OPEN

Extracting climate memory using Fractional Integrated Statistical Model: A new perspective on climate prediction

Naiming Yuan[1,2,3], Zuntao Fu[1] & Shida Liu[1]

[1]Lab for Climate and Ocean-Atmosphere Studies, Dept. of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, 100871, China, [2]Chinese Academy of Meteorological Science, Beijing, 100081, China, [3]Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus Liebig University Giessen, Senckenbergstrasse 1, 35390 Giessen, Germany.

Correspondence and requests for materials should be addressed to N.M.Y. (naimingyuan@ hotmail.com) or Z.T.F. (fuzt@pku.edu.cn)

Long term memory (LTM) in climate variability is studied by means of fractional integral techniques. By using a recently developed model, Fractional Integral Statistical Model (FISM), we in this report proposed a new method, with which one can estimate the long-lasting influences of historical climate states on the present time quantitatively, and further extract the influence as climate memory signals. To show the usability of this method, two examples, the Northern Hemisphere monthly Temperature Anomalies (NHTA) and the Pacific Decadal Oscillation index (PDO), are analyzed in this study. We find the climate memory signals indeed can be extracted and the whole variations can be further decomposed into two parts: the cumulative climate memory (CCM) and the weather-scale excitation (WSE). The stronger LTM is, the larger proportion the climate memory signals will account for in the whole variations. With the climate memory signals extracted, one can at least determine on what basis the considered time series will continue to change. Therefore, this report provides a new perspective on climate prediction.

C limate memory has been a well known concept ever since the middle of last century[1]. Affected by the slowly responding subsystems (such as the ocean), the climate variability usually exhibits long-term memory (LTM), which means the present climate states may have long-term influences on the states in far future. Normally, this kind of characteristic is considered as fractal properties or scaling behaviors in climate[2]. During the past few decades, benefited from the development of nonlinear techniques, many methods have been proposed for better diagnosing this characteristic, including spectral analysis[2,3], structure function method[4], wavelet analysis (WA)[5], as well as methods based on random walking theory, such as the rescaled-range (R/S) analysis[1], and detrended fluctuation analysis (DFA)[6], etc. With these methods, extensive researches have been done in the past few years, which were mainly focused on i) the diagnosis of fractal and multifractal properties in different climatic variables[7–13], ii) the influence of LTM on the internal stochastic trends and clustering of extremes[14–18], and iii) the capability of climate models in reproducing the scaling behaviors in different scales[19–24], as summarized by[25]. However, on the question of whether one can improve the climate predictability by using the diagnosed scaling behaviors, or more vividly, climate memory, only few explorations have been made[26].

Ever since the concept of stochastic climate model was first introduced by[27] in 1976, one begins to believe that the low frequency variability (LFV) of climate can be explained as the integral response to continuous random excitation by short period weather-scale disturbances (weather noise). The climate-weather system was summarized in terms of the Brownian motion analogy[28],

$$\frac{dx}{dt} = \varepsilon(t), \qquad (1)$$

where ε represents the "weather-scale" excitations, and x stands for "climate-scale" variability. However, after years of research, it has been noticed that calculations simply based on one-order integration sometimes cannot provide satisfactory simulations of many natural time series, especially when modeling time series with different fractal properties or LTM strengths. Therefore, by using fractional integral techniques, we recently established a new statistical model, Fractional Integral Statistical Model (FISM)[29]. FISM is actually based on fractional Brownian motion, which was first introduced by Kolmogorov in 1940 in[30]. With this model, one can reproduce LTM (or fractal properties) of any given climatic time series. By applying the Riemann-Liouville fractional integral formula, the model can be written explicitly as,

$$x(t) = \frac{1}{\Gamma(q)} \int_0^t \frac{\varepsilon(u)}{(t-u)^{1-q}} \, du, \tag{2}$$

where t stands for the present time point, u stands for a historical time point, $t - u$ represents the distance between the historical time point u and the present time t, and $\Gamma(q)$ denotes the Gamma function. The fractional integration starts from $u = 0$, which stands for the beginning of history. It is worth to note that normally people prefers to use negative infinity $(-\infty)$ as the beginning of history, since it is more convenient for understanding and interpreting. But considering further technology requirements (see the "Results" section and the "Methods" section), we choose to use $u = 0$, supposing that we know the starting point of history (more detailed discussion can be found in the "Results" section). By adjusting the integral-order q, one can simulate the climate-scale variability $x(t)$ with proper LTM. Different q-values determine different LTM strengths of $x(t)$. For quasi-Gaussian processes, it is easy to prove that q has one-to-one relations with other exponents such as the DFA exponent α (Detrended Fluctuation Analysis, a widely used method which can detect the strength of LTM[6,31],), or the PSD exponent β (Power Spectrum Density Analysis)[32], as

$$2q = \beta = 2\alpha - 1, \tag{3}$$

$q = 0$ means there is no integration procedures, the simulated climate-scale variability $x(t)$ has no memory and acts as white noise ($\alpha = 0.5$, $\beta = 0$). While $q > 0$ means the climate-scale variability is simulated by applying q-order integration to the weather-scale excitation $\varepsilon(u)$, which will simultaneously introduce the so-called long-term memory into $x(t)$ ($\alpha > 0.5$, $\beta > 0$) (see Fig. 2 in[29]). Therefore, the Fractional Integral Statistical Model (FISM) can be considered as a useful model in simulating LTM in climate. Although the statistics in this model are quasi-Gaussian, complex features such as intermittency, multifractality have not been taken into account, due to its reasonable physical framework as described above, we believe this model also useful in further studying LTM in climate.

Based on this model, combine with the concept of stochastic climate model proposed by[27], one can see clearly that, the "climate-scale" variability is considered as q-order integrations of "weather-scale" excitations. That is, to express it more explicitly, the WSE has long-lasting influences on the climate states. From this perspective, a new understanding of the climate variation can be reached, that is, any given climatic time series can be decomposed into two components: the cumulative climate memory signals (CCM) and the weather-scale excitation (WSE)[29],

$$x(t) = M(t) + \varepsilon(t). \tag{4}$$

$M(t)$ represents the cumulative climate memory signals (CCM), which is the cumulated long-lasting influences integrated from historical WSE, while $\varepsilon(t)$ stands for the present weather-scale excitation (WSE). Both the cumulative climate memory signals (CCM) and the present weather-scale excitations (WSE) composes the present climate states. If one can calculate the memory signals $M(t)$ and further estimate the influences of historical information quantitatively, our climate prediction skills will have a high probability to be improved. At least, we can determine on what basis the considered climate states will continue to be triggered by the present weather scale excitations. Therefore, how to extract the climate memory signals is our main focal point of this study.

Suppose we do not know the present state of $x(t)$, with FISM, theoretically the memory signals can be calculated from the historical WSE as,

$$M(t) = \frac{1}{\Gamma(q)} \int_{u=0}^{t-\delta} \frac{\varepsilon(u)}{(t-u)^{1-q}} \, du \tag{5}$$

Where $\varepsilon(0), \varepsilon(\delta), \cdots, \varepsilon(t-\delta)$ are the historical WSE, and δ is the sampling intervals of the observed time series. The difference between equation (5) and equation (2) is that we only use the historical WSE to calculate the memory signals $M(t)$. As for the integrations from $t - \delta$ to t, considering the sampling interval δ and the singular nature of equation (2), we choose to use the present WSE, $\varepsilon(t)$, to replace the integration from $t - \delta$ to t. From mathematical point of view, this procedure may be problematic, but from a practical point of view, it is reasonable since i) the integration will always need a cutoff due to the discrete sampling and the singularity; ii) even if we can divide the interval ($t - \delta$ to t) into small pieces with width of $\Delta\delta$, the estimated integrating factor (we also name it as memory kernel, see the following),

$$k = \sum_{t=\Delta\delta}^{\delta-\Delta\delta} \frac{dl}{\Gamma(q)(\delta-l)^{1-q}}, \tag{6}$$

is really around 1 (figures not shown here); and the most importantly, iii) it is in line with our above understandings, that any climatic time sereis is composed by historical cumulated influence $M(t)$ and the present WSE $\varepsilon(t)$. Therefore, equation (2) can be further written as,

$$x(t) = \frac{1}{\Gamma(q)} \int_{u=0}^{t-\delta} \frac{\varepsilon(u)}{(t-u)^{1-q}} \, du + \varepsilon(t), \tag{7}$$

or even in a discrete form,

$$x(t) = K(q)_t^\delta \otimes \Psi_0^{t-\delta} + \varepsilon(t), \tag{8}$$

where $K(q)_t^\delta \equiv \{ k(q; t), k(q; t-\delta), \cdots, k(q; t-u), \cdots, k(q; \delta) \}$, $k(q; t-u) = \dfrac{1}{\Gamma(q)(t-u)^{1-q}}$, represents the integrating factor of every single step, and $\Psi \equiv \{ \varepsilon(0), \varepsilon(\delta), \cdots, \varepsilon(u), \cdots, \varepsilon(t-\delta) \}$, represents the WSE at each historical time point. It is worth to note that the superscript and subscript of $K(q)_t^\delta$ denote the time distances between historical time points and the present time, like $t - u$ in equation (2), from $t - u = t$ (corresponding to $u = 0$) to $t - u = \delta$ (corresponding to $u = t - \delta$), while the superscript and subscript of $\Psi_0^{t-\delta}$ denote the historical time point, like u in equation (2), from $u = 0$ to $u = t - \delta$. Since the integrating factor $k(q; t - u)$ describes to what extent the historical WSE $\varepsilon(u)$ may affect the present state $x(t)$, we name it as the Memory Kernel (MK for short), see[29]. Obviously, MK is controlled by the integral-order q, or more precisely, by the strength of LTM. To extract the memory signals, one thus needs to make an accurate estimation of the historical WSE first. A general idea is that we may derive equation (8) reversely and obtain the historical WSE (time) point by (time) point from the historical observations $x(0), x(\delta), \cdots, x(t-\delta)$. With the estimated historical WSE, one thus can calculate the memory signals simply by using equation (5). Although we cannot make an estimation of the present WSE, the memory signals originated from the historical information are still important and useful for the climate prediction.

In the following, we will make a detailed discussion on this issue. We first make a simple test to determine whether the historical WSE can be estimated by deriving equation (8) reversely, and under what conditions the estimation is accurate enough for a further extraction of climate memory. Then two examples, the Northern Hemisphere monthly Temperature Anomalies (NHTA) and the Pacific Decadal Oscillation index (PDO), are used to illustrate how to extract climate memory by using FISM. Obviously, the stronger climate memory is, the larger proportion CCM should be in the whole variations of $x(t)$, therefore, this work suggest a new perspective on improving the climate prediction skills. For the whole procedures of how to extract the climate memory signals, a brief description can be found in the end of this paper.

Results

Estimation of the historical WSE. From the above discussion, a general idea to extract the historical WSE is to derive equation (8) reversely. Suppose $t = 0$ is the beginning of history, the climate state thus will not be influenced by memory signals, and $x(0) = \varepsilon(0)$. For the next time point $t = \delta$, from equation (8), $\varepsilon(\delta)$ will be calculated as $\varepsilon(\delta) = x(\delta) - K \otimes \Psi$. In this way, one can extract the historical WSE point by point exactly. However, before extracting, two crucial issues need to be addressed. Firstly, one needs to know exactly the starting point of history; and secondly, one should have the observations from the starting points of history till now. As for the first issue, it is impossible to find the starting point of history. Normally, people uses the negative infinite $(-\infty)$ to represent the far away history. When making calculations (such as integration, etc.), cutoffs are usually selected. Without affecting the calculation accuracy, we in our method choose $t = 0$ as the history starting point. It should be emphasized that we do not ignore the historical information before $t = 0$, but just define that the history starts at $t = 0$ (there are no historical information before $t = 0$). In this case, the first issue can be addressed and the reverse derivation of equation (8) becomes possible. But for the second issue, the reality is that we do not have the observation at the beginning of history. In fact, our observations only started about hundred years ago. Climate states before our observations will of course have influences on the present state, although the influences should be very small. In this case, if we starts our extraction at the beginning of "observation", does our method still work, or under what conditions the extraction of historical WSE is accurate enough for our further analysis, is the key issues need to be solved.

In the following, we will address this issue by analyzing artificially generated data. We generate two artificial time series by using FISM, Data I and Data II, see Fig. 1(b) and Fig. 1(c). They are integrated from the same white noise, $\{\varepsilon_0\}$, see Fig. 1(a), but with different integral order q. The white noise here can be considered as the weather-scale excitations (WSE), while Data I and Data II can be considered as the "observed" time series $x(t)$. For Data I, $q = 0.2$, indicates the generated time series is long-term correlated with DFA exponent $a = 0.7$. For Data II, $q = 0.5$, indicates the generated time series is long-term correlated with DFA exponent $\alpha = 1.0$. As discussed above, by deriving equation (2) reversely, we can obtain the historical WSE from both Data I and Data II, as shown in Fig. 2(a1) and Fig. 2(b1) (the red curves). By comparing the estimated WSE with the white noise $\{\varepsilon_0\}$ (the black curve in Fig. 2), one therefore can examine whether our estimation of WSE is reliable. In this study, we suppose the index of 5001 is the beginning time of "observation". Therefore, we begin our extraction of WSE at index of 5001, and "have to" ignore the values before. We choose a short window (index from 5001 to 5200) as our experimental subjects. As shown in Fig. 2, at the beginning of the estimation (index from 5001 to 5025), there are big errors, while as time goes by, the errors become smaller. The appearance of errors is understandable, since we start our estimation at the index of 5001. Historical information before (index before 5001) is ignored in our calculation. That is why big errors arise at the beginning. However, as time goes by, the influence of the ignored historical information (index before 5001) will decay considerably, which thus result in small errors. From Fig. 2, one can see that after 200 points, the relative error of the estimated WSE is only 1.7% (4.1%) for Data I (Data II). Therefore, we have confidence to believe that our estimation of historical WSE is reliable, but one has to sacrifice a certain number of data points at the beginning, to "wait" until the influence of unobserved history become very small, and make sure the estimation is accurate enough for the following procedures.

Extraction of climate memory. To the end of this section, we will show how to extract the so-called climate memory signals by using

FISM. We take two time series as example. One is the Northern Hemisphere monthly Temperature Anomalies (NHTA, from 1880 to 2011, downloaded from the National Aeronautics and Space Administration, NASA, http://data.giss.nasa.gov/gistemp/), and the other is the Pacific Decadal Oscillation index (PDO, from 1900 to 2011, downloaded from the Physical Sciences Division, National Oceanic and Atmospheric Administration, NOAA), as shown in Fig. 3. From the Detrended Fluctuation Analysis (here we use the second order, DFA-2, which can remove the effect of linear trends and has been widely used in climatic research, see[10,31,33]), it has been found that both time series are characterized by long-term memory. As shown in Fig. 4(a) and Fig. 4(c), the DFA exponent $\alpha = 0.85 \pm 0.01$ for NHTA, while $\alpha = 1.05 \pm 0.01$ for PDO (the black dots). We can also reach consistent results by using the Power Spectrum Density Analysis (PSD), as shown in Fig. 4(b) and Fig. 4(d), the PSD exponent $\beta = 0.7 \pm 0.05$ for NHTA, while $\beta = 1.1 \pm 0.05$ for PDO (the black curves). But considering the larger statistical uncertainties in PSD, we prefer using DFA as the method to detect the LTM strength. The appearance of LTM in these two time series is understandable. In fact, it is not new to find LTM in temperature records. Many previous researches have reported that temperature records from different regions are characterized by different LTM[33,34]. Normally, temperatures over ocean have strong LTM with $\alpha = 1.0$[35], over coastal area have a little weaker LTM with $\alpha = 0.65$[36], while over inner continents have the weakest LTM, even near to white noise with $\alpha = 0.5$[20]. Therefore, our DFA result for NHTA ($\alpha = 0.85 \pm 0.01$) looks reasonable since the Northern Hemisphere comprises both ocean and continent. As for the PDO, since it is defined as the leading EOF of monthly sea surface temperature anomalies (SSTA) over the North Pacific (poleward of $20°N$)[37], its PDO index (which is the standardized principal component time series) will of course be modulated by the slow changing effect of ocean. Therefore, strong LTM as we found ($\alpha = 1.05 \pm 0.01$) in PDO index is also reasonable.

Since the existence of LTM indicates the states from long-time ago may still have influences on the present states, how to estimate this long-lasting influence quantitatively is of great importance for our better understanding, or even predicting of the time variations of our considered time series. In the following, we will show our results on this issue. As discussed above, to estimate the memory signals quantitatively, one needs first to extract the historical WSE from the historical observations. See Fig. 3, we divide the time series into three periods: "Historical Info A (1880 ~ 1919)", "Historical Info B (1920 ~ 1991)", and "Test Zone (1992 ~ 2011)". By deriving equation (8) reversely, we can extract the historical WSE, as shown in Fig. 5. But the extracted WSE in "Historical Info A" will not be considered in our following procedures, since there might be big errors originated from the influences of unobserved historical information. As discussed in the last section, we have to sacrifice the extracted WSE in the period of "Historical Info A", and only use the WSE in "Historical Info B" for our following calculation.

It is worth to note that the dividing line between "Historical Info A" and "Historical Info B" is not arbitrarily determined. We choose it according to a numerical test. In the numerical test, we first generate plenty of artificial data by using FISM, as what we did in the previous section. For NHTA, we choose $q = 0.35$, to simulate the LTM in NHTA; while for PDO, we choose $q = 0.55$, to simulate the LTM in PDO index. By repeating the extraction procedures, we could tell how big the errors are at different time points. It is found that, if we choose the dividing line at the year of 1920, the errors of the extracted historical WSE for both time series will be smaller than 5% (detailed results are not shown). Therefore, we choose to sacrifice the extracted WSE before the year of 1920, as shown in Fig. 5.

Before we calculate the memory signals, the statistical properties of the extracted WSE in time period of "Historical Info B" are further checked by DFA and the power spectrum density analysis. As shown
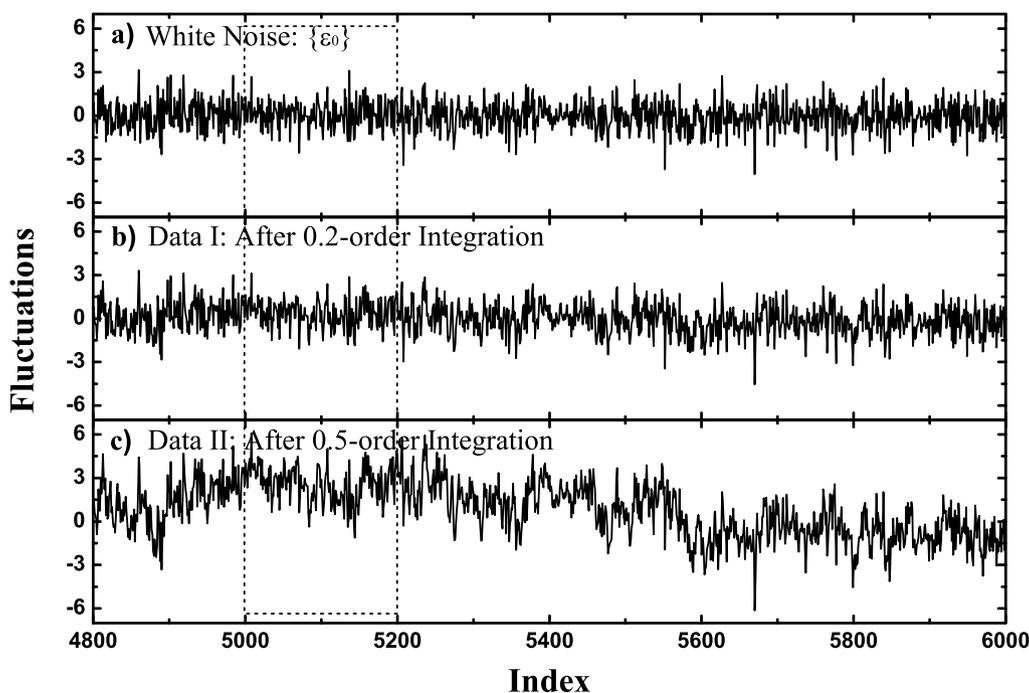
Figure 1 | Fragments of white noise $\{\varepsilon_0\}$ (a) and two artificially generated data: Data I (b) and Data II (c). Both data are generated from $\{\varepsilon_0\}$ by using FISM. For Data I, the integral order is $q = 0.2$, while for Data II, the integral order is $q = 0.5$. The dashed box shows the short window (index from 5001 to 5200) we choose for the detailed experiment in Fig. 2.

in Fig. 4 (the red dots and curves), it is clear to see that for both cases, the DFA exponent $\alpha = 0.5 \pm 0.01$ and the PSD exponent $\beta = 0 \pm 0.05$. The extracted WSE are not characterized by LTM. Therefore, we are more convinced that our extraction of the historical WSE is accurate and suitable for our further calculations.

With equation (5), we finally complete the extraction of the memory signals "hidden" in both the NHTA and the PDO time series. The period of 1992 ~ 2011 are chosen as the "Test Zone", to compare the extracted memory signals $M(t)$ with the observed time series. As shown in Fig. 6(a) and Fig. 6(c), the memory signals (red curves)
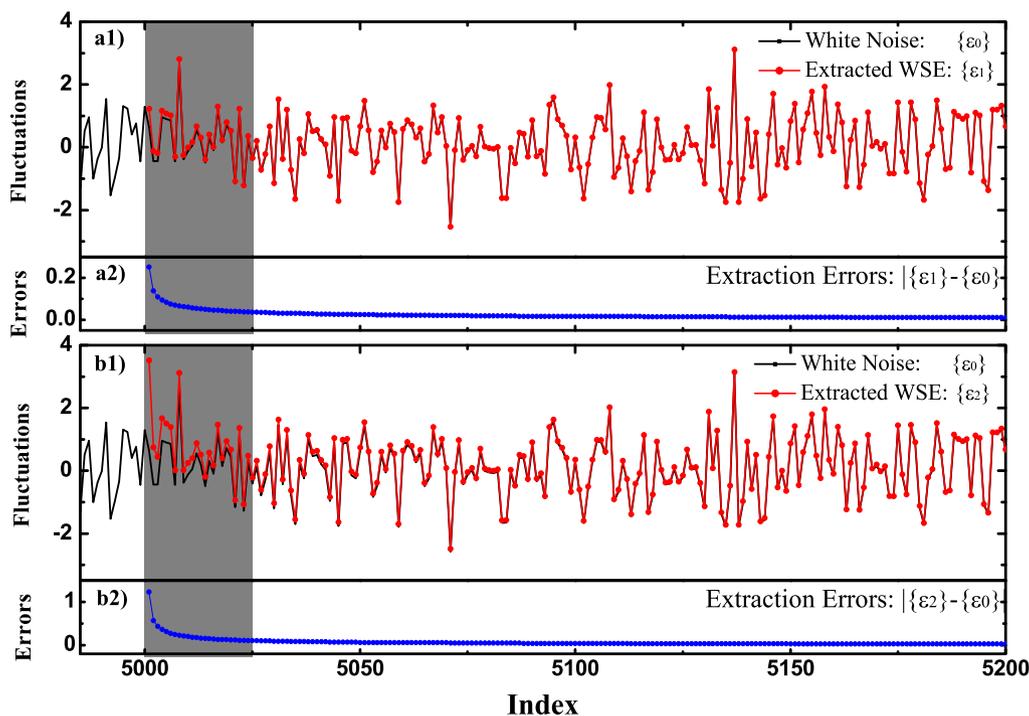
Figure 2 | Examination of the extracted "historical WSE". $\{\varepsilon_1\}$ is the extracted WSE from Data I, while $\{\varepsilon_2\}$ is the extracted WSE from Data II. (a1) and (b1) show the comparisons between the white noise $\{\varepsilon_0\}$ and the extracted WSE from both Data I and Data II. At the beginning (index = 5001), there are big errors, as shown in (a2) and (b2). But as time goes by, the errors become much smaller. At the index of 5200, the relative error of the extracted WSE is only 1.7% and 4.1%, for Data I and Data II respectively.
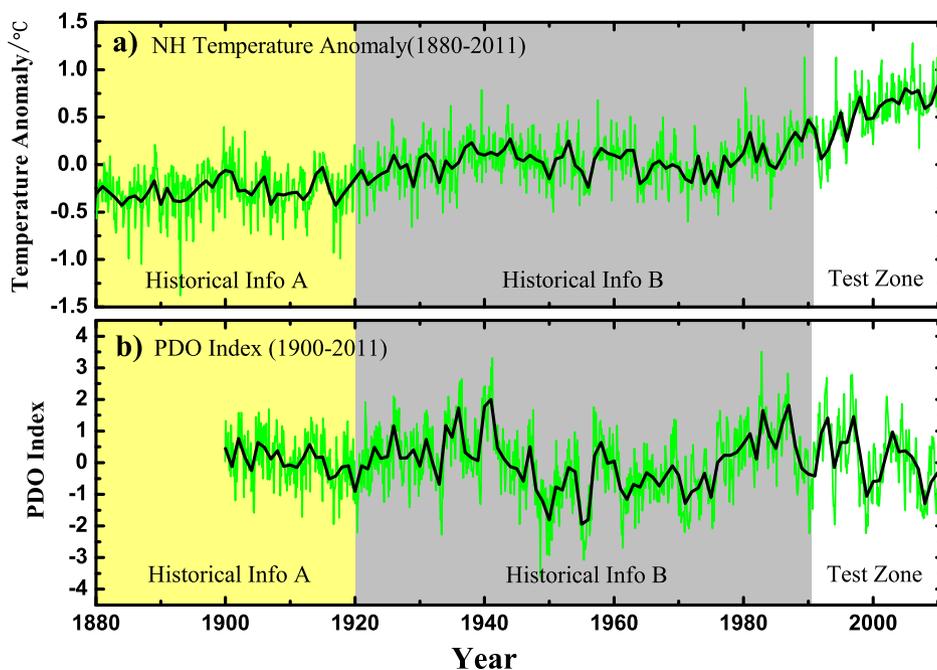
Figure 3 | The two time series we use for our analysis. (a) is the North Hemisphere monthly Temperature Anomalies (NHTA) from 1880 to 2011, while (b) is the Pacific Decadal Oscillation index (PDO) from 1900 to 2011. The green curves represent the monthly values, while the black curves is the annual fluctuations. The two time series are divided into three periods: "Historical Info A (1880 ~ 1919)", "Historical Info B (1920 ~ 1991)", and "Test Zone (1992 ~ 2011)".

account for a large proportion in the whole variations (black curves) for both time series. The stronger LTM is, the lager proportion will be found (the case of PDO). Of course, we like to admit that, the large proportion of climate memory found here is due to the strong LTM properties in these two time series. If the considered time series are characterized by very weak LTM, the proportion of the memory signals will be much smaller, and the whole variations will be mainly determined by the weather-scale excitations. For white noise (without LTM, $q = 0$, $\alpha = 0.5$), the memory signals calculated will be zero (figures are not shown here).

It is worth to note that the climate memory signals extracted in our study are not only integrated from the historical WSE of 1920 ~ 1991. For each time point t, we first update the historical WSE till $t - \delta$ (δ is the sampling intervals of the observed time series, in this study, it equals one month), and then estimate the memory signals accumulated till t. Therefore, our calculation can also be considered as a one-step-ahead estimation of the climate memory signals. Furthermore, we also need to note that our method is designed for anomaly series. Typical periodic signals such as annual cycles are removed before analysis. Like the time series used in our work, NHTA is temperature anomalies with annual cycle removed, while the PDO index is also calculated from the monthly sea surface temperature anomalies (SSTA) over the North Pacific. We do not take the annual cycle into account because it is a kind of periodic and relatively steady signals, while our target is to study the internal stochastic processes in climate system. Therefore, it is better to use anomaly series for the analysis.

Discussion

In this study, by using the Fractional Integral Statistical Model (FISM), we extracted the climate memory signals quantitatively. Essentially, this is a new method, with which one can estimate the long-lasting influences of historical climate states on the present time. We would like to emphasize that, the long-term climate memory we are discussing here is originated from the interactions

between fast-excitation subsystems and slow-response subsystems, such as the atmosphere-ocean interactions. Different from the short-term correlations originated from general weather-scale regimes, long-term climate memory is actually a characterization of scaling behavior in the whole climate system on time range from monthly to decadal[38,39]. Therefore, our method is better to be applied to monthly (or annual) data, where the effects of (short-term) general weather regimes have been averaged out. This is also the reason why this method may be useful for the climate-scale prediction.

From our research, we find the climate memory signals indeed can explain part of the variations if the considered time series are characterized by LTM. Such as the results shown in Fig. 6, the extracted climate memory signals account for a large proportion in the whole variations for both NHTA and PDO index. When simulating near future climate variations, with the memory signals $M(t)$ only, one can at least know part of information on the near future variations, as shown in equation (4). Therefore, for a given time series (or subsystem), $M(t)$ can be considered as the influences accumulated from the historical information. It is the climate memory, which determines on what basis the time series will continue to change. While $\varepsilon(t)$ stands for the continuous random excitation by short period weather-scale disturbances. It is a kind of external force, which determines in what direction the considered time series will change to. As shown in Fig. 6, besides the memory signals, we also calculated the corresponding WSE by using equation (4), and show them in Fig. 6(b) and Fig. 6(d). It is easy to find that for NHTA, which has a significant increasing trend during the past 20 years (as shown in Fig. 3), the calculated WSE are more likely to be positive. This indicates that WSE contains the external force, which triggered the NHTA taking on a warming trend during the past years. Although affected by the climate LTM, which is closely related with the LFV of climate, the considered time series may also take on a increasing (or decreasing) trend at some time scale (valleys and mountains, as discussed in[40]). But in our case, the memory signals obviously cannot explain the dramatically warming trend happened in the past years.
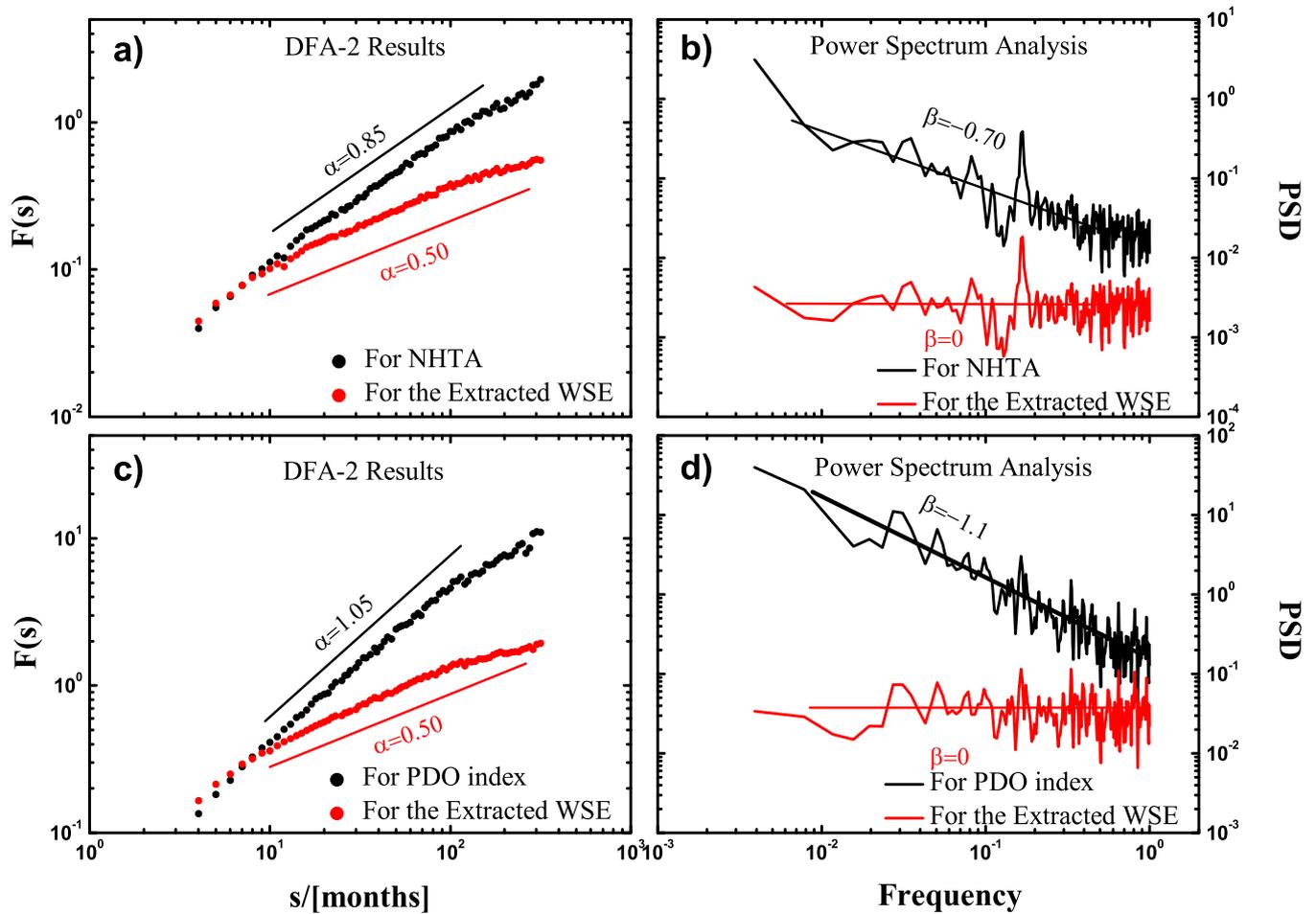
Figure 4 | The DFA and PSD analysis for the two considered time series. (a) and (b) are the results for NHTA, while (c) and (d) are the results for PDO. The black dots (curves) represent the results of the original time series in Fig. 3, while the red dots (curves) show the results of the extracted WSE in "Historical Info B (1920 ∼ 1991)" (see Fig. 5). It is found that the original time series (NHTA and PDO) are both characterized by strong LTM, with the DFA (PSD) exponent $\alpha = 0.85 \pm 0.01$ ($\beta = 0.7 \pm 0.05$) for NHTA and $\alpha = 1.05 \pm 0.01$ ($\beta = 1.1 \pm 0.05$) for PDO. While for the extracted historical WSE, as expected, no LTM property is found.

Therefore, our results also confirm that the warming trends observed during the past 20 years are more likely to be external-forcing induced, rather than caused by the natural variability.

In fact, in analogy to a famous traditional autoregressive fractionally integrated moving average (ARFIMA) model[41,42], our method can also be considered as a prediction model. In ARFIMA, the present climate states can be estimated according to

$$x(t) = \sum_{v=1}^{\infty} a(q;v)x(t-v) + \varepsilon(t), \qquad (9)$$

where $\varepsilon(t)$ denotes an independent and identically distributed (i.i.d) Gaussian noise, $a(q;v)$ is statistical weights defined by[38],

$$a(q;v) \equiv (-1)^{v+1}\frac{\Gamma(1+q)}{\Gamma(1+q-v)\Gamma(1+v)}, \qquad (10)$$

and $\Gamma(v)$ denotes the Gamma function. v stands for the distance between the present time t and the historical time $t-v$ (corresponding to the "$t-u$" in equation (2)), while q is a free parameter ranging from 0 to 0.5 (corresponding to the "q" used in our method). By calculating the statistical weights $a(q;v)$, along with the historical observation $x(t-v)$, one can make a simple prediction of $x(t)$. While in our method, the present climate states can be determined by,

$$x(t) = M(t) + \varepsilon(t) = K(q)_t^{\delta} \otimes \Psi_0^{t-\delta} + \varepsilon(t), \qquad (11)$$

where $K(q)_t^{\delta}$ is the memory kernel, represents the integrating factor of every single historical time point from $u = 0$ to $u = t - \delta$, and is governed by the LTM strength q. While $\Psi_0^{t-\delta}$ represents the historical WSE at each historical time point, and can be extracted from the historical observations $x(0), x(\delta), \cdots, x(t-\delta)$. By calculating the convolution between $K(q)_t^{\delta}$ and $\Psi_0^{t-\delta}$, the climate memory signal $M(t)$ is determined, with which a prediction of $x(t)$ can be made.

These two models seem to behave in a similar way and are both useful for the simulation of long-term correlated processes. Therefore, it is essential to make a careful comparison between them. To start the comparison, we would like to first recall the so-called fractional Brownian motion (fBm). It is well known that fBm describes physical processes that the successive increments (fractional Gaussian noise, fGn) are long-term correlated (not like the case in ordinary Brownian motion, Bm). Considering many natural-world processes are temporally long-term correlated, fBm is thus more powerful in understanding this kind of self-similar (or fractal) processes. From the work by Mandelbrot and Van Ness[28], fBm is derived from Bm by simply using fractional integral techniques. In other words, fractional integral technique can thus be used mathematically to simulate the behaviors of fBm, or more frequently, the temporal correlations of the corresponding successive increments (fGn). ARFIMA and FISM are such kind of simulation models. Since ARFIMA and FISM are both designed from fractional integral
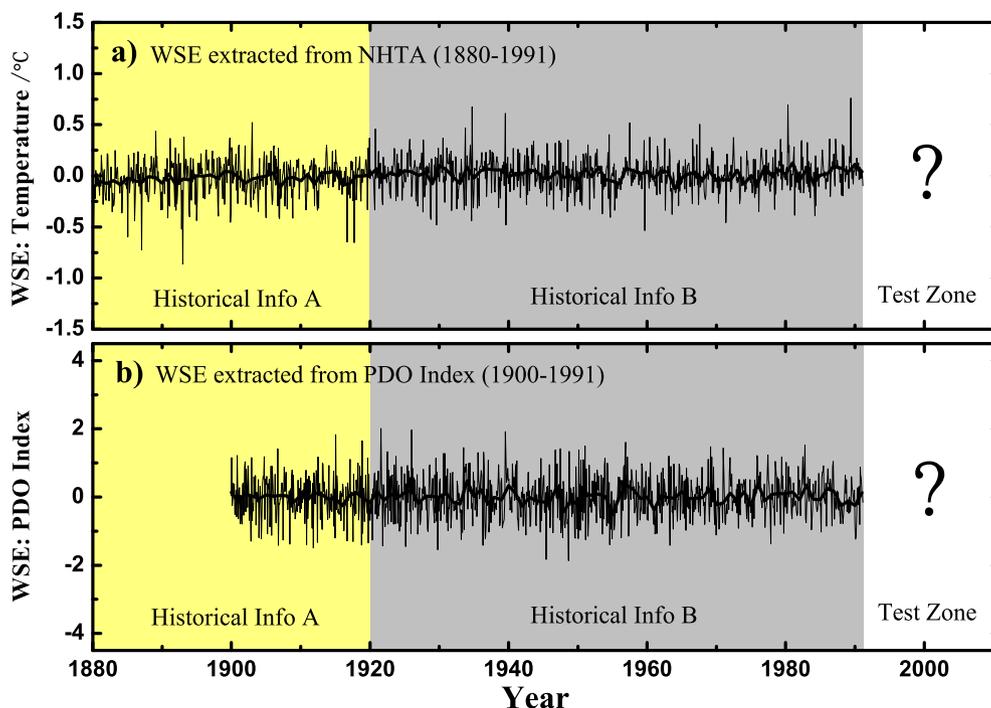
Figure 5 | **The historical WSE extracted from NHTA (a) and PDO (b).** The thin black curves represent the monthly values, while the thick black curves is the annual fluctuations. The historical WSE are only extracted and shown in the first two periods: "Historical Info A" and "Historical Info B". We use the last 20 years (1992 ~ 2011) as our test zone.
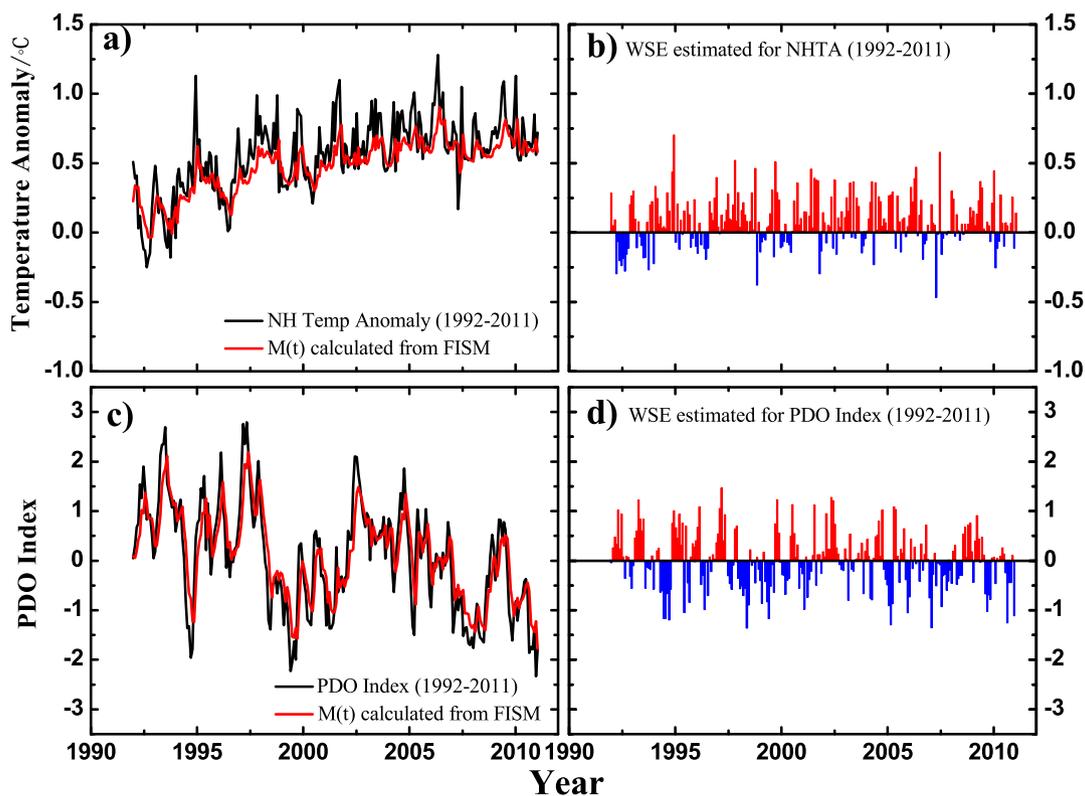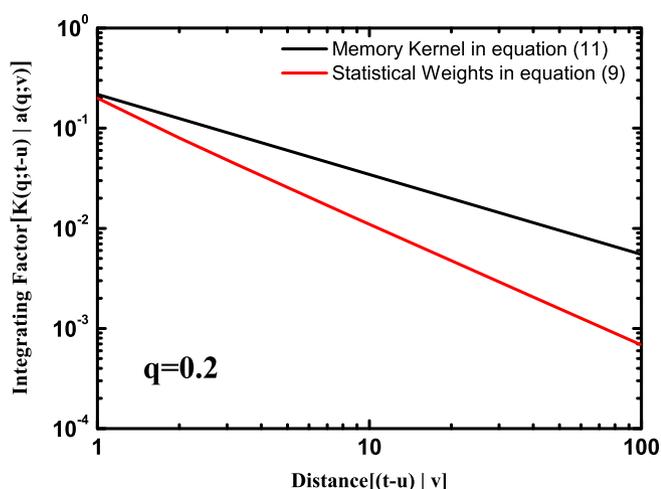

Figure 6 | **The calculated climate memory signals and their corresponding historical WSE.** (a) and (c) are the comparisons between the calculated climate memory signals $M(t)$ (red curves) and the original time series (black curves) in 1992 ~ 2011. (b) and (d) show the corresponding historical WSE in the same time period. (a),(b) represent the results of NHTA, while (c),(d) represent the results of PDO. One can see the memory signals $M(t)$ account for a large proportion in the whole variations for both time series, which indicates the extraction of $M(t)$ is helpful for our further climate prediction. Interestingly, the historical WSE for NHTA are asymmetrically distributed with more positive values, as shown in (b), which means the warming trends observed during the past 20 years are more likely to be external-forcing induced, rather than caused by the natural variability.

techniques, it is therefore not surprising that they have similar mathematical form and usage. But from equation (9) and equation (11), it should be noted that there still are significant differences, which mainly arise from the way of how the fractional integral is performed. To explain it more clearly, two key variables, the statistical weight $a(q; v)$ in equation (9) and the memory kernel $K(q)_t^\delta$ in equation (11) are considered. Suppose we have one long-term correlated time series (fGn) with DFA exponent $\alpha = 0.7$ (corresponding to the integral order $q = 0.2$). If we use ARFIMA to simulate this time series, the calculated statistical weight $a(q; v)$ is shown in Fig. 7 (the red curve), while if we use FISM to simulate this time series, the calculated memory kernel $K(q)_t^\delta$ is the black curve in Fig. 7. Obviously, both two variables decay along power function, which indicate long-term memory. But from the different decay rates, we believe there are fundamental differences between FISM and ARFIMA. From equation (9) and equation (11), one can easily find the statistical weight $a(q; v)$ in ARFIMA and the memory kernel $K(q)_t^\delta$ in FISM actually act on different variables. $a(q; v)$ acts on the historical observations $x(t - v)$ directly, while $K(q)_t^\delta$ acts on the historical WSE. We believe it is this difference that makes our method in this study seems to be superior, because we can easily decompose a present state $x(t)$ into two components: the cumulative historical memory part, and the present weather-scale excitations, and further give the variables in equation (11) physical meanings. Such as $\varepsilon(t)$, unlike equation (9), it not only denotes an independent and identically distributed (i.i.d) noise, but also represents the present WSE. Therefore, we believe our method is more than a fractionally integrated prediction model. Besides simulating long-term memory properties, it also tries to understand physical processes from an angle of why and how the long-term memory arise in climate. Therefore, a new perspective on climate prediction can be addressed. For example, one may improve our prediction skills by mainly focus on the weather-scale excitations. Combine with the cumulative climate memory signals, a reasonable prediction may be finally obtained.

However, it is worth to note that the proportion $M(t)$ accounts for is determined by the strength of LTM. In our research, the two cases are both characterized by strong LTM. Therefore, the extracted $M(t)$ accounts for a large proportion in the whole variations. But the reality is, not all the observed climate variables are characterized by strong

LTM. Some of them may only weakly related with the slow responding subsystems (such as the ocean), and therefore characterized by relatively weaker LTM. As a result, the whole variations may be mainly determined by WSE, which cannot be extracted statistically in our method. To improve the prediction in this case, we may need to build bridges between the considered variable with weak LTM and other variables with strong LTM. By making reliable predictions of "other variables", the prediction of the considered variable may be improved with the help of their links. However, to address this issue, detailed experiments are still needed.

In summary, this report shows a new method with which one can extract the climate memory signals quantitatively. We believe the results will provide us a new perspective on climate prediction, but many more detailed studies are still needed. In addition, it should be noted that not only the climatic time series are characterized by long-term memory. LTM is ubiquitous in nature, such as in economic records[43], physiological records[44], and even records of human activity[45,46]. Therefore, our method may even have broader application prospects. In a following paper, we will discuss it more thoroughly and show more results.

Methods

In this section, we will summarize the procedures on how to extract the climate memory signals. As discussed above, we believe any given climatic time series can be decomposed into two parts: the cumulative climate memory signals $M(t)$ and the present weather scale excitations $\varepsilon(t)$, as shown in equation (4). In this study, we proposed a method, which can be used to extract $M(t)$ by using FISM. But before applying it to observational records, we need first transfer equation (2) into discrete form, as equation (8), see also in the following,

$$x(t) = M(t) + \varepsilon(t)$$
$$= \frac{1}{\Gamma(q)} \int_{u=0}^{t-\delta} \frac{\varepsilon(u)}{(t-u)^{1-q}} \, du + \varepsilon(t) \qquad (12)$$
$$= K(q)_t^\delta \otimes \Psi_0^{t-\delta} + \varepsilon(t)$$

where $K(q)_t^\delta \equiv \{k(q; t), k(q; t-\delta), \cdots, k(q; t-u), \cdots, k(q; \delta)\}$, represents the integrating factor (we name it as Memory Kernel, MK for short, see[29]) of every single step, and $\Psi_0^{t-\delta} \equiv \{\varepsilon(0), \varepsilon(\delta), \cdots, \varepsilon(u), \cdots, \varepsilon(t-\delta)\}$, represents the WSE at each historical time point. The superscript and subscript of $K(q)_t^\delta$ denote all the time distances between historical time points and the present time, like $t - u$ in equation (2), from $t - u = t$ (corresponding to $u = 0$) to $t - u = \delta$ (corresponding to $u = t - \delta$), while the superscript and subscript of $\Psi_0^{t-\delta}$ denote the historical time point, like u in equation (2), from $u = 0$ to $u = t - \delta$. Obviously, to extract $M(t)$, one needs first determine both $K(q)_t^\delta$ and $\Psi_0^{t-\delta}$, and then calculate their convolution. For $K(q)_t^\delta$, with the strength of LTM (or the integral-order q), we can calculate it directly, while as for $\Psi_0^{t-\delta}$, it is extracted by deriving equation (12)(or equation (8)) reversely. Details of all the procedures are summarized as following. Suppose we have a climatic time series $\{x(t)\}$, to extract the climate memory signals at the present time t, one needs:

i) Determine the strength of LTM by DFA, obtain the DFA exponent α and further the integral-order q, according to equation (3).

ii) Calculate the memory kernel at each time distances according to its definition: $K(q)_t^\delta \equiv \{k(q; t), k(q; t-\delta), \cdots, k(q; t-u), \cdots, k(q; \delta)\}$, where $k(q; t-u) = \frac{1}{\Gamma(q)(t-u)^{1-q}}$.

iii) Using the calculated MK and the observed historical records $x(0), x(\delta), \cdots, x(t-\delta)$, reversely derive equation (12) and extract the historical WSE $\varepsilon(0), \varepsilon(\delta), \cdots, \varepsilon(t-\delta)$ step by step. For example, suppose $t' = 0$ is the beginning of our observation, ignore the influences before this time point, theoretically we have

$$\varepsilon(0) = x(0). \qquad (13)$$

For the next time point $t' = \delta$, since $u = 0$ is the corresponding historical time point, from equation (12), $\varepsilon(\delta)$ will be calculated as

$$\varepsilon(\delta) = x(\delta) - K(q)_\delta^\delta \otimes \Psi_0^0$$
$$= x(\delta) - k(q; \delta) \times \varepsilon(0). \qquad (14)$$

While for the third time point $t' = 2\delta$, similarly, $u = 0$ and $u = \delta$ become the corresponding historical time points, $\varepsilon(2\delta)$ will thus be determined as

Figure 7 | Difference between the memory kernel $K(q; t - u)$ (in our method) and the statistical weights $a(q; v)$ (in ARFIMA). We show the case when $q = 0.2$. We can see power-law behaviors for both $K(q; t - u)$ (black) and $a(q; v)$ (red), that is, as the distance between the present time and historical time increases, neither the memory kernel nor the statistical weights will reach zero. The memory kernel decreases slower than the statistical weights.

$$\varepsilon(2\delta) = x(2\delta) - K(q)_{2\delta}^{\delta} \otimes \Psi_0^{\delta}$$
$$= x(2\delta) - (k(q;2\delta) \times \varepsilon(0) + k(q;\delta) \times \varepsilon(\delta)). \qquad (15)$$

In this way, for time point $t' = t - \delta$, $\varepsilon(t - \delta)$ can be derived as

$$\varepsilon(t-\delta) = x(t-\delta) - K(q)_{t-\delta}^{\delta} \otimes \Psi_0^{t-2\delta}$$
$$= x(t-\delta) - (k(q;t-\delta) \times \varepsilon(0) + k(q;t-2\delta) \times \varepsilon(\delta) + \cdots + k(q;\delta) \times \varepsilon(t-2\delta)), \qquad (16)$$

and the historical WSE $\varepsilon(0), \varepsilon(\delta), \cdots, e(t-\delta)$ will finally be obtained. However, it should be noted that we must sacrifice a certain number of the extracted historical WSEs until the influences of unobserved historical information (before $u = 0$) become negligible, as discussed in Fig. 2.

iv) In the end, with the extracted historical WSE and corresponding memory kernel, the cumulative climate memory signals $M(t)$ can be calculated by using equation (5).

About this method, we would like to note that the most important procedure is to calculate the memory kernel $K(q)_t^{\delta}$. It not only related with the calculation of historical WSE, but also used in the extraction of climate memory signals. Therefore, to make a reliable extraction of $M(t)$, we need to ensure the accuracy of $K(q)_t^{\delta}$. Since $K(q)_t^{\delta}$ is determined by the integral-order q, or more precisely, by the strength of LTM, accurate diagnosis of LTM property is very essential. In other words, it is the LTM-estimation errors that determines the errors of our extraction of the climate memory signals. Actually, this is a problem not only faced by our method, but also faced by other models, like the ARFIMA, where one also need to make a estimation of the integral-order q[41], see equation (10). In this study, comparing with the power spectrum density (PSD) analysis, we find DFA may be a more reliable method in estimating the LTM strength, since the estimated DFA exponents α have small standard errors (± 0.01). However, we still suggest a further examination, such as applying DFA to the extracted WSE (as shown in Fig. 4) to make sure the accuracy of the calculation. Moreover, applying multi-methods together to improve the accuracy of estimated LTM strength may also be a good solution.

1. Hurst, H. E. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.* **116**, 770–808 (1951).
2. Malamud, B. D. & Turcotte, D. L. *Advances in Geophysics: Long Range Persistence in Geophysical Time Series*, Self-affine time series: I. Generation and analysis, Dmowska, R. & Saltzman, B. (ed.), pages 1–87 (Academic Press, San Diego, 1999).
3. Lovejoy, S. & Schertzer, D. Scale invariance in climatological temperatures and the local spectral plateau. *Ann. Geophys.* **4B**, 401–410 (1986).
4. Lovejoy, S. & Schertzer, D. *Extreme Events and Natural Hazards: The Complexity Perspective*, Low frequency weather and the emergence of the climate, Sharma, A. S., Bunde, A., Baker, D. & Dimri, V. P. (ed.), pages 231–254, DOI: 10.1029/2011GM001087, (AGU monographs, Washington, D. C., 2012).
5. Arneodo, A., Bacry, E., Graves, P. V. & Muzy, J. F. Characterizing long-range correlations in DNA sequences from wavelet analysis. *Phys. Rev. Lett.* **74**, 3293–3296 (1995).
6. Peng, C. K. *et al.* Mosaic organization of DNA nucleotides. *Phys. Rev. E* **49**, 1685–1689 (1994).
7. Schmitt, F., Lovejoy, S. & Schertzer, D. Multifractal analysis of the Greenland ice-core project climate data. *Geophys. Res. Lett.* **22**, 1689–1692 (1995).
8. Koscielny-Bunde, E. *et al.* Indication of a universal persistence law governing atmospheric variability. *Phys. Rev. Lett.* **81**, 729–732 (1998).
9. Ashkenazy, Y., Baker, D., Gildor, H. & Havlin, S. Nonlinearity and multifractality of climate change in the past 420,000 years. *Geophys. Res. Lett.* **30**, 2146 (2003).
10. Chen, X., Lin, G. & Fu, Z. Long-range correlations in daily relative humidity fluctuations: A new index to characterize the climate regions over China. *Geophys. Res. Lett.* **34**, L07804 (2007).
11. Vyushin, D. I. & Kushner, P. J. Power-law and long-memory characteristics of the atmospheric general circulation. *J. Climate* **22**, 2890–2904 (2009).
12. Bunde, A., Büntgen, U., Ludescher, J., Luterbacher, J. & von Storch, H. Is there memory in precipitation? *Nature Clim. Change* **3**, 174–175 (2013).
13. Yuan, N. & Fu, Z. Century-scale intensity-modulation of large-scale variability in long historical temperature records. *J. Climate* **27**, 1742–1750 (2014).
14. Bunde, A., Eichner, J. F., Kantelhardt, J. W. & Havlin, S. 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).
15. Rybski, D., Bunde, A., Havlin, S. & von Storch, H. Long-term persistence in climate and the detection problem. *Geophys. Res. Lett.* **33**, L06718 (2006).
16. Lennartz, S. & Bunde, A. Trend evaluation in records with long-term memory: Application to global warming. *Geophys. Res. Lett.* **36**, L16706 (2009).
17. Franzke, C. Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *J. Climate* **25**, 4172–4183 (2011).
18. Bunde, A., Ludescher, J., Franzke, C. & Büntgen, U. How significant is west Antarctic warming? *Nature Geosci.* **7**, 246–247 (2014).
19. Govindan, R. B. *et al.* Global climate models violate scaling of the observed atmospheric variability. *Phys. Rev. Lett.* **89**, 028501 (2002).
20. Blender, R. & Fraedrich, K. Long-time memory in global warming simulations. *Geophys. Res. Lett.* **30**, L017666 (2003).
21. Fraedrich, K., Luksch, U. & Blender, R. 1/f model for long-time memory of the ocean surface temperature. *Phys. Rev. E* **70**, 037301 (2004).
22. Vyushin, D., Zhidkov, I., Havlin, S., Bunde, A. & Brenner, S. Volcanic forcing improves Atmosphere-Ocean coupled General Circulation Model scaling performance. *Geophys. Res. Lett.* **31**, L10206 (2004).
23. Rybski, D., Bunde, A. & von Storch, H. Long-term memory in 1000-year simulated temperature records. *J. Geophys. Res.* **113**, D02106 (2008).
24. Lovejoy, S., Schertzer, D. & Varon, D. Do GCMs predict the climate…or macroweather? *Earth Syst. Dynam.* **4**, 439–454 (2013).
25. Fraedrich, K., Blender, R. & Zhu, X. Continuum climate variability: long-term memory, scaling, and 1/f noise. *Int. J. Mod. Phys. B* **23**, 5403–5416 (2009).
26. Zhu, X., Fraedrich, K., Liu, Z. & Blender, R. A demonstration of long-term memory and climate predictability. *J. Climate* **23**, 5021–5029 (2010).
27. Hasselmann, K. Stochastic climate models, Part I. Theory. *Tellus* **28**, 473–485 (1976).
28. Mandelbrot, B. B. & Van Ness, J. W. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422–437 (1968).
29. Yuan, N., Fu, Z. & Liu, S. Long-term memory in climate variability: A new look based on fractional integral techniques. *J. Geophys. Res.* **118**, 12962–12969 (2013).
30. Kolmogorov, A. N. Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C. R. (Doklady) Acad. URSS (N.S)* **26**, 115–118 (1940).
31. Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A. & Havlin, S. Detecting long-range correlations with detrended fluctuation analysis. *Physica A* **295**, 441–454 (2001).
32. Talkner, P. & Weber, R. O. Power spectrum and detrended fluctuation analysis: Application to daily temperature. *Phys. Rev. E* **62**, 150–160 (2000).
33. Király, A., Bartos, I. & Jánosi, I. M. Correlation properties of daily temperature anomalies over land. *Tellus* **58A**, 593–600 (2006).
34. Yuan, N., Fu, Z. & Mao, J. Different scaling behaviors in daily temperature records over China. *Physica A* **389**, 4087–4095 (2010).
35. Fraedrich, K. & Blender, R. Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.* **90**, 108501 (2003).
36. Eichner, J. F., Koscielny-Bunde, E., Bunde, A., Havlin, S. & Schellnhuber, H.-J. Power-law persistence and trends in the atmosphere: A detailed study of long temperature records. *Phys. Rev. E* **68**, 046133 (2003).
37. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.* **78**, 1069–1079 (1997).
38. Caballero, R., Jewson, S. & Brix, A. Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation. *Clim. Res.* **21**, 127–140 (2002).
39. Huybers, P. & Curry, W. Links between annual, Milankovitch and continuum temperature variability. *Nature* **441**, 329–332 (2006).
40. Lennartz, S. & Bunde, A. Distribution of natural trends in long-term correlated records: A scaling approach. *Phys. Rev. E* **84**, 021129 (2011).
41. Taqqu, M. S., Teverovsky, V. & Willinger, W. Estimators for long-range dependence: an empirical study. *Fractals* **03**, 785–798 (1995).
42. Baillie, R. T. & Chung, S.-K. Modeling and forecasting from trend-stationary long memory models with applications to climatology. *Int. J. Forecasting* **18**, 215–226 (2002).
43. Mantegna, R. N. & Stanley, H. E. Scaling behaviour in the dynamics of an economic index. *Nature* **376**, 46–49 (1995).
44. Bunde, A. *et al.* Correlated and uncorrelated regions in heart-rate fluctuations during sleep. *Phys. Rev. Lett.* **85**, 3736–3739 (2000).
45. Ivanov, P., Hu, K., Hilton, M., Shea, S. & Stanley, H. E. Endogenous circadian rhythm in human motor activity uncoupled from circadian influences on cardiac dynamics. *PNAS* **104**, 20702–20707 (2007).
46. Rybski, D., Buldyrev, S. V., Havlin, S., Liljeros, F. & Makse, H. A. Communication activity in a social network: relation between long-term correlations and inter-event clustering. *Sci. Rep* **2**, 560 (2012).

Author contributions

S.L. designed the study, N.Y. and Z.F. performed the study. N.Y. and Z.F. wrote the main manuscript text. All authors reviewed the manuscript. N.Y. and Z.F. contributed equally to this work.

Additional information