Hydrological Sciences Journal

ISSN: 0262-6667 (Print) 2150-3435 (Online) Journal homepage: www.tandfonline.com/journals/thsj20

The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data

K. H. Hamed

The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data

K. H. Hamed

Irrigation and Hydraulics Department, Faculty of Engineering, Cairo University, Egypt
hamedkhaled@hotmail.com

Received 1 April 2010; accepted 3 March 2011; open for discussion until 1 January 2012

Citation Hamed, K. H. (2011) The distribution of Kendall's tau for testing the significance of cross-correlation in persistent data. *Hydrol. Sci. J*. **56**(5), 841–853.

Abstract Kendall's tau (τ) has been widely used as a distribution-free measure of cross-correlation between two variables. It has been previously shown that persistence in the two involved variables results in the inflation of the variance of τ. In this paper, the full null distribution of Kendall's τ for persistent data with multivariate Gaussian dependence is derived, and an approximation to the full distribution is proposed. The effect of the deviation from the multivariate Gaussian dependence model on the distribution of τ is also investigated. As a demonstration, the temporal consistency and field significance of the cross-correlation between the North Hemisphere (NH) temperature time series in the period 1850–1995 and a set of 784 NH tree-ring width (TRW) proxies in addition to 105 NH tree-ring maximum latewood density (MXD) proxies are studied. When persistence is ignored, the original Mann-Kendall test gives temporally inconsistent results between the early half (1850–1922) and the late half (1923–1995) of the record. These temporal inconsistencies are largely eliminated when persistence is accounted for, indicating the spuriousness of a large portion of the identified cross-correlations. Furthermore, the use of the modified test in combination with a field significance test that is robust to spatial correlation indicates the absence of field significant cross-correlation in both halves of the record. These results have serious implications for the use of tree-ring data as temperature proxies, and emphasize the importance of utilizing the correct distribution of Kendall's τ in order to avoid the overestimation of the significance of cross-correlation between data that exhibit significant persistence.

Key words Kendall's tau; autocorrelation; cross-correlation; persistence; probability distribution; distribution-free; non-parametric; field significance; Gaussian dependence; copula

La distribution du tau de Kendall pour tester la significativité de la corrélation croisée dans des données persistantes

Résumé Le tau de Kendall (τ) a été largement utilisé comme mesure non paramétrique de corrélation croisée entre deux variables. Il a été précédemment montré que la persistance dans les deux variables considérées résulte en l'inflation de la variance de τ. Dans cet article, on dérive la distribution nulle complète du τ de Kendall pour des données persistantes avec une dépendance Gaussienne multivariée, et l'on propose une approximation de la distribution complète. L'effet de l'écart au modèle de dépendance Gaussienne multivariée sur la distribution de τ est également étudié. En guise de démonstration, la cohérence temporelle et le champ de significativité de la corrélation croisée entre des séries temporelles de températures de l'Hémisphère Nord (HN) pour la période 1850–1995, et un ensemble de 784 largeurs de cernes d'arbres ainsi que 105 valeurs de densité maximum des cernes du bois d'été sont étudiés. Lorsque la persistance est ignorée, le test original de Mann-Kendall donne des résultats temporellement incohérents entre la première (1850–1922) et la seconde moitié (1923–1995) des enregistrements. Ces incohérences temporelles sont en grande partie éliminées lorsque la persistance est prise en compte, indiquant qu'une grande partie des corrélations croisées identifiées sont fausses. De plus, l'utilisation du test modifié en combinaison avec un test du champ de significativité, qui est robuste par rapport à la corrélation spatiale, indique l'absence de corrélation croisée globalement significative au sein des deux moitiés des enregistrements. Ces résultats ont des implications sérieuses pour l'utilisation de données de cernes d'arbres comme indicateurs de températures, et soulignent l'importance d'utiliser la bonne distribution du τ de Kendall afin d'éviter la surestimation de la significativité de la corrélation croisée entre des données présentant une persistance significative.

Mots clefs tau de Kendall; autocorrélation; corrélation croisée; persistance; distribution de probabilité; distribution libre; non paramétrique; champ de significativité; dépendance Gaussienne; copule

ISSN 0262-6667 print/ISSN 2150-3435 online

doi: 10.1080/02626667.2011.586948
http://www.informaworld.com

INTRODUCTION

Recently, there has been an interest to investigate relationships between variables related to various natural phenomena, such as climatic teleconnections (e.g. Chiew *et al*. 1998, Ghanbari and Bravo 2008, Panarello and Dapeña, 2009, Soukup *et al*. 2009). Many natural data are characterized by skewed distributions, in which case distribution-free statistics, such as Kendall's tau (τ) or Spearman's rho, are preferred as a means to assess the significance of the cross-correlation between such variables. However, in testing the significance of cross-correlation, the important assumption of independent observations from each of the studied variables is often overlooked. The assumption of independence of the observations is violated by most natural data, which often exhibit various degrees of persistence, i.e. observations are positively autocorrelated over various ranges of time lags. The existence of autocorrelation affects the distribution of the test statistic, which results in erroneous results if such effect is not taken into consideration. Previous studies have investigated the effect of autocorrelation on the variance of Kendall's τ (Hamed 2009b) as well as the distribution of the Mann-Kendall trend test statistic (Hamed 2008, 2009a), which is a special case of Kendall's τ. This paper derives the full distribution of Kendall's τ under the null hypothesis that the two parents are independent, but when each is autocorrelated with a multivariate Gaussian (MVG) dependence model. An approximation to the full distribution is also suggested to reduce the computational effort.

Although classical time series modelling assumes that normal transformation of the data is sufficient to proceed with an MVG time series model (e.g. Salas *et al*. 1980), it has been recently shown that many natural data do not follow the MVG dependence model, regardless of normal transformation, which only affects their marginal distributions (e.g. Grimaldi and Serinaldi 2006). The effect of deviation from the MVG dependence model on the distribution of τ is also investigated in this paper. A case study illustrating the importance of considering the correct distribution of τ and the potential problems arising from ignoring the effects of persistence and spatial correlation is also presented.

BACKGROUND

Kendall's τ (Kendall 1955) is a measure of concordance between two observed variables. The statistic τ is defined as the difference between the probabilities of concordance and discordance between the two variables (Kendall and Gibbons 1990, Sec. 9.3 and 9.14):

$$\begin{aligned}\tau &= P(y_i < y_j | x_i < x_j) - P(y_i > y_j | x_i < x_j) \\ &= 2P(y_i > y_j | x_i < x_j) - 1 = 2\pi_1 - 1\end{aligned} \tag{1}$$

According to the definition in equation (1) above, τ is a population property. When X and Y are bivariate normal with correlation coefficient ρ, it is straight forward (Kendall and Gibbons, 1990) to show that:

$$\tau = \frac{2}{\pi} \sin^{-1} \rho \tag{2}$$

For a sample of n observations (x_1, y_1), (x_2, y_2), . . ., (x_n, y_n), an estimate t of τ can be calculated as:

$$t = \frac{2S}{n(n-1)} \tag{3}$$

The statistic S in equation (3) is given by:

$$S = \sum_{i<j} a_{ij}\, b_{ij} \tag{4}$$

where

$$\begin{aligned}a_{ij} &= sign(x_j - x_i) \\ &= sign(R_j - R_i) = \begin{cases} 1 & x_i < x_j \\ 0 & x_i = x_j \\ -1 & x_i > x_j \end{cases}\end{aligned} \tag{5}$$

and R_i and R_j are the ranks of observations x_i and x_j, and b_{ij} is defined similarly for y_i and y_j.

Under the null hypothesis that the two series X and Y are independent, and assuming that observations in each time series are independent, the mean and variance of S are given by (Kendall 1955)

$$E(S) = 0 \tag{6}$$

$$V_0(S) = n(n-1)(2n+5)/18 \tag{7}$$

Accordingly, the mean and variance of t are given by:

$$E(t) = 0 \tag{8}$$

$$V_0(t) = \frac{2(2n+5)}{9n(n-1)} \tag{9}$$

Kendall (1955) shows that the distribution of S, and accordingly that of t, tends to normality as the number of observations becomes large, and that for $n > 10$ the normal curve gives a satisfactory approximation of the distributions of S and t. Kendall (1955) also derives the full distribution of S (or equivalently t) based on the fact that all paired rankings of X and Y are equally probable under the same assumption of independence.

Kendall's τ has been classically used to test the significance of cross-correlation between two variables when their distributions significantly deviate from the normal distribution. In that case, a significance test based on the distribution-free τ, which is a function of the ranks of the variates rather than their actual values, offers more power than other parametric tests such as the Pearson product-moment correlation coefficient (Yue *et al*. 2002). Kendall's τ has also been used to test the significance of trends in univariate data by comparing the values of X with their time order. In that case, the test is known as the Mann-Kendall trend test (Mann 1945). The effect of persistence on the Mann-Kendall trend test statistic has been addressed by Hamed (2008). The effect of autocorrelation on the variance of S depends on the sign of the autocorrelations, where positive autocorrelation tends to inflate the variance, while negative autocorrelation tends to reduce it. The effect of autocorrelation on the distribution is consistent with its effect on the variance, where positive autocorrelation results in heavier tails of the distribution, and *vice versa* (Hamed 2009a).

DISTRIBUTION OF KENDALL'S τ FOR PERSISTENT DATA WITH GAUSSIAN DEPENDENCE

When the data follow the multivariate Gaussian dependence model, the derivation of the mean and variance of t, as well as its full distribution, can be facilitated by substituting the ranks of the observations with a corresponding set of normal variates (Hamed 2009a). This substitution would not change the distribution of τ, as the ranks remain unchanged, but it greatly facilitates calculations.

The mean and variance of Kendall's τ with autocorrelated parents have been derived by Hamed (2009b) as:

$$E(t) = 0 \tag{10}$$

$$V(t) = \frac{4\,V(S)}{n^2(n-1)^2} \tag{11}$$

where

$$V(S) = \frac{4}{\pi^2} \sum_{i<j} \sum_{k<l} \sin^{-1} r_X\,(i,j,k,l) \; \sin^{-1} r_Y\,(i,j,k,l) \tag{12}$$

$$r_X(i,j,k,l) = \frac{\rho_{jl} - \rho_{il} - \rho_{jk} + \rho_{ik}}{\sqrt{(2-2\rho_{ij})\,(2-2\rho_{kl})}} \tag{13}$$

and ρ_{pq} for different suffixes p and q are the correlation between the normal variates corresponding to the variable X, while $r_Y(i, j, k, l)$ is defined similarly for the variable Y.

It should be noted that the calculation of Kendall's τ allows for ties to exist in the data, as can be seen from equation (5) above, in which case the variance has to be modified to account for ties (Kendall 1955). However, the effect of ties is not easy to incorporate in the case of persistent data, and it will not be taken into account here. The variance in equation (11) applies to continuous hydrologic data, in which case the probability of ties existing in the data is theoretically equal to zero. However, rounding off of measurements due to instrument resolution may in some cases result in apparent ties, in which case the calculation of the variance may not be accurate. A discussion on the effect of ties on the variance of Kendall's τ in the special case of trend testing can be found in Hamed (2008).

Hamed (2009b) gives tables for the variance inflation factor for positively correlated first-order autoregressive (AR(1)) and fractional Gaussian noise (FGN) data as examples of the effect of short- and long-term persistence, respectively. It should also be noted that the independence of the observations of only one of the two series is sufficient to eliminate the effect of autocorrelation on the variance of Kendall's τ.

The basic idea in deriving the full distribution of S is to evaluate the probabilities of all possible rankings of the n involved observations (Kendall and Gibbons 1990, Sec. 5.2). The total number of such rankings for a sample of size n of the variable X is the factorial of n. The same is true for the other variable Y, and the total number of possible joint rankings is thus $(n!)^2$. Each of these joint rankings results in a different value of S depending on the ordering of the data. If the data are independent, all of these rankings are equally probable. The frequency distribution of S can thus be calculated as the number of times each value of S is repeated, divided by the total number of rankings. The distribution of S up to $n = 10$ for independent data is tabulated by Kendall and Gibbons (1990).

Following the same procedure of Hamed (2009a) for the Mann-Kendall trend test statistic, the full distribution of Kendall's τ can be calculated to a reasonable accuracy for small values of n. Methods and software for calculating the needed multivariate normal probabilities using stochastic integration (Genz and Bretz 1999, 2002) for values of n up to 100 can be found at the following internet site: http://www.math.wsu.edu/faculty/genz/homepage.

For illustration, Table 1 gives the distribution of the statistic S when the corresponding normal variables are first-order autocorrelated (AR(1)) for different combinations of the first-order autocorrelation coefficients ρ_X and ρ_Y, and for sample sizes up to $n = 8$. Figure 1 shows the distribution of τ for the case of $n = 8$. It is clear that as the values of autocorrelation increase, the distribution tails become heavier with a corresponding flattening of the peak, resulting in inflating the variance, as discussed earlier. The distribution can be calculated for any other autocorrelation structure (e.g. scaling, also known as the Hurst effect, which can be modelled for example as fractional Gaussian noise (FGN), see Koutsoyiannis 2003) by substituting the values of the autocorrelation function into equation (13).

Although calculating the full distribution for larger sample sizes is possible in principle, it is not practical due to the high computational requirements. Kendall (1955) gives a proof of the asymptotic normality of the distribution of τ based on the fact of symmetry of the distribution and an analysis of the number of terms involved in the evaluation of the even moments of the distribution. The same arguments hold for the case of persistent data in principle, since the number of involved terms does not change. However, the existence of autocorrelation results in slower convergence to the normal distribution. For example, Fig. 2 shows the change of the kurtosis coefficient of the distribution of S as a function of the sample size and the autocorrelation coefficients for two AR(1) series. It can be seen that the kurtosis coefficient approaches that of the normal distribution (kurtosis coefficient = 3) as the number of observations increases, but it becomes very slow when autocorrelation is high. As a result, the normal distribution may not give a good approximation unless the sample size n is sufficiently large. As an alternative for intermediate sample sizes, a more flexible alternative is suggested. Since the distribution of τ is bounded between -1 and $+1$, a symmetric discretized beta distribution (after proper shifting and scaling, adding 1 and dividing by 2) is suggested. Due to symmetry, the two parameters of the beta distribution will be equal, requiring the estimation of only one parameter. The required parameter, say a, can be estimated using a scaled version (divided by 4) of the variance of t calculated from equation (11) to give:

$$a = \frac{1}{2}\left(\frac{1}{V(t)} - 1\right) \tag{14}$$

The discretized probability mass function of t can thus be calculated as:

$$P(t = t_0) = \int_c^d f(x)\mathrm{d}x \tag{15}$$

where $c = (t_0 + 1)(N - 1)/2N$, $d = c + 1/N$, $N = n(n - 1)/2 + 1$, and

$$f(x) = \frac{\Gamma(2a)}{\Gamma^2(a)} x^{a-1}(1 - x)^{a-1}, \quad 0 < x < 1 \tag{16}$$

The integral in equation (15) involves the calculation of the incomplete beta function, which is available in many statistical software packages. Because of the slight difference between the variances of the continuous and discretized versions of the beta distribution, the resulting discretized beta distribution will not have exactly the required variance. However, a few iterations (usually less than three) by adjusting the variance used in equation (14) will result in a distribution with the required variance. The cumulative distribution function can accordingly be calculated as:

$$F(t_0) = P(t \le t_0) = \int_0^d f(x)\mathrm{d}x \tag{17}$$

where $d = (t_0 + 1)(N - 1)/[(2N + 1)N]$ and $N = n(n - 1)/2 + 1$, as before.

Table (1a) Exact distribution of S for AR(1) data with different values of n and unequal first order autocorrelation coefficients ρ_X and ρ_Y. The numbers in the table give the probability that S attains the given value or larger.

		ρ_X, ρ_Y						
	S	0, ρ	0.2, 0.4	0.2, 0.6	0.2, 0.8	0.4, 0.6	0.4, 0.8	0.6, 0.8
$n = 3$	1	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
	3	0.1667	0.1685	0.1694	0.1703	0.1720	0.1737	0.1770
$n = 4$	0	0.6250	0.6225	0.6215	0.6206	0.6184	0.6167	0.6133
	2	0.3750	0.3775	0.3785	0.3794	0.3816	0.3833	0.3867
	4	0.1667	0.1709	0.1726	0.1741	0.1780	0.1811	0.1876
	6	0.0417	0.0433	0.0441	0.0447	0.0464	0.0478	0.0510
$n = 5$	0	0.5916	0.5891	0.5881	0.5872	0.5848	0.5831	0.5796
	2	0.4083	0.4109	0.4120	0.4128	0.4151	0.4169	0.4205
	4	0.2417	0.2475	0.2499	0.2519	0.2573	0.2614	0.2701
	6	0.1167	0.1221	0.1243	0.1263	0.1316	0.1357	0.1447
	8	0.0417	0.0445	0.0457	0.0468	0.0498	0.0522	0.0577
	10	0.0083	0.0090	0.0093	0.0096	0.0104	0.0111	0.0126
$n = 6$	1	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
	3	0.3597	0.3642	0.3661	0.3677	0.3719	0.3751	0.3817
	5	0.2347	0.2419	0.2449	0.2475	0.2543	0.2595	0.2707
	7	0.1361	0.1433	0.1464	0.1491	0.1563	0.1619	0.1742
	9	0.0680	0.0734	0.0757	0.0777	0.0833	0.0878	0.0980
	11	0.0278	0.0306	0.0319	0.0330	0.0362	0.0389	0.0451
	13	0.0083	0.0094	0.0098	0.0103	0.0115	0.0126	0.0152
	15	0.0014	0.0016	0.0017	0.0018	0.0020	0.0022	0.0028
$n = 7$	1	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
	3	0.3863	0.3906	0.3923	0.3937	0.3979	0.4007	0.4071
	5	0.2810	0.2883	0.2914	0.2939	0.3012	0.3064	0.3179
	7	0.1907	0.1993	0.2030	0.2062	0.2150	0.2214	0.2359
	9	0.1195	0.1276	0.1312	0.1342	0.1428	0.1493	0.1643
	11	0.0681	0.0744	0.0773	0.0797	0.0868	0.0923	0.1053
	13	0.0345	0.0387	0.0406	0.0422	0.0471	0.0510	0.0606
	15	0.0151	0.0173	0.0183	0.0193	0.0220	0.0243	0.0301
	17	0.0054	0.0063	0.0067	0.0071	0.0083	0.0093	0.0121
	19	0.0014	0.0017	0.0018	0.0019	0.0023	0.0026	0.0035
	21	0.0002	0.0002	0.0003	0.0003	0.0003	0.0004	0.0005
$n = 8$	0	0.5476	0.5458	0.5449	0.5439	0.5423	0.5406	0.5376
	2	0.4524	0.4546	0.4554	0.4559	0.4580	0.4592	0.4623
	4	0.3598	0.3655	0.3679	0.3697	0.3754	0.3792	0.3880
	6	0.2742	0.2825	0.2861	0.2889	0.2973	0.3032	0.3166
	8	0.1994	0.2090	0.2131	0.2166	0.2265	0.2337	0.2501
	10	0.1375	0.1470	0.1512	0.1548	0.1649	0.1725	0.1900
	12	0.0894	0.0977	0.1014	0.1046	0.1139	0.1209	0.1377
	14	0.0543	0.0608	0.0637	0.0663	0.0738	0.0798	0.0943
	16	0.0305	0.0349	0.0370	0.0389	0.0444	0.0488	0.0601
	18	0.0156	0.0183	0.0196	0.0207	0.0243	0.0272	0.0350
	20	0.0071	0.0085	0.0092	0.0098	0.0118	0.0135	0.0181
	22	0.0028	0.0034	0.0037	0.0040	0.0049	0.0057	0.0080
	24	0.0009	0.0011	0.0012	0.0013	0.0016	0.0019	0.0028
	26	0.0002	0.0003	0.0003	0.0003	0.0004	0.0005	0.0007
	28	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001

An advantage of using the beta distribution is that in addition to the variance, the kurtosis is also a function of the distribution parameter a, offering more flexibility to approximate the actual distribution of t. Figure 3 shows a comparison between the approximations of the probability mass function of S using the normal and discretized beta distributions in relation to the exact distribution for a sample size of $n = 8$. It is clear that the discretized beta distribution offers a better approximation to the exact distribution for this small sample size as the effect of autocorrelation increases. As the number of observations increases, the differences between the normal and discretized beta approximations diminish, but the discretized beta

Table (1b) Exact distribution of S for AR(1) data with different values of n and equal first order autocorrelation coefficients ρ_X and ρ_Y. The numbers in the table give the probability that S attains the given value or larger.

		ρ_X, ρ_Y					
	S	0, 0	0.2, 0.2	0.4, 0.4	0.6, 0.6	0.8, 0.8	0.9, 0.9
$n=3$	1	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
	3	0.1667	0.1676	0.1703	0.1746	0.1803	0.1837
$n=4$	0	0.6250	0.6236	0.6203	0.6157	0.6102	0.6071
	2	0.3750	0.3764	0.3797	0.3843	0.3898	0.3929
	4	0.1667	0.1689	0.1746	0.1829	0.1938	0.2002
	6	0.0417	0.0425	0.0449	0.0487	0.0542	0.0577
$n=5$	0	0.5916	0.5903	0.5868	0.5820	0.5764	0.5732
	2	0.4083	0.4097	0.4132	0.4180	0.4236	0.4267
	4	0.2417	0.2448	0.2527	0.2640	0.2780	0.2860
	6	0.1167	0.1195	0.1270	0.1384	0.1533	0.1624
	8	0.0417	0.0431	0.0472	0.0538	0.0633	0.0696
	10	0.0083	0.0087	0.0097	0.0115	0.0144	0.0164
$n=6$	1	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
	3	0.3597	0.3620	0.3683	0.3771	0.3876	0.3933
	5	0.2347	0.2385	0.2485	0.2630	0.2808	0.2910
	7	0.1361	0.1399	0.1501	0.1656	0.1858	0.1979
	9	0.0680	0.0708	0.0785	0.0908	0.1081	0.1191
	11	0.0278	0.0292	0.0335	0.0407	0.0517	0.0594
	13	0.0083	0.0089	0.0104	0.0133	0.0181	0.0218
	15	0.0014	0.0015	0.0018	0.0024	0.0034	0.0044
$n=7$	1	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
	3	0.3863	0.3886	0.3944	0.4027	0.4126	0.4181
	5	0.2810	0.2848	0.2951	0.3100	0.3282	0.3384
	7	0.1907	0.1952	0.2075	0.2260	0.2493	0.2627
	9	0.1195	0.1237	0.1355	0.1540	0.1785	0.1933
	11	0.0681	0.0713	0.0808	0.0963	0.1183	0.1324
	13	0.0345	0.0366	0.0429	0.0539	0.0707	0.0822
	15	0.0151	0.0162	0.0196	0.0260	0.0366	0.0445
	17	0.0054	0.0058	0.0073	0.0101	0.0154	0.0197
	19	0.0014	0.0015	0.0020	0.0029	0.0047	0.0063
	21	0.0002	0.0002	0.0003	0.0004	0.0008	0.0011
$n=8$	0	0.5476	0.5465	0.5437	0.5397	0.5349	0.5323
	2	0.4524	0.4535	0.4562	0.4602	0.4650	0.4675
	4	0.3598	0.3627	0.3705	0.3820	0.3957	0.4032
	6	0.2742	0.2785	0.2902	0.3075	0.3286	0.3404
	8	0.1994	0.2043	0.2180	0.2388	0.2651	0.2802
	10	0.1375	0.1424	0.1562	0.1779	0.2066	0.2235
	12	0.0894	0.0936	0.1059	0.1261	0.1542	0.1716
	14	0.0543	0.0576	0.0673	0.0842	0.1091	0.1254
	16	0.0305	0.0327	0.0396	0.0522	0.0722	0.0861
	18	0.0156	0.0169	0.0212	0.0295	0.0438	0.0545
	20	0.0071	0.0078	0.0101	0.0148	0.0237	0.0310
	22	0.0028	0.0031	0.0041	0.0063	0.0110	0.0152
	24	0.0009	0.0010	0.0013	0.0022	0.0041	0.0060
	26	0.0002	0.0002	0.0003	0.0005	0.0011	0.0017
	28	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003

distribution remains closer to the exact distribution. This is shown in Fig. 4 for a sample size of $n = 25$, where the distribution of t is simulated using 10^6 replicas of two independent AR(1) series. Similar results have also been obtained for simple scaling stochastic processes (e.g. FGN) with different scaling (Hurst) parameters.

NON-GAUSSIAN DEPENDENCE: MULTIVARIATE COPULAS

The derivation of the distribution of τ in the previous section assumes an MVG dependence structure for each of the two involved variables. However, as noted by one reviewer, the use of the corresponding normal

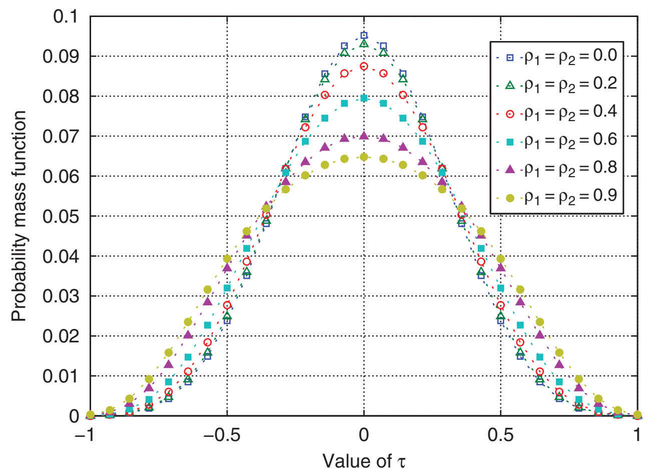

Fig. 1 Distribution of τ for sample size $n = 8$ and AR(1) autocorrelated parents with different first-order correlation coefficients. The distribution of tau is discrete, but dotted lines are used for clarity.

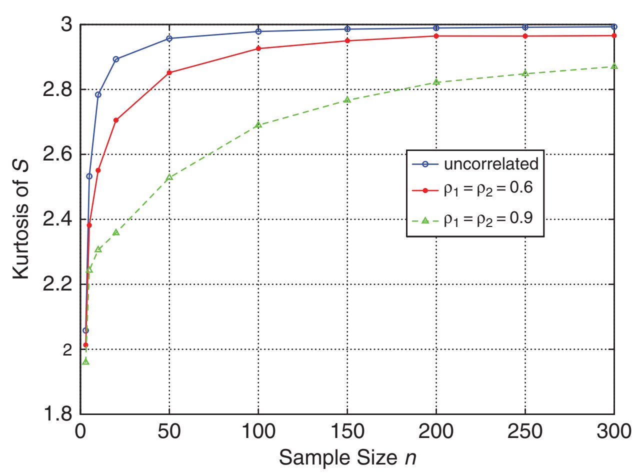

Fig. 2 Convergence of the kurtosis coefficient of S for different sample sizes and different degrees of autocorrelation.

variates does not guarantee an MVG dependence structure. In fact, although it changes the marginal distributions into Gaussian ones, the normal transformation cannot change the temporal dependence structure of the ranks, a fact that has been overlooked in previous studies (e.g. Hamed 2009a). It has been recently shown that many natural data do not follow the MVG temporal dependence model (e.g. Grimaldi and Serinaldi 2006, Salvadori *et al*. 2007, Schölzel and Friederichs 2008). It is therefore important to investigate the effect of deviation from the MVG dependence model on the distribution of τ.

The procedure for calculating the variance of Kendall's τ (e.g. Hamed 2009b) involves first calculating the cross-product autocorrelation function ρ_{ij}^{R} between the ranks of the observations, which corresponds in this case to Spearman's rho, then converting the autocorrelations to Normal autocorrelations using the relationship (Kendall 1955, Sec. 9.15):

$$\rho_{ij} = 2\sin\left(\frac{\pi}{6}\rho_{ij}^{R}\right) \tag{18}$$

However, as noted by the reviewer, this relationship holds only for MVG dependence. To investigate the extent of the effect of deviation from the MVG model on the distribution of τ, Monte Carlo simulation was used. For this investigation, three non-Gaussian copulas of the Archimedean family are considered. These are the Clayton copula, the Frank copula, and the Gumbel copula (Grimaldi and Serinaldi 2006, Schölzel and Friederichs 2008). The Archimedean family allows for a wide variety of dependence structures ranging from lower tail dependence (Clayton), through no tail dependence (Frank), to upper tail dependence (Gumbel), and seems to be well applicable to many problems in geosciences (Genest and Favre 2007, Schölzel and Friederichs 2008). First-order Markov chains from each of the three copulas with sample sizes of 50 and 100 observations were generated.

Figure 5 shows the ratio of the variance of τ calculated from 100 000 replicas of the data to the variance calculated assuming MVG dependence with the same value of first-order autocorrelation coefficient between the ranks (Spearman's rho). It can be seen in Fig. 5 that the variance may be overestimated or underestimated depending on the type of copula and the value of the first-order autocorrelation coefficient. For this range of sample sizes, the relative error ranges from -10% to $+16\%$, with the Gumbel copula being the least affected (–1% to $+4\%$). Although the relative error is not small for the Clayton copula, one should consider that as the sample size and autocorrelation coefficient increase, the variance inflation due to autocorrelation also increases rapidly (Hamed 2009b). For example, when $n = 100$ and $\rho = 0.8$, the variance inflation factor is 4.2 for the MVG case (4.8 for the Clayton case). Compared to this 420% variance inflation, the overall contribution of the $+16\%$ error in calculating the inflated variance, for example, becomes relatively small (around 3.8% only). This suggests that the assumption of MVG dependence may, in general, provide a reasonable approximation of the true variance of τ.

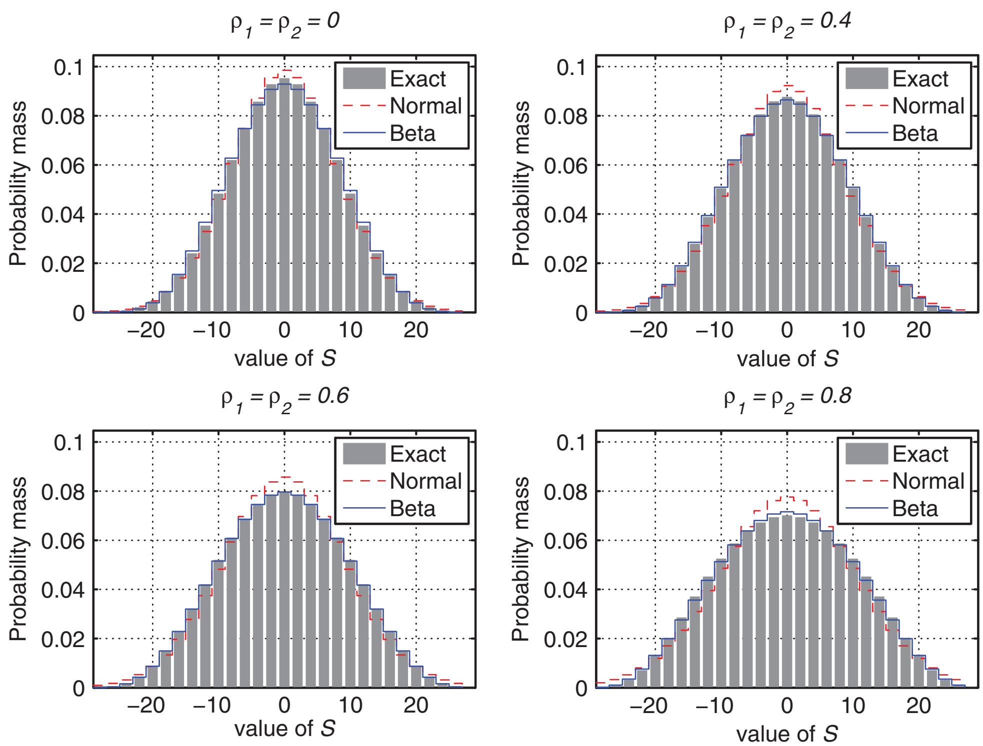

Fig. 3 Comparison of the exact probability mass function of S for sample size $n = 8$ with approximations using the normal and discretized beta distributions.

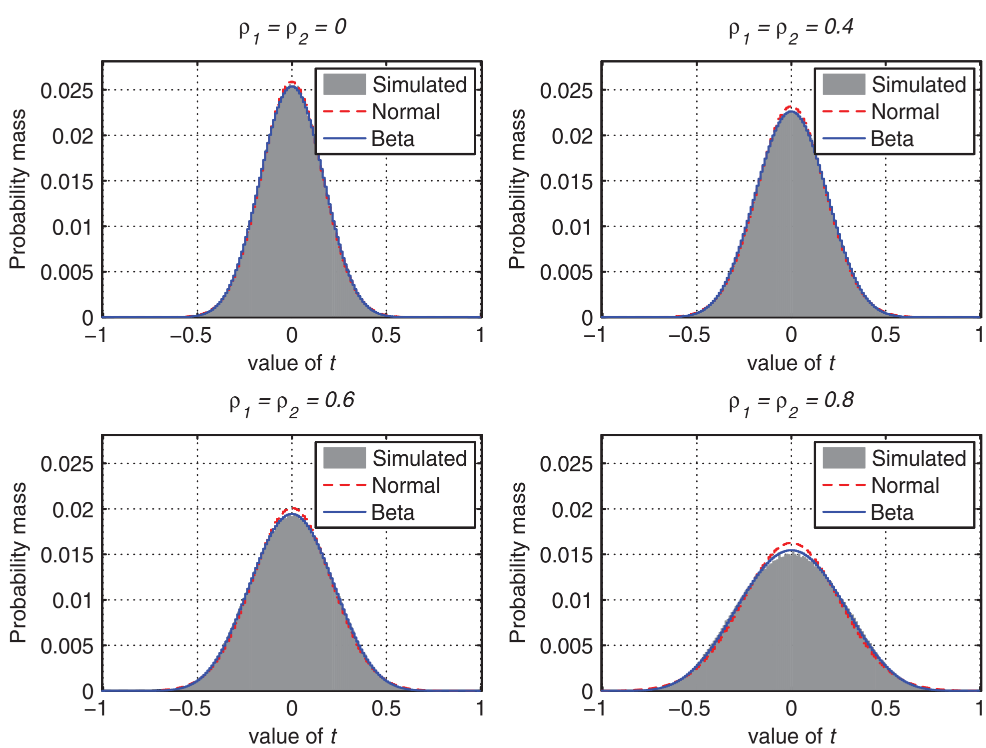

Fig. 4 Comparison of simulated (10^6 samples) probability mass function of t for sample size $n = 25$ with approximations using the normal and discretized beta distributions.

To further investigate this issue, the overall effect of assuming MVG dependence for Archimedean copulas on the results of testing the significance of autocorrelation between two persistent variables based on Kendall's τ is assessed. The false rejection rate results of testing the independence of 100 000

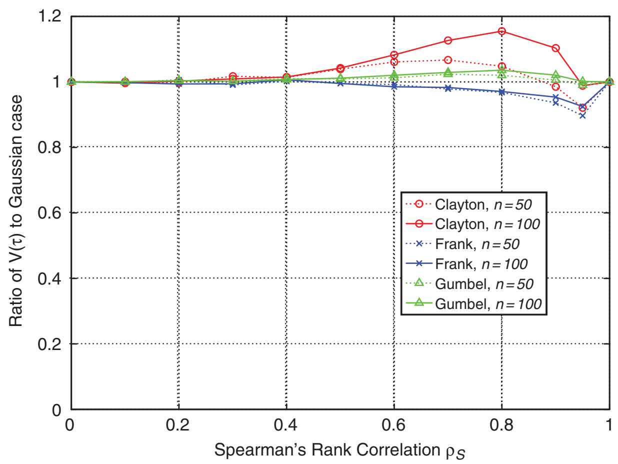

Fig. 5 Ratio of the actual variance of τ to the variance assuming MVG dependence for different values of the first-order cross-product autocorrelation between ranks and different copulas.

sets of first-order Markov chain data from the three considered Archimedean copulas at the 5% significance level are summarized in Table 2. It can be seen from Table 2 that, from a practical point of view, assuming MVG dependence as a working hypothesis results in a relatively very small error in the false rejection rates compared to the case of not making the correction at all.

Although the above analysis applies only to Archimedean copulas, the flexibility of the Archimedean family suggests that the effect of misspecification of the dependence structure is relatively small compared to the effect of dependence itself, especially if the deviation from multivariate Gaussianity is not large. Of course this does not eliminate the need for exploring other dependence structures, or attempting a more exact treatment by considering the statistical properties of copulas. However, such detailed analysis is beyond the scope of this paper.

CASE STUDY

To illustrate the importance of applying the correct distribution of Kendall's τ, we investigate the cross-correlation between the global North Hemisphere (NH) temperature record and tree-ring width (TRW) as well as maximum latewood density (MXD) proxies, which were used in a study by Mann *et al.* (2008) to reconstruct the global temperature record. The data was obtained online from http://www.meteo.psu.edu/~mann/supplements/MultiproxyMeans07/data. This presentation aims at illustrating how conclusions based on statistical evidence alone may be inconsistent and/or misleading if the correct distribution of Kendall's τ is not taken into account.

The temporal consistency and field significance (i.e., the joint significance of individual test results) of cross-correlation between 784 TRW series as well as 105 MXD time series and the corresponding global NH temperature record in the period 1850–1995 is studied. For proxy data, the Doornik-Hansen test of multivariate normality (Doornik and Hansen 1994, Trujillo-Ortiz *et al*. 2007a) and the Royston test (Royston 1982, Trujillo-Ortiz *et al*. 2007b) were applied to the normalized proxy data (by inverse normal transformation of rank-based plotting positions). The results of both tests indicate that multivariate normality of the TRW series as well as the MXD series cannot be rejected at the 5% level for several time lags. For temperature data, the two multivariate normality tests give conflicting results. On the other hand, it has been noticed that the temperature data shows stronger upper tail dependence, and that a Gumbel copula may fit the lag-1 data pairs reasonably well, as shown in Fig. 6. At any rate, Gumbel dependence, as discussed in the in the previous section, is the least affected by the misspecification of the dependence model, and therefore calculations based on MVG dependence can be expected to yield a satisfactory approximation.

Both the global temperature record and tree-ring data, among many other natural time series, have been shown to exhibit significant scaling in several studies in the literature (e.g. Koutsoyiannis and Montanari 2007, Hamed 2007). The scaling (or Hurst) parameter for temperature data was estimated at 0.91 using a maximum likelihood (ML) estimator developed by McLeod and Hipel (1978). For TRW data, the values of the scaling parameter ranged from 0.40 to 0.99. In both cases, the scaling parameters were estimated from the ranks of normalized detrended data in order to avoid the effect of possibly deterministic trends on the estimated value of the scaling coefficient. Although the observed trends may well be stochastic, it is intended to show that the results of this analysis strongly hold even if the values of the scaling coefficient were slightly underestimated, since it can be argued that large values of the scaling coefficient are due to the existence of a real trend. With these values of the scaling coefficients, the corresponding autocorrelation between data values results in the inflation of the variance of τ by a factor of 2.0 on average. Table 3 gives the number of significant positive as well as negative series, each at the 5% significance level for

Table 2 False rejection rates, at 5% significance, of the null hypothesis of independence of two variables. Each variable is a first-order Markov chain with MVG in addition to three Archimedean copula dependence structures. The rejection rates are given with and without variance correction, where correction assumes MVG dependence in all cases.

Sample size (n)	ρ_1	% False rejection rate (nominal value is 5%)							
		Without variance correction				With variance correction			
		Gauss	Clayton	Frank	Gumbel	Gauss	Clayton	Frank	Gumbel
50	0.1	5.1	5.3	5.3	5.2	4.9	5.1	5.1	5.0
	0.2	5.7	5.7	5.6	5.7	5.0	5.0	4.9	5.0
	0.3	6.7	6.7	6.5	6.6	5.0	5.1	4.9	5.0
	0.4	8.0	8.2	7.7	7.9	4.9	5.2	4.8	5.0
	0.5	9.7	10.3	9.7	9.9	5.0	5.5	5.0	5.1
	0.6	12.2	12.6	12.2	12.6	4.9	5.6	4.9	5.2
	0.7	16.0	16.1	15.5	15.9	5.0	5.7	4.8	5.2
	0.8	20.3	19.9	19.9	20.6	5.0	5.7	4.7	5.3
	0.9	26.7	25.2	25.6	26.3	5.0	5.2	4.4	5.1
100	0.1	5.1	5.1	5.1	5.2	5.0	4.9	5.0	5.0
	0.2	5.6	5.6	5.6	5.6	5.0	5.0	5.0	5.0
	0.3	6.5	6.7	6.5	6.6	5.0	5.2	5.0	5.1
	0.4	8.0	8.0	7.8	8.0	5.0	5.2	4.9	5.0
	0.5	9.9	10.5	9.8	10.2	5.0	5.5	4.9	5.2
	0.6	12.5	13.5	12.6	13.0	5.0	5.9	4.9	5.4
	0.7	16.3	17.4	16.2	16.6	5.0	6.3	4.8	5.3
	0.8	21.3	22.2	21.0	21.7	5.1	6.5	4.7	5.5
	0.9	28.3	27.7	27.9	28.2	5.0	6.2	4.6	5.2

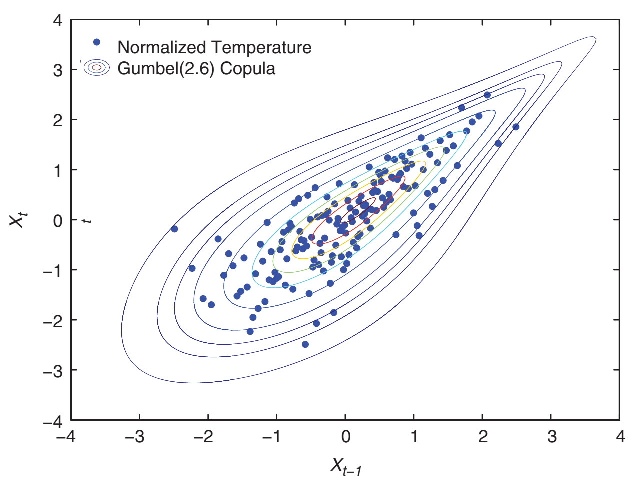

Fig. 6 Normalized temperature data fit by a Gumbel copula.

the recent (1923–1995) and early (1850–1922) halves of the record, each being 73 years long. It should be noted that the significance level of any statistical test represents the proportion of the tested cases expected to be found significant only due to chance. Thus, for the detected cross-correlations to be field significant, the number of significant cross-correlations should be significantly larger than the expected number due to chance, which is commonly known as the counting approach (Wilks 2006). The global null hypothesis is thus rejected if the probability of having obtained the observed number of locally significant series, or more, is no larger than a chosen global test level, which we choose to be equal to the local significance level of 5%. This probability, which we will call the field p value, can be readily calculated by noting that the number of significant cases is distributed as binomial with $n = 784$ and $p = 0.05$ (Wilks 2006). Field p values for the studied series are shown in parentheses in Table 3.

Considering the results for the late half of the record in Table 3, it is clear that the number of tree-ring series that are negatively correlated with the late half of the NH temperature series is not field significant under both the original and modified tests. However, significant positive cross-correlation is found in 202 series under the original test, which drops to 103 when accounting for persistence, which is still highly field significant at the 5% field significance level, with a p value approaching zero.

The picture is, however, different for the early half of the data in Table 3. Using the original test, negatively cross-correlated series seem to have much higher field significance (p value approaching zero) than positively cross-correlated ones (p value $=$ 0.050, only marginally field significant). Given that this is the same group of series, such a result indicates temporal inconsistency of the cross-correlation. Going back to the detailed results, it turns out that, out

Table 3 Numbers of 784 tree-ring width (TRW) series that show significant positive or negative cross-correlation with the NH temperature series at the 5% significance level. Field p values are given in parentheses, and field significant results are indicated in bold face.

Test	Data set	Significant positive	Significant negative
Original	Early (1850–1922)	**50 (0.050)**	**111 (0.000)**
	Late (1923–1995)	**202 (0.000)**	37 (0.664)
Modified	Early (1850–1922)	20 (1.000)	37 (0.664)
	Late (1923–1995)	**103 (0.000)**	14 (1.000)

of the 784 tested TRW series, 373 series (more than 47%) had changed from negative cross-correlation in the early half to positive cross-correlation in the recent half. Of these 373 series, 142 series were found significant by the original test in the late half. In other words, around 70% (142 out of 202) of the series declared significantly positively cross-correlated with the late half of the temperature series using the original test had negative cross-correlation with the early half, which is an obvious temporal inconsistency. Furthermore, out of these 142 series, 20 series changed from significant negative cross-correlation with the early half of the temperature record to significant positive cross-correlation with the late half when the original test is used, which is again an anomalous result.

In contrast, the modified test indicates no field significant positive or negative cross-correlation in the early half of the data, while it indicates field significant positive cross-correlation in 103 series (p value approaching zero) in the late half. Furthermore, the number of time series changing from significant negative to significant positive cross-correlations drops from 20 series to only 5 series, but we note that negative cross-correlation in the early half is not field significant at the 5% level in this case.

Table 4 presents the results for 105 MXD time series. The MXD data have scaling coefficient values ranging from 0.46 to 0.84, which result in an average variance inflation of the test statistic of 1.35. We note in Table 4 the absence of temporal inconsistency between the early and late halves, which is consistent with the findings of many studies indicating that MXD series are more powerful proxies of temperature than TRW (e.g. Grudd 2008). However, we also note that positively cross-correlated series using the modified test in the late half (nine series) become field insignificant at the 5% significance level (p value $= 0.08$) after accounting for persistence in the data.

In the above analysis, no account was made for the effect of spatial correlation between the tested tree-ring series. Similar to the effect of positive autocorrelation, spatial correlation between tested series is known to result in a more liberal test, i.e. the null hypothesis of independence will tend to be rejected more frequently than it should be, falsely implying higher field significance (Douglas *et al*. 2000). A number of methods have been suggested in the literature to take the effect of spatial correlation into account (e.g. Douglas *et al*. 2000, Yue and Wang 2002). Most of these methods involve the estimation of the empirical distribution of a regional test statistic using Monte Carlo simulation of a similar size region with the same spatial correlation pattern or by re-sampling techniques. Wilks (2006) suggests two simpler procedures to test field significance that are more robust than the counting approach in the presence of spatial correlation. These are the Walker test and the False Discovery Rate (FDR) test. The Walker test compares the smallest local p value with the critical value p_{Walker} given by:

$$p_{\text{Walker}} = 1 - (1 - \alpha_{\text{Global}})^{1/K} \tag{19}$$

Table 4 Numbers of 105 tree-ring maximum latewood density (MXD) series that show significant positive or negative cross-correlation with the NH temperature series at the 5% significance level. Field p values are given in parentheses, and field significant results are indicated in bold face.

Test	Data set	Significant positive	Significant negative
Original	Early (1850–1922)	**25 (0.000)**	3 (0.901)
	Late (1923–1995)	**13 (0.002)**	1 (0.995)
Modified	Early (1850–1922)	**15 (0.000)**	0 (1.000)
	Late (1923–1995)	9 (0.080)	1 (0.995)

where α_{Global} is the global significance level and K is the number of tested series. Alternatively, the field p value corresponding to the smallest local p value $p_{(1)}$ can be calculated by inverting equation (19) to give:

$$\alpha_{\mathrm{Field}} = 1 - (1 - p_{(1)})^K \tag{20}$$

The FDR test, on the other hand, considers not only the smallest local p value, but all local p values by establishing a sliding scale for smallness of the local p values that depends on their placement among the full sorted collection of local p values (Wilks 2006).

Tables 5 and 6 give the field p values for TRW and MXD series, respectively, that are locally significantly positively or negatively cross-correlated with the NH temperature record, as given by the Walker test based on equation (20). According to the results in Table 5, only the positive cross-correlation in the late half of the TRW series is field significant when the original test is used. On the other hand, no field significant cross-correlation is detected for TRW series in either half when the modified test is used. Similarly, in Table 6, only the negative cross-correlation in the late half of MXD series is field significant under the original test. Again, no field significant cross-correlation is detected for MXD series in either half when the modified test is used. Identical results for both TRW and MXD series were also obtained by applying the FDR test. These results indicate that spatial correlation is suspected for inflating the field significance results obtained by the counting approach.

Thus, by accounting for the effect of autocorrelation, the modified test removed the apparent inconsistency that field significant positive cross-correlation prevailed in the late half of the TRW record, while significant negative cross-correlation prevailed in the early half. However, the detailed results of the tests on TRW and MXD series, as well as the results of the modified test after accounting for spatial correlation between the data, raise serious questions about the suitability of tree-ring time series as global temperature proxies, and whether the relationship between global temperature and tree-ring data is actually unique.

SUMMARY AND CONCLUSIONS

The full distribution of Kendall's τ in the case of persistent data has been derived and an approximation for moderate sample sizes has been suggested. It was shown that persistence causes the distribution of τ to have heavier tails thus increasing the chance of falsely rejecting the null hypothesis of independence in favour of cross-correlation. Using the modified distribution in significance testing eliminates that effect. Although the derived distribution is based on the assumption of multivariate Gaussian (MVG) dependence between the observations in each of the two involved variables, it was shown that the error resulting from the misspecification of the dependence model is relatively small for the case of the flexible Archimedean copula family. However, it is interesting to note that the variance of Kendall's τ depends on the type of the dependence structure of the data for the same autocorrelation values between the ranks, which does not fit with the classical "distribution-free" or "nonparametric" labels of rank-based tests. However, this aspect needs further detailed investigation.

Table 5 Results of the Walker test for TRW data. Given are the field p values corresponding to the minimum local p values. Field significant results at the 5% level are indicated in bold face.

Test	Data set	Positive	Negative
Original	Early (1850–1922)	0.549	0.058
	Late (1923–1995)	**0.002**	0.056
Modified	Early (1850–1922)	0.967	0.945
	Late (1923–1995)	0.312	0.999

Table 6 Results of the Walker test for MXD data. Given are the field p values corresponding to the minimum local p values. Field significant results at the 5% level are indicated in bold face.

Test	Data set	Positive	Negative
Original	Early (1850–1922)	0.119	0.900
	Late (1923–1995)	0.069	**0.042**
Modified	Early (1850–1922)	0.195	1.000
	Late (1923–1995)	0.504	0.584

A case study has been presented where the effects of persistence were illustrated and the merits of using the modified distribution were demonstrated. The significance of cross-correlation between NH temperature data and tree-ring proxies has been investigated under the modified distribution of Kendall's τ and considering field significance of the test results. It has been shown that most of the observed cross-correlations are probably not field significant. The application of the modified test using the correct distribution of Kendall's τ is therefore recommended in order to avoid erroneous and/or anomalous conclusions about the significance of cross-correlation between natural data that exhibit significant persistence.

Acknowledgements The author is deeply grateful to the Co-Editor, Demetris Koutsoyiannis, Francesco Serinaldi and another anonymous reviewer for their helpful comments and discussion that resulted in this improved manuscript.

REFERENCES

Chiew, F. H. S., Piechota, T. C., Dracup J. A., and McMahon T. A., 1998. E1 Nino/Southern Oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting. *J. Hydrol*., 204, 138–149

Doornik, J. A., and Hansen, H., 1994. An omnibus test for univariate and multivariate normality. Oxford: Nuffield College, Discussion Paper W4&91.

Douglas, E. M., Vogel, R. M., and Kroll, C. N., 2000. Trends in flood and low flows in the United States: impact of spatial correlation. *J. Hydrol*., 240, 90–105.

Genest, C., and Favre, A. C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *J. Hydrol. Eng*., (124), 347–368.

Genz, A., and Bretz, F., 1999. Numerical computation of multivariate τ probabilities with application to power calculation of multiple contrasts. *J. Statist. Comput. Simul*., 63, 361–378.

Genz, A., and Bretz, F., 2002. Comparison of methods for the computation of multivariate τ probabilities. *J. Comp. Graph. Stat*., 11 (4), 950–971.

Ghanbari, R. N., and Bravo, H. R., 2008. Coherence between atmospheric teleconnections, Great Lakes water levels, and regional climate. *Adv. Water Resour*., 31, 1284–1298.

Grimaldi, S., and Serinaldi, F., 2006. Design hyetograph analysis with 3-copula function. *Hydrol. Sci. J*., 51 (2), 223–238.

Grudd, H., 2008. Torneträsk tree-ring width and density AD 500–2004: a test of climatic sensitivity and a new 1500-year reconstruction of north Fennoscandian summers. *Clim. Dyn*., 31, 843–857.

Hamed, K. H., 2007. Improved finite-sample Hurst exponent estimates using rescaled range analysis. *Water Resour. Res*., 43, W04413, doi:10.1029/2006WR005111.

Hamed, K. H., 2008. Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *J. Hydrol*., 349 (3–4), 350–363.

Hamed, K. H., 2009a. Exact distribution of the Mann-Kendall trend test statistic for persistent data. *J. Hydrol*., 365 86–94.

Hamed, K. H., 2009b. Effect of persistence on the significance of Kendall's tau as a measure of correlation between natural time series. *Eur. Phys. J. Special Topics*, 174, 65–79.

Kendall, M. G., 1955. *Rank Correlation Methods*. London: Griffin.

Kendall, M. G., and Gibbons, J. D., 1990. *Rank Correlation Methods*. Fifth edn. London: Griffin.

Kendall, M. G., and Stuart, A., 1976. *The Advanced Theory of Statistics*, vol. I: *Distribution Theory*. London: Griffin.

Koutsoyiannis, D., 2003. Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrol. Sci. J*. 48 (1), 3–24.

Koutsoyiannis, D., and Montanari, A., 2007. Statistical analysis of hydroclimatic time series: Uncertainty and insights *Water Resour. Res*. 43, W05429.

Mann, H. B., 1945. Nonparametric tests against trend. *Econometrica*, 13, 245–259.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F., 2008. Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *PNAS*, 105 (36), 13252–13257.

McLeod, A. I, and Hipel, K. W., 1978. Preservation of the rescaled adjusted range, 1. A reassessment of the Hurst phenomenon. *Water Resour. Res*., 14 (3), 491–508.

Panarello, H. O., and Dapeña, C., 2009. Large scale meteorological phenomena, ENSO and ITCZ, define the Parana River isotope composition, *J. Hydrol*., 365, 105–112

Royston, J. P. 1982. An extension of Shapiro and Wilk's W test for normality to large samples. *Appl. Statist*. 31 (2), 115–124.

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L., 1980. *Applied Modeling of Hydrologic Time Series*. Littleton, CO: Water Resources Publications, first edition.

Salvadori, G., De Michele, C., Kottegoda, N. T., and Rosso, R. 2007. *Extremes in Nature: An Approach Using Copulas*, Dordrecht: Springer.

Schölzel, C., and Friederichs, P., 2008. Multivariate non-normally distributed random variables in climate research—introduction to the copula approach. *Nonlinear Processes Geophys*., 15, 761–772.

Soukup, T. L., Aziz, O. A., Tootle, G. A., Piechota, T. C., and Wulff, S. S., 2009. Long lead-time streamflow forecasting of the North Platte River incorporating oceanic–atmospheric climate variability. *J. Hydrol*., 368, 131–142.

Trujillo-Ortiz, A., Hernandez-Walls, R., Barba-Rojo, K., and Cupul-Magana, L., 2007a. DorHanomunortest: Doornik-Hansen Omnibus Multivariate (Univariate) Normality Test, A MATLAB File [online], Available from: www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=17530.

Trujillo-Ortiz, A., Hernandez-Walls, R., Barba-Rojo, K., and Cupul-Magana, L., 2007b. Roystest: Royston's Multivariate Normality Test, A MATLAB file. [online]. Available from: http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=17811.

Wilks, D. S., 2006. On field significance and the false discovery rate. *J. Appl. Met. Climatol*., 45, 1181–1189.

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

Yue, S., Pilon, P., and Cavadias, G., 2002. Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *J. Hydrol*., 259, 254–271.