

## Trend analysis in the presence of short- and long-range correlations with application to regional warming

Ewan T. Phillips <sup>1,\*</sup>, Marc Höll <sup>1</sup>, Holger Kantz <sup>1</sup> and Yu Zhou <sup>1,2,†</sup>

<sup>1</sup>Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany

<sup>2</sup>Institute for Global Innovation and Development and School of Urban & Regional Science, East China Normal University, Shanghai 200062, China



(Received 17 May 2023; accepted 21 July 2023; published 5 September 2023)

Many real-world time series exhibit both significant short- and long-range temporal correlations. Such correlations enhance the errors of linear trend analysis. In this paper, we provide a general framework for trend analysis under the consideration of such correlations. We propose a parsimonious model containing both a single short-range autoregressive parameter and long-range fractional parameter. We derive analytical closed-form results for the error bars of the least-squares estimate of the trend for such time series, highlighting the different effects of short- and of long-range correlations. We employ an ensemble method for the automated extraction of scaling regions to estimate the fractional parameter of the data model together with its error bar, and the Grünwald-Letnikov derivative for the identification of the autoregressive parameter. We apply this framework to the study of warming trends on gridded temperature data in central Europe. We make use of the redundancy of the trend signal in adjacent grid points using methods of spatial averaging and the first principal component of empirical orthogonal function analysis. We find good agreement between the results of these two methods. We find a statistically significant decadal warming trend in central Europe over the past 70 years, which shows a particularly dramatic increase over the past 20 years.

DOI: [10.1103/PhysRevE.108.034301](https://doi.org/10.1103/PhysRevE.108.034301)

### I. INTRODUCTION

Linear trend analysis plays a prominent role in geoscience [1], where it is used to quantify the temperature increases relevant to global warming [2–9], trends in stream flow [10–13], and precipitation [14–16], among other things [17]. It also has applications in many other fields such as medicine, where trend analysis is used to study mortality rates [18], and cognitive neuroscience [19], where linear modeling has become increasingly ubiquitous, being used, for example, to model neural responses to speech and language [20,21].

Typically, in trend analysis the time series is considered a (weak) trend plus a (large amplitude) fluctuation noise term. The statistical significance of an observed trend depends crucially on the null hypothesis, which in turn reflects assumptions about the nature of the underlying system. If the noise, for example, has a persistent nature, which is not correctly identified in the null hypothesis, then persistent patterns in the noise may be incorrectly interpreted as trends. Such correlations in the noise when correctly taken into account thus generally weaken the trend significance [22]. It is, for

example, a challenging problem to determine to what extent global warming may be a consequence of an external forcing such as increased CO<sub>2</sub> levels, and to what extent it is an intrinsic feature of the fluctuations themselves.

Most often this noise is assumed to be a simple kind of *short-range correlation* (SRC) process with a converging autocorrelation time [23,24]. However, much evidence has emerged that geophysical time series generally exhibit *long-range correlations* (LRCs), having a diverging autocorrelation time [25,26]. LRC has been widely reported in time series of surface temperature records [27–29] and is well described by a fractional Gaussian process [30,31]. The effects of LRC on trend detection has been considered, for example, in Refs. [32–39].

Some recent studies have suggested that temperature anomalies, i.e., daily deviations of temperature from its climatological seasonal cycle, exhibit both *short- and long-range correlations* (SLRCs) and that a combination of both an autoregressive and a fractional parameter is most appropriate for describing such anomalies [40,41]. An analytical treatment of trend estimates for such a series has been an open problem. Such a treatment is desirable, since in this way significantly more accurate statements can be made about the statistical likelihood of the trend compared to when using Monte Carlo methods. We present here analytically derived expressions for the uncertainty in trend estimation in the presence of SLRC which show how both correlation types contribute to estimation errors. Our results reproduce the already known results for SRC and for pure LRC as special cases. Further, we provide robust methods for obtaining the short- and long-range parameters.

\*tphillips@pks.mpg.de

†Corresponding author: zhouyu@pks.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

Generally, in climate science interest is not only in the trend of a single time series, but more generally the trend of a spatial region [42]. Thus, beyond the considerations of temporal correlations in the noise, considerations of spatial correlations should also be taken into account. The use of spatially distributed data allows for more rigorous results. The analysis of such data is, however, nontrivial, since neighboring time series are highly correlated with each other, thus the information gained from each new station is less than it would be if the stations were independent.

A straightforward approach to spatially distributed data is to create an averaged signal by performing a spatial average of the local time series. For areas, which are not too large, however, we will argue that a more natural way to obtain the overall regional trend is to use *empirical orthogonal functions* (EOFs), commonly also referred to as *principal component analysis* (PCA) [43,44]. EOF analysis offers an objective way to reduce noise in the data by filtering out regional anomalies that distort the overall trend (for example, mountain ranges) as higher order components [45]. EOFs being based on the spatial covariance structure of the data are better for large-scale analysis than analytical correlation-versus-distance functions as these cannot take account of geographically varying teleconnections [46]. Both methods of averaging increase the statistical significance of the trend, since they reduce the fluctuations whilst at the same time conserving the trend.

This paper is organized as follows. After introducing the trend model and the notion of short- and of long-range correlated processes, we provide a rigorous argumentation for the existence of SLRC in a time series of station temperature anomalies and show that an auto-regressive fractionally integrated moving average model, i.e., ARFIMA(1, $d$ ,0), is a sufficient model as well as how to estimate the parameters of this model. We propose a method to obtain the Hurst exponent  $H$ , where  $d = H - 1/2$  by combining detrended fluctuation analysis (DFA) with a recently introduced method of automated extraction, which provides values for the uncertainty of  $d$ , and then proceed to derive the autoregressive parameter  $\phi$  by first removing the LRCs. In the following section, we provide an analytical formula for the variance of the least-squares trend estimator for finite samples of an ARFIMA(1, $d$ ,0) process. With this, both the  $p$  value and the confidence interval of an observed trend can be easily obtained for any SLRC time series. In the second half of this paper, we turn our attention to spatially distributed time series. We show that while the trends of the grid points are statistically likely for the estimated values of  $d$ , this result depends very sensitively on  $d$ . To obtain a single value for the spatial region, we compare the results for both the spatially averaged time series and the first principal component (PC), using a variation of EOF analysis. We then estimate the warming trend of central Europe and its statistical significance, showing that while for some grid points the warming trend is insignificant, the regional average is nevertheless significant.

## II. UNCERTAINTY OF TREND ESTIMATES IN SLRC TIME SERIES

The key idea of linear trend analysis is that the time series  $T_t$  can first be preprocessed to remove periodic variations and

that the resulting anomalies can then be partitioned into a deterministic linear trend component and a fluctuating component such that

$$T_t = n + mt + \epsilon_t, \quad (1)$$

where  $n$  is the offset,  $m$  is the trend coefficient, and  $\epsilon_t$  represents the noise, which is assumed to be a stationary process of zero mean.

Most commonly, when SRC is considered, the fluctuations are assumed to be of the form of a stochastic autoregressive AR(1) process,

$$\epsilon_{t+1} = \phi \epsilon_t + \xi_t, \quad (2)$$

where  $\phi \in (-1, 1)$  is the autoregressive parameter and  $\xi_t$  is Gaussian white noise (WN). Clearly, when  $\phi = 0$  this process corresponds to discrete WN. The AR(1) process can also be written with the backshift operator  $B$  as  $(1 - \phi B)\epsilon_t = \xi_t$  [23]. Such noise has an autocorrelation, which decays exponentially with a decay time  $\tau = -1/\ln \phi$ . Notice that the variance of the AR(1) process depends on the parameter  $\phi$ , so we later use an additional normalization.

When, on the other hand, an LRC process is considered, the standard one-parameter time-discrete model is the ARFIMA(0, $d$ ,0) process [47,48], which is a time discrete version of fractional Gaussian noise [49]. The ARFIMA process can be written as  $\epsilon_t = (1 - B)^{-d}\xi_t$ , where  $d \in (-0.5, 0.5)$  and  $(1 - B)^{-d}$  is known as the time-discrete fractional differencing operator [48,50], related to the Hurst exponent as  $d = H - 0.5$ . For the purpose of generating numerical time series data, it is more appropriate to consider the expansion [23]

$$\epsilon_t = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} B^j \xi_t, \quad (3)$$

where in practice one truncates the sum over  $j$  at a large but finite number. Such noise  $\epsilon_t$  has an autocorrelation, which decays asymptotically for large time lags  $\tau$  as a power law  $c(\tau) \simeq \tau^{-\gamma}$  with the power  $\gamma = 1 - 2d$ .

In discrete time, we consider SLRC to be a combination of both these effects, known as an ARFIMA(1, $d$ ,0) process [48], given by replacing the WN  $\xi_t$  in Eq. (2) by the power-law correlated noise on the right-hand side of Eq. (3):

$$\epsilon_t = \phi \epsilon_{t-1} + \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \xi_{t-j}. \quad (4)$$

This is the most parsimonious model of SLRC noise, containing only two parameters  $\phi$  and  $d$  corresponding to SRC and LRC, respectively.

### A. An ARFIMA(1, $d$ ,0) model for temperature anomalies

We now demonstrate both the existence and implications of an SLRC process in a daily 2 meter above ground air temperature time series using the Potsdam station in Germany [51] as an example. As common in meteorological data analysis, we first remove the periodicity of the annual cycle by subtracting from every value of the time series the averaged value for the given calendar day of the year to obtain the anomalies. For example, from the temperature at Jan. 1, 1970, we subtract

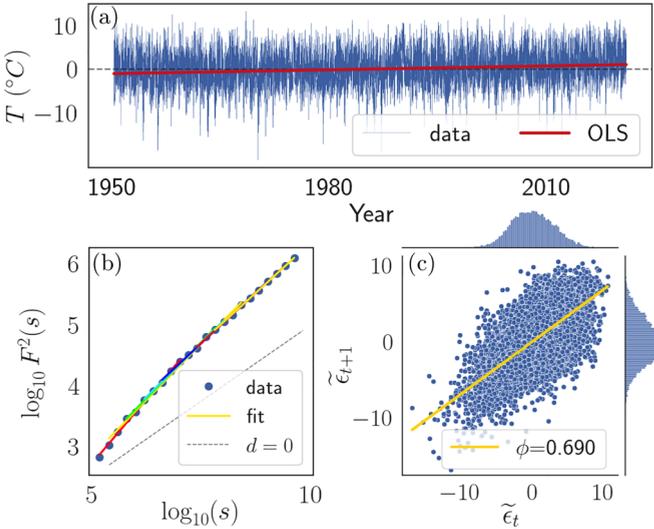


FIG. 1. (a) Anomalies of Potsdam temperature time series with the OLS fit of the trend. (b) DFA2 method to obtain parameter  $d$ . Some of the different fitted power laws are shown (each with a different color). (c) Plot to obtain  $\phi$  using a linear fit.

the average temperature of Jan. 1, averaged over all the years of our time series. It has been verified that another method involving fitting a sinusoidal function to the data series as the annual cycle and subtracting this yields anomalies with the same statistical properties. The resulting time series is then assumed to be of the form of Eq. (1). The anomalies are to a good approximation Gaussian distributed [40].

The estimated trend  $\hat{m}$  is then obtained by fitting a linear model to the data using ordinary least squares (OLS) [52,53], which is justified, since noise has been shown to be Gaussian [40] in good approximation, and thus the estimators are also Gaussian, since it relies on a linear combination. We note that even for the case that the noise is not Gaussian, the OLS estimate of the slope parameter  $m$  has been shown to be asymptotically normal as long as  $d \neq 0.5$  [54].

We restrict the data set to the years 1950–2020, the period over which most of the significant global warming is believed to have happened. We note [as can be seen in Fig. 1(a)] that even in this timeframe, the temperature increase is somewhat stronger than linear. While techniques such as ensemble empirical mode decomposition have been introduced to address this, a linear fit is most commonly chosen in the literature as it is a well-justified first-order approximation and is amenable to analytical treatment.

To test for LRC properties  $d$  of the Potsdam time series, we use DFA [55–57]. In DFA( $k$ ), long-term correlations can be measured while removing trends of order  $k$  by dividing the cumulative sum of the time series  $T_t$  into time windows of length  $s$  samples each, and in each window fitting that part of the series with a polynomial fit  $Y_t$  of order  $k$ . The squared differences between polynomial fit and the cumulative signal are then accumulated to form the fluctuation function  $F^2(s)$  as a function of the scale  $s$  (window length), which obeys a power law [see Fig. 1(b)].

This scaling region is typically fitted by eye, which is subjective and often challenging due to curvatures from crossover

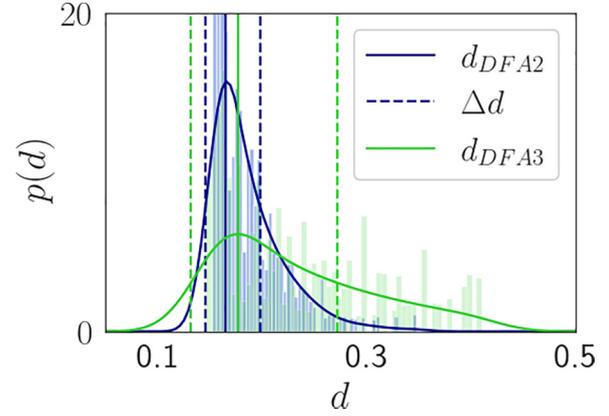


FIG. 2. Histograms and corresponding kernel smoothed probability distributions of  $d$  with errors (for both DFA2 and DFA3) for the Potsdam temperature anomalies time series.

effects [58]. Here we apply a recently introduced general automated method for fitting scaling regions to the task of accurately determining the Hurst exponent. This method was introduced by Deshmukh *et al.* in a different context of estimation of fractal dimensions and Lyapunov exponents [59]. The idea is to first create an ensemble of intervals of  $s$  values by considering all possible combinations of end points, generating a distribution of slopes from least-squares fits weighted by the length of the fitting line and the inverse square of the fit error [shown in Fig. 1(b)]. This set of values is then smoothed using a nonparametric Gaussian kernel to yield a PDF of the slopes. The mode of this distribution gives an estimate of the slope of the scaling region (shown in Fig. 2).

A major advantage of this method is that from the distribution of slopes, we can derive an estimate of the error of  $d$ . The upper and lower errors are given by the respective half-width half heights of this distribution. It can be seen that the upper error is significantly larger than the lower error. This is due to DFA tending to have a range of small  $s$  values where the increase is more steep, which is due to exponential correlations in the data. It can be seen that for higher order the histogram becomes much less sharply peaked, and the error for  $d$  becomes correspondingly bigger. This will be very important for calculation of the  $p$  values later in this paper.

For DFA2, we obtain a value of  $d = 0.173^{+0.053}_{-0.040}$  while for DFA3 we obtain a value of  $d = 0.19^{+0.18}_{-0.06}$ . The value of  $d$  for DFA3 lies within the error bounds of  $d$  obtained by DFA2. The value of  $d$  for DFA3 is observed to be larger than that of DFA2, presumably because as previously mentioned the trend of the temperature anomalies is observed to increase more strongly than linear.  $k = 4$  like  $k = 3$  also lies comfortably within the margin of error of  $k = 2$ . We note that this technique for obtaining an accurate estimation of  $d$  with an error can also be used in conjunction with other methods, which rely on scaling to obtain  $d$  such as wavelet analysis [60,61] in exactly the same way.

To quantify the autoregressive component responsible for purely SRCs in the noise, we first rewrite Eq. (3) as

$$(1 - B)^d \epsilon_t = \sum_{j=0}^{\infty} (-1)^j \binom{d}{j} \epsilon_{t-j}. \quad (5)$$

The term on the right-hand side is known as the first-order finite-difference approximation of the Grünwald-Letnikov derivative  $D_t^d \epsilon_t \equiv \tilde{\epsilon}_t$ . We can now remove LRC from a general ARFIMA process by applying the Grünwald-Letnikov derivative to both sides to obtain an ARIMA process. In our case,  $(1 - \phi B)\epsilon_t = (1 - B)^{-d}\xi_t$  becomes  $(1 - \phi B)D_t^d \epsilon_t = \xi_t$ , and thus we obtain a purely autoregressive process  $\tilde{\epsilon}_t = \phi\tilde{\epsilon}_{t-1} + \xi_t$ . This is a general method for Gaussian noise to derive an SRC time series from an SLRC time series [62]. We use the value of  $d$  obtained before by DFA.

The resulting time series  $\tilde{\epsilon}$  is observed to have a close to exponential decay. The Akaike information criterion suggests that an autoregressive short memory model AR(1) [as given by Eq. (2)] is a significantly better fit than WN. Increasing the order of the AR process further still is not observed to significantly improve AIC further. The autoregressive parameter  $\phi$  is obtained by plotting the data in the  $(T_{t+1}, T_t)$  plane and using a linear fit [shown in Fig. 1(c)], which yields  $\phi = 0.679$ . It can be seen that the clustering of points corresponds nicely to a bivariate Gaussian distribution, as expected for a discrete autoregressive process.

With the parameters  $\phi$  and  $d$  thus obtained, we run the ARFIMA(1, $d$ ,0) model and can verify that its autocorrelation function agrees well with the one of the temperature anomalies  $T_t$  on time lags up to 50. This agreement is far superior to that of an AR(1) model or ARFIMA(0, $d$ ,0) model.

The Potsdam temperature anomalies time series together with its linear fit is shown in Fig. 1(a), where the estimated trend is  $m = 0.29 \pm 0.09$  K/decade ( $p$  value of 0.0011) for this data set, where the error estimate refers to one standard deviation.

### B. Statistical significance

To obtain the statistical significance of the trend, we use the correlation properties of the noise to study trend estimates of samples of the null hypothesis that there is no trend. We denote the distribution of estimated as trends  $P(m; N)$ , where  $m$  is the trend and  $N$  is the length of the time series. On samples satisfying the null hypothesis, this distribution can be assumed to be Gaussian with mean zero, since it is an OLS estimator, and as such is unbiased and consistent.

Since the distribution of trends  $P(m; N)$  of the null hypothesis is approximately Gaussian and of mean zero, it is clear that only knowledge of the variance  $\sigma^2[\hat{m}]$  is further needed to obtain not only a measure of the error of the trend but also the statistical significance of the *observed trend*. This is given by the  $p$  value, defined as the probability that the trend estimate for data satisfying the null hypothesis is larger than this observed trend (one-sided test) or that it differs more from zero than the observed trend (two-sided test). Using a two-tailed test and considering symmetry of  $P(m; N)$ , this is given by

$$p(\hat{m}; N) = 2 \int_{\hat{m}}^{\infty} P(m'; N) dm'. \quad (6)$$

As is common in the literature, we consider the observed trend to be statistically likely if the  $p$  value is less than 5%. For a Gaussian distribution  $P(m; N)$ , this is the case when

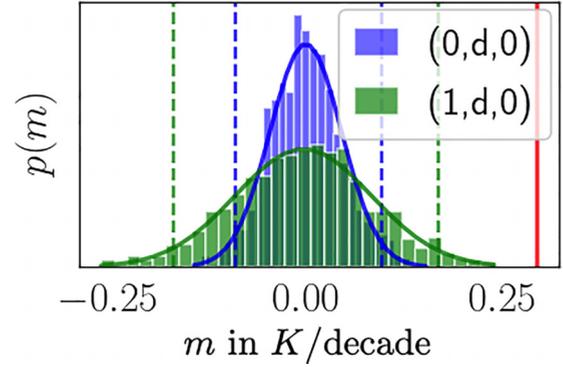


FIG. 3. Histograms show estimated trends of Monte Carlo numerically generated trendless ARFIMA(0, $d$ ,0) and ARFIMA(1, $d$ ,0) time series (representing the null hypothesis) for Potsdam station data. Fitted using Gaussian distribution. Dashed lines are 95% confidence interval, and red line is measured trend of the Potsdam series.

the observed trend is approximately more than two standard deviations away from zero ( $\mu \pm 2\sigma$ ) [63].

Increasing the parameters  $\phi$  and  $d$  increases the noise persistence, which may give the appearance of a trend, even where none exists. Indeed, numerical trend estimates of ARFIMA time series without trend can yield both positive and negative trend values. We create the distributions of estimated trend values numerically for samples of 1000 time series of equal length as the Potsdam data and with the parameters  $\phi$  and  $d$  obtained from the data. It can be seen that ARFIMA(1, $d$ ,0) has a wider distribution than that of ARFIMA(0, $d$ ,0), due to the larger persistency of these noise anomalies (see Fig. 3). The measured trend of the Potsdam time series is significantly larger than the upper 95% confidence interval of both and is thus considered to be statistically significant. When using Monte Carlo estimates, we obtain  $p_{(0,d,0)} = 3.4 \times 10^{-9}$  and  $p_{(1,d,0)} = 0.00081$ .

Due to the fact that the  $p$  value depends only on the tails of the distribution, it is very sensitive to differences in the value of  $d$ . If we use the upper error bar of DFA2  $d + \Delta d = 0.216$  instead of  $d = 0.173$ , then we obtain, for example,  $p_{(1,d+\Delta d,0)} = 0.011$ , more than an order of magnitude larger. The  $p$  value for this reason also varies a lot if we repeat the Monte Carlo simulations with a different ensemble of trajectories. A more reliable estimate of the  $p$  value can be obtained using analytics, which will be the focus of the next section.

### C. Analytical results for trend significance

We now give an analytical expression for how the variance of the OLS estimator  $\hat{m}$  for SLRC time series depends on the autoregressive parameter  $\phi$ , the fractional parameter  $d$  and the time-series length  $N$ .

The result of the OLS fit  $\hat{m}$  of the linear trend can be rewritten as a linear filter on the time series for which the fit is performed. Assuming Eq. (1) as the proper time series model we find (see Appendix) that  $\hat{m} = m + \sum_{t=1}^N a_t \epsilon_t$ , where the weight is given by

$$a_t = \frac{12t - 6(N + 1)}{N^3 - N}. \quad (7)$$

We see that  $\hat{m}$  is a Gaussian random variable if  $\epsilon_t$  is Gaussian. Importantly, even with auto correlations in  $\epsilon_t$ , the estimation is unbiased, since  $\langle \hat{m} \rangle = m$ . The variance of the distribution of the estimator of the trend  $\sigma^2[\hat{m}]$  depends on the autocovariance of the noise  $C(s) = \langle \epsilon_t \epsilon_{t+s} \rangle$ , as [64]

$$\sigma^2[\hat{m}] = 12/(N^3 - N) + 2 \sum_{s=1}^{N-1} C(s)L(s), \quad (8)$$

where  $L(s)$  is a weighting function defined as

$$L(s) = \sum_{t=1}^{N-s} a_t a_{t+s}. \quad (9)$$

Using these results, we analytically calculate  $\sigma^2[\hat{m}]$  for the ARFIMA(1,  $d$ , 0) model. The full expression of  $\sigma^2[\hat{m}]$  for ARFIMA(1,  $d$ , 0) is derived in the Appendix. The asymptotic behavior of  $\sigma^2[\hat{m}]$  for a large length  $N$  of the time series follows the power law

$$\sigma^2[\hat{m}] \sim \sigma^2[T] f(\phi, d) N^{2d-3}, \quad (10)$$

where  $\sigma^2[T]$  is the variance of the analyzed time series and  $f(\phi, d)$  is calculated to be

$$f(\phi, d) := \frac{1 + \phi}{(1 - \phi)(2 {}_2F_1(1, d, 1 - d, \phi) - 1)} \times \frac{36(1 - 2d)\Gamma(1 - d)}{d(1 + 2d)(3 + 2d)\Gamma(d)}, \quad (11)$$

where  $\Gamma(\cdot)$  and  ${}_2F_1(\cdot)$  are the gamma and hypergeometric functions. As a special case of this, we also obtain both the known asymptotics of  $\sigma^2[\hat{m}]$  [65–67] for AR(1),

$$\sigma^2[\hat{m}] \sim \sigma^2[T] 12 \left( \frac{1 + \phi}{1 - \phi} \right) N^{-3}, \quad (12)$$

from which the WN limit can also be easily seen by setting  $\phi = 0$ , and ARFIMA(0,  $d$ , 0):

$$\sigma^2[\hat{m}] \sim \sigma^2[T] \frac{36(1 - 2d)\Gamma(1 - d)}{d(1 + 2d)(3 + 2d)\Gamma(d)} N^{2d-3}. \quad (13)$$

A comparison of the analytical results of the variance to those of the numerically generated time series is shown in Fig. 4. It can be seen that the asymptotic formulas [shown in Fig. 4(a)] are a good fit for  $N \gtrsim 200$ . For purely LRC noise, the asymptotic fits are a good approximation for any  $N$ .

It can be seen that for time series with small  $N$  and SRC, the asymptotic formulas tend to somewhat overestimate  $\sigma^2[\hat{m}]$  and consequently weaken the significance of the estimated trends. Thus, in cases where  $N$  is less than 200 and there exists a significant presence of SRC, we suggest that the full closed formulas may be more appropriate. This is often the case, for example, in climate studies, which make use of monthly data.

More importantly, our closed form expressions and their asymptotics allow us to identify the different effects of SLRCs in detail. It can be seen from Eqs. (12) and (13) that the autoregressive parameter  $\phi$  does not alter the asymptotics of the decay of the variance  $\sigma^2[\hat{m}]$  as a function of  $N$ , which is like  $N^{-3}$  for both WN and an AR(1) process. However, the short-range correlations increase the error bars by a  $\phi$ -dependent prefactor, which can be interpreted as reduction of

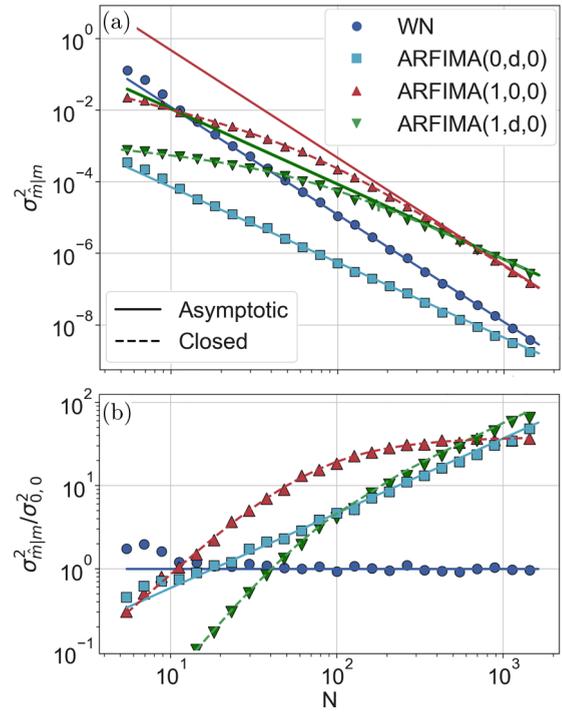


FIG. 4. (a) Closed formulas, asymptotic behaviors [Eqs. (10)–(13)] and the numerically simulated results of the variance of the OLS trend estimator for the indicated models with  $d = 0.45$ ,  $\phi = 0.95$  as function of time series length  $N$ .  $\sigma^2[T] = 1$  for all data except ARFIMA(0,  $d$ , 0), where  $\sigma^2[T] = 0.01$  (shifting the entire curve downward by two orders of magnitude) for more clear presentation. (b) Variances of the models of (a) divided by the variance assuming WN on top of the trend.

the sample size by measuring it in numbers of the correlation time interval, and they introduce a crossover regime in  $N$ . The situation is very different for long-range correlated ARFIMA noise: the parameter  $d \neq 0$  changes the power by which  $\sigma^2[\hat{m}]$  drops in  $N$  from  $-3$  to  $2d - 3$ . Interestingly, for very short data sets,  $N \simeq 10$ , the noise model with largest uncertainty is the WN, while the model with smallest uncertainty is the ARFIMA(1,  $d$ , 0) model. But for large  $N$  (over 150), we observe the variance of ARFIMA(1,  $d$ , 0) to be the largest. This, in turn, means that for large  $N$  it is essential to use the appropriate noise model in order not to underestimate the error bars.

A quite unexpected finding is the reduction of uncertainty by correlations for very short time series, as compared to WN [most clearly visible in Fig. 4(b)]. This can be understood quite easily: For strongly correlated noise, the empirical variance of a small sample,  $N \approx 10$ , is much smaller than unity (the variance of the marginal distribution) and hence such noise affects more so the estimate of the offset  $n$  than that of the slope  $m$  in Eq. (1).

For the Potsdam anomalies time series of Fig. 1, using our parameters  $d = 0.173^{+0.043}_{-0.029}$  (from DFA2),  $\phi \approx 0.683$ , and  $N = 25567$  (70 years), the  $p$  value is obtained analytically for LRC as  $p_{\text{an},(0,d,0)} = 5.5 \times 10^{-9}$  and for SLRC as  $p_{\text{an},(1,d,0)} = 0.00149$ . These values are comparable to those of the Monte Carlo method, however, much more accurate. As previously

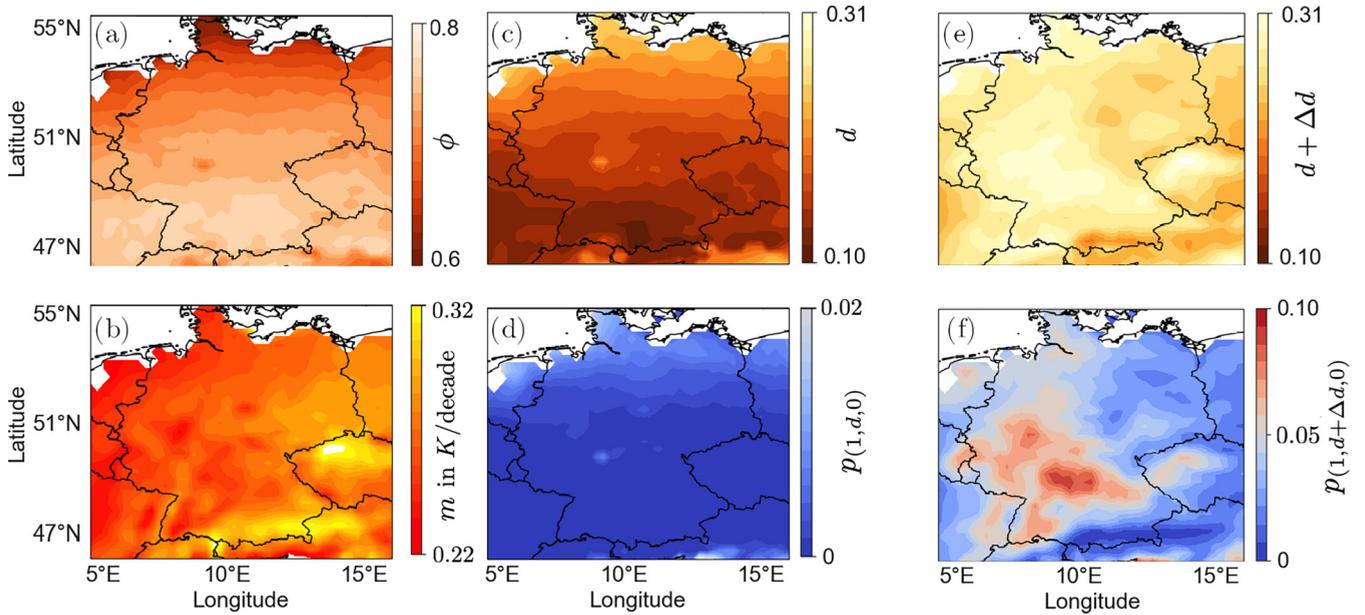


FIG. 5. Values at grid points in central Europe (ERA5 reanalysis data). (a) AR(1) exponents  $\phi(x, y)$ , (b) trends  $m(x, y)$ , (c) fractional exponents  $d(x, y)$ , and (d)  $p$  values obtained with an ARFIMA(1, $d$ ,0) ansatz. (e)–(f) Upper bounds  $(d + \Delta d)(x, y)$  and resulting  $p$  values obtained with ARFIMA(1, $d + \Delta d$ ,0) ansatz. Red indicates the region where the  $p$  value is statistically insignificant.

mentioned, the Monte Carlo method fluctuates particularly strongly even for large numbers of simulations, since the  $p$  value depends on the amount of simulations with values higher than the measured trend, which is even for large numbers of simulations not very many.

The difference of the  $p$  values of the observed trend under two different model assumptions emphasizes the importance of taking into account the SRCs in the data by using the ARFIMA(1, $d$ ,0) model. In the example of the Potsdam time series of the previous section, the difference between the  $p$  values with and without consideration of SRCs are orders of magnitude  $p_{\text{an},(1,d,0)}/p_{\text{an},(0,d,0)} \sim 10^6$ .

### III. SPATIALLY DISTRIBUTED TIME SERIES

We now apply our results to spatially distributed data. We use daily grid land data of central Europe from 1950 to 2020 collected from the reanalysis ERA5 data set with a grid separation of  $0.5^\circ$  [68]. The values of  $\phi$ ,  $m$ ,  $d$ , and the  $p$  values of the different grid points are shown in Figs. 5(a)–5(d), respectively, where  $d$  has been obtained by DFA2. It can be seen in Fig. 5(a) that the autoregressive parameter  $\phi$  assumes large values over the whole region, although the values are somewhat smaller nearer the coast. It can also be clearly seen in Fig. 5(c) that the fractional exponent  $d$  is significantly larger near the coast. This causes the  $p$  value to be larger near the coast [Fig. 5(d)], although we note that all trends are statistically significant.

The error bars of  $d$  have a dramatic effect on the reliability of the  $p$  value. The spatially dependent fractional exponent and corresponding  $p$  value for the case that the upper bound of  $d$  (using DFA2) is used are shown in Figs. 5(e)–5(f). Interestingly, the error  $\Delta d$  seems to generally be largest in the Southwest of Germany [compare Figs. 5(c) and 5(e)]. It can be seen that the trend in this case is not statistically significant

in all regions of central Europe. This raises the question of whether the overall trend of central Europe is statistically significant?

To answer this question, we need to go beyond our analysis of the temporal correlations on the variance  $\sigma^2[\hat{m}]$  of the time series [making use of Eq. (10)] and further consider the effect of spatial correlations on the variance  $\sigma^2[T]$ . As can be seen from Eqs. (10) and (12), the uncertainty of the trend estimation is proportional to the magnitude of the fluctuations of the signal  $\sigma^2[T]$ . We now wish to exploit the redundancy of the trend information contained in spatially distributed data to construct a signal with lower amplitude fluctuations.

To explore the redundancy in such data to obtain more accurate trend estimates we compare two different methods. Both involve deriving a single time series from the gridded data and then evaluating that time series using the previously discussed methods. The first method is to calculate the time series as a spatial average of the gridded time series, and the second method is to use PCA. While the first method of taking the spatial average over all time series may be conceptually simple, it can obscure important spatial detail and may well obscure much of the physics associated with the majority of the observed variation. PCA offers a more elegant alternative by partitioning the series into series representing different geographical patterns, where higher order patterns may be discarded as noise.

In both methods, the fact that grid points of greater latitude are closer together in the longitudinal direction and thus each have a smaller surface area than those nearer the equator is respected by using spatial weighting [69].

#### A. Averaged signal

We first explore the method of defining the averaged signal as the average over the time series at each grid point weighted by the area of the grid cells. In this case, the area weighted

average series is given by

$$\bar{T}_t := \langle T_t(x, y) \rangle_{\text{aw}} = \frac{1}{A} \sum_{(x,y)} A(x, y) T_t(x, y), \quad (14)$$

where  $\langle \cdot \rangle_{\text{aw}}$  is the area weighted average,  $A$  is the total spatial area, and  $A(x, y)$  is the area of the grid cell at location  $(x, y)$ . We now once again use Eq. (1) as an ansatz for our system,

$$\bar{T}_t = \bar{n} + \bar{m}t + \bar{\epsilon}_t, \quad (15)$$

where  $\bar{m} := \langle m(x, y) \rangle_{\text{aw}}$  and  $\bar{\epsilon}_t := \langle \epsilon_t(x, y) \rangle_{\text{aw}}$  are the area-weighted averages. Since averaging over the different fluctuations leads to a reduction of the variance, it holds that

$$\sigma^2[\hat{m}] \leq \sigma^2[\hat{m}]. \quad (16)$$

If the  $\epsilon_t(x, y)$  were all independent (which they are not!) the variance would decrease proportional to the number of time series which we average over  $\sigma^2[\hat{m}] = \sigma^2[\hat{m}]N^{-1}$ . Since time series which are spatially close to each other are more correlated, we find that the variance decreases more gradually.

We find exactly the same trend if we perform an area-weighted average over the trend values of the single grid points, since both averaging the time series, averaging the trends, and calculation of the OLS trend are linear operations and hence commute. However, using the averaged time series, we can determine the statistical significance of the trend  $\bar{m}$  right away using Eq. (10).

## B. Principal component analysis

The idea of EOF analysis, otherwise known as PCA, is to find a relatively small number of independent variables, which account for as much of the original signal as possible [70,71]. It uses a set of EOFs and PCs to represent the time series in the following way:

$$T_t(x, y) = \sum_{k=1}^N c_k \mathcal{P}_{t,k} \frac{1}{c_k} \mathcal{E}_k(x, y), \quad (17)$$

where  $\mathcal{E}_k(x, y)$  represents the  $k$ th spatial pattern (EOF $_k$  where  $k = 1, 2, \dots$ ) and  $\mathcal{P}_{t,k}$  represents the principal components (PC $_k$ ), which describe how the amplitude of each EOF varies with time.  $c_k$  is a scaling constant.

The basic procedure for EOF analysis is that the spatially distributed time series is first arranged into a matrix  $T$ , which in turn is used to calculate the covariance matrix  $R = T^T T$ . From this, an eigenvalue problem is solved to obtain the EOFs (which correspond to the eigenvectors), the EOF variances (which correspond to the eigenvalues), and the PCs (which are computed from the projection of  $T$  onto the EOFs).

Here, for purposes which will become clear later in this paper, we rescale the PCs and the EOF in Eq. (17) by coefficients  $c_k$ . At this point, we deviate from the traditional literature on PCA and introduce a different normalization  $c_k := \langle \mathcal{E}_k \rangle$  over the spatial region to normalize the EOFs to one. By introducing this normalization, the scaled PCs  $\mathcal{P}_{t,k} \rightarrow c_k \mathcal{P}_{t,k}$  become normalized to the magnitude of the spatial average of the original time series  $T_t$ . Thus, given this normalization, a simple summation of the PC series  $\mathcal{P}_k$  is sufficient to construct

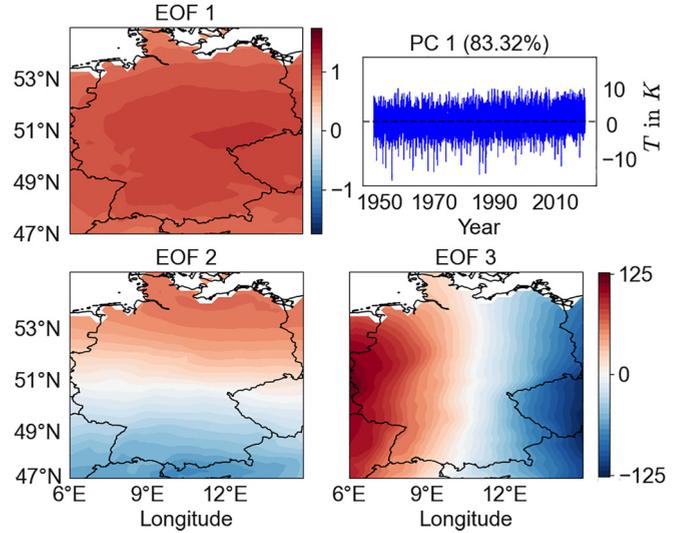


FIG. 6. First three EOFs and PC $_1$  in units of °C. The mean of the EOFs have all been normalized to 1.

spatially averaged time series:

$$\langle T_t(x, y) \rangle = \sum_{k=1}^N \mathcal{P}_{t,k}. \quad (18)$$

This property is desirable, since the combined effect of several PCs can be studied. It will mean in simple cases that the first PC is simply the spatially homogeneous part of the spatiotemporal time series. This can be underlined by the fact that the PCs can all be assigned the physical unit K (Kelvin), while the EOFs are dimensionless.

In contrast to our choice of  $c_k$ , typically in the literature  $\mathcal{P}_k$  is made comparable to other time series by defining  $c_k$  as the square root of the  $k$ th singular value [24]. In that case, the renormalized patterns carry the physical units and the coefficients are dimensionless (zero mean and unit standard deviation).

Our introduced method of dividing by the mean has however two advantages for the purposes of analyzing a spatially averaged signal. First, dividing by the mean rewards homogeneity by making PCs with more homogeneous EOFs larger. Second, it allows the averaged signal to be written simply as a sum of the PCs as in Eq. (18). This means that we can construct a sequence of increasingly better approximations of the averaged signal by adding more and more PCs, starting from the  $\mathcal{P}_{t,1}$  as opposed to treating the PCs separately as is typically done in the literature.

PCA can smooth out details which are not necessary to the overall trend of a system, and can reduce grid data to a single time series. For a topologically simple region, which is not too large, the first PC-EOF pair describes the overall trends over the entire region, while higher order terms tend to describe fluctuations around this trend. The second and third pairs, for example, describe the fluctuations around this overall trend in the north-south and east-west directions, respectively. This is a straightforward consequence of the orthogonality condition of EOF analysis. An example of this is shown in Fig. 6 for the case of central Europe.

TABLE I. Comparison of the properties of the averaged signal  $\langle T(x, y) \rangle$  and the first principal component for central Europe (data of Figs. 5 and 6). Potsdam is the single time series of the Potsdam station [Fig. 1(a)].

	$\langle T(x, y) \rangle$	$\mathcal{P}_1$ (PC <sub>1</sub> )	Potsdam
$m$ (K/decade)	$0.27 \pm 0.08$	$0.27 \pm 0.08$	$0.29 \pm 0.09$
$d$	$0.16^{+0.11}_{-0.06}$	$0.16^{+0.11}_{-0.06}$	$0.173^{+0.053}_{-0.04}$
$\phi$	0.804	0.800	0.679
$p_{\text{an},(0,d,0)}$	$1 \times 10^{-14}$	$1 \times 10^{-14}$	$5.5 \times 10^{-9}$
$p_{\text{an},(1,d,0)}$	0.0011	0.0011	0.00149
$p_{\text{an},(1,d+\Delta d,0)}$	0.026	0.030	0.011

The higher order EOFs have a mean value close to zero, which in turn means that  $c_k \mathcal{P}_{t,k}$  are very small for  $k \in \{2, 3\}$ . Normalization thus reveals that the trend of the first PC is dominant while the higher order PCs have trends, which are several orders of magnitude smaller. The size of these PCs are thus not only a function of the variance of each PC (given by the eigenvalue of the covariance matrix) but also of  $c_k$ . This is because when we are exploring the overall trend we are not interested in geographical fluctuations around the mean and normalizing the PCs discounts the effect of these fluctuations in a very simple way. The first PC is thus more significant than the higher not only because it is by design the PC which describes the largest proportion of the variance but also because the first PC describes the homogeneous component, which is most relevant to the trend.

Related to this is the fact that unusual topographical features in a spatial region (such as mountain ranges) typically only affect higher PCs and thus this version of PCA offers an objective way to filter out such effects. This is in contrast to the spatial mean where such features may distort the spatial mean. The effective variance of the PCs for the purposes of determining the overall trend of a spatial region is thus even smaller than that suggested by the eigenvalues. This method, its validity, and its effects will be the topic of a separate paper.

The reasons that the trend is almost completely contained in  $\mathcal{P}_{t,1}$  are described in the Appendix. The basic idea is that if  $\mathcal{E}_1$  (EOF<sub>1</sub>) is spatially homogeneous and the trend almost the same for all grid points, then it will only affect  $\mathcal{P}_1$  (PC<sub>1</sub>) and leave all other PCs unchanged. Since this homogeneity is to a good approximation satisfied by  $\mathcal{E}_1$ ,  $\mathcal{P}_1$  carries the trend. In the case of central Europe, both conditions are well satisfied (see Fig. 6).

The tendency for PCA to produce simple patterns such as a homogeneous component, an east-west component, etc. is reliant on two things. The first is the simplicity of the topology, water, ice coverage, etc. of the system and the second is that the geographical region should not be too large. Both of these conditions are met in the case of the land data of central Europe. These things will be explored in more detail in a separate paper.

### C. Results for central Europe

All main results are summarized in Table I, where we compare the trends obtained (including the standard deviation

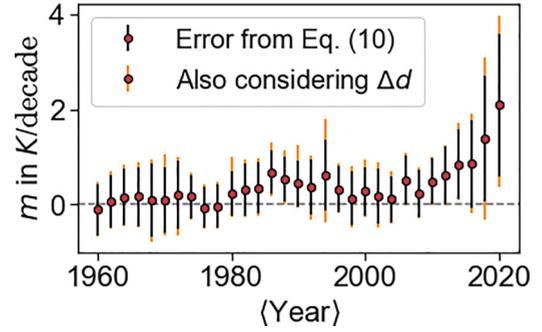


FIG. 7. The time-resolved trend analysis for PC<sub>1</sub>( $t$ ) of the ERA5 temperature anomalies. Linear trend estimated on overlapping 21-year intervals windows together with error bars (standard deviation). Orange error bars additionally consider error  $\Delta d$  with Gaussian error propagation.

tions of the OLS estimator which can be considered as error bars), the parameters of the optimal ARFIMA(1, $d$ ,0) model, and the analytical  $p$  values of the estimated trends for the averaged signal  $\bar{T}_t$  and for the first PC  $\mathcal{P}_1$  of central Europe. Results of the Potsdam single time series are also included for comparison. We see that for this spatial region of central Europe there is only minimal variation between the methods of spatial averaging and PCA. In terms of climate change, it is evident that the estimated trend for this region is statistically significant, even when taking a reasonable upper bound for  $d$ . It is, however, nevertheless important to note that the  $p$  value is orders of magnitude larger when taking the upper bound of the error of  $d$ . In comparison to the single station time series of Potsdam, the error bars of the trend value have reduced in size only slightly. This is due to the strong spatial correlations—the noise cancellation effect of spatial averaging is small: the variance  $\sigma^2[T]$  of the signal only drops from 14.6 to 9.8, while at the same time the SRCs got slightly stronger with  $\phi$ , increasing from 0.679 to 0.80. The significance of  $p_{\text{an},(1,d,0)}$  is somewhat larger for Potsdam compared to central Europe, principally due to  $d$  being larger in Potsdam (due to Potsdam being nearer to the coast).

As a final application of the results of this paper, we explore change in the trend of central Europe over time and the corresponding error bars. This is shown in Fig. 7. The trends have been obtained by performing linear trend analysis on moving time windows of 21 years length for the PC<sub>1</sub>( $t$ ) time series. For years (Year) later than 2012, where no ten-year intervals into the future exist, we estimate the trend on the interval [(Year) - 10, 2022]. With the estimated values of  $d$  and  $\phi$  and the actually used time window length, we then calculate the error bars. As the figure shows, trend values are usually found to be generally positive but small until the time window reaches around 2006, after which the trend increases dramatically. The increase of error bars in recent years is due to the shorter-than-20-years time windows. These rather dramatic results of temperature increase make evident that the long-term temperature change is not linear. Trend models that display trends comparable to an exponential trend are generally not considered in the literature and seem to be more dramatic than predictions made using climate simulations.

### D. Conclusions

This paper provides important tools for linear trend analysis both in the presence of SLRCs and of spatial correlations. We demonstrate under the consideration of climatological spatial regions (of gridded data) that neither short nor long temporal correlations can be disregarded without deriving misleading conclusions with regard to the statistical significance the trend. This is significant, since in the literature generally only consideration is made of either short-range or long-range temporal correlations. We expect such effects may also be important in other complex systems such as brain activity.

We have considered a realistic and still parsimonious model for the fluctuating part of such a signal, the ARFIMA(1, $d$ ,0) process. For this model, we have analytically derived the calculation of the variances of OLS estimators of the trend on data. Simple asymptotic approximations are found to be valid for time series of length  $N \gtrsim 200$ . This enables quick and accurate calculation of the  $p$  value, which is otherwise observed to fluctuate dramatically when obtained using (even very long and time consuming) Monte Carlo simulations. From this, both the error bars of the estimated trend value and the  $p$  value can be obtained to quantify the statistical significance of the estimated trend. These formulas also highlight the different impacts of the SLRCs on the accuracy of the trend estimate. We have applied a particularly accurate automated method to the task of obtaining the long-range fractional parameter (Hurst exponent) together with its error. This automation is particularly necessary in the context of big data sets, where visual inspection of each individual time series would not be practical. We have shown that the error bars of this parameter are important in that within the margin of error, the  $p$  value can fluctuate by orders of magnitude, causing trends of many time series to become statistically unlikely.

For dealing with spatial data, we have argued that the most natural way to consider spatial correlations is with the use of PCA. We have introduced a practical normalization technique such that the spatially averaged time series can be written simply as the sum of PCs. We have shown that given the not-too-large size and homogeneity of the region, an analysis of the first PC is sufficient to obtain the overall trend for the region of central Europe. As a final result, we have obtained a warming trend over the past 70 years in central Europe of  $0.27 \pm 0.08$  K/decade, which is statistically significant with a  $p$  value of 0.0011 in a two-sided test. We have also applied the techniques of this paper to demonstrate that the magnitude of the warming trend of central Europe has increased more dramatically over the last 20 years than previously assumed. From the point of view of trend analysis methodology, this result suggests that an important next step could be going beyond OLS methods to consider nonlinear trends such as exponential trends.

### APPENDIX: ANALYTICAL CALCULATION OF THE DISTRIBUTION OF THE OLS ESTIMATED TREND

In this Appendix, we derive both the full closed form and its asymptotics for large  $N$  of the distribution of the OLS

estimated trend values of time series  $T_t$ , using the ansatz of Eq. (1),

$$T_t = n + mt + \epsilon_t, \quad t = 1, \dots, N,$$

for different types of correlated noises  $\epsilon_t$ . Without loss of generality and to highlight the effects of other model parameters, we assume  $\{\epsilon_t\}$  to have unit variance. For the trend estimation, we calculate the linear fitting polynomial

$$p_t = \hat{m}t + \hat{n} \quad (\text{A1})$$

for  $T_t$  by OLS. Minimizing the sum of the squared residuals  $\sum_{t=1}^N (T_t - p_t)^2$  with parameters  $\hat{n}$  and  $\hat{m}$ , we obtain two equations:

$$\begin{aligned} \sum_{t=1}^N (p_t - T_t) &= 0, \\ \sum_{t=1}^N t(p_t - T_t) &= 0. \end{aligned} \quad (\text{A2})$$

We solve these two equations for  $\hat{m}$  and  $\hat{n}$ , where  $\hat{m}$  is the estimated trend and reads

$$\hat{m} = \sum_{t=1}^N \frac{S_0 t - S_1}{S_0 S_2 - S_1^2} T_t. \quad (\text{A3})$$

Here,  $S_j = \sum_{t=1}^N t^j$  for  $j = 0, 1, 2$ . For the Gaussian process  $\{\epsilon_t\}_{t=1}^N$ , the conditional distribution  $P(\hat{m}|m; N)$  of the estimated trend  $\hat{m}$  given  $m$  is a Gaussian itself and hence is completely characterized by the first two conditional moments  $\langle \hat{m}|m \rangle$  and  $\langle \hat{m}^2|m \rangle$ . In the following, we use the short hand notation:

$$\begin{aligned} a_t &:= \frac{S_0 t - S_1}{S_0 S_2 - S_1^2} \\ &= \frac{12t - 6(N+1)}{N^3 - N}, \end{aligned} \quad (\text{A4})$$

where the second equation follows from  $S_0 = N$ ,  $S_1 = N(N+1)/2$ , and  $S_3 = N(N+1)(2N+1)/12$ . The first conditional moment can now be obtained as the ensemble average of Eq.(A3) for given  $m$ :

$$\begin{aligned} \langle \hat{m}|m \rangle &= \sum_{t=1}^N a_t \langle T_t|m \rangle \\ &= \sum_{t=1}^N a_t (n + mt + \langle \epsilon_t \rangle). \end{aligned} \quad (\text{A5})$$

Since  $\epsilon_t$  has zero mean, the first conditional moment is found to be

$$\langle \hat{m}|m \rangle = m. \quad (\text{A6})$$

This is in agreement with common knowledge that OLS estimators are unbiased.

The second conditional moment is correspondingly given by

$$\langle \hat{m}^2|m \rangle = \left\langle \left( \sum_{t=1}^N a_t T_t \right)^2 \right\rangle. \quad (\text{A7})$$

Under the assumption of statistical independence between noise and trend, i.e.,  $\langle \epsilon_k(mt+n) \rangle = \langle \epsilon_k \rangle (mt+n) = 0$  for any  $t$  and  $k$ , we can split the expanded square of (A7) into two parts:

$$\left\langle \left( \sum_{t=1}^N a_t T_t \right)^2 \right\rangle = \left\langle \left( \sum_{t=1}^N a_t \epsilon_t \right)^2 \right\rangle + \left\langle \left( \sum_{t=1}^N a_t (mt+n) \right)^2 \right\rangle. \quad (\text{A8})$$

The second term gives  $m^2$  since  $\sum a_t = 0$  and  $\sum t a_t = 1$ . The first part, denoted by  $\langle \hat{m}^2 | m \rangle_\epsilon$ , reflects the correlations of the noise  $\epsilon_t$ :

$$\begin{aligned} \langle \hat{m}^2 | m \rangle_\epsilon &= \left\langle \left( \sum_{t=1}^N a_t \epsilon_t \right)^2 \right\rangle = \sum_{t,k=1}^N \langle \epsilon_t \epsilon_k \rangle a_t a_k \\ &= \sum_{s=0}^{N-1} \sum_{t=1}^{N-s} \langle \epsilon_t \epsilon_{t+s} \rangle a_t a_{t+s}. \end{aligned} \quad (\text{A9})$$

Thus the second conditional moment is

$$\langle \hat{m}^2 | m \rangle = \langle \hat{m}^2 | m \rangle_\epsilon + m^2. \quad (\text{A10})$$

Consequently, the variance of the estimated trend  $\sigma^2[\hat{m}|m]$  can be obtained as  $\langle \hat{m}^2 | m \rangle - \langle \hat{m} | m \rangle^2 = \langle \hat{m}^2 | m \rangle_\epsilon$ . In summary, the conditional distribution of the estimated trend  $\hat{m}$  given  $m$  is

$$p(\hat{m}|m) = \mathcal{N}(m, \langle \hat{m}^2 | m \rangle_\epsilon). \quad (\text{A11})$$

To obtain this distribution, we only need to know the variance  $\sigma^2[\hat{m}|m] = \langle \hat{m}^2 | m \rangle_\epsilon$ , which completely depends on the autocovariance of the noise  $\langle \epsilon_t \epsilon_{t+s} \rangle$  as shown in Eq. (A9).

### 1. Trend variance for SRC or LRC correlated noise

In this subsection, we consider the cases of  $\epsilon$  as WN, AR(1) noise (parameter  $\phi$ ), and fractional Gaussian noise (parameter  $d$ ), corresponding to ARFIMA(0,0,0), ARFIMA(1,0,0), and ARFIMA(0,  $d$ , 0), respectively. The most general case of an ARFIMA(1,  $d$ , 0) noise will be treated separately, but being based on these results.

In the following, we derive the asymptotic dependency of  $\sigma^2[\hat{m}|m]$  on  $\phi$  and  $d$ , which we will denote by the function  $f(\phi, d)$ , as well as on  $N$ . As notation, we use the subscript  $\{\phi, d\}$  to denote the noise model used, and we distinguish the noise processes as follows:  $\epsilon_{\{0,0\}}$  is WN with unit variance,  $\epsilon_{\{\phi,0\}}$  is AR(1) noise but normalized to unit variance, and  $\epsilon_{\{0,d\}}$

is ARFIMA(0,  $d$ , 0)-noise normalized to unit variance. This normalization is part of our trend model Eq. (1) and is used to remove the trivial proportionality between the variance of the trend estimator and the square of the noise amplitude, and to be able to study the effect of correlations of the noise independent of its variance. For WN,

$$\langle \epsilon_{\{0,0\}t} \epsilon_{\{0,0\}t+s} \rangle = \delta_{s,0}. \quad (\text{A12})$$

Using Eq. (A9), we obtain

$$\sigma^2[\hat{m}|m]_{\{0,0\}} = \frac{12}{N^3 - N} \sim f(0, 0)N^{-3}, \quad (\text{A13})$$

where  $f(0, 0) = 12$ . For a normalized AR(1) noise,

$$\epsilon_{\{\phi,0\}} = (1 - \phi B)^{-1} \frac{\xi}{\sqrt{1 - \phi^2}},$$

where  $\phi \in (-1, 1)$ . The autocovariance function is

$$\langle \epsilon_{\{\phi,0\}t} \epsilon_{\{\phi,0\}t+s} \rangle = \phi^s \sigma_{\{\phi,0\}}^2, \quad (\text{A14})$$

with the variance of  $\epsilon_{\{\phi,0\}} = 1$  due to the rescaling of  $\xi$ . We can now, as previously, obtain the variance of the trend estimator as

$$\begin{aligned} \sigma^2[\hat{m}|m]_{\{\phi,0\}} &= \frac{-1}{(\phi - 1)^4 (N^6 - 2N^4 + N^2)} \\ &\times (12(\phi - 1)^3 (1 + \phi)N^3 + 72\phi(\phi - 1)^2 N^2 \\ &- 12(\phi - 1)^3 (\phi + 1)N - 72\phi(1 + \phi)^2 \\ &+ 72\phi^{N+1}((1 - \phi)N + 1 + \phi)^2) \\ &\sim f(\phi, 0)N^{-3}, \end{aligned} \quad (\text{A15})$$

where

$$f(\phi, 0) = 12 \left( \frac{1 + \phi}{1 - \phi} \right).$$

In the case of ARFIMA(0,  $d$ , 0) noise before normalization:

$$\epsilon_{\{0,d\}} = (1 - B)^{-d} \xi.$$

Thus, the autocovariance function is given by

$$\langle \epsilon_{\{0,d\}t} \epsilon_{\{0,d\}t+s} \rangle = \frac{\Gamma(s+d)\Gamma(1-2d)}{\Gamma(s+1-d)\Gamma(1-d)\Gamma(d)}, \quad (\text{A16})$$

and the variance of  $\epsilon_{\{0,d\}}$  is given as

$$\sigma_{\{0,d\}}^2 = \frac{\Gamma(1-2d)}{\Gamma^2(1-d)}. \quad (\text{A17})$$

We can now calculate the till now unpublished closed form of  $\sigma^2[\hat{m}|m]_{\{0,d\}}$  for normalized ARFIMA(0,  $d$ , 0) by renormalizing  $\epsilon_{\{0,d\}}$  by  $\sigma_{\{0,d\}}$  and get

$$\begin{aligned} \sigma^2[\hat{m}|m]_{\{0,d\}} &= \frac{1}{d(1+2d)(3+2d)} \frac{1}{\Gamma(1-2d)\Gamma(d)\Gamma(N-d)(N^6 - 2N^4 + N^2)} \\ &\times [-36\Gamma(N+d)\Gamma(1-d)\Gamma(1-2d)((-1+2d)N^3 + d(-1+2d)N^2 + N + d) \\ &+ 12\Gamma(N-d)\Gamma(1-2d)((3+8d+4d^2)(d\Gamma(d) - \Gamma(1+d))N^3 \\ &+ 3d(3+2d)\Gamma(1+d)N^2 - (3+8d+4d^2)(d\Gamma(d) - \Gamma(1+d))N - 3d\Gamma(1+d))] \\ &\sim f(0, d)N^{2d-3}, \end{aligned} \quad (\text{A18})$$

where

$$f(0, d) = \frac{36(1 - 2d)\Gamma(1 - d)}{d(1 + 2d)(3 + 2d)\Gamma(d)}. \quad (\text{A19})$$

We see that not only the prefactor is affected by  $d$  but, more importantly, also the power in  $N$  by which the variance drops when  $N$  is increased is reduced. This was to be expected because the effective sample size is reduced by correlations. For consistency, it is relevant to observe that both results, for AR(1) and ARFIMA(0,  $d$ , 0), converge to the WN result for  $\phi \rightarrow 0$  or  $d \rightarrow 0$ , respectively.

## 2. Trend variance for SLRC correlated noise

A short-long-range correlated noise generated by an ARFIMA(1,  $d$ , 0) process has an autocovariance function which is too complicated to be inserted into Eq. (A9). We therefore use prewhitening so we can relate the ARFIMA(1,  $d$ , 0) process to the ARFIMA(0,  $d$ , 0) process. More precisely, we perform the operation  $(1 - \phi B)$  on both sides of Eq. (1) and use the substitutions  $\tilde{\epsilon}_t = (1 - \phi B)\epsilon_t$ ,  $\tilde{T}_t = (1 - \phi B)T_t$ ,  $\tilde{m} = (1 - \phi)m$  and  $\tilde{n} = \phi m + (1 - \phi)n$  to obtain a new time series:

$$\tilde{T}_t = \tilde{n} + \tilde{m}t + \tilde{\epsilon}_t. \quad (\text{A20})$$

This way, we transform the ARFIMA(1,  $d$ , 0) process into an ARFIMA(0,  $d$ , 0) process. The estimated trend  $\tilde{m}$  of  $\tilde{T}_t$  has the distribution

$$P(\tilde{m}|\tilde{m}; N) = \mathcal{N}(\tilde{m}, \langle \tilde{m}^2 | \tilde{m} \rangle). \quad (\text{A21})$$

As previously,  $\langle \tilde{m}^2 | \tilde{m} \rangle$  depends only on the autocovariance  $\langle \tilde{\epsilon}_t \tilde{\epsilon}_{t+s} \rangle$ . Given the stationarity, the autocovariance of WN, AR(1), ARFIMA(0,  $d$ , 0), and ARFIMA(1,  $d$ , 0) should only depend on the time lag  $s$ . Hence, we denote the autocovariances of  $\epsilon$  and  $\tilde{\epsilon}$  as  $C(s) = \langle \epsilon_t \epsilon_{t+s} \rangle$  and  $\tilde{C}(s) = \langle \tilde{\epsilon}_t \tilde{\epsilon}_{t+s} \rangle$ , respectively. The relationship between  $C(s)$  and  $\tilde{C}(s)$  is easily found to be

$$\tilde{C}(s) = (1 + \phi^2)C(s) - \phi C(s - 1) - \phi C(s + 1). \quad (\text{A22})$$

As a special case,  $C(0)$  and  $\tilde{C}(0)$  are, respectively, the variance of  $\epsilon$  and  $\tilde{\epsilon}$ . We define

$$\begin{aligned} L(s) &:= \sum_{t=1}^{N-s} a_t a_{t+s} \\ &= \frac{24s^3 + (12 - 36N^2)s + (12N^3 - 12N)}{N^2(N^2 - 1)^2}, \end{aligned} \quad (\text{A23})$$

where  $a_t$  was defined in Eq. (A4), and reexpress the variance of the corresponding estimated trends in  $T_t$  and  $\tilde{T}_t$  according to Eq. (A9) as

$$\sigma^2[\hat{m}|m] = \sum_{t=1}^N C(0)a_t^2 + 2 \sum_{s=1}^{N-1} C(s)L(s) \quad (\text{A24})$$

and

$$\sigma^2[\hat{m}|\tilde{m}] = \sum_{t=1}^N \tilde{C}(0)a_t^2 + 2 \sum_{s=1}^{N-1} \tilde{C}(s)L(s). \quad (\text{A25})$$

We now focus on the term  $\sum_{s=1}^{N-1} \tilde{C}(s)L(s)$ . Given the general relationship between  $C(s)$  and  $\tilde{C}(s)$  in Eq. (A22), we can use  $C(s)$  to reexpress  $\sum_{s=1}^{N-1} \tilde{C}(s)L(s)$  as follows:

$$\begin{aligned} \sum_{s=1}^{N-1} \tilde{C}(s)L(s) &= \sum_{s=1}^{N-1} (1 + \phi^2)C(s)L(s) \\ &\quad - \phi \sum_{s=1}^{N-1} C(s-1)L(s) \\ &\quad - \phi \sum_{s=1}^{N-1} C(s+1)L(s). \end{aligned} \quad (\text{A26})$$

This equation can be further simplified as

$$\sum_{s=1}^{N-1} \tilde{C}(s)L(s) = (1 - \phi)^2 \sum_{s=1}^{N-1} C(s)L(s) - A, \quad (\text{A27})$$

with  $A$  as

$$\begin{aligned} A &= \phi \left( \frac{144}{N^2(N^2 - 1)^2} \left( \sum_{s=1}^{N-1} \tilde{C}(s)s \right. \right. \\ &\quad \left. \left. + \phi(C(0) - C(N-1)N + C(N)(N-1)) \right) \right. \\ &\quad \left. + C(0)L(1) - C(N-1)L(N) \right. \\ &\quad \left. + C(N)L(N-1) - C(1)L(0) \right). \end{aligned} \quad (\text{A28})$$

Here,  $C(0)$ ,  $C(1)$ ,  $C(N-1)$ , and  $C(N)$  are already known. Therefore,  $A$  can be calculated directly and we find the following expression:

$$\begin{aligned} \sigma^2[\hat{m}|\tilde{m}] &= \sum_{t=1}^N \tilde{C}(0)a_t^2 + 2 \sum_{s=1}^{N-1} \tilde{C}(s)L(s) \\ &= \sum_{t=1}^N \tilde{C}(0)a_t^2 + 2(1 - \phi)^2 \sum_{s=1}^{N-1} C(s)L(s) - 2A. \end{aligned} \quad (\text{A29})$$

Given the relationship between  $\tilde{C}(0)$  and  $C(0)$ , we can obtain  $\sigma^2[\hat{m}|m]$  from the corresponding  $\sigma^2[\hat{m}|\tilde{m}]$ . Usually, the calculation of  $\sigma^2[\hat{m}|\tilde{m}]$  is much easier due to its simpler autocovariance structure. Therefore, such calculation based on the relationship Eq. (A22) can provide an alternative way to obtain  $\sigma^2[\hat{m}|m]$  for more complicated processes.

In particular, for AR(1) noise  $\epsilon_{\phi,0}$  and the corresponding WN  $\tilde{\epsilon}_{\phi,0} = \epsilon_{0,0}$ ,

$$\tilde{C}_{\{\phi,0\}}(0) = (1 - \phi^2)C_{\{\phi,0\}}(0), \quad (\text{A30})$$

while for ARFIMA(1,  $d$ , 0) noise  $\epsilon_{\phi,d}$  and the corresponding ARFIMA(0,  $d$ , 0) noise  $\tilde{\epsilon}_{\phi,d}$ ,

$$\tilde{C}_{\{\phi,d\}}(0) = (1 + \phi^2 - 2\phi\rho)C_{\{\phi,d\}}(0), \quad (\text{A31})$$

where  $\rho$  is the lag 1 autocorrelation of ARFIMA(1,  $d$ , 0) given by

$$\rho = \frac{(1 + \phi^2)F(1, d; 1 - d; \phi) - 1}{\phi[2F(1, d; 1 - d; \phi) - 1]}, \quad (\text{A32})$$

where  $F(a, b; c, z)$  is the hypergeometric function.

Before tackling the calculation of  $\sigma^2[\hat{m}|\tilde{m}]_{\{\phi,d\}}$ , we first show how the relationship Eq. (A22) can be used to derive  $\sigma^2[\hat{m}|m]_{\{\phi,0\}}$  for AR(1) noise  $\epsilon_{\phi,0}$  from  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,0\}}$  for WN,  $\tilde{\epsilon}_{\phi,0} = \epsilon_{0,0}$ . From the unit variance of  $\epsilon$ , it follows that  $C(0) = 1$ . Therefore,  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,0\}}$

should be accordingly renormalized:  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,0\}} = (1 - \phi^2)\sigma^2[\hat{m}|m]_{\{0,0\}}$ .

For normalized AR(1) noise  $C_{\{\phi,0\}}(s) = \phi^s$ , and for WN  $\tilde{C}_{\{0,0\}}(s) = 0$  if  $s \geq 1$ . In this sense, Eq. (A29) can be specified as

$$\begin{aligned} \sigma^2[\hat{m}|\tilde{m}]_{\{0,0\}} &= \sum_{t=1}^N \tilde{C}_{\{0,0\}}(0)a_t^2 + 2(1 - \phi)^2 \sum_{s=1}^{N-1} C_{\{\phi,0\}}(s)L(s) - 2A \\ &= (1 - \phi)^2 \sum_{t=1}^N C_{\{\phi,0\}}(0)a_t^2 + 2(1 - \phi)^2 \sum_{s=1}^{N-1} C_{\{\phi,0\}}(s)L(s) - 2A \\ &= (1 - \phi)^2 \left( \frac{1 - \phi^2}{(1 - \phi)^2} \sum_{t=1}^N C_{\{\phi,0\}}(0)a_t^2 + 2 \sum_{s=1}^{N-1} C_{\{\phi,0\}}(s)L(s) \right) - 2A \\ &= (1 - \phi)^2 \left( \sum_{t=1}^N C_{\{\phi,0\}}(0)a_t^2 + 2 \sum_{s=1}^{N-1} C_{\{\phi,0\}}(s)L(s) + \frac{2\phi}{1 - \phi} \sum_{t=1}^N C_{\{\phi,0\}}(0)a_t^2 \right) - 2A \\ &= (1 - \phi)^2 \left( \sum_{t=1}^N C_{\{\phi,0\}}(0)a_t^2 + 2 \sum_{s=1}^{N-1} C_{\{\phi,0\}}(s)L(s) \right) - 2A + \frac{2\phi}{1 - \phi} \sum_{t=1}^N \tilde{C}_{\{0,0\}}(0)a_t^2 \\ &= (1 - \phi)^2 \sigma^2[\hat{m}|m]_{\{\phi,0\}} - 2A + \frac{2\phi}{1 - \phi} \sum_{t=1}^N \tilde{C}_{\{0,0\}}(0)a_t^2. \end{aligned} \tag{A33}$$

Consequently, we can obtain  $\sigma^2[\hat{m}|m]_{\{\phi,0\}}$  from  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,0\}}$  as

$$\begin{aligned} \sigma^2[\hat{m}|m]_{\{\phi,0\}} &= \frac{1}{(1 - \phi)^2} \sigma^2[\hat{m}|\tilde{m}]_{\{0,0\}} + \frac{2}{(1 - \phi)^2} A - \frac{2\phi}{(1 + \phi)(1 - \phi)^2} \sum_{t=1}^N \tilde{C}_{\{0,0\}}(0)a_t^2, \\ &= \frac{1 - \phi^2}{(1 - \phi)^2} \sigma^2[\hat{m}|m]_{\{0,0\}} + \frac{2}{(1 - \phi)^2} A - \frac{2\phi}{(1 + \phi)(1 - \phi)^2} \sum_{t=1}^N \tilde{C}_{\{0,0\}}(0)a_t^2, \end{aligned} \tag{A34}$$

We can further simplify Eq. (A34) and obtain as expected the result Eq. (A15). Asymptotically,

$$\begin{aligned} &\frac{2}{(1 - \phi)^2} A - \frac{2\phi}{(1 + \phi)(1 - \phi)^2} \sum_{t=1}^N \tilde{C}_{\{0,0\}}(0)a_t^2 \\ &\sim -72 \frac{\phi_1(1 + \phi^N)}{(1 - \phi)^2} N^{-4}. \end{aligned} \tag{A35}$$

Therefore, the asymptotic behavior of  $\sigma^2[\hat{m}|m]_{\{\phi,0\}}$  is dominated by

$$\frac{1 - \phi^2}{(1 - \phi)^2} \sigma^2[\hat{m}|m]_{\{0,0\}} \sim 12 \left( \frac{1 + \phi}{1 - \phi} \right) N^{-3},$$

consistent with that previously obtained in Eq. (A15).

For ARFIMA(1,  $d$ , 0) noise  $\epsilon_{\{\phi,d\}}$ , it is too complicated to calculate the corresponding variance  $\sigma^2[\hat{m}|m]_{\{1,d\}}$  directly based on Eq. (A9). However, the above indirect calculation of  $\sigma^2[\hat{m}|m]_{\{\phi,0\}}$  from  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,0\}}$  based on the relationship Eq. (A22) provides us an alternative way to calculate  $\sigma^2[\hat{m}|m]_{\{\phi,d\}}$  from  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,d\}}$ , where  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,d\}}$  has already been obtained in the previous subsection. For ARFIMA( $\phi$ ,  $d$ , 0)  $\epsilon_{\{\phi,d\}}$  and the corresponding

ARFIMA(0,  $d$ , 0)  $\tilde{\epsilon}_{\{\phi,d\}}$ ,

$$C_{\{0,d\}} = \tilde{C}_{\{\phi,d\}}(s) = \frac{\Gamma(s + d)\Gamma(1 - d)}{\Gamma(s + 1 - d)\Gamma(d)} (1 + \phi^2 - 2\phi\rho), \tag{A36}$$

and

$$\begin{aligned} C_{\{\phi,d\}}(s) &= \frac{\tilde{C}_{\{\phi,d\}}(s)}{1 - \phi^2} (F(1, d + s; 1 - d + s; \phi) \\ &\quad + F(1, d - s; 1 - d - s; \phi) - 1). \end{aligned} \tag{A37}$$

It should be noted that  $\tilde{C}_{\{0,d\}}(s)$  is different from Eq. (A16), because it has to be renormalized to ensure the unit variance of  $\epsilon$ . Accordingly,  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,d\}} = (1 + \phi^2 - 2\phi\rho)\sigma^2[\hat{m}|m]_{\{0,d\}}$ . Based on the relationship from Eq. (A22), the closed form of  $\sigma^2[\hat{m}|m]_{\{\phi,d\}}$  can be calculated from  $\sigma^2[\hat{m}|\tilde{m}]_{\{0,d\}}$  as

$$\begin{aligned} \sigma^2[\hat{m}|m]_{\{\phi,d\}} &= \frac{(1 + \phi^2 - 2\phi\rho)}{(1 - \phi)^2} \sigma^2[\hat{m}|m]_{\{0,d\}} + \frac{2}{(1 - \phi)^2} A \\ &\quad - \frac{2\phi(1 - \rho)}{(1 + \phi^2 - 2\phi\rho)(1 - \phi)^2} \sum_{t=1}^N \tilde{C}_{\{0,d\}}(0)a_t^2, \end{aligned} \tag{A38}$$

with  $\sigma^2[\hat{m}|m]_{\{0,d\}}$ ,  $A$ , and  $\tilde{C}_{\{0,d\}}(0)$  given in Eqs. (A18), (A28), and (A31), respectively. We note that

$$\begin{aligned} & \frac{2}{(1-\phi)^2}A - \frac{2\phi(1-\rho)}{(1+\phi^2-2\phi\rho)(1-\phi)^2} \sum_{i=1}^N \tilde{C}_{\{0,d\}}(0)a_i^2 \\ & \sim -\frac{24\phi}{(1-\phi)^2}((\phi-\rho)N^{-3} + (3\phi^N + \phi - \rho + 3)N^{-4}). \end{aligned} \quad (\text{A39})$$

The asymptotic behavior of  $\sigma^2[\hat{m}|m]_{\{\phi,d\}}$  is thus given by

$$\begin{aligned} \sigma^2[\hat{m}|m]_{\{\phi,d\}} & \sim \frac{(1+\phi^2-2\phi\rho)}{(1-\phi)^2} \sigma^2[\hat{m}|m]_{\{0,d\}} \\ & = \frac{(1+\phi^2-2\phi\rho)}{(1-\phi)^2} f(0,d)N^{2d-3} \\ & = f(\phi,d)N^{2d-3}, \end{aligned} \quad (\text{A40})$$

where  $\rho$  and  $f(0,d)$  are given by Eqs. (A32) and (A19), respectively.

### 3. Criteria under which the trend is in PC<sub>1</sub>

In the following, we show that if the first EOF is homogeneous and the trend is the same for all grid points, then the trend will be completely contained within the first PC. Without loss of generality, we assume evenly spaced Cartesian grid points,

$$T_i(x,y) = n(x,y) + m(x,y)t + \epsilon_i(x,y), \quad (\text{A41})$$

where, as usual, the anomalies  $T_i(x,y)$  are constructed to have zero mean on the considered time interval. We use the notation  $A_i = (x_j, y_k)$ , where  $i, j, k$  are integers and each  $i$  corresponds to a unique pair  $j, k$ .

Correlations between the noise are quantified by the covariance matrix  $\mathbb{C}$  with  $(\mathbb{C})_{ij} = \langle \epsilon(A_i)\epsilon(A_j) \rangle$ . We first as-

sume that all time series have the same deterministic trend  $m(x,y) = m$  and  $n(x,y) = n$  for all grid points. We now calculate the covariance matrix of the time series with these trends,

$$\begin{aligned} \langle T_i(A_i)T_j(A_j) \rangle & = \langle (n+mt + \epsilon_i(A_i))(n+mt + \epsilon_j(A_j)) \rangle \\ & = (\mathbb{C})_{ij} + \langle (n+mt)^2 \rangle \\ & = (\mathbb{C})_{ij} + m^2(\langle t^2 \rangle - \langle t \rangle^2), \end{aligned} \quad (\text{A42})$$

where we have used that  $n = -m\langle t \rangle$  due to the series having zero mean. We write this in matrix form as  $(\mathbb{M})_{ij} = m^2(\langle t^2 \rangle - \langle t \rangle^2)$ . We make use of a decomposition of the covariance matrix, whose eigenvectors are orthogonal to each other because of its symmetry, into matrices formed by its eigenvectors, which we call the set of eigenmatrices:  $\mathbb{C} = \sum_{i=1}^n \lambda_i \mathbb{V}_i$ , where  $\mathbb{V}_i = \vec{v}_i \otimes \vec{v}_i^\dagger$ , where  $\vec{v}_i$  is the eigenvector corresponding to eigenvalue  $\lambda_i$ . Notice that each individual  $\mathbb{V}_i$  is a real symmetric matrix. Evidently, if we add to  $\mathbb{C}$  the matrix  $a\mathbb{V}_l$  for given  $l$ , then the eigenvalues of the new matrix are  $\lambda_l + a$  and all other eigenvalues unchanged. This is true because  $\mathbb{V}_i \vec{v}_j = \delta_{ij} \vec{v}_i$  due to orthogonality of the eigenvectors.

In the case of strong positive correlations among the noises  $\epsilon_i(t)$  as for our gridded temperature data, the first EOF,  $\vec{v}_1$ , is usually close to a constant, i.e.,  $v_{1,i} \approx c \forall i$ , and hence  $\mathbb{V}_1$  is a good approximation to the fully degenerate matrix  $\mathbb{M}$  (up to some scalar factor). If so, then the trend  $m$  in the time series should affect the eigenvalue  $\lambda_1$  of the covariance matrix of  $\vec{T}$ , and also the nonstationarity of  $T_i(A_i)$  due to the term  $m_i t$  should predominantly be visible in the first PC  $\mathcal{P}_{t,1}$ .

This reasoning becomes exact if two requirements are fulfilled: The EOF<sub>1</sub> of the covariance matrix, i.e., its eigenvector  $\vec{v}_1$ , is precisely a multiple of the vector  $(1, 1, 1, \dots, 1)$  and the trend is exactly the same for all grid points. If either of these conditions is violated, the additional trend matrix  $\mathbb{M}$  in the covariances of Eq. (1) cannot be mapped onto a single EOF, and hence the trend will not be fully contained in  $\mathcal{P}_{t,1}$ . In this (more realistic) case, the trend contained in  $\mathcal{P}_{t,1}$  will be an underestimation of the true trend.

- 
- [1] T. Stocker, D. Qin, G. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, *Climate Change 2013: The Physical Science Basis* (Cambridge University Press, Cambridge, UK, 2014).
- [2] S. Lovejoy, Scaling fluctuation analysis and statistical hypothesis testing of anthropogenic warming, *Clim. Dyn.* **42**, 2339 (2014).
- [3] S. Lovejoy, Return periods of global climate fluctuations and the pause, *Geophys. Res. Lett.* **41**, 4704 (2014).
- [4] P. Y. Groisman, R. W. Knight, T. R. Karl, D. R. Easterling, B. Sun, and J. H. Lawrimore, Contemporary changes of the hydrological cycle over the contiguous united states: Trends derived from in situ observations, *J. Hydrometeorol.* **5**, 64 (2004).
- [5] M. Becker, M. Karpytchev, and S. Lennartz-Sassinek, Long-term sea level trends: Natural or anthropogenic? *Geophys. Res. Lett.* **41**, 5571 (2014).
- [6] M. E. Mann, R. S. Bradley, and M. K. Hughes, Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophys. Res. Lett.* **26**, 759 (1999).
- [7] R. P. Allan and B. J. Soden, Atmospheric warming and the amplification of precipitation extremes, *Science* **321**, 1481 (2008).
- [8] G. A. Meehl, J. M. Arblaster, J. T. Fasullo, A. Hu, and K. E. Trenberth, Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods, *Nat. Clim. Change* **1**, 360 (2011).
- [9] J. Ludescher, A. Bunde, C. L. E. Franzke, and H. J. Schellnhuber, Long-term persistence enhances uncertainty about anthropogenic warming of antarctica, *Clim. Dyn.* **46**, 263 (2016).
- [10] H. F. Lins and J. R. Slack, Streamflow trends in the United States, *Geophys. Res. Lett.* **26**, 227 (1999).
- [11] M. Mudelsee, M. Borngen, G. Tetzlaff, and U. Grunewald, No upward trends in the occurrence of extreme floods in central Europe, *Nature* **425**, 166 (2003).
- [12] H. E. Hurst, Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.* **116**, 770 (1951).
- [13] H. E. Hurst, A suggested statistical model of some time series which occur in nature, *Nature* **180**, 494 (1957).

- [14] K. E. Trenberth, A. Dai, R. M. Rasmussen, and D. B. Parsons, The changing character of precipitation, *Bull. Amer. Meteor. Soc.* **84**, 1205 (2003).
- [15] K. E. Trenberth, J. Fasullo, and L. Smith, Trends and variability in column-integrated atmospheric water vapor, *Climate Dyn.* **24**, 741 (2005).
- [16] S. Kao and A. R. Ganguly, Intensity, duration, and frequency of precipitation extremes under 21st-century warming scenarios, *J. Geophys. Res. Atmos.* **116**, D16119 (2011).
- [17] J. W. Dippner and C. Pohl, Trends in heavy metal concentrations in the western and central baltic sea waters detected by using empirical orthogonal functions analysis, *J. Mar. Syst.* **46**, 69 (2004).
- [18] L. X. Clegg, B. F. Hankey, R. Tiwari, E. J. Feuer, and B. K. Edwards, Estimating average annual per cent change in trend analysis, *Stat. Med.* **28**, 3670 (2009).
- [19] M. J. Crosse, N. J. Zuk, G. M. Di Liberto, A. R. Nidiffer, S. Molholm, and E. C. Lalor, Linear modeling of neurophysiological responses to speech and other continuous stimuli: methodological considerations for applied research, *Front. Neurosci.* **15**, 705621 (2021).
- [20] L. S. Hamilton and A. G. Huth, The revolution will not be controlled: Natural stimuli in speech neuroscience, *Lang. Cognit. Neurosci.* **35**, 573 (2020).
- [21] M. C. K. Wu, S. V. David, and J. L. Gallant, Complete functional characterization of sensory neurons by system identification, *Annu. Rev. Neurosci.* **29**, 477 (2006).
- [22] T. A. Cohn and H. F. Lins, Nature's style: Naturally trendy, *Geophys. Res. Lett.* **32**, L23402 (2005).
- [23] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control* (John Wiley & Sons, Hoboken, New Jersey, USA, 2015), p. 440.
- [24] F. W. Zwiers and H. Von Storch, Taking serial correlation into account in tests of the mean, *J. Clim.* **8**, 336 (1995).
- [25] P. Bloomfield and D. Nychka, Climate spectra and detecting climate change, *Clim. Change* **21**, 275 (1992).
- [26] A. Kulkarni and H. von Storch, Monte Carlo experiments on the effect of serial correlation on the Mann-Kendall test of trend, *Meteorol. Z.* **4**, 82 (1995).
- [27] E. Koscielny-Bunde, A. Bunde, S. Havlin, H. E. Roman, Y. Goldreich, and H. J. Schellnhuber, Indication of a universal persistence law governing atmospheric variability, *Phys. Rev. Lett.* **81**, 729 (1998).
- [28] R. Blender and K. Fraedrich, Long time memory in global warming simulations, *Geophys. Res. Lett.* **30**, 1769 (2003).
- [29] K. Fraedrich and R. Blender, Scaling of atmosphere and ocean temperature correlations in observations and climate models, *Phys. Rev. Lett.* **90**, 108501 (2003).
- [30] J. W. Kantelhardt, E. Koscielny-Bunde, D. Rybski, P. Braun, A. Bunde, and S. Havlin, Long-term persistence and multifractality of precipitation and river runoff records, *J. Geophys. Res. Atmos.* **111**, D01106 (2006).
- [31] A. Bunde, J. F. Eichner, J. W. Kantelhardt, and S. Havlin, Long-Term Memory: A Natural Mechanism for the Clustering of Extreme Events and Anomalous Residual Times in Climate Records, *Phys. Rev. Lett.* **94**, 048701 (2005).
- [32] C. Franzke, On the statistical significance of surface air temperature trends in the eurasian arctic region, *Geophys. Res. Lett.* **39**, L23705 (2012).
- [33] P. F. Craigmile, P. Guttorp, and D. B. Percival, Trend assessment in a long memory dependence model using the discrete wavelet transform, *Environmetrics* **15**, 313 (2004).
- [34] A. Bunde, J. Ludescher, C. L. E. Franzke, and U. Büntgen, How significant is West Antarctic warming? *Nat. Geosci.* **7**, 246 (2014).
- [35] E. Zorita, T. F. Stocker, and H. von Storch, How unusual is the recent series of warm years? *Geophys. Res. Lett.* **35**, L24706 (2008).
- [36] D. Rybski, A. Bunde, S. Havlin, and H. von Storch, Long-term persistence in climate and the detection problem, *Geophys. Res. Lett.* **33**, L06718 (2006).
- [37] D. Rybski and A. Bunde, On the detection of trends in long-term correlated records, *Physica A* **388**, 1687 (2009).
- [38] S. Lennartz and A. Bunde, Trend evaluation in records with long-term memory: Application to global warming, *Geophys. Res. Lett.* **36**, L16706 (2009).
- [39] A. Tamazian, J. Ludescher, and A. Bunde, Significance of trends in long-term correlated records, *Phys. Rev. E* **91**, 032806 (2015).
- [40] M. Massah and H. Kantz, Confidence intervals for time averages in the presence of long-range correlations, a case study on earth surface temperature anomalies, *Geophys. Res. Lett.* **43**, 9243 (2016).
- [41] S. Fatichi, S. M. Barbosa, E. Caporali, and M. E. Silva, Deterministic versus stochastic trends: Detection and challenges, *J. Geophys. Res. Atmos.* **114**, D18121 (2009).
- [42] E. Hawkins and R. Sutton, The potential to narrow uncertainty in regional climate predictions, *Bull. Am. Meteorol. Soc.* **90**, 1095 (2009).
- [43] M. Begert and C. Frei, Long-term area-mean temperature series for Switzerland-combining homogenized station data and high resolution grid data, *Int. J. Climatol.* **38**, 2792 (2018).
- [44] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences* (Academic Press, Cambridge, MA 02139, USA, 2011), Vol. 100.
- [45] M. Widmann and C. Schär, A principal component and long-term trend analysis of daily precipitation in Switzerland, *Int. J. Climatol.* **17**, 1333 (1997).
- [46] C. K. Folland, N. A. Rayner, S. J. Brown, T. M. Smith, S. S. P. Shen, D. E. Parker, I. Macadam, P. D. Jones, R. N. Jones, N. Nicholls *et al.*, Global temperature change and its uncertainties since 1861, *Geophys. Res. Lett.* **28**, 2621 (2001).
- [47] C. W. J. Granger and R. Joyeux, An introduction to long-memory time series models and fractional differencing, *J. Time Ser. Anal.* **1**, 15 (1980).
- [48] J. R. M. Hosking, Modeling persistence in hydrological time series using fractional differencing, *Water Resour. Res.* **20**, 1898 (1984).
- [49] B. B. Mandelbrot and J. W. Van Ness, Fractional Brownian motions, fractional noises and applications, *SIAM Rev.* **10**, 422 (1968).
- [50] M. S. Peiris and B. J. C. Perera, On prediction with fractionally differenced ARIMA models, *J. Time Ser. Anal.* **9**, 215 (1988).
- [51] Historical daily station observations for germany, version v004, Deutscher Wetterdienst (DWD), Offenbach, Germany. Climate Data Center (CDC). Station: Potsdam, ID 3987, data file tageswerte\_KL\_03987\_18930101\_20221231\_hist.zip,

- [https://opendata.dwd.de/climate\\_environment/CDC/observations\\_germany/climate/daily/kl/historical/](https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/historical/).
- [52] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis* (John Wiley & Sons, Hoboken, New Jersey, USA, 2021).
- [53] S. Seabold and J. Perktold, Statsmodels: Econometric and statistical modeling with Python, in *9th Python in Science Conference*, edited by S. van der Walt and J. Millman (SciPy, 2010), pp. 92–96.
- [54] R. S. Deo and C. M. Hurvich, Linear trend with fractionally integrated errors, *J. Time Ser. Anal.* **19**, 379 (1998).
- [55] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Mosaic organization of DNA nucleotides, *Phys. Rev. E* **49**, 1685 (1994).
- [56] M. Höll, K. Kiyono, and H. Kantz, Theoretical foundation of detrending methods for fluctuation analysis such as detrended fluctuation analysis and detrending moving average, *Phys. Rev. E* **99**, 033305 (2019).
- [57] M. Höll and H. Kantz, The relationship between the detrended fluctuation analysis and the autocorrelation function of a signal, *Eur. Phys. J. B* **88**, 327 (2015).
- [58] P. G. Meyer and H. Kantz, Inferring characteristic timescales from the effect of autoregressive dynamics on detrended fluctuation analysis, *New J. Phys.* **21**, 033022 (2019).
- [59] V. Deshmukh, E. Bradley, J. Garland, and J. D. Meiss, Toward automated extraction and characterization of scaling regions in dynamical systems, *Chaos* **31**, 123102 (2021).
- [60] P. Abry and D. Veitch, Wavelet analysis of long-range-dependent traffic, *IEEE Trans. Inf. Theory* **44**, 2 (1998).
- [61] P. F. Craigmile, P. Guttorp, and D. B. Percival, Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes, *IEEE Trans. Signal Process.* **53**, 3151 (2005).
- [62] J. A. Kassel and H. Kantz, Statistical inference of one-dimensional persistent nonlinear time series and application to predictions, *Phys. Rev. Res.* **4**, 013206 (2022).
- [63] E. C. Weatherhead, G. C. Reinsel, G. C. Tiao, C. H. Jackman, L. Bishop, S. M. H. Frith, J. DeLuisi, T. Keller, S. J. Oltmans, E. L. Fleming *et al.*, Detecting the recovery of total column ozone, *J. Geophys. Res. Atmos.* **105**, 22201 (2000).
- [64] J. D. Cryer and K. Chan, *Time Series Analysis With Applications in R*, 2nd ed. (Springer, Berlin, 2008).
- [65] G. C. Tiao, G. C. Reinsel, D. Xu, J. H. Pedrick, X. Zhu, A. J. Miller, J. J. DeLuisi, C. L. Mateer, and D. J. Wuebbles, Effects of autocorrelation and temporal sampling schemes on estimates of trend and spatial correlation, *J. Geophys. Res. Atmos.* **95**, 20507 (1990).
- [66] E. C. Weatherhead, G. C. Tiao, G. C. Reinsel, J. E. Frederick, J. J. DeLuisi, D. Choi, and W. Tam, Analysis of long-term behavior of ultraviolet radiation measured by Robertson-Berger meters at 14 sites in the united states, *J. Geophys. Res. Atmos.* **102**, 8737 (1997).
- [67] E. C. Weatherhead, G. C. Reinsel, G. C. Tiao, X.-L. Meng, D. Choi, W. Cheang, T. Keller, J. DeLuisi, D. J. Wuebbles, J. B. Kerr *et al.*, Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res. Atmos.* **103**, 17149 (1998).
- [68] J. Muñoz-Sabater, E. Dutra, A. Agustí-P, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach *et al.*, ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth System Science Data* **13**, 4349 (2021).
- [69] K. Cowtan, P. Jacobs, P. Thorne, and R. Wilkinson, Statistical analysis of coverage error in simple global temperature estimators, *Dyn. Stat. Clim. Syst.* **3**, dzy003 (2018).
- [70] E. N. Lorenz, *Empirical Orthogonal Functions and Statistical Weather Prediction* (Massachusetts Institute of Technology, Department of Meteorology, Cambridge, 1956), Vol. 1.
- [71] B. C. Weare and J. S. Nasstrom, Examples of extended empirical orthogonal function analyses, *Mon. Weather Rev.* **110**, 481 (1982).