Effects of the prewhitening method, the time granularity and the time segmentation on the Mann-Kendall trend detection and the associated Sen's slope

By Martine Collaud Coen, Elisabeth Andrews, Alessandro Bigi, Gonzague Romanens, Giovanni Martucci and Laurent Vuilleumier

Submitted to Atmospheric Measurement Techniques
amt-2020-178

REFEREE REPORT

General comments

The authors propose a new algorithm of trend analysis on autocorrelated meteorological data via incorporating the merit of three prewhitening techniques. The effect of time granularity, time segmentation and time series length on trend analysis are also evaluated on the basis of real meteorological observations.

The proposed algorithm is a good trial to purse the ideal goal of trend detection methods, that is high power with controllable Type I error, and accurate slope estimates. I think this algorithm is practically sound.

But still I have reservations about some statements in this paper. The manuscript and the quality of figures should be improved before it is formally published.

Specific comments

1. Line 120-122. "*These approaches (variance correction approaches) appear not able to preserve the significance level and the power of the MK-test in the case of correlated time series with a trend*"

Comment: Both the variance correction approach and the prewhitening approach can preserve the pre-assigned significance level when there is no trend. Because detecting trends with known statistical confidence is the primary goal of trend analysis, either on independent data or autocorrelated data. The power of trend identification may be different for distinct methods.

2. Line 139-140: *it (PW method) reduces the power of the test due to an over-/underestimation of* ak_1^{data} *in the case of a positive/negative trend.*

Comment: The existence of real trend, either positive or negative, can lead to an overestimation of lag-1 autocorrelation coefficient.

3. Comment on Line 151-164: The brief introduction on the TFPW-WS method (Wang and Swail's 2001) includes some mistakes. I suggest rephrase this paragraph.

The original idea of Wang and Swail's (2001) was intended to implement the MK test on the prewhitened series, rather than on the prewhitened detrended series, as it was given by Eq.(8). If the prewhitened series are detrended, then we will never identify any trends.

The critical value to stop iteration should be a tiny number, e.g. 0.0001, instead of 0.05.

The primary consideration of iteration procedure was to mitigate the adverse effect of trend on the accuracy of lag-1 autocorrelation coefficient estimate.

4. Line 164-167: *The PW-cor method refers to the preliminary step of the first iteration in the TFPW-WS method and consequently corrects the prewhitened data by the same factor. To the knowledge of the authors, this PW-cor method is not referenced in the literature but is a potential method tested in this study.*

Comment: After rephrasing the TFPW-WS method, please describe the PW-cor method more clearly.

5. Line 184-185: *VCTFPW preserves to some extent the power of the test, but only mitigates the type 1 errors.*

Comment: Similar to other prewhitening methods, the VCTFPW method mitigates the inflationary type 1 errors raised by autocorrelation as its priority. Then the method preserves the power of the trend test to some extent.

6. Line 203-204: *If PW is ss but TFPW-Y is not, then the trend is considered as a false negative due to the lower test power of PW and the trend has to be considered as ss.*

Comment: If we consider the trend to be statistically significant, then we cannot say the detected trend is a false negative result. It is illogical to report a trend and meanwhile state this is an error.

Figure 1 should be revised accordingly.

7. Line 225-228: *Trend analyses were applied on several periods. For all the data sets,* ***a 10-year period*** *is considered first and then further possible* ***multi-decadal periods up to 60 y*** *for the radio-sounding time series. For the in-situ aerosol properties,* ***tests with 4 to 9 y periods*** *are also computed in order to illustrate the problems of trend analysis on very short time series.*

Line 781-782: *lag-1 autocorrelation of the observations (ak1data) and number of ss partial autocorrelations* ***for the 10y period (order)****, number of data in the* ***10y period*** *(N) and reference.*

Comment: I think in this section "3 Experimental", the authors should clarify how to analyze the measure data, in order to support the coming results. The meaning of "*a 10-year period*" or "*multi-decadal periods up to 60 y*" are unclear and obscure.

In table 2, the meaning of "*number of ss partial autocorrelations* ***for the 10y period (order)****, number of data in the* ***10y period*** *(N) and reference"* is unclear either.

8. Line 275-276: *CL of MK, PW and TFPW-Y, which remove the lag-1 autocorrelation without compensation for the* ***mean values and the variances****…*

Comment: Does that mean "mean and variances of the slope estimate"?

9. Line 278-280: *The* ***ss*** *often decreases for coarser time granularities occasionally leading to not* ***ss trends*** *for some of the prewhitening methods. PW, TFPW-WS and VCTFPW methods become* ***not ss*** *at finer time granularities than TFPW-Y and MK due to their lower number of false positives.*

Comment: It's hard to identify the relationship between the significance of trend and the time granularity from Fig. 2.

10. Line 281-282: *The discrepancies between prewhitening methods are larger than the discrepancies that occur when different temporal segmentations (months or meteorological seasons) are applied.*

Comment: Fig.2 does not support this finding.

11. Line 284-285: *the similarity of MK slopes with TFPW slopes*.

Line 350-352: *Due to the detrending procedure, the absolute values of the TFPW-Y* *slope are larger than the PW slopes and similar to the MK slope values (Fig. 2), even if a tendency to have larger TFPWY than MK slopes are observed.*

Line 367: *TFPW-Y slopes tend to be larger than MK slopes (Fig. 4b), with larger differences at high ak1 data leads.*

Comment: If my understanding is right, the MK and TFPW-Y should yield exactly the same slope of trend. The MK test does not estimate the slope of trend directly. It usually reports the magnitude of trend by the use of Sen's slope. The TFPW-Y also estimates Sen's slope as its first step. It will reinstall this trend to the prewhitened series without any modification. So these two slopes should be equal to each other.

12. Line 285-287: *For example, the number of data points in the AOD time series (about 65 per year) induces higher CL for time granularities finer than the measurement frequency (about 10 days).*

Line 372-373: *Removing the lag-1 autocorrelation increases the variance, but decreases the mean.*

Line 391-395: *The spread of the slopes of the aerosol number concentration for the one-year aggregation on Fig. 2c shows that the yearly data still have a ss ak1 data for the longest periods of 20 and 24 years (see similar cases in Fig. 2). For shorter periods (5 to 9 years), the ak1 data decreases rapidly for averaging longer than 10 days and even becomes negative for yearly averages.*

Comment: These sentences are difficult to be understood. Please rephrase.

13. Line 294-296: *The yearly trend was computed for all periods (from 5y to 24y) at all considered time granularities (1 day to 1 month for the meteorological season temporal segmentation), leading to* ***40 trends****.*

Comment: Please clarify what is the 40 trends?

14. Line 323-325: *PW is used as the reference for false positives because it is the prewhitening method with the lowest type 1 error, while TFPW-Y is the reference for false negatives because it is the most powerful test.*

Comment: It's inappropriate to state that the TFPW-Y is the most powerful test. The TFPW-Y tends to report significant trends at the expense of committing high type 1 error. This finding has been verified by many literatures. So we can say the TFPW-Y tends to identify significant trends more frequently than other methods, but we cannot say it is the most powerful test.

15. Line 338-342: *For the time series considered in this study, the following conclusions can be made: 1) PW performs very well with an almost vanishingly small (⩽0.3%) number of false negatives and the ss of PW-cor is similar to that for PW; 3) VCTFPW has a very high type 1 and 2 errors and should consequently not be used to determine the ss; and 4) it is not possible to determine whether MK or TFPW-Y is the most powerful method.*

Line 786-788: *Table 3*

Comment: The three conclusions made here do not align with the consensus about the prewhitening method among the community. I suggest to recheck the results.

1) The PW tends to overestimate the lag-one autocorrelation coefficient without trend removal, see Hamed (2009). In addition, the PW reduces a portion of real trend, see Yue and Wang (2002). That's the reason why Yue et al. (2002) suggest to remove trend before whitening. So if the TFPW-Y is the reference for false negatives, the PW is less likely to miss only 0.2% significant trends.

3) As it was stated by the authors, e.g. Line 265-266, Table 1, Figure 4(a). The VCTFPW slopes lies between the TFPW and the PW slope values. So no matter one takes the PW or the TFPW-Y as the reference, the VCTFPW is less likely to commit the highest error among all the prewhitening methods.

4) For the autocorrelated data, the MK and TFPW-Y are not really powerful method. They only tend to report significant trends more frequently than other PW methods. However, both of them commit high type I error as a price.

I have to say, the above opinions are given by Monte-Carlo simulation results. They may not suitable to every real-world series. This study deals with measured data. So I suggest to recheck your results again.

16. Line 345: *The slope of the trend is always enhanced by the positive ak1data*.

Comment: I think it should be "the slope estimates of the trend is influenced by the positive lag-one autocorrelation". The autocorrelation increases the difficulty of an accurate slope estimation. But it does not increase or decrease the real slope of the trend.

17. Line 628-629: *Consistent with the literature, the use of MK, TFPW-Y and*

VCTFPW results in a large amount of false positive results while TFPW-WS results in less than 2% of false positives.

Comment: After recheck your results, e.g. table 3, this conclusion should be revised accordingly.

18. Line 637: *The confidence limits are much broader for coarser time granularities and the ss is lower.*

Comment: Fig. 8 supports this conclusion but Fig. 10 does not. As the time granularity becomes coarser, the confidence limits are much narrower in Fig. 10.

19. Comment on Figure 2: it is hard to distinguish the time segmentation.

20. Comment on Figure 7: it is not easy to identify different PW methods.

21. Comment on Figure 8 and 10: it is unclear how to analyze the slope of trend as well as the confidence limit within each time segmentation. It should be well explained.

22. I suggest to improve the quality of the figures, to make them self-explaining.

Technical corrections

Line 109. Zwang and Zwiers (2004) does not given in the reference list.

Line 168. I think the correct citation about the VCTFPW method should be "Wang, W., et al., 2015. Variance correction pre-whitening method for trend detection in auto-correlated data. Journal of Hydrologic Engineering, 04015033. doi:10.1061/(ASCE)HE.1943-5584.0001234."

Line 272: "*aerosol absorption coefficient*" should be "aerosol scattering coefficient".

Fig 2 Caption: “*Scattering coefficient, 10y*” and “Tropopause level, 50y”. Should it be “24y” and “60y” ?

References

Hamed, K.H., 2009. Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. Journal of Hydrology, 368(1-4): 143-155.

Yue, S., Pilon, P., Phinney, B., Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. Hydrological Processes, 16(9): 1807-1829.

Yue, S., Wang, C.Y., 2002. Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test. Water Resources Research, 38(6): 1068.