



1 **Reconstructing coupled time series in climate systems by machine learning**

2 Yu Huang¹, Lichao Yang¹, Zuntao Fu^{1*}

3 ¹Lab for Climate and Ocean-Atmosphere Studies, Dept. of Atmospheric and Oceanic Sciences,
4 School of Physics, Peking University, Beijing, 100871, China

5 *Correspondence to:* Zuntao Fu (fuzt@pku.edu.cn)

6 **Abstract.**

7 Despite the great success of machine learning, its applications in climate dynamics have not been
8 well developed. One concern might be how well the trained neural networks could learn a dynamical
9 system and what can be the potential applications of this kind of learning. Detailed studies show that
10 the coupling relations or dynamics among variables in linear or nonlinear systems can be well learnt
11 by reservoir computer (RC) and long short-term memory (LSTM) machine learning, and these learnt
12 coupling relations can be further applied to reconstruct one series from the other dominated by
13 common coupling dynamics. In order to validate the above conclusions, toy models are applied to
14 address the following three questions: (i) what can be learnt from different dynamical time series by
15 machine learning; (ii) what factors significantly influence machine learning reconstruction; and (iii)
16 how to select suitable explanatory or input variables for the reconstructed variable for machine
17 learning. The results from these toy models show that both RC and LSTM can indeed learn coupling
18 relations among variables, and the learnt implicit coupling relation can be applied to accurately
19 reconstruct one series from the other. Both linear and nonlinear coupling relations between variables
20 can influence the quality of the reconstructed series. If there is a strong linear coupling between
21 variables, all of variables can be taken as explanatory variables for the reconstructed variable, and the



22 reconstruction can be bi-directional. However, when the linear coupling among variables is much
23 weaker, but with stronger nonlinear causality among variables, the reconstruction quality is direction-
24 dependent and it may be only uni-directional. We propose using convergent cross mapping causality
25 (CCM) index $\rho_{a \rightarrow b}$ to determine which variable can be taken as the reconstructed one and which can
26 be taken as the explanatory variable. For example, the Pearson correlation between the average
27 Tropical Surface Air Temperature (TSAT) and the average Northern Hemispheric SAT (NHSAT) is as
28 weak as 0.08, but the CCM index of NHSAT cross maps TSAT is $\rho_{N \rightarrow T} = 0.70$, it means that NHSAT
29 could be taken as the explanatory variable. Then we find that TSAT can be well reconstructed from
30 NHSAT by means of RC. However, the reconstruction quality in the opposite direction is poor, because
31 the CCM index of TSAT cross maps NHSAT is only $\rho_{T \rightarrow N} = 0.24$. These results also provide insights
32 on machine learning approaches for paleoclimate reconstruction, parameterization scheme, and
33 prediction in related climate studies.

34 **Key words:** Reconstruction, Climate time series, Machine learning, Causality, Reservoir computer,
35 Surface air temperature



36 **Highlights:**

- 37 i) Learnt coupling dynamics between series by machine learning can be used to reconstruct series.
- 38 ii) Reconstruction quality is direction- and variable-dependent for nonlinear systems.
- 39 iii) The CCM index is a potential indicator to choose reconstructed and explanatory variables.
- 40 iv) The tropical average SAT can be well reconstructed from the average Northern Hemispheric SAT.

41



42 1 Introduction

43 Making use of observed climatic time series is very important to solve climate problems, such
44 as paleoclimate reconstruction (Brown, 1994; Emile-Geay and Tingley, 2016), interpolation for the
45 missing points in measurements (Hofstra et al., 2008), parameterization schemes (Wilks, 2005; Vissio
46 and Lucarini, 2018), and seasonal climate prediction (Comeau et al., 2017; Wang et al., 2017). Neural
47 network-based machine learning, which recently attracts great attention in climate studies (Reichstein
48 et al., 2019), is useful for these problems. First, machine learning has been widely applied to
49 downscaling and data mining (Mattingly et al., 2016; Racah et al., 2017). Second, the machine
50 learning frameworks for time series were used for predicting climate variables, such as temperature,
51 humidity, runoff and air pollution (Zaytar and Amrani, 2016; Biancofiore et al., 2017; Kratzert et al.,
52 2019; Feng et al., 2019). Recently it is demonstrated the large potentials for machine learning to
53 simulate the temporal dynamics of complex systems (Pathak et al., 2017; Du et al., 2017; Watson,
54 2019), while the physics of systems is suggested for consideration. For example, the results of
55 applying machine learning to Lorenz system (Lorenz, 1963) and Rossler models showed that their
56 chaotic attractors were able to be well simulated (Pathak et al., 2017; Lu et al., 2018; Carroll, 2018).
57 Chaos is the characteristic of climatic series (Lorenz, 1963; Patil et al., 2001), so that success in chaotic
58 time series offers deep understanding for the application of machine learning.

59 Applying machine learning to climatic series attracts much attention, but it is still unclear what
60 can be learnt by machine learning during the training process, and if there exist physical characteristics
61 in climatic series, what is the key factor determining the performance of machine learning. This is
62 crucial for investigating why machine learning performs not well with some datasets, and how to
63 improve the performance for them. The cross-correlation of climatic time series might be first thought



64 of, because linear correlation is the implicit assumption for traditional statistical methods, and it is
65 known that they often fail if linear correlation is weak (Brown, 1994; Sugihara et al., 2012; Emile-
66 Geay and Tingley, 2016). However, previous studies (Sugihara et al., 2012; Emile-Geay and Tingley,
67 2016) suggest that a nonlinear correlation could be useful even though the linear correlation is weak.
68 For instance, the linear cross-correlations of sea surface temperature series observed in different
69 tropical areas are unstable and vary with time, which leads to an overall weak correlation, but this
70 nonlinear correlation could result in better El Niño predictions (Ludescher et al., 2014; Conti et al.,
71 2017). The phase plots of the ENSO/PDO index and some proxy variables are not linear lines but
72 nonlinear trajectories, which contributes greatly to reconstructing longer climate series (Mukhin et al.,
73 2018). These studies indicate that nonlinear relations can contribute to better analysis, reconstruction,
74 and prediction (Hsieh et al., 2006; Donner, 2012; Schurer et al., 2013; Badin et al., 2014; Drótos et
75 al., 2015; Van Nes et al., 2015; Comeau et al., 2017; Vannitsem and Ekkelmans, 2018). Accordingly,
76 when applying machine learning to reconstruct climatic series, is it necessary to give attention to the
77 linear or nonlinear relations induced by the physical couplings? This is an open question.

78 To this question, we first discuss a real-world example from climate system. As we know, there
79 exists coupling in the atmospheric motions between the tropics and the Northern Hemisphere, which
80 is realized through the transfer of atmospheric energy (Farneti and Vallis, 2013). Due to the underlying
81 complicated processes, it is difficult to use a formula to cover this coupling between the tropical
82 average surface air temperature (TSAT) series and the Northern Hemispheric surface air temperature
83 (NHSAT) series. In the result part, it will be shown that the linear Pearson correlation between the
84 TSAT and NHSAT is very weak, but we can still use machine learning to well reconstruct the TSAT
85 series from the NHSAT series. However, the NHSAT series cannot be reconstructed from the TSAT



86 series. Such a contrasting result means that the machine learning approach is not the same as the
87 traditional statistical methods. Accordingly, is there any physical feature hidden in the NHSAT-TSAT
88 coupling influencing this result? Furthermore, one might wonder if the similar phenomenon will be
89 universal for some other coupled climate series.

90 In this paper, to make progress towards understanding how machine learning approach is
91 influenced by the physical couplings of climatic series, we will investigate the machine learning
92 performances and their dependence on different coupling dynamics in climatic conceptual models. It
93 will be demonstrated that machine learning is able to “learn” these couplings, and the “learnt” coupling
94 relation are very useful for reconstructing climatic series. Moreover, our study also analyzes the
95 features of linear relations and nonlinear relations in climate systems, and a method to select
96 explanatory variables for machine learning is proposed. Hence, the underlying mechanism of the
97 abovementioned NHSAT-TSAT reconstruction will be also explained.

98 Our paper is organized as follows. In section 2, the methods about reconstructing time series and
99 data are introduced. In section 3, the ability of machine learning to “learn” different coupling relations
100 of climate system, and the potential application of machine learning to reconstruct climate series, are
101 shown. Then the performance dependence of machine learning on different factors is presented. The
102 underlying coupling strength is demonstrated to play a key role. In section 4, the application to real-
103 world climate series is shown. Summary is made in section 5.

104 **2 Methods and data**

105 **2.1 Learning coupling relations and reconstructing coupled time series**

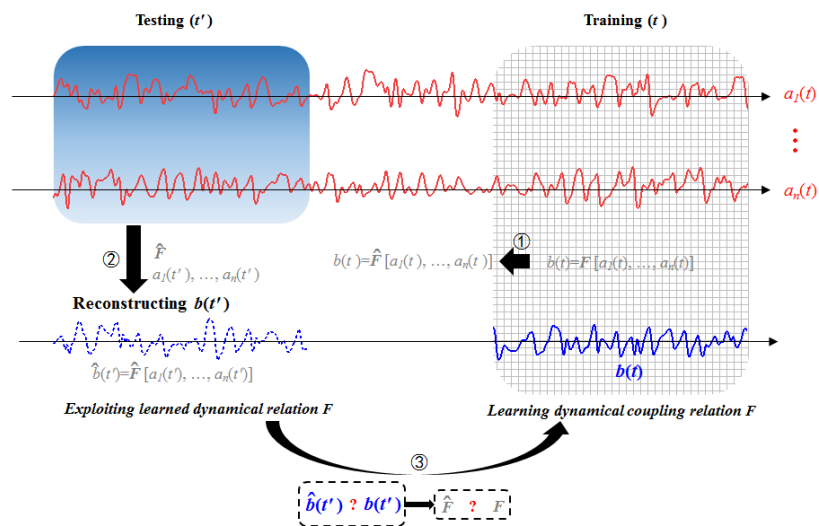
106 Firstly, we introduce our workflow for learning couplings of dynamical systems by machine



107 learning, and reconstructing the coupled time series. The total time series can be divided into two parts:
108 the training series (time lasting denoted as t) and the testing series (time lasting denoted as t'). For the
109 systems of toy models, the coupling relation or dynamics is stable and unchanged with time, i.e., there
110 is the stable coupling or dynamic relation $b(t) = F[a_1(t), a_2(t), \dots, a_n(t)]$ among inputs
111 $a_1(t), a_2(t), \dots, a_n(t)$ and output $b(t)$. If this inherent coupling relation can be “learnt” by machine
112 learning in the training series, the “learnt” coupling relation should be reflected by machine learning
113 in the testing series. Therefore, the workflow of our study can be summarized as follows (see Fig. 1):
114 (i) During the training period, $a_1(t), a_2(t), \dots, a_n(t)$ and $b(t)$ are input into the machine learning
115 frameworks to learn the coupling or dynamic relation $b(t) = F[a_1(t), a_2(t), \dots, a_n(t)]$. The “learnt”
116 coupling relation is denoted as $b(t) = \hat{F}[a_1(t), a_2(t), \dots, a_n(t)]$. Then it is tested whether this coupling
117 relation can be learnt by machine learning.
118 (ii) The second step is accomplished with the testing series to apply the learnt coupling relation \hat{F}
119 together with only $a_1(t'), a_2(t'), \dots, a_n(t')$ to derive $b(t')$, denoted as $\hat{b}(t')$. $\hat{b}(t')$ is called “the
120 reconstructed $b(t')$ ” since only $a_1(t'), a_2(t'), \dots, a_n(t')$ and the learnt coupling relation \hat{F} have been
121 taken into account.
122 (iii) The first objective of this study is to answer whether the coupling relation
123 $b(t) = F[a_1(t), a_2(t), \dots, a_n(t)]$ can be learnt by machine learning, i.e., whether the learnt coupling
124 relation \hat{F} can well approximate the real coupling relation F . Since we do not intend to reach an
125 explicit formula of the learnt coupling relation \hat{F} , we will answer this question indirectly by
126 comparing the reconstructed series $\hat{b}(t')$ with the original series $b(t')$. If $\hat{b}(t') \approx b(t')$, then it can be
127 regarded as $\hat{F} \approx F$, and the machine learning framework can indeed learn the intrinsic coupling
128 relation among $a_1(t), a_2(t), \dots, a_n(t)$ and $b(t)$.



129 (iv) If the machine learning framework can learn the intrinsic coupling relation between
 130 $a_1(t), a_2(t), \dots, a_n(t)$ and $b(t)$, the learnt coupling relation \hat{F} can be applied to reconstruct output
 131 $b(t')$ even if only $a_1(t'), a_2(t'), \dots, a_n(t')$ are available.



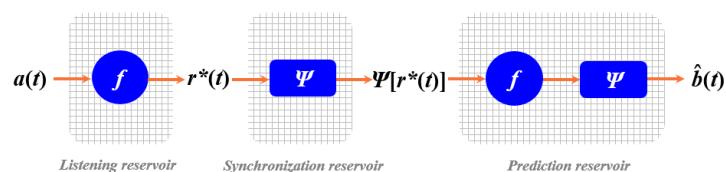
132
 133 **Figure 1** Diagram illustration for reconstructing time series by machine learning. (1) The available part of the
 134 dataset $\{a_1(t), \dots, a_n(t), b(t)\}$ is used to train the neural network ($a_1(t), \dots, a_n(t)$ and $b(t)$ are the time series of the
 135 variables a_1, \dots, a_n, b). So that the inherent coupling relation F among these variables can be learnt by the neural
 136 network, and the learnt coupling relation is noted as \hat{F} . (2) $b(t)$ is unknown, but the dataset $\{a_1(t'), a_2(t'), \dots,$
 137 $a_n(t')\}$ is available which is input into the trained neural network, and the unknown series $b(t')$ can be
 138 reconstructed, denoted as $\hat{b}(t')$. (3) If $\hat{b}(t') \approx b(t')$, then $\hat{F} \approx F$ can be derived, and it indicates that the
 139 machine learning framework have learnt the intrinsic coupling relation.

140 2.2 Machine learning frameworks

141 Inspired by several recent studies (As discussed in the introduction), we focus on employing the
 142 commonly-used machine learning frameworks for time series to accomplish our study: reservoir



143 computer (RC), back propagation (BP), and long short-term memory (LSTM) neural networks. The
 144 RC (Du et al., 2017; Lu et al., 2017; Pathak et al., 2018) has three layers: listening reservoir,
 145 synchronization reservoir and prediction reservoir (see Fig. 2).



146
 147 **Figure 2** Schematic of the RC neural network: the three parts of the RC neural network are the listening reservoir,
 148 the synchronization reservoir and the prediction reservoir. A time series $a(t)$ is input into the RC neural network.
 149 After the training processes of the three parts, the time series of b variable can be reconstructed by machine learning,
 150 denoted as $\hat{b}(t)$.

151 **The first layer is the listening reservoir.** A time series $a(t)$ (here for brevity, only one
 152 dimensional series is considered) is first input into the listening reservoir. The function “ f ” governing
 153 the listening reservoir training process is defined as follows:

$$154 \quad f : r^*(t) = \tanh[M \cdot r(t) + W_{in} \cdot a(t) + E]. \quad (1)$$

155 Here $r(t)$ represents the initial state of the listening reservoir, and $r^*(t)$ is the updated state of $r(t)$.
 156 The listening reservoir has N neurons, and these neurons are sparsely connected in the Erdős-Rényi
 157 matrix graph configuration (Lu et al., 2017). The connectivity of these neurons is represented by the
 158 adjacency matrix M of size $N \times N$, whose values are drawn from a uniform random distribution on the
 159 interval $[-1, 1]$. The typical reservoir size used in this study is $N = 1000$. The other two components
 160 of the listening reservoir are an input-to-reservoir weight matrix W_{in} , and a unit matrix E . The
 161 elements of W_{in} are randomly chosen from a uniform distribution in $[-1, 1]$. In the training process,
 162 and E is crucial for modulating the bias.



163 **The second layer is the synchronization reservoir.** After the training process of listening
164 reservoir, the updated state $r^*(t)$ is transported to a synchronization function “ Ψ ”. The function is
165 as follows:

$$166 \quad \Psi: \Psi[r^*(t)] = W_{out} r^*(t). \quad (2)$$

167 When using the trained RC neural network make prediction or reconstruction, the role of Eq. (2) will
168 be making $\Psi[r^*(t)]$ close to the target series $b(t)$. This is a procedure to make $\Psi[r^*(t)]$ close to the
169 target variable $b(t)$. The reservoir-to-output matrix W_{out} denotes a linear fitting process. Here it
170 should be noted that when using training data to train the RC neural network, W_{out} needs to be trained
171 by the input data through Eq. (3) so that the elements of W_{out} can be determined.

$$172 \quad W_{out} = \arg \min_{W_{out}} \|W_{out} r^*(t) - b(t)\| + \alpha \|W_{out}\| \quad (3)$$

173 Here the symbol $\|\cdot\|$ is the L_2 -norm of a vector (L_2 denotes the least square method) and α is the L_2
174 regularization (ridge regression) coefficient.

175 **The third layer is the prediction reservoir.** As Fig.2 shows, the prediction reservoir consists
176 of the function “ f ” and “ Ψ ”. Actually this is a feedback loop to the former two layers. In this reservoir
177 layer, the input data is $\Psi[r^*(t)]$ rather than $a(t)$. Finally, the output of this layer is the reconstruction
178 for the target series $b(t)$, which is denoted as $\hat{b}(t)$.

179 Both BP and LSTM have widely used by many fields (Reichstein et al., 2019), and their
180 algorithms can be open access in <https://ww2.mathworks.cn/help/deeplearning>. The LSTM requires
181 to update its network every step, for both training and testing (Zaytar and Amrani, 2016; Kratzert et
182 al., 2019). Regarding this point, we stop the LSTM updates after training has been accomplished, and
183 this case is denoted as LSTM*.

184 **2.3 Evaluation of reconstruction quality**



185 To evaluate the quality of reconstruction by machine learning, the root mean squared error
186 (RMSE) of residual series (Hyndman and Koehler, 2006) is adopted, which represents the difference
187 between the original series $b(t')$ and the reconstructed series $\hat{b}(t')$. In order to fairly compare the
188 errors of reconstructing different processes (Pennekamp et al., 2018), we normalize the RMSE. The
189 RMSE and normalized RMSE (nRMSE) are defined as

$$190 \quad RMSE = \sqrt{\frac{1}{k} \sum_t [b(t') - \hat{b}(t')]^2}, \quad (4)$$

$$191 \quad nRMSE = \frac{RMSE}{\max[b(t')] - \min[b(t')]} \quad (5)$$

192 In the results section, it can be observed that the good reconstruction quality corresponds to the
193 nRMSE value which is smaller than 0.1.

194 2.4 Data preparation

195 2.4.1 The time series from conceptual climate models

196 A **linearly coupled model**: The autoregressive fractionally integrated moving average
197 (ARFIMA) model (Granger and Joyeux, 1980) maps a Gaussian white noise $\varepsilon(t)$ into a correlated
198 sequence $x(t)$ (Eq. (6)), which could simulate the linear dynamics of oceanic-atmospheric coupled
199 system (Hasselmann, 1976; Franzke, 2012; Massah and Kantz, 2016; Cox et al., 2018).

$$200 \quad \varepsilon(t) \xrightarrow{ARFIMA(p,d,q)} x(t) \quad (6)$$

201 The parameters p , d and q denote the order p of the autoregressive process (AR (p) process), the Hurst
202 exponent $H=d+0.5$ of long-term memory process, and the order q of the moving average process (MA
203 (q) process). The coefficients in this model are set as follows: (1) $p=3$ (we set the AR(3) coefficients
204 as 0.6, 0.2 and 0.1 respectively); (2) $d=0.2$ (the Hurst exponent is 0.7); (3) $q=3$ (we set the MA(3)



205 coefficients as 0.3, 0.2 and 0.1 respectively). Finally $x(t)$ is a complex time series which is composited
 206 from all of the above three processes.

207 **A nonlinearly coupled model:** The Lorenz 63 (in the following referred as to Lorenz63) chaotic
 208 system (Lorenz, 1963) depicts the nonlinear coupling relation in a low-dimensional chaotic system.

209 The system reads

$$\begin{aligned}
 \frac{dx}{dt} &= -\sigma(x - y) \\
 \frac{dy}{dt} &= \mu x - xz - y \\
 \frac{dz}{dt} &= xy - Bz
 \end{aligned}
 \tag{7}$$

211 When the parameters are fixed at $(\sigma, \mu, B) = (10, 28, 8/3)$, the state in the system is chaotic. We
 212 employed the Runge-Kutta integral of the fourth order to acquire the series output from Lorenz63.
 213 The time steps were 0.01.

214 **A high-dimensional model:** The two-layer Lorenz 96 (in the following referred as to Lorenz96)
 215 model (Lorenz, 1996) is a high-dimensional chaotic system, which is generally employed to mimic
 216 mid-latitude atmospheric dynamics (Chorin and Lu, 2015; Hu and Franzke, 2017; Vissio and Lucarini,
 217 2018; Chen and Kalnay, 2019; Watson, 2019). It reads

$$\begin{aligned}
 \frac{dX_k}{dt} &= X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F - \frac{h_1}{J} \sum_{j=1}^J Y_{j,k} \\
 \frac{dY_{k,j}}{dt} &= \frac{1}{\theta} [Y_{k,j+1}(Y_{k,j-1} - Y_{k,j+2}) - Y_{k,j} + h_2 X_k]
 \end{aligned}
 \tag{8}$$

219 In the first layer of the Lorenz 96 there are 18 variables marked as X_k (k is any integer of interval [1,
 220 18]), and each X_k is coupled with $Y_{k,j}$ ($Y_{k,j}$ is from the second layer). The parameters are set as follows:

221 $J = 20$, $h_1 = 1$, $h_2 = 1$, and $F = 10$. The scale parameter θ controls the scale separation of the two layers.

222 When $\theta > 1$, processes in the second layer will be slower than processes in the first layer because



223 the increment of $Y_{k,j}$ is decreased by the term of θ . The time scale of $Y_{k,j}$ can be also close to that of
224 X_k by modulating the value of θ ; especially, the coupling strength will be amplified when θ is much
225 smaller than 1. The Runge-Kutta integral of the fourth order and periodic boundary condition are
226 adopted (that is: $X_0 = X_K$ and $X_{K+1} = X_1$; $Y_{k,0} = Y_{k-1,J}$ and $Y_{k,J+1} = Y_{k+1,1}$), and the integral time unit was
227 taken as 0.05.

228 **2.4.2 Real-world climatic series**

229 TSAT, NHSAT and the Nino3.4 index are chosen to represent real-world climatic time series.
230 The original data is from National Centers for Environmental Prediction (NCEP,
231 <https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html>) and KNMI Climate Explorer
232 (<http://climexp.knmi.nl/selectdailyindex.cgi?id=someone@somewhere>). The series of TSAT and
233 NHSAT are from the regional average of gridded daily data in NCEP Reanalysis 2. The selected
234 spatial range is 20°N – 20°S for the tropics and 20°N – 90°N for the Northern Hemisphere. The selected
235 temporal range is from 1981/09/01 to 2018/12/31.

236 **Training and testing the datasets:** Before analysis, all the time series are standardized to take
237 zero mean and unit standard deviation. We divide the total series into two parts: 60% of the time series
238 training the neural network and 40% being the testing series. Specific data lengths of the training
239 series and testing series will be also listed in the results section.

240 **3 Results**

241 **3.1 Coupling relation learning**

242 **3.1.1 Linear coupling relation and machine learning**



243 First of all, we consider the simplest case: the linear coupling relation between two variables.

244 **Quantifying the linear coupling strength:** To quantify the linear correlation between two series

245 $a(t)$ and $b(t)$, the Pearson correlation coefficient is adopted and is defined as

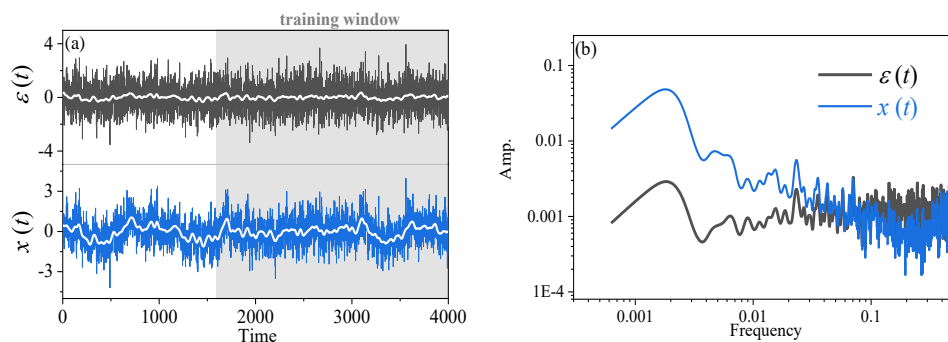
$$246 \quad corr. = \frac{mean[(a - \bar{a}) \cdot (b - \bar{b})]}{std(a) \cdot std(b)}. \quad (9)$$

247 In Eq. (9), the symbols “*mean*” and “*std*” denote the average and standard deviation for series $a(t)$ and
248 $b(t)$, respectively.

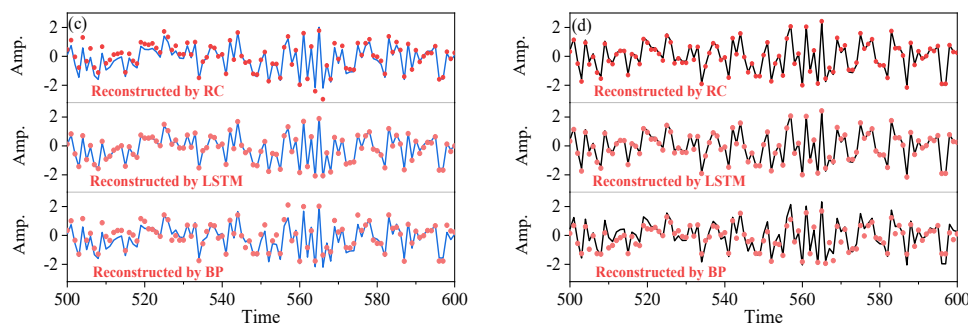
249 Here, the ARFIMA (3, 0.2, 3) model is taken as an example for “learning” strong linear coupling
250 between two variables. Obviously, there are different temporal structures in $x(t)$ and $\varepsilon(t)$, especially
251 for their large-scale trends (Fig. 3a) and power spectra (Fig. 3b). The marked difference between $x(t)$
252 and $\varepsilon(t)$ is in their low-frequency variations, and there are more low-frequency and larger-scale
253 structures in $x(t)$ than in $\varepsilon(t)$. We employ machine learning (RC, LSTM, LSTM*, and BP) to learn the
254 dynamics of this model (Eq. (6)) by the procedure shown in Fig. 1. The training parts of $\varepsilon(t)$ are
255 selected from the gray shadow in Fig. 3a. RC, LSTM, LSTM*, and BP are used to learn the coupling
256 relation between $x(t)$ and $\varepsilon(t)$. Then, the learnt coupling together with $\varepsilon(t')$ is used to reconstruct $x(t')$.
257 The reconstruction results and the performance of different machine learning frameworks are listed in
258 Table 1. It shows that there is a strong linear correlation (0.88) between $x(t')$ and $\varepsilon(t')$. This
259 reconstruction result indicates that strong linear coupling can be well captured by all kinds of machine
260 learning frameworks since the values of nRMSE are low for all of the machine learning frameworks.



261



262



263 **Figure 3** (a) The $x(t)$ time series (blue) and the $\varepsilon(t)$ time series (black) of the ARFIMA(3,0.2,3) model. White lines
264 depict the large-scale trends of these time series acquired by 50-step smoothing average. (b) Comparison of the
265 power spectrum of $x(t)$ (blue) with the power spectrum of $\varepsilon(t)$ (black). (c) Comparison of the reconstructed time
266 series of $x(t)$ by RC, LSTM and BP respectively (red dots), and the original $x(t)$ time series are presented by the blue
267 lines. (d) Comparison of the reconstructed time series of $\varepsilon(t)$ by RC, LSTM and BP respectively (red dots), and the
268 original $\varepsilon(t)$ time series are presented by the black lines. Only partial segments of the reconstructed series are shown.

269 Detailed comparisons between the original and reconstructed series are shown in Fig. 3c. When
270 inputting $\varepsilon(t)$, the learnt linear coupling can be applied to accurately reconstruct the original $x(t)$ by
271 the RC or LSTM. When $x(t)$ is reconstructed from $\varepsilon(t)$ by LSTM, the minimum value of nRMSE
272 (0.01) is reached; all of the reconstructed data are nearly overlapped with the original ones and cannot
273 be visually differentiated (see Fig. 3c). When reconstructing $x(t)$ from $\varepsilon(t)$ by the RC, the



274 reconstruction quality is also well. The best performance of LSTM among the given machine learning
 275 frameworks benefits from its network updating all the time. When updating is stopped, then the
 276 reconstruction error of LSTM* is no longer better than that of the RC, and is just a little bit better than
 277 that of BP (see Table 1). Therefore, if the training of machine learning frameworks is only carried out
 278 on the training series and there is no further updating, the RC is the best machine learning framework.

279 **Table 1** Details of reconstructing ARFIMA (3, 0.2, 3)

Input (a)	Output (b)	<i>corr.</i>	Data length (training/testing)	Neural network	RMSE	nRMSE
$\varepsilon(t')$	$x(t')$	0.88	2400/1600	RC	0.31	0.04
				LSTM	0.07	0.01
				LSTM*	0.46	0.06
				BP	0.52	0.07
$x(t')$	$\varepsilon(t')$	0.88	2400/1600	RC	0.09	0.01
				LSTM	0.08	0.01
				LSTM*	0.45	0.06
				BP	0.50	0.07

280 3.1.2 Nonlinear coupling relation and machine learning

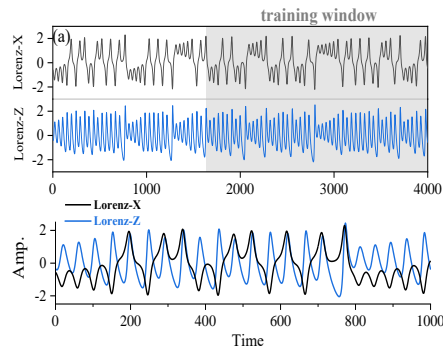
281 From the above results, it is known that a strong linear correlation is of great importance for
 282 series reconstruction and coupling learning. For a nonlinear coupled system, it has been reported that
 283 the coupling strength between two variables cannot be quantified by linear correlation (Brown, 1994;
 284 Sugihara et al., 2012). Hence, the convergent cross mapping (CCM, Sugihara et al., 2012) method is
 285 adopted.

286 **Quantification of the nonlinear coupling strength:** CCM index can quantify the coupling or
 287 causality between two variables (Sugihara et al., 2012; Tsonis et al., 2018; Zhang et al., 2019). When
 288 two series are output from the same dynamical system or coupled to each other, there will be shared
 289 information in their dynamical phase spaces (Takens, 1981). The stronger the coupling or forcing, the

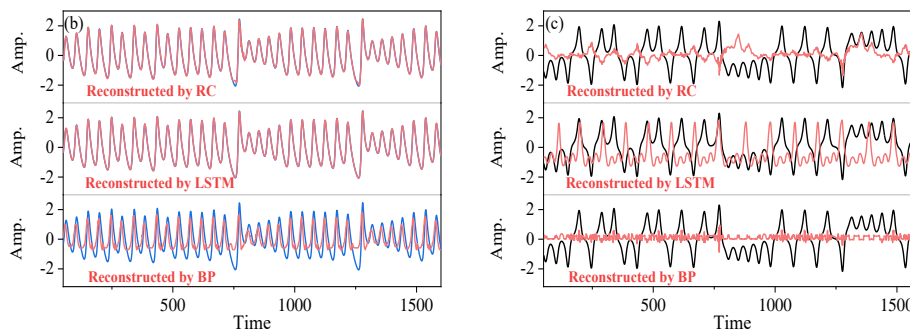


290 more information they share with each other. The shared information can be quantified by CCM index
291 ρ . For two time series $a(t)$ and $b(t)$, there are two CCM indices (Tsonis et al., 2018): $\rho_{a \rightarrow b}$ and $\rho_{b \rightarrow a}$:
292 (i) If variable b does influence variable a , the information of $b(t)$ will be stored in $a(t)$, and thus we
293 will acquire a high value of cross-mapping skill $\rho_{a \rightarrow b}$; (ii) if variable a does influence variable b , the
294 information of $a(t)$ will be stored in $b(t)$, and thus we will acquire a high value of cross-mapping skill
295 $\rho_{b \rightarrow a}$. The implementation of that $a(t)$ cross maps $b(t)$ is based on a nonlinear operator known as the
296 empirical dynamics model (Sugihara, 1994), and detailed algorithm for CCM can be found in a
297 previously reported study (Tsonis et al., 2018).

298 When the linear correlation between variables is very weak, could machine learning frameworks
299 still be applied to learn the underlying coupling dynamics? To address this question, we turn to the
300 Lorenz 63 system (Lorenz, 1963). There is a very weak linear correlation between variables X and Z
301 (with a Pearson correlation coefficient of 0.002) in the Lorenz63 model (see Table 2), and such a weak
302 linear correlation is induced by the unstable local correlation between variables X and Z (see Fig. 4a):
303 For example, X and Z are negatively correlated in the interval of 0–200, but positively correlated in
304 200–400. This alternation of negative and positive correlation appears over the whole processes of X
305 and Z , which leads to an overall weak linear correlation. In this case, it cannot obtain a feasible linear
306 regression model between X and Z to reconstruct one from the other, since there is no such good linear
307 dependency as found in the ARFIMA (p, d, q) model (see Figs. 5a and 5b). However, there is indeed
308 a well-defined coupling relation between X and Z in the Lorenz63 system (with a high CCM index
309 $\rho_{X \rightarrow Z} = 0.91$), so that it can be possible to reconstruct Z from X by some tools.

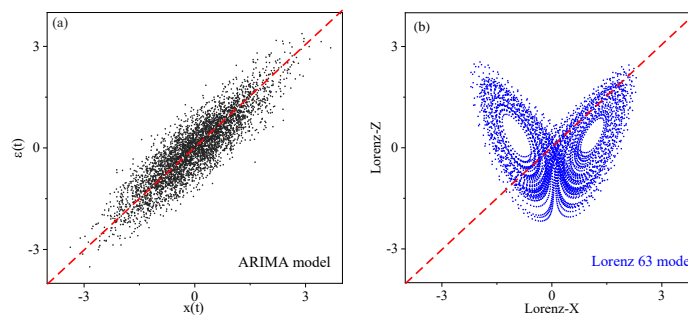


310



311

312 **Figure 4** (a) The X time series (black) and the Z time series (blue) of the Lorenz 63 model. (b) Comparison of the
 313 reconstructed time series of Z by RC, LSTM and BP respectively (red lines), and the original Z time series are
 314 presented by the blue lines. (c) Comparison of the reconstructed time series of X by RC, LSTM and BP respectively
 315 (red lines), and the original X time series are presented by the black lines.



316

317 **Figure 5** (a) Scatter plot of $x(t)$ versus $\varepsilon(t)$ from ARFIMA(3,0.2,3) model (black dots). (b) Scatter plot of X time



318 series and Z time series of the Lorenz 63 model (blue dots).

319 **Table 2** Details of Lorenz63 system reconstruction

Input (a)	Output (b)	<i>corr.</i>	$\rho_{a \rightarrow b}$	Data length (training/testing)	Neural network	RMSE	nRMSE
Lorenz -X	Lorenz-Z	0.002	0.91	2400/1600	RC	0.04	0.008
					LSTM	0.02	0.004
					LSTM*	1.02	0.24
					BP	0.77	0.17
Lorenz -Z	Lorenz-X	0.002	0.03	2400/1600	RC	1.13	0.34
					LSTM	1.03	0.31
					LSTM*	1.08	0.33
					BP	1.01	0.31

320 Different from the case of linear system with strong linear coupling, the reconstruction for the
 321 time series of the Lorenz63 system depends on the used machine learning frameworks (Fig. 4b). The
 322 series reconstructed by LSTM nearly overlaps with the original series, and has the minimum nRMSE
 323 (0.004); moreover, the RC performs quite well, with only a little difference found at some peaks and
 324 dips. These reconstruction results indicate that, even though the linear correlation is very weak, the
 325 strong nonlinear correlation will allow RC and LSTM to fully capture the underlying coupling
 326 dynamics. However, BP and LSTM* perform poorly, and their reconstruction results have large errors
 327 (nRMSE = 0.17 for BP, and nRMSE = 0.24 for LSTM*). The reconstructed series heavily departs
 328 from the original series, especially for all peaks and dips, and the reconstructed values for each
 329 extreme point are underestimated (Fig. 4b). This means that both of BP and LSTM* cannot learn the
 330 nonlinear coupling.

331 **3.2 Reconstruction quality and impact factors**

332 From the above results, it is found that machine learning framework is able to learn different



333 (linear or nonlinear) coupling relations well. Just as mentioned in the introduction section, under the
334 condition that machine learning framework can learn the coupling relation between or among
335 variables within the same system, the learnt coupling relation can be applied to reconstruct one
336 variable's time series from others (see Fig.1). We further investigate the ability of such time series
337 reconstruction, including what factors might influence the reconstruction quality.

338 **3.2.1 Direction dependence and variable dependence**

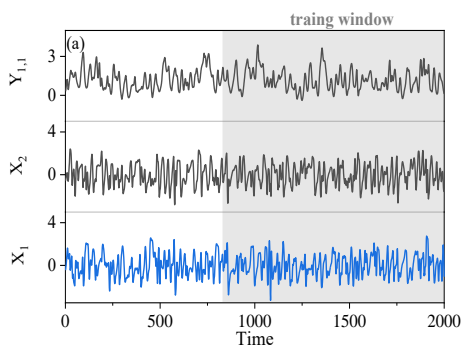
339 When reconstructing the time series of linear model of Eq. (6), it can be found that the
340 reconstruction is invertible (see Fig. 3d and Table 1): one variable can be taken as explanatory variable
341 to reconstruct another variable's time series well; oppositely, it can be also well reconstructed by
342 another variable. In fact, when there is a strong linear correlation between variables, the invertible (or
343 bi-directional) construction can also be accomplished by using a traditional regression approach
344 (Brown, 1994). Hence, when the linear coupling is weak but the nonlinear coupling is strong, will the
345 bi-directional reconstruction still be allowed? The answer is usually no. For example, when comparing
346 the reconstruction quality of reconstructing variable Z from variable X (Fig. 4b) with that of
347 reconstructing variable X from variable Z (Fig. 4c), it is obvious that all of the machine learning
348 frameworks fail (large values of nRMSE are all close to 0.3) for reconstructing X from Z . The
349 reconstruction direction is no longer invertible in this nonlinear system, where the reconstruction
350 quality is direction-dependent and variable-dependent.

351 Hence, we further discuss how to select the suitable explanatory variable or reconstruction
352 direction. Tables 1 and 2 show that the reconstruction quality in a linear coupled system highly
353 depends on the Pearson correlation, however it might be different for a nonlinear system. For the

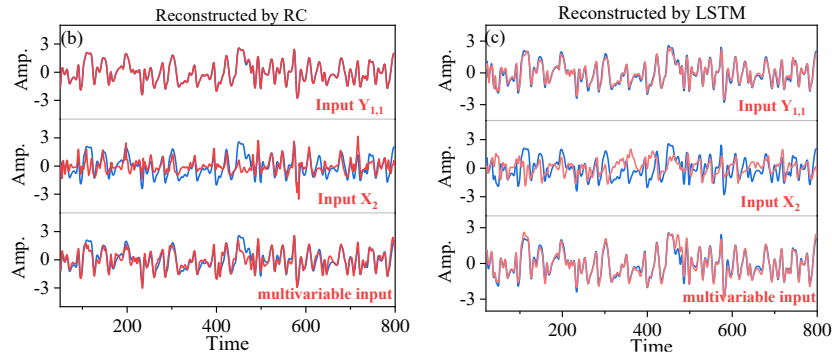


354 Lorenz 63 system, the two-direction CCM coefficients between the variable X and variable Z is
355 asymmetric (with a stronger $\rho_{X \rightarrow Z} = 0.91$ and weaker $\rho_{Z \rightarrow X} = 0.03$), and then variable Z can be well
356 reconstructed from variable X by machine learning but variable X cannot be reconstructed from
357 variable Z (Fig. 4b and 4c). CCM index can be taken as a potential precursor to determine the
358 explanatory variable and reconstructed variable for this nonlinear system. Actually, for most of
359 nonlinearly coupled variables (Here called a and b), the two-direction causal forces might be
360 asymmetric, and the values of $\rho_{a \rightarrow b}$ and $\rho_{b \rightarrow a}$ are often unequal to each other (Tosnis et al., 2019).
361 Here the asymmetric reconstruction quality is just caused by the asymmetric causality between
362 variables.

363 **The cases of high-dimensional system:** It is noted that the above conclusions and results are
364 derived from the low-dimensional chaotic system (Eq. (7)). It is also interesting to know whether they
365 are also applicable to a high-dimensional chaotic system of Lorenz 96 model. Since BP and LSTM*
366 fully fail in reconstructing for Lorenz63 system (Not shown here), which suggests that they are not
367 suitable for handling nonlinear systems, so the results from LSTM and RC will be shown next.



368



369

370 **Figure 6** (a) The $Y_{l,l}$ time series (black), X_2 time series (black) and X_l time series (blue) of the Lorenz 96 model. (b)

371 By means of the RC machine learning, when using $Y_{l,l}$, X_2 and multivariate to be the explanatory variable

372 respectively, the corresponding reconstructed X_l time series are showed respectively from the top panel to the bottom

373 panel (red lines), and the original X time series are presented by the blue lines. (d) By means of the LSTM machine

374 learning, when using $Y_{l,l}$, X_2 and multivariate to be the explanatory variable respectively, the corresponding

375 reconstructed X_l time series are showed respectively from the top panel to the bottom panel (red lines), and the

376 original X time series are presented by the blue lines.

377 Firstly, we use variables X_1 and $Y_{l,l}$ in Eq.(8) to illustrate the direction dependence in the high-

378 dimensional system. Details of X_1 and $Y_{l,l}$ are shown in Fig. 6a, and the Pearson correlation between

379 X_1 and $Y_{l,l}$ is weak (only -0.11, see Table 3). For the given value of each parameter in Eq. (8), the

380 forcing from X_1 to $Y_{l,l}$, is much stronger than the forcing from $Y_{l,l}$ to X_1 . Hence, CCM index shows

381 the asymmetric causality: $\rho_{Y_{l,l} \rightarrow X_1} = 0.98$ and $\rho_{X_1 \rightarrow Y_{l,l}} = 0.61$. It means that reconstructing X_1 from $Y_{l,l}$

382 might obtain a better quality than the opposite direction. Hence, by means of RC, the error of

383 reconstructing X_l from $Y_{l,l}$ is nRMSE=0.01, and in the opposite direction it is nRMSE=0.06 (Table

384 3), which is consistent with the indication of CCM index (Similar results can be found in Figs. 6b and

385 6c).



386

Table 3 Details of reconstructing the Lorenz 96 model

Input (a)	Target (b)	<i>corr.</i>	$\rho_{a \rightarrow b}$	Data length (training/testing)	Neural network	RMSE	nRMSE
$Y_{1,1}$	X_1	-0.11	0.98	1200/800	RC	0.03	0.01
					LSTM	0.34	0.05
X_1	$Y_{1,1}$	-0.11	0.61	1200/800	RC	0.35	0.06
					LSTM	0.42	0.08
X_2	X_1	-0.06	0.37	1200/800	RC	0.69	0.13
					LSTM	1.09	0.20
X_1	X_2	-0.06	0.25	1200/800	RC	0.95	0.17
					LSTM	0.84	0.16
X_2, X_{17}, X_{18}	X_1	-0.06, -0.24,	0.37, 0.29,	1200/800	RC	0.41	0.08
		0.06	0.41		LSTM	0.32	0.06

387 The reconstruction between X_1 and X_2 in the same layer of Lorenz96 is also shown. There is an
 388 asymmetric causal relation ($\rho_{X_2 \rightarrow X_1} = 0.37$ and $\rho_{X_1 \rightarrow X_2} = 0.25$) between X_1 and X_2 , and their linear
 389 correlation is very weak (see Table 3). The RC gives better result of reconstructing X_1 from X_2
 390 (nRMSE=0.13) than reconstructing X_2 from X_1 (nRMSE=0.17). LSTM also has different results for
 391 X_1 and X_2 (Table 3), where the quality of reconstructing from X_1 to X_2 (nRMSE=0.16) is better than
 392 reconstructing from X_2 to X_1 (nRMSE=0.20). The reconstruction quality of LSTM is worse than the
 393 RC, and the reconstruction results by LSTM are not consistent with the coupling strengths. This might
 394 indicate that LSTM will perform worse in some cases than RC, and the best choice of machine learning
 395 framework in data reconstruction is RC.

396 Secondly, it is also found that the reconstruction quality is influenced by the chosen explanatory
 397 variables: The quality of reconstructing X_1 from $Y_{1,1}$ is better than the quality of reconstructing X_1 from
 398 X_2 by (see Fig. 6b and 6c). Then, if more than one explanatory variable in the same layer are considered,
 399 the reconstruction of X_1 from X_2 can be greatly improved (see Figs. 6b and 6c). For example, when all
 400 of X_2, X_{17} and X_{18} are regarded as the explanatory variables, the nRMSE of reconstructed X_1 is reduced
 401 from 0.13 to 0.08 (Table 3). For both of RC and LSTM machine learning, the multivariable

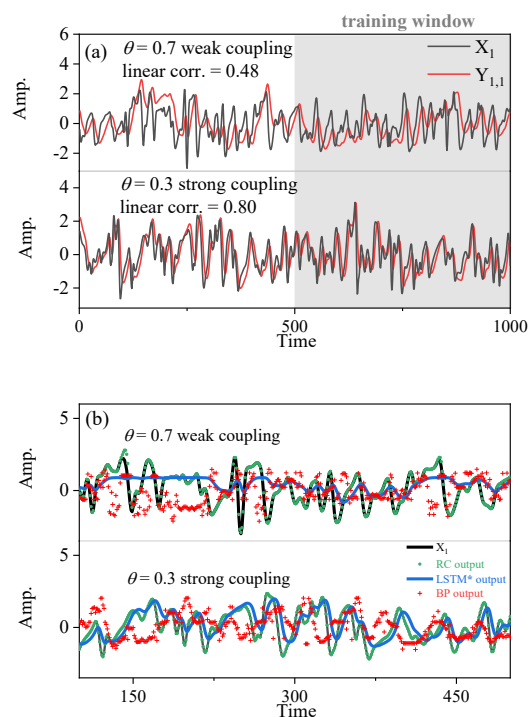


402 reconstruction reaches lower error than those from unit-variable reconstruction.

403 **3.2.2 Explanation of the reconstruction quality dependency: Role of coupling** 404 **strength**

405 From Table 1 to Table 3, it is obvious that the value of nRMSE varies with the Pearson correlation
406 or CCM index. Furthermore, the lower nRMSE often corresponds to the higher correlation or CCM
407 index. Compared to RC and LSTM, it is revealed that BP and LSTM* are much more sensitive to the
408 Pearson correlation; meanwhile, the RC is sensitive to CCM index. In the following, we demonstrate
409 that the above reconstruction quality dependencies are determined by the coupling strength of
410 variables.

411 The setting of Eq. (8) is as follows: the value of h_l is set as 0, and the value of θ is decreased
412 from 0.7 to 0.3. When θ is equal to 0.7, the forcing from X_l to $Y_{l,l}$ is weak. At that time, the Pearson
413 correlation between X_l and $Y_{l,l}$ is only 0.48, and the performances of BP and LSTM* are not good.
414 When θ is equal to 0.3, the forcing is dramatically magnified. As the second panel of Fig. 7a shows,
415 this strong forcing makes $Y_{j,i}$ synchronized to X_i , and the linear correlation between X_l and $Y_{l,l}$ is
416 greatly increased to 0.8. When the forcing strength is magnified, the performance of machine learning
417 is also enhanced (Fig. 7b): red dots (BP output) and the blue line (LSTM* output) are much closer to
418 the black line (original target series). This means, that the reconstruction quality of BP and LSTM* is
419 sensitive to this linear coupling variation, and it is greatly improved when the linear correlation is
420 increased.



421

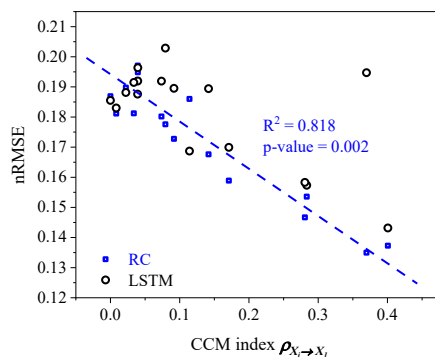
422

423 **Figure 7** Influence of strong coupling on series' linear correlation and machine learning performances. (a)
424 Comparison of the linear correlation when the coupling strength is different. The top panel corresponds to the weak
425 coupling strength, and the bottom panel corresponds to the strong coupling. The red lines present the input
426 explanatory variable and the black lines present the target series of machine learning. (b) Comparison of the machine
427 learning performances when the coupling strength is different. The top panel corresponds to the weak coupling
428 strength, and the bottom panel corresponds to the strong coupling. The black lines are the original series; the
429 reconstructed series by RC (green lines), LSTM*(blue lines) and BP (red dots) are shown respectively.

430 However, the RC is not so much restrained by the Pearson correlation. When θ is equal to 0.7
431 or 0.3, the values of CCM index are both higher than 0.9, that is to say, the nonlinear coupling strength
432 is not changed by θ . Then, it can be found that the quality of reconstructed X_1 by RC is always good.
433 As Fig. 7b shows, the green dots (RC output) in Fig. 7b always overlap with the black line (original



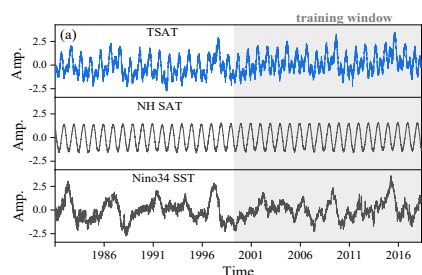
434 target series). Actually, the reconstruction quality of the RC is determined more by the nonlinear
435 coupling strength. The values of CCM index are calculated between X_1 and X_2, X_3, \dots, X_{18} ; meanwhile,
436 X_1 is reconstructed from X_2, X_3, \dots, X_{18} , respectively. Then, a significant correspondence exists
437 between the nRMSE and CCM index (Fig. 8), especially for the results of RC. This indicates that the
438 reconstruction quality is dependent on the coupling strength between the reconstructed variable and
439 different explanatory variables.



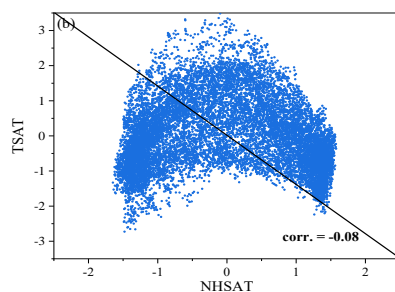
440
441 **Figure 8** Scatter plot of nRMSE values and CCM index values. The blue boxes are results of the RC machine
442 learning, and the black cycles are results of the LSTM machine learning. The blue dashed line is the fitted linear
443 trend of the blue cycles, and this dependency trend is significant because the p-value is much smaller than 0.05.

444 3.3 Application to real-world climate series: reconstructing SAT

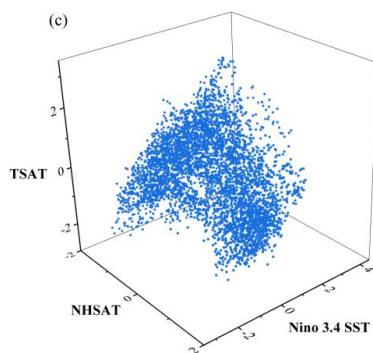
445 The natural climate series are usually nonstationary, and are encoded with the information of
446 many physical processes in the earth system. In the following, we illustrate the utility of the above
447 methods and conclusions by investigating a real-world example mentioned in the introduction, by
448 showing that: whether the NHSAT time series can be reconstructed from the TSAT time series, and
449 whether the TSAT time series can be also reconstructed from the NHSAT time series.



450



451



452

453 **Figure 9** (a) Daily time series of TSAT, NHSAT and Nino 3.4 index. (b) Scatter plot of normalized NHSAT and
454 normalized TSAT. (c) Three-dimensional scatter plot of normalized NHSAT, normalized TSAT and normalized Nino
455 3.4 SST.

456 The daily NHSAT and TSAT time series are shown in Fig. 9a. It is quite different for the
457 oscillation shapes of the NHSAT and TSAT series, and there is a weak linear correlation (0.08, see
458 Table 4) between them. In the scatter plot for the NHSAT and TSAT (Fig.9b), the marked nonlinear
459 structure is observed between NHSAT and TSAT. Such a weak linear correlation will make the linear

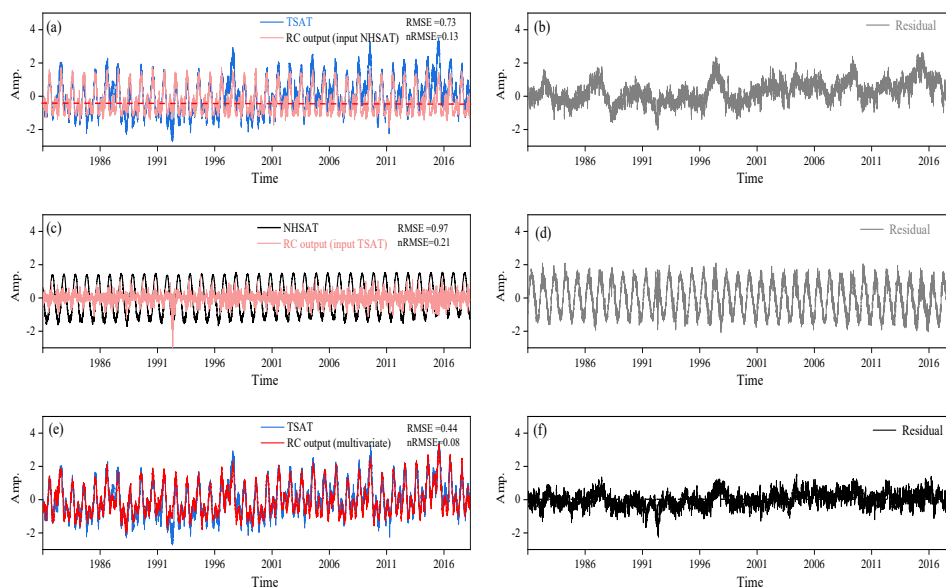


460 regression model fail to reconstruct one series from the other. Likewise, there is no explicit physical
 461 expression that can transform TSAT and NHSAT to each other. Now we try to using machine learning
 462 to reconstruct these climate series. CCM index of that NHSAT cross maps TSAT is $\rho_{N \rightarrow T} = 0.70$, and
 463 CCM index of that TSAT cross maps NHSAT is $\rho_{T \rightarrow N} = 0.24$ (Table 4). CCM index indicates that
 464 the reconstruction from NHSAT to TSAT will obtain a better quality than the opposite direction.

465 **Table 4** Details of temperature records' reconstruction

Input (a)	Output (b)	<i>corr.</i>	$\rho_{a \rightarrow b}$	Data length (training/testing)	Neural network	RMSE	nRMSE
NHSAT	TSAT	0.08	0.70	8182/5454	RC	0.73	0.13
					LSTM	1.14	0.20
					BP	1.45	0.26
					RC	0.97	0.21
TSAT	NHSAT	0.08	0.24	8182/5454	LSTM	1.04	0.23
					BP	1.23	0.37

466 By means of RC machine learning, TSAT can be described by the reconstructed time series
 467 (Fig.10a). But the corresponding nRMSE is equal to 0.13, this is because some extremes of the TSAT
 468 time series have not been described (Fig.10b). When using TSAT to reconstruct the time series of
 469 NHSAT, the reconstructed time series cannot describe the original time series of NHSAT (Fig.10c),
 470 and the corresponding nRMSE is equal to 0.21. Besides, we also use the LSTM and BP to reconstruct
 471 these natural climate series, the performances of these two neural networks are worse than RC (Table
 472 4). For BP, this might be due to its inability to deal with nonlinear coupling. As for LSTM, the unstable
 473 variance might influence its performance because it is heavily impacted by the unstable variance.



474

475

476

477

478

479

480

481

482

Figure 10 (a) Reconstructed TSAT time series (red) when NHSAT is the explanatory variable; (b) Residual series given by the original TSAT series and the reconstructed TSAT series of (a). (c) Reconstructed NHSAT time series (red) when TSAT is the explanatory variable. (d) Residual series given by the original NHAST series and the reconstructed NHSAT series of (c). (e) Reconstructed TSTA time series (red) when NHSAT and Nino3.4 index are the explanatory variables. (f) Residual series given by the original TSAT series and the reconstructed TSAT series of (e).

483

484

485

486

487

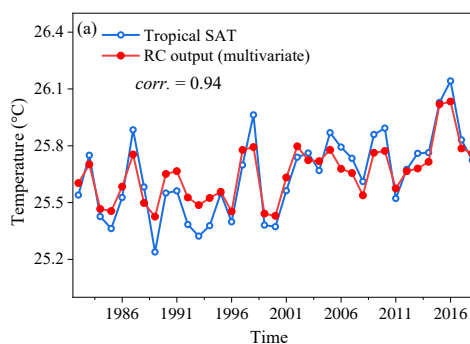
488

489

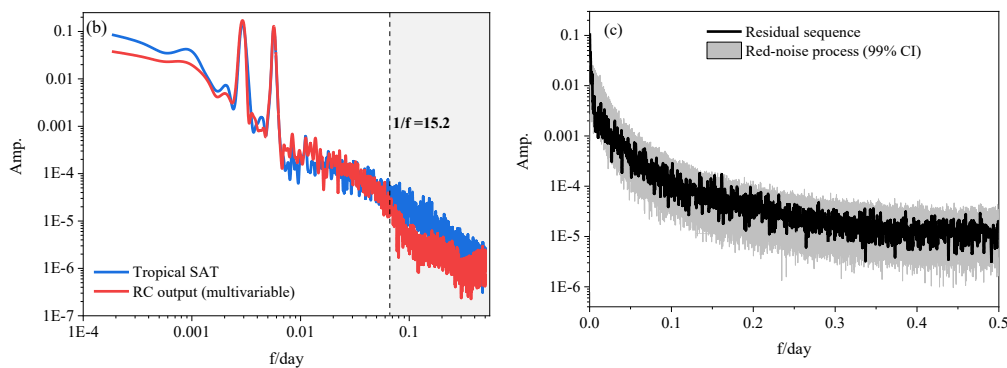
We can also improve the reconstruction quality of TSAT. Considering that the tropics climate system do not only interact with the Northern Hemisphere climate system, we can use the information of other subsystems to improve the reconstruction. Looking at the time series of Nino 3.4 index (Fig. 9), some of its extremes occur in the same time regions as the extremes of TSAT. Moreover, when Nino3.4 is included into the scatter plot (Fig. 9c), a nonlinear attractor structure is revealed. We combine NHSAT with Nino 3.4 index to reconstruct the time series of TSAT, and the reconstruction is by means of the RC machine learning. The reconstructed TSAT (Fig. 10e) is much closer to the



490 original TSAT series, and the corresponding nRMSE has been improved to 0.08.



491



492

493 **Figure 11** (a) Comparison between the annual mean values of reconstructed TSAT (blue) and the annual mean values
494 of original TSAT (red). (b) Comparison between the power spectrum of reconstructed TSAT (blue) and the power
495 spectrum of original TSAT (red). (c) Red-noise test for residual series, the gray shaded area is the 99% CI of red-
496 noise process.

497 Finally, we further compare the original TSAT and the reconstructed TSAT: (i) the annual
498 variations of TSAT and the reconstructed TSAT are close to each other (Fig. 11a). (ii) The power
499 spectrum of TSAT and the reconstructed TSAT are compared in Fig. 11b, and it can be seen that the
500 main deviation is in the frequency bands corresponding to around 0-15 days. We guess the reason



501 might be the local weather processes are not input into this RC reconstruction. This conjecture can be
502 further confirmed by red-noise test with response time 15 days for the residual series (here the red-
503 noise test is the same as the method used in Roe, 2009). The gray shadow in Fig. 11c gives the 99%
504 confidence interval of red-noise process, and the black line is the residual series. All data points of the
505 residual series lie within the confidence intervals, and this means, the residual is possibly induced by
506 local weather processes that is not input into the RC.

507 **4 Conclusions and discussions**

508 In this study, three kinds of machine learning frameworks is used to reconstruct the time series
509 of toy models and real-world climate systems. One series can be reconstructed from the other series
510 by machine learning when they are governed by the common coupling relation. For the linear system,
511 variables are coupled by the linear mechanism, and a strong Pearson correlation benefits to machine
512 learning with bi-directional reconstruction. For a nonlinear system, the time series often have a weak
513 Pearson correlation, but the machine learning can still well reconstruct the time series when two
514 variables share the common information through their interactions; moreover, the reconstruction
515 quality is direction-dependent and variable-dependent, which is determined by the coupling strength
516 and causality between the dynamical variables.

517 Considering the reconstruction quality dependency, selecting the suitable explanatory variables
518 is crucial for obtaining a good reconstruction quality. Hence, we propose using CCM index to select
519 explanatory variables. Especially for the time series of nonlinear systems, when the Pearson
520 correlation is weak, CCM index might be strong enough, and then the corresponding variable can be
521 selected as an explanatory variable. It is well known that atmospheric or oceanic motions are



522 nonlinearly coupled over most of scales, and therefore, in the natural climate series, there would be
523 similar nonlinear correlations to the Lorenz 63 and Lorenz 96 system (the linear correlation is weak
524 but CCM indices are of high values). However, if only Pearson correlation is used to select the
525 explanatory variable, some useful nonlinearly correlated variables might be left out.

526 Besides, it is very important to select the correct direction for reconstructing one variable from
527 other variable within a nonlinear system. For example, there are three Eigen directions in the Lorenz
528 63 system, and only one of them has a positive Lyapunov exponent (Kantz and Schreiber, 2005;
529 Strogatz, 2018). It is known that the direction of reconstruction must correspond to the one with
530 positive Lyapunov exponent, so that the reconstruction could succeed (Lu et al., 2017 and 2018;
531 Strogatz, 2018). Such phenomenon might also occur in some natural nonlinear systems. But the
532 limited data length and red noise are common issues in natural time series, which makes the correct
533 Lyapunov exponent estimating difficult (Kantz and Schreiber, 2005). In this study, it has been shown
534 that the correct direction for reconstruction can be selected by CCM index without estimating the
535 correct Lyapunov exponent.

536 In section 3.3, the TSAT series is reconstructed from the NHSAT series by means of RC, where
537 CCM index plays a key role in guiding the reconstruction variable selection. Moreover, the result also
538 suggests that: (i) the coupling mechanism between the tropics and Northern Hemisphere is nonlinear
539 since their reconstructions are direction-dependent. (ii) It is noted that CCM index that TSAT cross
540 maps NHSAT is $\rho_{N \rightarrow T} = 0.70$, and CCM index that TSAT cross maps NHSAT is $\rho_{T \rightarrow N} = 0.24$.
541 According to the causality theory (Sugihara et al., 2012; Tsonis, 2018; Zhang et al., 2019), the
542 asymmetric CCM index indicates that the main energy transport direction might be mainly from the
543 tropics to the Northern Hemisphere. (iii) If there are enough historical measured series of the Northern



544 hemisphere and ENSO, the unmeasured climate series (such as SAT) in the tropics might be well
545 reconstructed.

546 Finally, it is worth noting once more that there are still more potential applications for machine
547 learning in the climate studies. For instance, a series $b(t)$ is unmeasured during some periods for the
548 measuring instrument failure, but there are other kinds of variables without missing observations.
549 Moreover, CCM can be applied to select the suitable variables coupled with $b(t)$, and then the RC can
550 be employed to reconstruct the unmeasured part of $b(t)$ (following Fig. 1). This is very useful to some
551 important climate studies, such as paleoclimate reconstruction (Brown, 1994; Donner 2012; Emile-
552 Geay and Tingley, 2016), interpolation for the missing points in measurements (Hofstra et al., 2008),
553 and the parameterization schemes (Wilks, 2005; Vissio and Lucarini, 2018). Our study in this article
554 is only a beginning for reconstructing climate series by machine learning, and more detailed
555 investigations will be reported soon.

556



557 ***Code and data availability.*** The code for CCM causality coefficient and the relevant time series are
558 available upon request to the authors. The data will be made available on zenodo.org once the
559 manuscript is accepted for discussion.

560 ***Author contribution.*** Yu Huang and Zuntao Fu designed this study. All of the authors contributed to
561 the preparation and writing of the manuscript.

562 ***Competing interests.*** The authors declare no competing interest.

563 ***Acknowledgement.*** We acknowledge the supports from National Natural Science Foundation of China
564 through Grants (No. 41675049 and No. 41475048).

565



566 **References**

- 567 Badin, G., Domeisen, D. I.: A search for chaotic behavior in stratospheric variability: comparison between the
568 Northern and Southern Hemispheres. *J. Atm. Sci.*, 71(12), 4611-4620, 2014.
- 569 Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., ... Di Carlo, P.: Recursive
570 neural network model for analysis and forecast of PM10 and PM2.5. *Atmos. Pollut. Res.*, 8(4), 652-659, 2017.
- 571 Brown, P. J.: *Measurement, Regression, and Calibration*, vol. 12 of Oxford Statistical Science Series, Oxford
572 University Press, USA, 216 pp, 1994.
- 573 Carroll, T. L.: Using reservoir computers to distinguish chaotic series. *Phys. Rev. E* 98(5), 052209, 2018.
- 574 Chen, T. C., Kalnay, E.: Proactive quality control: observing system simulation experiments with the Lorenz'96
575 Model. *Mon. Wea. Rev.*, 147(1), 53-67, 2019.
- 576 Chorin, A. J., Lu, F.: Discrete approach to stochastic parameterization and dimension reduction in nonlinear
577 dynamics. *P. Natl. Acad. Sci.*, 112(32), 9804-9809, 2015.
- 578 Comeau, D., Zhao, Z., Giannakis, D., Majda, A. J.: Data-driven prediction strategies for low-frequency patterns of
579 North Pacific climate variability. *Clim. Dyn.*, 48(5-6), 1855-1872, 2017.
- 580 Conti, C., Navarra, A., Tribbia, J.: The ENSO Transition Probabilities. *J. Clim.*, 30 (13), 4951-4964, 2017.
- 581 Cox, P. M., Huntingford, C., Williamson, M. S.: Emergent constraint on equilibrium climate sensitivity from global
582 temperature variability. *Nature*, 553(7688), 319, 2018.
- 583 Donner, R. V.: Complexity concepts and non-integer dimensions in climate and paleoclimate research. *Fractal
584 Analysis and Chaos in Geosciences*, Nov 14:1, 2012.
- 585 Drótos, G., Bódai, T., Tél, T.: Probabilistic concepts in a changing climate: A snapshot attractor picture. *J. Clim.*,
586 28(8), 3275-3288, 2015.
- 587 Du, C., Cai, F., Zidan, M. A., Ma, W., Lee, S. H., Lu, W. D.: Reservoir computing using dynamic memristors for
588 temporal information processing. *Nat. Commun.*, 8(1), 2204, 2017.
- 589 Emile-Geay, J., Tingley, M.: Inferring climate variability from nonlinear proxies: application to paleo-ENSO studies.
590 *Clim. Past.*, 12(1), 31-50, 2016.
- 591 Farneti, R., Vallis, G. K.: Meridional energy transport in the coupled atmosphere–ocean system: Compensation and
592 partitioning. *J. Clim.*, 26(18), 7151-7166, 2013.
- 593 Feng, X., Fu, T. M., Cao, H., Tian, H., Fan, Q., Chen, X.: Neural network predictions of pollutant emissions from
594 open burning of crop residues: Application to air quality forecasts in southern China. *Atmos. Environ.*, 204, 22-
595 31, 2019.
- 596 Franzke, C.: Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *J. Clim.*,
597 25(12), 4172-4183, 2012.
- 598 Granger, C. W., Joyeux, R.: An introduction to long-memory time series models and fractional differencing. *J. Time.
599 Ser. Anal.*, 1(1), 15-29, 1980.



- 600 Hasselmann, K.: Stochastic climate models part I. Theory. *Tellus*, 28(6), 473-485, 1976.
- 601 Hofstra, N., Haylock, M., New, M., Jones, P., Frei, C.: Comparison of six methods for the interpolation of daily
602 European climate data. *J. Geophys. Res.*, 113(D21), 2008.
- 603 Hsieh, W. W., Wu, A., Shabbar, A.: Nonlinear atmospheric teleconnections. *Geophys. Res. Lett.*, 33(7): L07714,
604 2006.
- 605 Hu, G., Franzke, C. L.: Data assimilation in a multi-scale model. *Mathematics of Climate and Weather Forecasting*,
606 3(1), 118-139, 2017.
- 607 Hyndman, R. J., Koehler, A. B.: Another look at measures of forecast accuracy. *Int. J. Forecasting.*, 22(4), 679-688,
608 2006.
- 609 Kantz, H., Schreiber, T.: *Nonlinear time series analysis (Vol. 7)*. Cambridge university press, 2004.
- 610 Kratzert, F., Hernegger, M., Klotz, D., Hochreiter, S., Klambauer, G. *Neural Hydrology-Interpreting LSTMs in*
611 *Hydrology*. arXiv:1903.07903, 2019.
- 612 Lorenz, E. N.: Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20(2), 130-141, 1963.
- 613 Lorenz, E. N.: Predictability: a problem partly solved. *Proc. ECMWF Seminar on Predictability*, vol I, Reading,
614 United Kingdom, ECMWF, pp 40–58, 1996.
- 615 Lu, Z., Pathak, J., Hunt, B., Girvan, M., Brockett, R., Ott, E.: Reservoir observers: Model-free inference of
616 unmeasured variables in chaotic systems. *Chaos*, 27(4), 041102, 2017.
- 617 Lu, Z., Hunt, B. R., Ott, E.: Attractor reconstruction by machine learning. *Chaos*, 28(6): 061104, 2018.
- 618 Ludescher, J., Gozolchiani, A., Bogachev, M. I., Bunde, A., Havlin, S., Schellnhuber, H. J.: Very early warning of
619 next El Niño. *P. Natl. Acad. Sci.*, 111(6), 2064-2066, 2014.
- 620 Massah, M., Kantz, H.: Confidence intervals for time averages in the presence of long-range correlations, a case
621 study on Earth surface temperature anomalies. *Geophys. Res. Lett.*, 43(17), 9243-9249, 2016.
- 622 Mattingly, K. S., Ramseyer, C. A., Rosen, J. J., Mote, T. L., Muthyala, R.: Increasing water vapor transport to the
623 Greenland Ice Sheet revealed using self-organizing maps. *Geophys. Res. Lett.*, 43(17), 9250-9258, 2016.
- 624 Mukhin, D., Gavrilov, A., Loskutov, E., Feigin, A., Kurths, J.: Nonlinear reconstruction of global climate leading
625 modes on decadal scales. *Clim. Dyn.*, 51(5-6), 2301-2310, 2018.
- 626 Pathak, J., Lu, Z., Hunt, B. R., Girvan, M., Ott, E.: Using machine learning to replicate chaotic attractors and
627 calculate Lyapunov exponents from data. *Chaos*, 27(12), 121102, 2017.
- 628 Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A., Ott, E.: Local low dimensionality of atmospheric dynamics. *Phys*
629 *Rev Lett* 86(26): 5878, 2001.
- 630 Pennekamp, F., Iles, A. C., Garland, J., Brennan, G., Brose, U., Gaedke, U., Novak, M.: The intrinsic predictability
631 of ecological time series and its potential to guide forecasting. *Ecol. Monogr.*, e01359, 2019.
- 632 Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, M., Pal, C.: ExtremeWeather: A large-scale climate
633 dataset for semi-supervised detection, localization, and understanding of extreme weather events. In *Advances*
634 *in Neural Information Processing Systems* (pp. 3402-3413), 2017.



- 635 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N.: Deep learning and process
636 understanding for data-driven Earth system science. *Nature*, 566(7743), 195, 2019.
- 637 Roe, G.: Feedbacks, timescales and seeing red. *Ann. Rev. Earth. Plan. Sci.*, 37: 93-115, 2009.
- 638 Schurer, A. P., Hegerl, G. C., Mann, M. E., Tett, S. F., Phipps, S. J.: Separating forced from chaotic climate
639 variability over the past millennium. *J. Clim.*, 26(18), 6954-6973, 2013.
- 640 Strogatz, S. H. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*.
641 CRC Press, 2018.
- 642 Sugihara, G., May, R., Ye, H., Hsieh, C. H., Deyle, E., Fogarty, M., Munch, S.: Detecting causality in complex
643 ecosystems. *Science*, 338(6106), 496-500, 2012.
- 644 Takens, F.: Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence, Lecture Notes in*
645 *Mathematics*, 898, 366–381 (Springer Berlin Heidelberg), 1981.
- 646 Tsonis, A. A., Deyle, E. R., Ye, H., Sugihara, G.: Convergent cross mapping: theory and an example. In *Advances*
647 *in Nonlinear Geosciences* (pp. 587-600), Springer, Cham., 2018.
- 648 Vallis, G. K., Farneti, R.: Meridional energy transport in the coupled atmosphere–ocean system: Scaling and
649 numerical experiments. *Q. J. Roy. Meteor. Soc.*, 135(644), 1643-1660, 2009.
- 650 Van, Nes, E. H., Scheffer, M., Brovkin, V., Lenton, T. M., Ye, H., Deyle, E., Sugihara, G.: Causal feedbacks in
651 climate change. *Nat. Clim. Change*, 5(5): 445, 2015.
- 652 Vannitsem, S., Ekelmans, P. Causal dependences between the coupled ocean–atmosphere dynamics over the tropical
653 Pacific, the North Pacific and the North Atlantic. *Earth Syst. Dyn.*, 9(3), 1063-1083, 2018.
- 654 Vissio, G., Lucarini, V.: A proof of concept for scale-adaptive parameterizations: the case of the Lorenz 96 model.
655 *Q. J. Roy. Meteor. Soc.*, 144(710), 63-75, 2018.
- 656 Wang, L., Ting, M., Kushner, P. J.: A robust empirical seasonal prediction of winter NAO and surface climate. *Sci.*
657 *Rep.*, 7(1), 279, 2017.
- 658 Watson, P. A.: Applying machine learning to improve simulations of a chaotic dynamical system using empirical
659 error correction. *J. Adv. Model Earth. Sys.*, doi.org/10.1029/2018MS001597, 2019.
- 660 Wilks, D. S.: Effects of stochastic parametrizations in the Lorenz'96 system. *Q. J. Roy. Meteor. Soc.*, 131(606), 389-
661 407, 2005.
- 662 Zaytar, M. A., El, Amrani, C.: Sequence to sequence weather forecasting with long short-term memory recurrent
663 neural networks. *Int. J. Comput. Appl.*, 143(11), 7-11, 2016.
- 664 Zhang, N. N., Wang, G. L., Tsonis, A. A.: Dynamical evidence for causality between Northern Hemisphere annular
665 mode and winter surface air temperature over Northeast Asia. *Clim. Dyn.*, 52, 3175-3182, 2019.