



# On memory and non-memory parts of surface air temperatures over China: can they be simulated by decadal hindcast experiments in CMIP5?

Feilin Xiong<sup>1</sup> · Naiming Yuan<sup>2</sup> · Xiaoyan Ma<sup>1</sup> · Zhenghui Lu<sup>2</sup> · Jinhui Gao<sup>1</sup>

Received: 26 March 2018 / Accepted: 7 August 2018 / Published online: 17 August 2018  
© The Author(s) 2018

## Abstract

It has been well recognized that, for most climatic records, their current states are influenced by both past conditions and current dynamical excitations. However, how to properly use this idea to improve the climate predictive skills, is still an open question. In this study, we evaluated the decadal hindcast experiments of 11 models (participating in phase 5 of the Coupled Model Intercomparison Project, CMIP5) in simulating the effects of past conditions (memory part,  $M(t)$ ) and the current dynamical excitations (non-memory part,  $\varepsilon(t)$ ). Poor skills in simulating the memory part of surface air temperatures (SAT) are found in all the considered models. Over most regions of China, the CMIP5 models significantly overestimated the long-term memory (LTM) of SAT. While in the southwest, the LTM was significantly underestimated. After removing the biased memory part from the simulations using fractional integral statistical model (FISM), the remaining non-memory part, however, was found reasonably simulated in the multi-model means. On annual scale, there were high correlations between the simulated and the observed  $\varepsilon(t)$  over most regions of the country, and for most cases they had the same sign. These findings indicated that the current errors of dynamical models may be partly due to the unrealistic simulations of the impacts from the past. To improve predictive skills, a new strategy was thus suggested. As FISM is capable of extracting  $M(t)$  quantitatively, by combining FISM with dynamical models (which may produce reasonable estimations of  $\varepsilon(t)$ ), improved climate predictions with the effects of past conditions properly considered may become possible.

**Keywords** Long-term memory · Memory part · Non-memory part · CMIP5

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00382-018-4395-9>) contains supplementary material, which is available to authorized users.

---

✉ Naiming Yuan  
yuannm@tea.ac.cn

<sup>1</sup> Key Laboratory of Meteorological Disaster, Ministry of Education (KLME)/Joint International Research Laboratory of Climate and Environment Change (ILCEC)/ Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD)/Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science and Technology, Nanjing 210044, China

<sup>2</sup> CAS Key Laboratory of Regional Climate Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

## 1 Introduction

Recently, long-term memory (LTM) has become a well-known concept in climate community. With the development of many advanced approaches such as the structure function method (Lovejoy and Schertzer 2012), the wavelet analysis (Arneodo et al. 1995; Abry and Veitch 1998), as well as the detrended fluctuation analysis (Peng et al. 1994; Kantelhardt et al. 2001), etc., a number of studies have been stimulated and many climate variables (e.g. temperatures, relative humidity, wind, atmospheric general circulations, etc.) were found to be characterized by LTM (Koscielny-Bunde et al. 1998; Eichner et al. 2003; Monetti et al. 2003; Kantelhardt et al. 2006; Chen et al. 2007; Rybski et al. 2008; Vyushin and Kushner 2009; Feng et al. 2009; Franzke 2010; Dangendorf et al. 2014; Massah and Kantz 2016). As the name implies, LTM measures the connections of climate states observed at different time points. But different from short-term persistence that exists in weather systems with

time scales of several days to weeks, LTM describes the scaling behavior of climate variables on time scales ranging from months to decades (Fraedrich and Blender 2003; Yuan et al. 2010; Luo et al. 2015; Jiang et al. 2017). Accordingly, for a system with LTM, its current climate conditions can have long-lasting influences on the conditions in future (Kantelhardt et al. 2001). Therefore, this phenomenon is considered as a kind of “climate inertia”, and the ocean in climate system, with huge heat capacity, may be a main contributor (Yuan et al. 2013).

During the past years, besides detecting LTM in different climate variables, there are growing efforts focusing on the applications of LTM, such as (1) developing new theories for trend evaluation (Lennartz and Bunde 2009; Franzke 2012; Kumar et al. 2013; Ludescher and Bunde 2016; Yuan et al. 2017); (2) designing early warning systems for extreme events (Bunde et al. 2005; Bogachev and Bunde 2011); as well as (3) evaluating model simulations and reanalysis/proxy datasets using LTM as a test bed (Govindan et al. 2002; Vyushin et al. 2004; Bunde et al. 2013; Zhao et al. 2018). However, among all the potential applications of LTM, climate prediction is the most appealing one that has not been studied systematically (Zhu et al. 2010). Can we improve the climate predictive skill by properly considering the effects of long-term climate memory? How to integrate this concept into current climate prediction models? These are important questions that deserves further studies.

As LTM is ubiquitous in climate system, it has been recognized that the current climate state is influenced by (1) the past conditions and (2) the current dynamical excitations (Yuan et al. 2013, 2014). For variables with strong LTM, the influences from past can explain a high variance of the total climate variability (Yuan et al. 2014), which, in most cases, are non-negligible in climate predictions. Using a recently developed fractional integral statistical model (FISM), it has become possible to extract the influences from past quantitatively, and further decompose the current climate state into two parts:

$$x(t) = M(t) + \varepsilon(t), \quad (1)$$

where  $x(t)$  represent the current climate state,  $M(t)$  stands for the influences from past (the memory part) and  $\varepsilon(t)$  the current dynamical excitations (the non-memory part) (Yuan et al. 2014). Accordingly, a new way for climate prediction is suggested. That is, one may focus on the memory part  $M(t)$  and the non-memory part  $\varepsilon(t)$  separately, to implement climate prediction (Yuan et al. 2018).

In this study, we follow this research idea and evaluate the performance of state-of-the-art models in simulating the memory and non-memory part of surface air temperatures (SAT) over China. With the development of climate science and the improvement of computing resources, current dynamical models have been improved considerably during the past years.

However, due to the lack of perfect initial conditions and the inability to perfectly model the climate system, there are unavoidable uncertainty and inadequacy in the model simulations (Palmer 2000; Slingo and Palmer 2011), and even the internal variability is so far not fully captured by current models (Doblas-Reyes et al. 2013; Meehl et al. 2014). In this situation, are current models able to reproduce the LTM that was detected from observational data, and further capture the memory part  $M(t)$ ? After removing  $M(t)$  using statistical models (such as FISM), are the remaining non-memory part  $\varepsilon(t)$  comparable with the observed  $\varepsilon(t)$ ? Which part is more responsible for the biases of model simulations? To improve the current climate predictive skills, these are relevant questions that need to be addressed.

In this work, we employed the detrended fluctuation analysis of the second-order (DFA2) (Kantelhardt et al. 2001) and the fractional integral statistical model (FISM) (Yuan et al. 2014) to evaluate the memory and non-memory parts simulated from the decadal hindcast experiments (1981–2010) in 11 CMIP5 models (Taylor et al. 2012). Using DFA2, the ability of the CMIP5 models in reproducing the observed LTM was studied. While using FISM, the memory part was removed and the remaining non-memory part was further assessed. It is worth noting that, discussing the month-to-month simulations in a decadal hindcast experiment makes little sense. Therefore, we mainly focus on the simulations on annual scale and evaluated the year by year variations of the non-memory part. But to ensure the accuracy of the DFA2 study, monthly data were used and the strength of LTM was determined on time scales from a few months to a few years (inter-annual time scale). We found poor skills of the CMIP5 models in simulating the observed LTM, but regarding of the non-memory part, the CMIP5 models seem to have better performance. Based on these findings, a new strategy for climate prediction was suggested, which may improve the predictive skills by properly considering the impacts of the past.

The rest of this paper is organized as follows. In Sect. 2, we will make a brief introduction of the data and the methods we use for analysis. Model simulated LTM is estimated in Sect. 3, and the results are compared with those obtained from observational data. After decomposing the surface air temperatures into the memory and non-memory parts, we further evaluated the models' capacity in capturing the non-memory part. In Sect. 5, we discuss the results and conclude this work with a further outlook.

## 2 Data and methods

### 2.1 Data

#### 2.1.1 Observational data

In this study, observational grid dataset of Chinese monthly surface air temperature is used for analysis. It is downloaded from the China Meteorological Data Service Center (CMDC) ([http://data.cma.cn/data/cdcdetail/dataCode/SURF\\_CLI\\_CHN\\_TEM\\_MON\\_GRID\\_0.5.html](http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_TEM_MON_GRID_0.5.html)). The spatial resolution of the observational grid data is  $0.5^\circ \times 0.5^\circ$ , and the data from 1981 to 2010 are analyzed. Besides the grid dataset, four in-situ surface air temperature records observed in Beijing ( $116.28^\circ\text{E}$ ,  $39.48^\circ\text{N}$ ), Harbin ( $126.34^\circ\text{E}$ ,  $45.56^\circ\text{N}$ ), Kunming ( $102.39^\circ\text{E}$ ,  $25^\circ\text{N}$ ) and Nanchang ( $115.55^\circ\text{E}$ ,  $28.36^\circ\text{N}$ ) are also analyzed. They are downloaded from CMDC at the address [http://data.cma.cn/data/cdcdetail/dataCode/SURF\\_CLI\\_CHN\\_MUL\\_MON\\_CES.html](http://data.cma.cn/data/cdcdetail/dataCode/SURF_CLI_CHN_MUL_MON_CES.html). The length of the in-situ data is also 30

years, from 1981 to 2010. Before analysis, seasonal trend is removed by subtracting annual cycle from the observed data, as  $x(t) = \tau(t) - \langle \tau(t) \rangle$ , where  $\tau(t)$  is the observed data,  $\langle \tau(t) \rangle$  is the long-time climatological average for each calendar, and  $x(t)$  is the anomalies we use for analysis.

#### 2.1.2 CMIP5 decadal hindcast simulations

Monthly surface air temperatures simulated from the Decadal hindcast experiments in 11 CMIP5 models are used in this study (Taylor et al. 2012). The data are downloaded from the Earth System Grid Federation (ESGF) at the address <https://esgf-data.dkrz.de/projects/esgf-dkrz/>. Information about the 11 models are summarized in Tables 1 and 2, and more details can be found in (Meehl et al. 2014). To ensure the accuracy of the LTM detection, we use the long lead time series at 30 years. That is, each model have been initialized by the end of 1980 and the lead time is 30 years from Jan 1981 to Dec 2010. Only data from the region  $72^\circ\text{N}$ – $136^\circ\text{N}$ ,  $18^\circ\text{E}$ – $54^\circ\text{E}$  (China) are extracted for analysis. To keep consistent with the observational grid dataset, the

**Table 1** Detailed information of the 11 CMIP5 models used in this study

No.	Models	Resolution	Ensemble member	Initialization method	Lead time
1	CCSM4	$1.25^\circ \times 0.9^\circ$	r1i2p1	Full field	1981–2010
2	CMCC-CM	$0.75^\circ \times 0.75^\circ$	r1i1p1		
3	CanCM4	$1.875^\circ \times 1.875^\circ$			
4	FGOALS-g2	$2.8125^\circ \times 2.8125^\circ$			
5	HadCM3	$3.75^\circ \times 2.5^\circ$	r1i2p1	Anomaly	
6	IPSL-CM5A-LR	$3.75^\circ \times 1.9^\circ$	r1i1p1		
7	MIROC4h	$0.5625^\circ \times 0.5625^\circ$			
8	MIROC5	$1.40625^\circ \times 1.40625^\circ$			
9	MPI-ESM-LR	$1.875^\circ \times 1.875^\circ$			
10	MRI-CGCM3	$1.125^\circ \times 1.125^\circ$			
11	bcc-csm1-1	$1.875^\circ \times 1.875^\circ$		Full field	

**Table 2** Skill scores of the 11 CMIP5 models (model name expansions included) in simulating the LTM of surface air temperatures

Model	Expansion	Skill score
CCSM4	Community Climate System Model, version 4.0	0.54
CMCC-CM	The Centro Euro-Mediterraneo sui Cambiamenti Climatici Climate Model	0.35
CanCM4	The Canadian fourth generation coupled Climate Model	0.43
FGOALS-g2	Flexible Global Ocean–Atmosphere–Land System Model: Grid-point Version 2	0.65
HadCM3	Hadley Centre Coupled Model, version 3	0.41
IPSL-CM5A-LR	L’Institut Pierre-Simon Laplace Coupled Model, version 5, coupled with NEMO (low resolution)	0.45
MIROC4h	Model for Interdisciplinary Research on Climate, version 4 (high resolution)	0.40
MIROC5	Model for Interdisciplinary Research on Climate, version 5	0.29
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution	0.39
MRI-CGCM3	Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 3	0.54
bcc-csm1-1	Beijing Climate Center Climate System Model, version 1	0.57
Multi-Model	Multiple Models Mean	0.43

simulated surface air temperatures are bi-linearly interpolated to the spatial resolution of  $0.5^\circ \times 0.5^\circ$ . That is, linear interpolation is first performed in the zonal direction, and then again in the meridional direction [for more information about the bi-linear interpolation, we refer to Press et al. (1992)]. Only the anomalies with annual cycle removed (see Sect. 2.1.1) are analyzed. Besides analyzing the simulations from the 11 CMIP5 models, we also compute the Multi-Model Means (MMM) over the 11 models, and compare the results with those obtained from observational dataset.

## 2.2 Methods

### 2.2.1 Detrended fluctuation analysis

In this study, we employed the detrended fluctuation analysis of the second-order (DFA2) to detect the LTM (Kantelhardt et al. 2001). Suppose we have a time series  $\{x_i\}$ ,  $i = 1, \dots, N$ . In DFA2, one mainly considers the cumulated sum  $Y_k = \sum_{i=1}^k \{x_i - \langle x \rangle\}$  and divides it into non-overlapping windows of size  $s$ . In each window, one determines the variance  $F_s^2(j)$  of the  $Y_k$  around the best polynomial fit of order 2, where  $j$  points to the  $j$ th window. By averaging over all windows and taking the square root, the desired fluctuation function  $F(s)$  is obtained. If  $F(s)$  increases with  $s$  as  $F(s) \sim s^\alpha$ , and the scaling exponent  $\alpha$  is larger than 0.5, we say the time series  $\{x_i\}$  is characterized by LTM. The bigger  $\alpha$  is, the stronger the LTM will be. While if the scaling exponent  $\alpha$  equals 0.5, the time series  $\{x_i\}$  is considered as white noise with no memory.

In this work, we analyzed the LTM in SAT. It has been shown by previous works that the scaling range of SAT normally spans multiple scales from months to decades (Lovejoy 2015; Yuan et al. 2018). Accordingly, as long as the  $\alpha$  values are measured in this scaling range, the estimated LTM should stay unchanged and is independent of the data resolution (e.g. monthly, annually, etc.) (see also Fig. S1 in the supporting information, SI). Therefore, although we mainly focus on the model simulations on annual scale, to reduce the uncertainties of the estimated  $\alpha$  values, we used the monthly data for the DFA2 calculations.

### 2.2.2 Fractional integral statistical model

To decompose the time series  $\{x_i\}$  and obtain the non-memory part, we employed the fractional integral statistical model (FISM) (Yuan et al. 2014). Previous studies have pointed out that, for processes with LTM, climate state at present time may have long-lasting influences on the climate states in future, and the decay of the influences (with time) can be simulated using fractional integral (Yuan et al. 2013). Accordingly, FISM was designed using Riemann-Liouville

Fractional integral formula, as shown below (Yuan et al. 2014),

$$x(t) = \frac{1}{\Gamma(q)} \int_{u=0}^{t-\delta} \frac{\varepsilon(u)}{(t-u)^{1-q}} du + \varepsilon(t), \quad (2)$$

where  $\varepsilon(u)$  denotes the historical dynamical excitation,  $\Gamma$  is the gamma function,  $q$  is the integral order,  $t-u$  represents the distance between historical time point  $u$  and the present time  $t$ , and  $\delta$  is the sampling time interval. In FISM, the current climate state  $x(t)$  consist of two parts. On the right side of Eq. (2), the first part represents the influences of the past accumulated till the present time, while the second part is the current dynamical excitations. Since the integral order  $q$  can be calculated from the DFA exponent  $\alpha$  as  $q = \alpha - 0.5$ , suppose the historical observations  $x(u) = x(0), x(\delta), \dots, x(t-\delta)$  are known, one can derive reversely the historical  $\varepsilon(u)$ , with which the memory part at present time  $t$  can be calculated as following,

$$M(t) = \frac{1}{\Gamma(q)} \int_{u=0}^{t-\delta} \frac{\varepsilon(u)}{(t-u)^{1-q}} du, \quad (3)$$

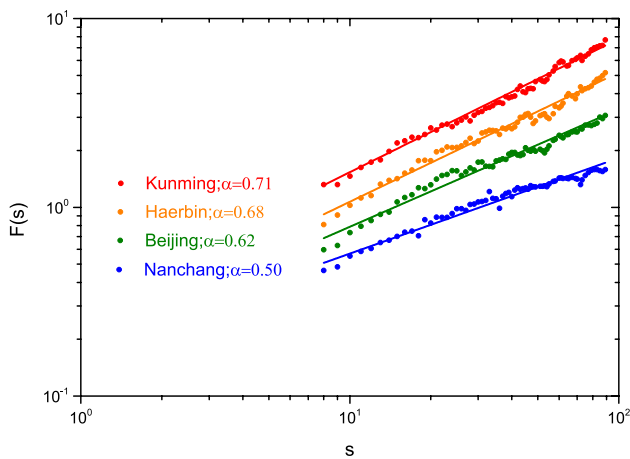
If  $x(t)$  is known, from Eq. (1) one can easily calculate the non-memory part  $\varepsilon(t)$  at the present time  $t$ . Therefore, using FISM, both memory part  $M(t)$  and the non-memory part  $\varepsilon(t)$  can be extracted.

## 3 Evaluation of the simulated memory part

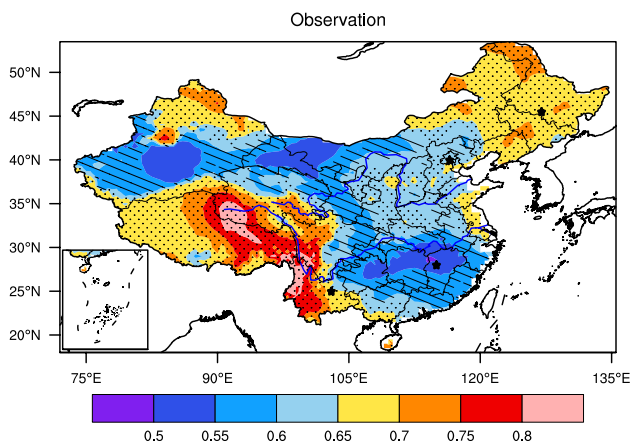
In order to evaluate the simulated memory part of the surface air temperatures over China, we employed the DFA2 to check whether the decadal hindcast simulations can reproduce the observed LTM. It has been well recognized that nearly all the SAT records over different regions of China are characterized by LTM (Yuan et al. 2010). As shown in Fig. 1, we applied DFA2 to four randomly selected in-situ SAT records. In this log-log plot, straight lines ranging from 9 to 90 months are observed, which means power-law increase of the fluctuation function  $F(s)$  with the time scale  $s$ . The slope of the straight line is the DFA2 exponent  $\alpha$ . For Kunming (red), Harbin (orange), and Beijing (green), the  $\alpha$  values are all larger than 0.5, while for Nanchang (blue), the DFA2 exponent  $\alpha = 0.5$ .

To better show the LTM of observed SAT records over China, we further analyzed the observational grid dataset, and the distribution of  $\alpha$  values are shown in Fig. 2. As one can see, higher  $\alpha$  values are found in the northeast and southwest of China. But in the south of the Yangtze River, the SAT records have much lower  $\alpha$  values. In some regions, the  $\alpha$  values are even close to 0.5. Considering that only 360 months were used for the DFA calculations, there are





**Fig. 1** DFA results of surface air temperatures (SAT) observed from the four randomly selected stations. From top to bottom, they are Kunming (red), Haerbin (orange), Beijing (green), and Nanchang (blue). In this double-logarithmic plot, the slope of each line represents the DFA exponent  $\alpha$ . Since only 30 years (1981–2010) data are analyzed,  $\alpha$  was measured using the scaling range of 9–90 months. As one can see, the SAT in Kunming, Haerbin, and Beijing are all characterized by LTM ( $\alpha > 0.5$ ), while in Nanchang, the SAT is not long-term correlated



**Fig. 2** Geographical distribution of the DFA exponent  $\alpha$  calculated from the observational grid SAT dataset. Higher  $\alpha$  values are found in the northeast and southwest of China, while in the south of Yangtze River, the  $\alpha$  values are much lower. The dotted area represent the regions with  $\alpha > 0.616$ , while the slash area show the regions with  $0.558 < \alpha < 0.616$ . The four stars shown in the map are the locations of the four stations analyzed in Fig. 1

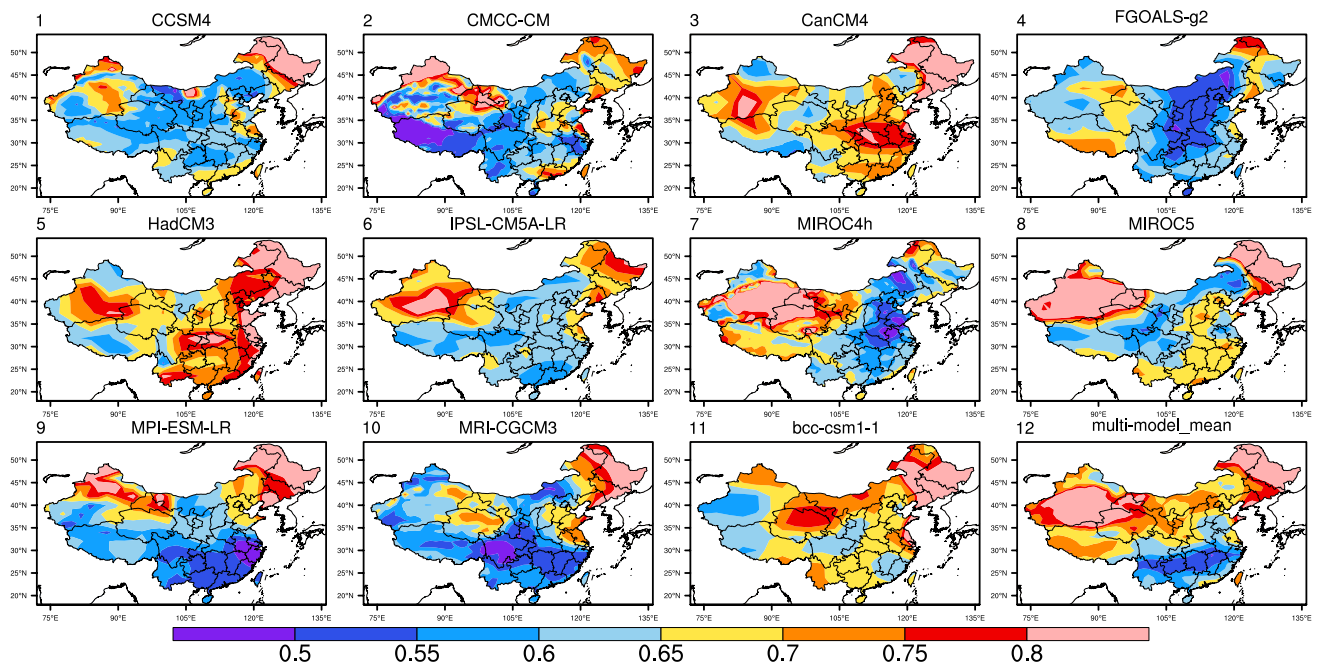
unavoidable uncertainties in the estimation of  $\alpha$ . To rule out that the observed  $\alpha$  value is an artifact of this uncertainty, it is necessary to estimate how  $\alpha$  varies when a large number of white noises (with the same length  $l = 360$ ) are analyzed. By (1) generating two long white noises (i.e. length  $L = 2,000,000$ ), (2) dividing the two long white noises into 10,000 short data of length 360, and (3) applying DFA2 to

these 10,000 short data, we found that the standard deviation of the 10,000  $\alpha$  values is  $sd = 0.058$ . Accordingly, the  $\alpha$  values larger than 0.616 ( $0.5 + 2sd$ ) indicates the existence of significant LTM (see the dotted area in Fig. 2). For regions with  $0.558 < \alpha < 0.616$  (see the slash area in Fig. 2), one can only roughly consider the SATs have weak LTM. For the very small area with  $\alpha < 0.558$ , one cannot rule out the possibility that the SATs behave as white noises. This distribution is in line with many previous reports, even though different datasets are analyzed (Yuan et al. 2010).

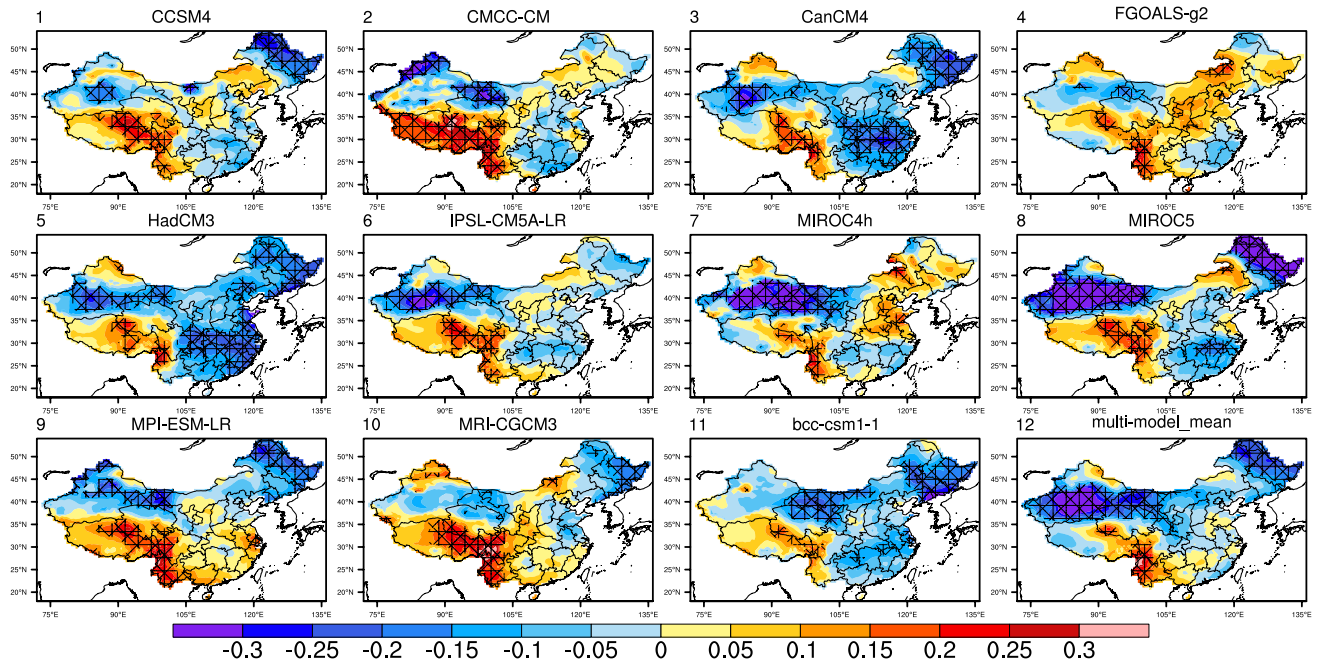
However, if we apply DFA2 to the model simulations, quite different results are obtained. Figure 3 shows the distributions of  $\alpha$  for the 11 CMIP5 models. There is no identical pattern among the models. The surface air temperatures simulated by some models (such as CanCM4, HadCM3) have stronger LTM in the south of China, while the simulated LTM by some other models (such as IPSL-CM5A-LR, MIROC4h, MIROC5) is stronger in the northwest. Most models except FGOALS-g2 and MIROC4h simulated the strong LTM in the northeast of China, but nearly no model reproduced the strong LTM in the southwest of China. By calculating the differences between the observed  $\alpha$  values (Fig. 2) and the simulated  $\alpha$  values (Fig. 3), very poor skills of the 11 models in reproducing LTM are found. As shown in Fig. 4, after taking the uncertainties of the  $\alpha$  values into account (see Fig. S2 in the SI), the LTM in the southwest of China are found significantly underestimated in all the models, while in the northeast and the northwest, most models significantly overestimated the LTM. The simulations are slightly better in the south of the Yangtze River, but there are still unexpected strong LTM simulated in 4 of the 11 models.

Besides comparing the observed LTM with individual models, we also computed Multi-Model Means (MMM) over the 11 CMIP5 models. But the simulated LTM does not improve much. As shown in Figs. 3 and 4 (the sub-figure no. 12), the LTMs in the northeast and the northwest of China are significantly overestimated, while over the southwest of China, the LTMs are significantly underestimated. Only in the south of the Yangtze River, slightly better LTMs of the surface air temperatures are simulated.

To better evaluate the capability of the CMIP5 models in simulating LTM, we further presented the Taylor diagram in Fig. 5 (Taylor 2001). At first glance, one can find very poor spatial correlation coefficients (SCC) for all the models. Except FGOALS-g2, the SCCs of all the other 10 models are smaller than 0.3, and some models (6 out of 11) even have negative SCCs. This indicates poor similarity between the observed and the simulated LTM pattern (Figs. 2, 3). Regarding of the normalized standard deviations (NSD), only two models (FGOALS-g2 and IPSL-CM5A-LR) have similar standard deviations as that obtained from the observational dataset. All the other models including the MMM overestimated the spatial variability of the LTM in surface

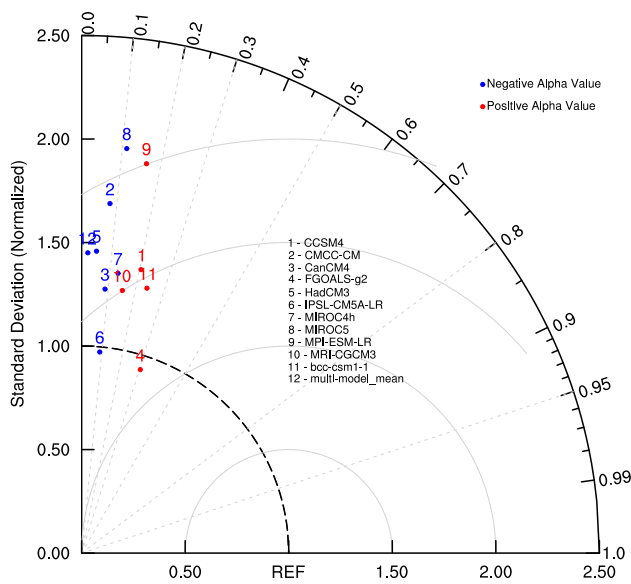


**Fig. 3** Geographical distribution of the DFA exponent  $\alpha$  calculated from the decadal hindcast simulations of the 11 CMIP5 models, as well as the multi-model means (MMM). As one can see, no model reproduced the observed LTM as shown in Fig. 2



**Fig. 4** Differences between the observed  $\alpha$  in Fig. 2 and the simulated  $\alpha$  in Fig. 3. The values shown in this figure are obtained by subtracting the simulated  $\alpha$  from the observed  $\alpha$ . The hatching area represents significantly different  $\alpha$  values at 95% confidence level, and the bounds of the 95% confidence intervals are estimated as follows: for each grid point in Fig. 2, we (1) generated two long ( $L = 2,000,000$ )

artificial data with the  $\alpha$  value the same as the observed  $\alpha$  (at this grid point), by using Fourier Filtering technique (Turcotte 1997). After (2) dividing the two long artificial data into 10,000 short data of length  $l = 360$ , we (3) applied DFA2 to these short data. From the 10,000  $\alpha$  values, the bounds of the 95% confidence interval were determined (see Fig. S2 in the SI)



**Fig. 5** Taylor diagram of  $\alpha$  values over China between observations (Fig. 3) and model simulations (Fig. 4). The azimuthal positions represent the spatial correlation coefficients between the observed  $\alpha$  and simulated  $\alpha$ . The radial distances from origin to the dots stand for the standard deviations, which are normalized according to that of the observed  $\alpha$  values. With the observed  $\alpha$  values serving as a reference, the radial distances from “REF” to the dots are the root mean square errors (RMSE). The blue and red dots represent negative and positive correlations, respectively. Each dot represents a model (incl. multi-model mean), which is identified by its number on the right

air temperatures. Due to the small SCCs and the large NSDs, all the models have big root mean square errors (RMSE). As shown in Fig. 5, there are long distances from the model simulations to the Reference (REF).

To quantify the model performance in simulating LTM of surface air temperatures, we further calculate the skill score, which is defined as below (Taylor 2001),

$$S = \frac{4(1 + R)}{(\hat{\sigma}_f + 1/\hat{\sigma}_f)^2(1 + R_0)}, \tag{4}$$

where  $R$  is the spatial correlation coefficient,  $\hat{\sigma}_f$  is the spatial standard deviation of the simulated LTM divided by that of the observed LTM, and  $R_0$  is the maximum correlation attainable (Kusunoki and Arakawa 2015). Here we assumed that  $R_0 = 1$ . As shown in Table 2, the scores are low for all the models. Many models (7 of 11) even have skill scores that are smaller than 0.5. Among the 11 models, FGOALS-g2 has the highest skill score, which is only 0.65. While MIROC5 has the lowest skill score, which is 0.29. These low skill scores again indicate that the decadal hindcast experiment in CMIP5 models cannot reproduce the observed LTM in surface air temperatures over China. In other words, the memory part  $M(t)$  cannot be properly simulated from the decadal hindcast experiments in the CMIP5 models.

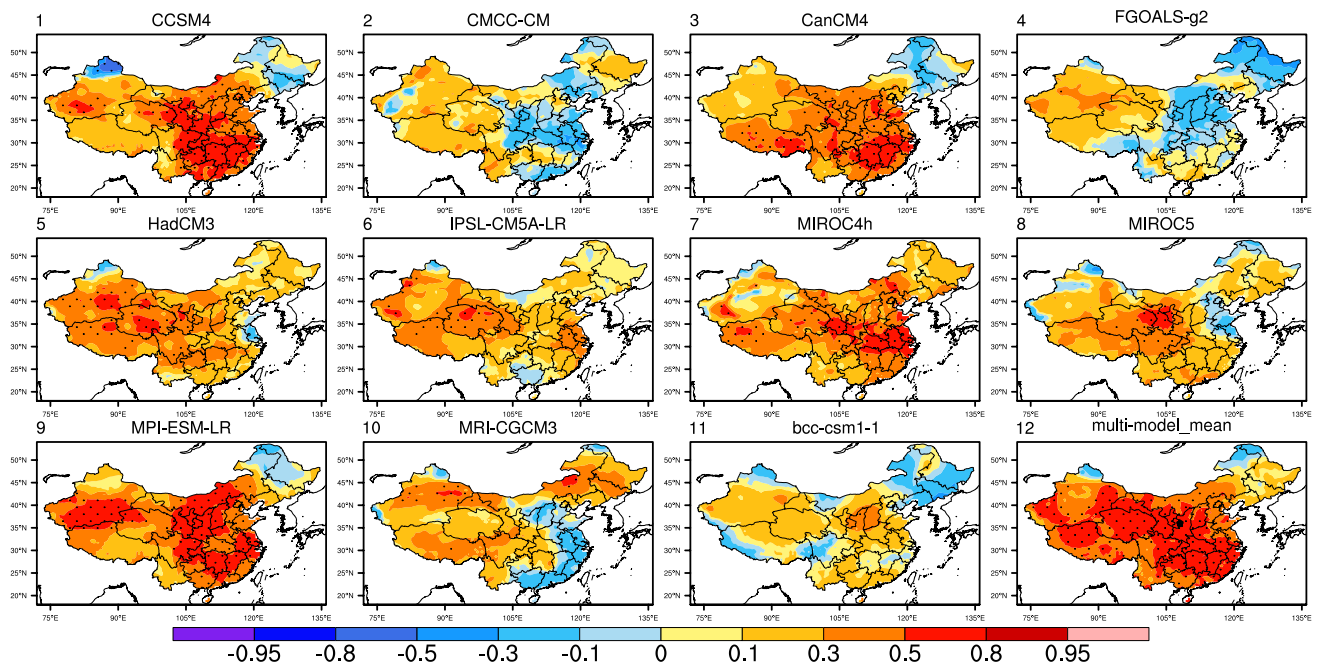
### 4 Evaluation of the simulated non-memory part

Although the 11 CMIP5 models are not capable of reproducing the observed LTM of surface air temperatures in the decadal hindcast experiments, one can remove the biased memory part  $M(t)$  and only focus on the non-memory part  $\epsilon(t)$ . In this section, we applied FISM to both the observational grid dataset and the model simulated surface air temperatures. As described in Sect. 2.2, using the  $\alpha$  values of each time series, the corresponding non-memory part  $\epsilon(t)$  can be calculated quantitatively. Then we evaluated the simulated  $\epsilon(t)$  for each model. Since it makes little sense to discuss the month-to-month simulations in a decadal hindcast experiment, we calculated annual  $\epsilon(t)$  from both observational dataset and model simulations. By studying their similarity, the performance of CMIP5 models in simulating  $\epsilon(t)$  was evaluated on annual scale.

Figure 6 shows the correlations of the simulated  $\epsilon(t)$  with that obtained from observational dataset. As one can see, most models (e.g. CMCC-CM, FGOALS-g2, HadCM3, IPSL-CM5A-LR, MIROC4h, MIROC5, MRI-CGCM3, and bcc-csm1-1) failed in simulating proper  $\epsilon(t)$ . For these models, low correlations are found over most regions of the country. In some regions (e.g. the northeast of China), there are even negative correlations. Only a few models show some skills, such as in CCSM4, CanCM4, and MPI-ESM-LR, there are significant correlations between the simulated and observed  $\epsilon(t)$  in the middle and eastern part of China (except the northeast). However, if we consider the multi-model means (MMM), considerable regions with significant correlations are found. As shown in the sub-figure no. 12, the simulated  $\epsilon(t)$  is highly correlated with the observed  $\epsilon(t)$  in most regions. In some specific regions, the correlations are even higher than 0.5, indicating reasonable simulations of  $\epsilon(t)$  for the time period of being studied (1981–2010).

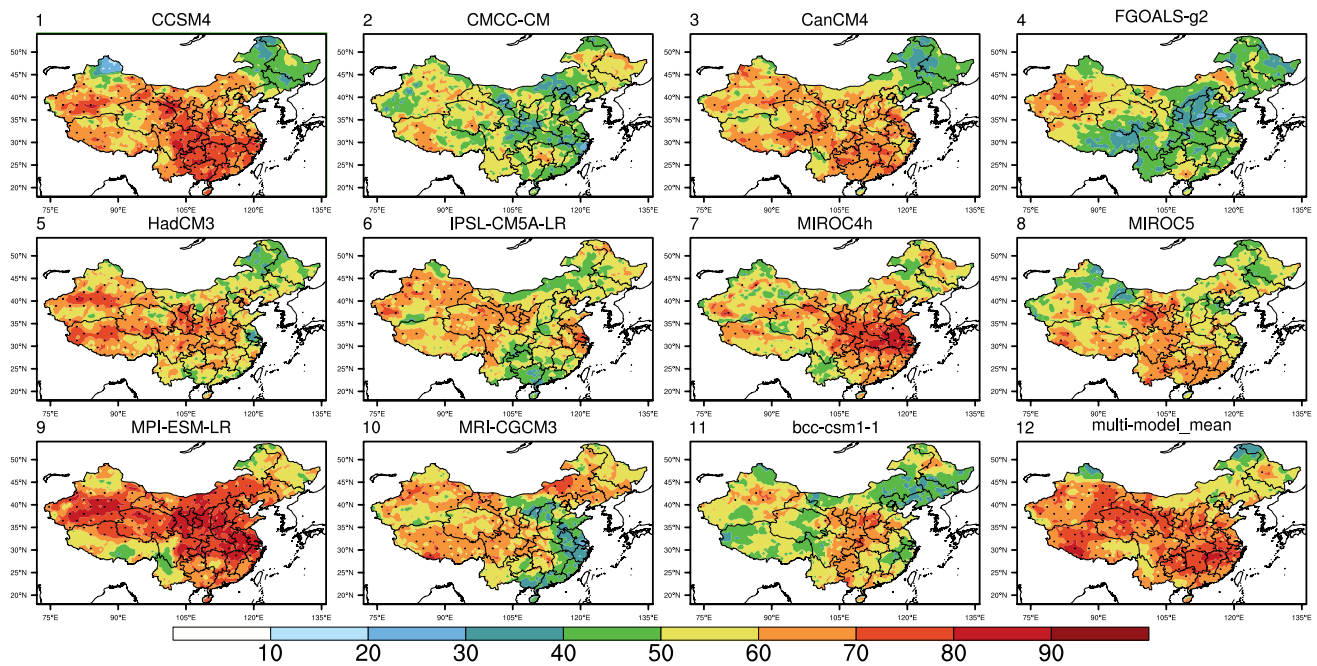
Different from the past influences  $M(t)$ , since  $\epsilon(t)$  determines the further changing directions of the climate states (Yuan et al. 2014), it is important that the simulated  $\epsilon(t)$  has the same sign as the observed  $\epsilon(t)$ . Therefore, besides studying the correlations, we also calculated the percentage of the simulated  $\epsilon(t)$  that has the same sign as the observed  $\epsilon(t)$ . Figure 7 shows the geographical distribution of the percentage for each model. Better than Fig. 6, most models except CMCC-CM and FGOALS-g2 more or less have some regions where the percentage is higher than 60%. In some models such as the CCSM4 and the MPI-ESM-LR, the percentage is higher than 60% over most regions of the country. In some specific regions, it can be even higher than 80%. Under this situation, if we consider multi-model means (the sub-figure no. 12 in Fig. 7),





**Fig. 6** Correlation coefficients ( $r$ ) between the observed and the simulated non-memory part  $\varepsilon(t)$ . Annual  $\varepsilon(t)$  from 1981 to 2010 are used for the calculation of correlation. After obtaining the critical value from the student's  $t$  test [df (degrees of freedom) = 28,  $t_{0.05/2,28} = 2.048$ ,  $r_{0.05/2,28} = 0.361$ ], areas with statistically significant

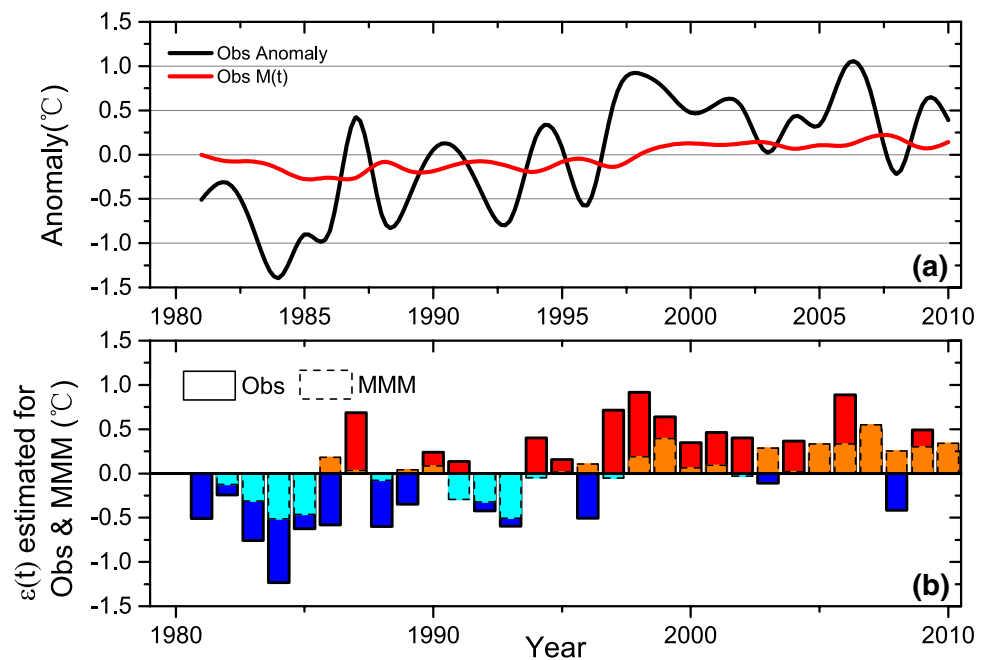
correlations are marked by small dots. Besides the results for individual models, correlations between the observed  $\varepsilon(t)$  and the simulated  $\varepsilon(t)$  from multi-model means (MMM) are also shown (see the sub-figure no. 12), and significant correlations are found over most regions



**Fig. 7** Percentage of the simulated  $\varepsilon(t)$  that has the same sign as the observed  $\varepsilon(t)$ . The small dots in the map mark the areas where percentages are higher than 60%. As in Fig. 6, besides individual models,

the results of the multi-model means (MMM) are also shown as the sub-figure no. 12

**Fig. 8** Taking the surface air temperature of a randomly selected grid point (the bigger black points shown in the sub-figure no. 12 of Fig. 6) as an example, **a** the annual temperature anomalies (black) and the  $M(t)$  (red) extracted by FISM; **b** the observed  $\epsilon(t)$  and the simulated  $\epsilon(t)$  from multi-model means (MMM)



large areas (except the northeast of China and the Inner Mongolia) with high percentage can be identified, which again indicates potential predictability of  $\epsilon(t)$  in the decadal hindcast experiments.

### 5 Discussion and conclusion

In this study, the memory and non-memory parts of the simulated surface air temperatures (SAT) over China are evaluated. In the decadal hindcast experiments of 11 CMIP5 models, we find significantly overestimated long-term memory (LTM) over most regions of China (e.g., northeast and northwest). While in the southwest, the LTMs are significantly underestimated. The poor skills in the LTM simulation indicate that current state-of-the-art models have difficulties in simulating the memory part  $M(t)$ . In other words, it is still challenging for current dynamical models to fully capture the influences from past on the current climate states.

However, after removing the biased  $M(t)$  from the simulated SAT, the remaining non-memory part  $\epsilon(t)$  is found reasonably modeled in the multi-model means. Over most regions, the simulated annual  $\epsilon(t)$  are significantly correlated with the observed annual  $\epsilon(t)$ . In some specific regions, the correlations can be larger than 0.5. Considering the percentage of the simulated  $\epsilon(t)$  that has the same sign as the observed  $\epsilon(t)$ , high percentages are widely found over the country. In some regions, the percentages are even as high as 80%. Therefore, it is evident to say that the current state-of-the-art models have a certain ability in simulating the non-memory part  $\epsilon(t)$  on annual scale.

Since nearly all the SAT records observed over different regions of China are characterized by LTM, their current states depend on (1) the influences from past ( $M(t)$ ) and (2) the current dynamical excitations ( $\epsilon(t)$ ) (Yuan et al. 2013, 2014). From our findings, obviously the biased simulations of  $M(t)$  are responsible for the errors of dynamical models. Therefore, to improve the model simulations, one needs to better capture the memory part of the surface air temperatures.

In fact, using fractional integral statistical model (FISM), the memory part  $M(t)$  at the present time can be extracted quantitatively as long as the past non-memory part  $\epsilon(u)$  are known. As shown in Fig. 8a, we took the observational SAT (black curve) of one grid point from central China as an example (see the black point in the sub-figure no. 12 of Fig. 6). Using FISM, the memory part  $M(t)$  (red curve) were calculated. However, these  $M(t)$  values are not predictions as they are calculated using the past non-memory part  $\epsilon(u)$ . To continue the calculation of  $M(t)$  till future, one needs to first predict the future non-memory part  $\epsilon(t)$ , which is beyond the ability of FISM. In this study, we found current state-of-the-art models have a certain ability in simulating annual  $\epsilon(t)$ . As shown in Fig. 8b, the simulated  $\epsilon(t)$ s (dashed histogram) are quite close to the true  $\epsilon(t)$ s. Accordingly, a new strategy for climate prediction comes out. That is, one may combine FISM with dynamical models. Using the simulated  $\epsilon(t)$  from dynamical models, we are able to drive FISM, which in turn provides more reliable estimations of  $M(t)$ . By coupling the simulated  $\epsilon(t)$  from dynamical models and the estimated  $M(t)$  from FISM, improved climate predictions (e.g. on annual scale) with the effects of LTM properly considered may become possible.



**Acknowledgements** Many thanks are due to supports from National Key R&D Program of China (2016YFA0600404 and 2016YFA0601504) and from National Natural Science Foundation of China (nos. 41675088 and 41475005). N. Y. thanks also the supports from CAS Pioneer Hundred Talents Program. We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Abry P, Veitch D (1998) Wavelet analysis of long-range-dependent traffic. *IEEE Trans Inf Theory* 44(1):2–15
- Arneodo A, Bacry E, Graves PV, Muzy JF (1995) Characterizing long-range correlations in DNA sequences from wavelet analysis. *Phys Rev Lett* 74:3293–3296
- Bogachev MI, Bunde A (2011) On the predictability of extreme events in records with linear and nonlinear long-range memory: efficiency and noise robustness. *Phys A* 390:2240–2250
- Bunde A, Eichner JF, Kantelhardt JW, Havlin S (2005) Long-term memory, a natural mechanism for the clustering of extreme events and anomalous residual times in climate records. *Phys Rev Lett* 94(4):048701
- Bunde A, Büntgen U, Ludescher J, Luterbacher J, von Storch H (2013) Is there memory in precipitation? *Nat Clim Change* 3:174–175
- Chen X, Lin G, Fu Z (2007) Long-range correlations in daily relative humidity fluctuations: a new index to characterize the climate regions over China. *Geophys Res Lett* 34:L07804
- Dangendorf S, Rybski D, Mudersbach C, Müller A, Kaufmann E, Zorita E, Jensen J (2014) Evidence for long-term memory in sea level. *Geophys Res Lett* 41:5530–5537
- Doblas-Reyes FJ, Garcia-Serrano J, Lienert F, Biescas AP, Rodrigues LRL (2013) Seasonal climate predictability and forecasting: status and prospects. *WIREs Clim Change* 4:245–268
- Eichner JF, Koscielny-Bunde E, Bunde A, Havlin S, Schellnhuber H-J (2003) Power-law persistence and trends in the atmosphere: a detailed study of long temperature records. *Phys Rev E* 68:046133
- Feng T, Fu Z, Deng X, Mao J (2009) A brief description to different multi-fractal behaviors of daily wind speed records over China. *Phys Lett A* 373:4134–4141
- Fraedrich K, Blender R (2003) Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys Rev Lett* 90:108501
- Franzke C (2010) Long-range dependence and climate noise characteristics of Antarctic temperature data. *J Clim* 23:6074–6081
- Franzke C (2012) Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *J Clim* 23:4172–4183
- Govindan RB, Vyushin D, Bunde A, Brenner S, Havlin S, Schellnhuber H-J (2002) Global climate models violate scaling of the observed atmospheric variability. *Phys Rev Lett* 89:028501
- Jiang L, Li N, Zhao X (2017) Scaling behaviors of precipitation over China. *Theor Appl Climatol* 128:63–70
- Kantelhardt JW, Koscielny-Bunde E, Rego HHA, Havlin S (2001) Detecting long-range correlations with detrended fluctuation analysis. *Phys A* 295:441–454
- Kantelhardt JW, Koscielny-Bunde E, Rybski D, Braun P, Bunde A, Havlin S (2006) Long-term persistence and multifractality of precipitation and river runoff records. *J Geophys Res* 111:D01106
- Koscielny-Bunde E, Bunde A, Havlin S, Roman HE, Goldreich Y, Schellnhuber HJ (1998) Indication of a universal persistence law governing atmospheric variability. *Phys Rev Lett* 81:729–732
- Kumar S, Merwade V, Kinter JL III, Niyogi D (2013) Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. *J Clim* 26:4168–4185
- Kusunoki S, Arakawa O (2015) Are CMIP5 models better than CMIP3 models in simulating precipitation over East Asia? *J Clim* 28:5601–5621
- Lennartz S, Bunde A (2009) Trend evaluation in records with long-term memory: application to global warming. *Geophys Res Lett* 36:L16706
- Lovejoy S (2015) A voyage through scales, a missing quadrillion and why the climate is not what you expect. *Clim Dyn* 44:3187–3210
- Lovejoy S, Schertzer D (2012) Extreme events and natural hazards: the complexity perspective, low frequency weather and the emergence of the climate. In: Sharma AS, Bunde A, Baker D, Dimri VP (eds) *AGU monographs*. Washington, D. C., pp 231–254
- Ludescher J, Bunde A (2016) Long-term persistence enhances uncertainty about anthropogenic warming of Antarctica. *Clim Dyn* 46(1):263–271
- Luo M, Leung Y, Zhou Y, Zhang W (2015) Scaling behaviors of global sea surface temperature. *J Clim* 28:3122–3132
- Massh M, Kantz H (2016) Confidence intervals for time averages in the presence of long-range correlations, a case study on Earth surface temperature anomalies. *Geophys Res Lett* 43(17):9243–9249
- Meehl GA et al (2014) Decadal climate prediction: an update from the Trenches. *Bull Am Meteorol Soc* 95:243–267
- Monetti RA, Havlin S, Bunde A (2003) Long-term persistence in the sea surface temperature fluctuations. *Phys A* 320:581–589
- Palmer TN (2000) Predicting uncertainty in forecasts of weather and climate. *Rep Prog Phys* 63:71–116
- Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. *Phys Rev E* 49:1685–1689
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) *Numerical recipes in C: the art of scientific computing*, 2nd edn. Cambridge University Press, Cambridge
- Rybski D, Bunde A, von Storch H (2008) Long-term memory in 1000-year simulated temperature records. *J Geophys Res* 113:D02106
- Slingo J, Palmer TN (2011) Uncertainty in weather and climate prediction. *Philos Trans R Soc A* 369:4751–4767
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res* 106:7183–7192
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498
- Turcotte D (1997) *Fractals and Chaos in geology and geophysics*, 2nd edn. Cambridge University Press, Cambridge
- Vyushin DI, Kushner PJ (2009) Power-law and long-memory characteristics of the atmospheric general circulation. *J Clim* 22(11):2890–2904

- Vyushin DI, Zhidkov I, Havlin S, Bunde A, Brenner S (2004) Volcanic forcing improves atmosphere–ocean coupled general circulation model scaling performance. *Geophys Res Lett* 31:L10206
- Yuan N, Fu Z, Mao J (2010) Different scaling behaviors in daily temperature records over China. *Phys A* 389:4087–4095
- Yuan N, Fu Z, Liu S (2013) Long-term memory in climate variability: A new look based on fractional integral techniques. *J Geophys Res* 118:12962–12969
- Yuan N, Fu Z, Liu S (2014) Extracting climate memory using fractional integrated statistical model: a new perspective on climate prediction. *Sci Rep* 4:6577
- Yuan N, Ding M, Ludescher J, Bunde A (2017) Increase of the Antarctic sea ice extent is highly significant only in the Ross Sea. *Sci Rep* 7:41096
- Yuan N, Huang Y, Duan J, Zhu C, Xoplaki E, Luterbacher J (2018) On climate prediction: how much can we expect from climate memory? *Climate Dyn* (**in press**)
- Zhao S, He W, Jiang Y (2018) Evaluation of NCEP-2 and CFSR reanalysis seasonal temperature data in China using detrended fluctuation analysis. *Int J Climatol* 38:252–263
- Zhu X, Fraedrich K, Liu Z, Blender R (2010) A demonstration of long-term memory and climate predictability. *J Clim* 23:5021–5029