Water Resour* 53:33–44
- Rybski D, Neumann J (2011) A review on the Pettitt test. In: Kropp J, Schellnhuber HJ (eds) *In extremis*. Springer, Dordrecht, pp 202–213
- Sagarika S, Kalra A, Ahmad S (2014) Evaluating the effect of persistence on long-term trends and analyzing step changes in streamflows of the continental United States. *J Hydrol* 517:36–53
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 63(324):1379–1389
- Serinaldi F (2010) Use and misuse of some Hurst parameter estimators applied to stationary and non-stationary financial time series. *Phys A Stat Mech Appl* 389(14):2770–2781
- Shumway RH, Stoffer DS (2011) *Time series analysis and its applications: with R examples*, 3rd edn. Springer, New York
- Tramblay Y, El Adlouni S, Servat E (2013) Trends and variability in extreme precipitation indices over Maghreb countries. *Nat Hazards Earth Syst Sci* 13(12):3235–3248
- Tyralis H, Koutsoyiannis D (2011) Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. *Stoch Environ Res Risk Assess* 25(1):21–33
- van Giersbergen NPA (2005) On the effect of deterministic terms on the bias in stable AR models. *Econ Lett* 89(1):75–82
- Villarini G, Serinaldi F, Smith JA, Krajewski WF (2009) On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resour Res* 45(8):W08417
- Villarini G, Smith JA, Serinaldi F, Ntelekos AA (2011) Analyses of seasonal and annual maximum daily discharge records for central Europe. *J Hydrol* 399(3–4):299–312
- von Storch H (1999) Misuses of statistical analysis in climate research. In: von Storch H, Navarra A (eds) *Analysis of climate variability*. Springer, Dordrecht, pp 11–26
- Wallis JR, O'Connell PE (1972) Small sample estimation of  $\rho_1$ . *Water Resour Res* 8(3):707–712
- Wang XL, Swail VR (2001) Changes of extreme wave heights in northern hemisphere oceans and related atmospheric circulation regimes. *J Clim* 14(10):2204–2221
- Wuertz D, many others, see the SOURCE file (2013) fArma: ARMA Time Series Modelling. <http://www.CRAN.R-project.org/package=fArma,rpackageversion3010.79>
- Xie H, Li D, Xiong L (2014) Exploring the ability of the Pettitt method for detecting change point by Monte Carlo simulation. *Stoch Environ Res Risk Assess* 28(7):1643–1655
- Yue S, Pilon P (2004) A comparison of the power of the t test, Mann–Kendall and bootstrap tests for trend detection. *Hydrol Sci J* 49(1):21–37
- Yue S, Wang C (2002) Applicability of prewhitening to eliminate the influence of serial correlation on the Mann–Kendall test. *Water Resour Res* 38(6):41–47
- Yue S, Wang C (2004a) Reply to comment by Mehmetcik Bayazit and Bihret Önöz on “Applicability of prewhitening to eliminate the influence of serial correlation on the Mann–Kendall test”. *Water Resour Res* 40(8):W08802
- Yue S (2004b) Reply to comment by Xuebin Zhang and Francis W. Zwiers on “Applicability of prewhitening to eliminate the influence of serial correlation on the Mann–Kendall test”. *Water Resour Res* 40(3):W03806
- Yue S, Wang CY (2004c) The Mann–Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resour Manag* 18(3):201–218
- Yue S, Pilon P, Cavadias G (2002a) Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *J Hydrol* 259(1–4):254–271
- Yue S, Pilon P, Phinney B, Cavadias G (2002b) The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrol Process* 16(9):1807–1829
- Zhang X, Zwiers FW (2004) Comment on “Applicability of prewhitening to eliminate the influence of serial correlation on the Mann–Kendall test” by Sheng Yue and Chun Yuan Wang. *Water Resour Res* 40(3):W03805
- Zhang X, Vincent LA, Hogg WD, Niitsoo A (2000) Temperature and precipitation trends in Canada during the 20th century. *Atmos Ocean* 38(3):395–429