Response to Reviewers' Comments

We thank the Editors very much for handling of our manuscript entitled "**Inverse Differential Equation Modeling of ENSO Prediction Based on Memory Kernel Functions**" (egusphere-2024-2181) by Wan et al. First and foremost, I would like to express my sincere gratitude for the thorough review and insightful comments provided on our manuscript. Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our research. We have carefully considered each point raised and have made the following revisions to our manuscript. Please see our point-by-point response (in blue normal font) below.

Reply to Reviewer#2:

We thank the reviewer for recognizing the value of our work and providing helpful comments. In the following, the text with normal font are the reviewer's original comments, and those with blue normal font are the authors' response.

General comments:

This paper performs data-driven predictions of ENSO by learning a function of past values. Unfortunately, the paper as it stands has severe problems, and I cannot recommend publication. Namely, (a) the method is described very vaguely; (b) I am unclear about whether this method can actually be applied for forecasting or is just a regression from SST to the ENSO index; (c) I am unclear about what novelty this paper has over past ones from the same group; (d) there is no comparison to other methods in the literature; (e) the verification of the method is inadequate; and (f) the overall paper is quite unclear.

Response: Thank you for your insightful comments and suggestions We have added a description of the method in Section 2.2. Several indices are used to monitor the tropical Pacific Ocean, all of which are based on the mean sea surface temperature (SST) anomaly for a given region. The Niño 3.4 index is the most common index used to define El Niño and La Niña events. The Niño 3.4 anomalies represent the average equatorial SSTs across the Pacific, from approximately the dateline to the South

American coast. This index typically employs a 5-month running mean, and El Niño or La Niña events are defined when the Niño 3.4 SST anomalies exceed +/- 0.4°C for a period of 6 months or more. In our study of ENSO, we primarily use the Niño 3.4 index, calculated based on actual SSTs, to characterize ENSO variations.

In our previous article, we focused on independently modeling local processes of complex systems for accurate simulation and prediction. In this paper, our work includes three main aspects: studying the period of SST time series using wavelet transform to obtain a time-frequency graph and extract specific periods, which lays the foundation for constructing the periodic memory kernel function (MKF); conducting inertial memory analysis of aperiodic components to establish the theoretical basis for the inertial MKF; and modeling the SST dynamic system with the MKF by presupposing its basic differential form and investigating whether the actual performance of the ODE-MKF aligns with the expected performance.

Although ENSO forecasts are critical, current business forecasts typically do not extend beyond 6 months. Our results, however, can predict trends up to 26 months in advance. Despite this, there is still a need to improve forecasts for years with extremely high or low values. We have provided a detailed comparison with existing ENSO forecast results in the following reply.

We also considered the influence of different initial values on the prediction effect and improved the overall description of the paper to enhance its clarity and coherence.

Major comments:

1. The authors state on lines 182 and 188 that their functions take as input *observed* SST from previous and current months. Unless this was misstated and the authors meant the *predicted* SST (or observed if the lead time is less than the lag of SST one wants to use), this means that the method cannot actually be applied as a forecast method without the use of future information, since it can only be applied up to a lead time of the minimum lag of observed SST taken as input (in this case it would only be able to predict for one step, since it needs the current observed SST). If this is

true, then this is not a forecast method, but rather a regression model mapping from observed SST to the ENSO index.

Response: Thanks for the reviewer's insightful comments. Several indices are used to monitor the tropical Pacific Ocean, all based on the mean SST anomaly for a given region. The Niño 3.4 index is the most common index used to define El Niño and La Niña events. The Niño 3.4 anomaly represents the mean equatorial SST in the Pacific Ocean from the approximate date line to the coast of South America. The index typically uses a 5-month operating average to define an El Niño or La Niña event when a Niño 3.4 SST anomaly exceeds +/- 0.4°C and persists for 6 months or more. In our ENSO study, we mainly use Niño 3.4 index based on actual SST calculation to characterize ENSO changes. The index is calculated based on key SST regions, so similar results can be calculated using observed SST data. Moreover, Ham et al. (2019) achieve an effective 18-month advance prediction of the Niño 3.4 SST index. In our results, we were able to predict ENSO trends about 26 months in advance. Such longer-lead forecasts could be of high value, because decision-making and management in a number of key socio-economic sectors could be greatly improved

2. The description of the genetic algorithm is very unclear. The authors say that a genetic algorithm was used to pick the functions comprising the model and their parameters. First of all, the actual set of functions that can possibly be selected, as well as what the parameters represent, is not stated anywhere. The authors say that "a subset of the parameter vectors was selected based on their fitness for crossover and mutation", but the selection criteria for this subset is not stated. The authors then state that "parts of the two parameter vectors were combined through crossover operations, and genes were randomly altered through mutation operations to generate new parameter vectors", but don't state how the crossover nor mutation operations work.

Response: Your valuable feedback is greatly appreciated. Genetic algorithm is an optimization algorithm that simulates the principles of natural selection and genetics. We use genetic algorithms to minimize the error function $E(m, f, p, q)$, Where p and q are unknown, their values are estimated by evolutionary algorithm. It should be noted that m and f are not fixed functions, but mathematical functions containing self-memorizing information (MKF or basis function), which are also dynamically determined by evolutionary algorithms. Specific steps of genetic algorithm are as follows:

a) Initialize: Generate an initial population containing multiple parameter vectors p, q, and function gene expressions m, f.

b) Fitness evaluation: for each parameter vector p_j, q_j, m_j, f_j in the population, the corresponding simulated state vector x is computed by numerical integration to simulate *(t)*, and the error function $E(m_i, f_i, p_i, q_i)$ is computed. The fitness of each parameter vector is evaluated according to the values of $E(m_i, f_i, p_i, q_i)$.

c) Selection: Select some parameter vectors for crossover and mutation according to their fitness.

d) Crossover and Mutation: Combine some genes of two parameter vectors through crossover operations and randomly change the values of some genes through mutation operations to generate new parameter vectors. The crossover and mutation steps in the evolutionary modeling process are detailed as follows:

For crossover, after completing an evolutionary algorithm iteration, randomly select a certain proportion of individuals from the total population M to form a new population. From this new population, select the best individuals to be part of the next generation population. Form another subgroup by randomly selecting a corresponding proportion of M individuals and again choose the best individual to include in the next generation. Repeat this process until half of the population size ($M/2$) individuals are selected. For the Mutation, randomly select an individual from the new population obtained through the evolutionary operations. Designate target F as the next generation individual. Repeat the mutation operation until the combined number of individuals obtained through mutation and replication operations reaches the preset population size.

e) Iteration: repeat the process until the stop criteria are met.

f) Result output: outputs the parameter vector with the highest fitness as the estimation result.

Please refer in the revised version. As follows the section of 2.2 Inverse differential equation modeling of MKFs.

3. What is the novelty of this paper over a similar paper from the same group, Ma et al. (2023)? That paper seems to use the same method and applies it to ENSO prediction. Similarly, that paper is quite vague about the algorithm.

Response: We sincerely appreciate the reviewer's constructive feedback. In our previous article, we focused on independently modeling local processes in complex systems for accurate simulation and prediction. However, random disturbances or stochastic processes inherent in complex systems are challenging to predict and can obscure primary periodic characteristics, interfering with accurate forecasting. This paper addresses these challenges by using wavelet transform to study the period of SST time series, creating a time-frequency graph to extract specific periods and laying the foundation for constructing a periodic MKF. Additionally, we analyze the inertial memory of aperiodic components to establish the theoretical basis for the inertial MKF. We then implement the MKF to model the SST dynamic system, presetting its basic differential form and investigating whether the actual performance of ODE-MKF aligns with the expected performance.

4. The validation of this method is insufficient. The authors only evaluate it on two different initialization dates. More initial dates should be considered, and the skill measures averaged over these. Moreover, the authors do not look at how the skill decays with lead time, which is crucial.

Response: Thanks for the reviewer's insightful comments. To reduce prediction uncertainties and skill decay, ensemble methods are commonly employed to improve prediction skill. In our previous research the ensemble Scheme-C prediction of 27 models with a backtracking scale of m=24, 14 etc. The ensemble prediction became smoother, with reduced random fluctuations. Although the ability to predict subtle changes in ENSO was reduced, the ability to capture trends improved. The transition from La Niña to El Niño events also accurately predicted. Furthermore, Scheme-C generated many prediction members by using the initial perturbation method within the same mode. Thus, in this research we selected the most suitable and stable initial dates. Besides, Figure 4 (Figure 3 of previous versions) also deposes our model avoiding rapid skill decays. We have added necessary explanations in lines 267-269: The high-frequency component primarily comprised retrospective initial values that included certain nonlinear terms, which also avoids rapid skill decays.

Ma Q, Sun Y, Wan S, et al. An ENSO Prediction Model Based on Backtracking Multiple Initial Values: Ordinary Differential Equations–Memory Kernel Function[J]. Remote Sensing, 2023, 15(15): 3767.

5. I did not understand the point of the whole section on transfer entropy, or actually how it was applied. Later the authors use the autocorrelation to determine relevant timescales, which seems to be sufficient.

Response: Thank you for your comprehensive review and helpful comments. The purpose of discussing transfer entropy and the correlation coefficient here is to analyze the system's memory characteristics. This discussion aims to simplify calculations using the MKF and evolutionary algorithm and to build a prediction model that incorporates memory effects.

The mathematical essence of transfer entropy is to determine the relationship between three variables using conditional probability. Entropy, as a state function, originally emerged from thermodynamics and statistical mechanics, describing the disorder of a material system. For closed systems, entropy generally increases, meaning the system tends toward maximum disorder. Transfer entropy's physical meaning involves energy exchange between systems, seeking maximum information exchange to achieve equilibrium. In contrast, the correlation coefficient is a descriptive statistic indicating the degree to which one variable's change corresponds to another. However, it only reflects synchronous changes between two variables without indicating causal relationships. To discuss the system's memory, we use both transfer entropy and the correlation coefficient. This dual approach simplifies the time required for modeling calculations and ensures a rational understanding of entropy transfer.

Given the observation sequence X, Y, and Z, the transfer entropy is defined as:

$$T_{Y\to X} = \sum_i p(x_i, y_i, z_i) \log \frac{p(x_i|y_i, z_i))}{p(x_i|, z_i)} \qquad (1),$$

equation (1) represents the transfer of joint entropy to *X* by *Y* based on *Z*. Unlike mutual information or conditional entropy, *X*, *Y*, *Z* is asymmetric and non-static, that is, the information transfer between *X*, *Y*, and *Z* is unidirectional, irreversible, and non-shared. Where $p(x_i|y_i, z_i)$ represents conditional probability, $p(x_i|y_i, z_i) = p(x_i, y_i, z_i)/p(y_i, z_i)$. Three scale control parameters are used in the calculation, ε_x, ε_y, ε_z, namely, $p(x_i, y_i, z_i) = p(\varepsilon_x, \varepsilon_y, \varepsilon_z)$ and $p(y_i, z_i) = p(\varepsilon_y, \varepsilon_z)$. Since the effect on *X* is calculated, the other two parameters are determined in a fixed way. The method is to look at $p(x_i|y_i)$ as a function of ε_y, and take the value of m at its maximum:

$$\varepsilon_y = argmax \cdot p(x_i|y_i) \qquad (2),$$

similarly, we can determine the value of ε_z.

Studies have shown that $T_{*\to X}$ is generally not affected by noise, has strong anti-interference ability, is suitable for analyzing observational sequences, and is not limited by sample data length, and can analyze shorter sequences, which is much smaller than the required sample size for computing association and correlation.

Please refer in the revised version. As follows in section 2.3 Inertial MKF.

6. There is no comparison of skill to other methods, either dynamical or statistical/machine learning, to see if this one is competitive. For a recent benchmark, see Ham et al (2019).

Response: Thank you for the detailed and thoughtful feedback We add a comparison with the existing conclusions. Please refer to:

Lines 406-419: ENSO forecasts are critical, yet most operational forecasts typically do not extend beyond 6 months. Our results capture the system's nonlinear characteristics and reflect its trend changes. From January 1951 to July 2022, we successfully identified the SST change trend and forecasted ENSO trend changes up to 26 months in advance. For comparison, Petrova et al. (2024) employed a statistical unobserved dynamic components model to predict ENSO events 11 to 17 months in advance. Additionally, the extended nonlinear recharge oscillator model demonstrates proficient ENSO forecasts with lead times of up to 16–18 months (Zhao et al., 2024). Ham et al. (2019) used a CNN model to achieve an effective 18-month advance prediction. Hu et al. (2021) enhanced model performance and stability by incorporating dropout technology and residual connection modules, leading to the development of the Res-CNN model, which refines prediction capability up to 20 months in advance. Overall, our results demonstrate a superior capability to capture ENSO trend changes, allowing us to predict ENSO trends up to 26 months in advance. However, the prediction accuracy for some anomalously high and low SST values needs improvement, which we aim to address in future work.

7. I don’t understand what "inertial memory" means, or what Figure 6 is showing.

Response: Your thorough review and valuable suggestions are much appreciated. Complex systems have their own memory, expressed in two ways: periodic "reciprocating motion," known as "periodic memory," and inertial motion, called "inertial memory." Specifically, inertial memory, which reflects the current state of the system, contains information from previous states. We first analyzed the periodicity of the SST time series and constructed periodic MKF, with the related results shown in Figure 3 (Figure 2 in previous versions). Second, we performed inertial memory analysis on the aperiodic component to establish the theoretical basis for constructing the inertial MKF, as shown in Figure 7 (Figure 6 in previous versions).

According to the results of correlation coefficient and transfer entropy, we find that the memory cycle of ENSO is 6 months, so the upper limit of the number of initial values considered when constructing MKF is 7 (including the current initial value). Therefore, compared with the original method of randomly selecting the past 24 initial values, this initial value scale can greatly reduce the range of numerical solutions searched by evolutionary modeling, and is expected to significantly improve the modeling efficiency. Similar to Figure 3 (in previous versions of Figure 2), we set the parameters (The training dataset consisted of 180 months (sample range: 650–838, from

April 2005 to December 2018), while the testing dataset comprised 26 months (830–856, from January 2019 to July 2022)).

Different from MKF which is composed of time variables considering periodic memory, MKF of inertial memory is mainly composed of initial value vectors, and the mathematical form of kernel function is uncertain, which may have multiple expressions. So we use a tanh function similar to the sigmoid activation function, which compresses the output in the interval form (-1,1). $\tanh(x)$ is the memory switch function, which can be expressed as:

$$\tanh(x) = T(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad (3),$$

The specific expression for the inertial MKF(t,x) obtained from evolutionary modeling is as follows:

$$MKF(t, sst) = (\tanh(255.27 \times sst[-3]) - (265.89 \times sst[-1])) \qquad (4),$$

The high-frequency component f(t,X) is expressed as follows:

$$f(sst) = 0.99 \times sst[0] \times sst[-1] \times sst[-5] + (sst[-4] - 0.61) \times 5.69 \times sst[-3] \times sst[0] - (13.41 - sst[-2] \times sst[0] + sst[-1] \times sst[-6]) - ((sst[-1] + 3.01) \times sst[-2]) + sst[-5] \times (sst[-1] + sst[-5] + sst[-6]) \times \sin(3.24 - sst[-3]) + 8.755 \times \cos(e^{sst[0]} \times \cos(sst[-2] \times (sst[-5] + 0.85))) \qquad (5).$$

As can be seen from the MKF expression, the kernel function is in a relatively simple form but captures the memory of the system. Moreover, the high-frequency component is composed of traceback value vector and trigonometric function, and also has some nonlinear characteristics. The strength of our approach is that it combines these two parts to build ODE-MKF. Compared with the aforementioned periodic memory, the ODE-MKF still has a big difference. The former has more small random fluctuations, while the latter reflects smooth periodicity. Both of them have better ability to predict the trend.

Other issues:

1. What numerical method is used to integrate the model? Presumably a time-step of 1 month is used?

Response: The integration method we use is the Runge-Kutta method, and the time-step of the integration is 1 month based on the scale of the data used.

2. Eq. 1 (and 2) is written as a regular ordinary differential equation, but if it a function of previous points in its own history then it should be written as a delay differential equation.

Response: Your valuable feedback is greatly appreciated. In this paper, we propose the concept of constructing a system MKF based on multiple initial values (iterative output values) and utilizing a more flexible evolutionary algorithm for parameter estimation. In our approach, m (x, p, t) is dynamically defined by the evolutionary algorithm. It is not a fixed function but a mathematical function containing self-memorizing information. This means that the rate of change of the current state also depends on the state at a previous point in time, which can be expressed by m (x, p, t).

3. The axes of Figure 1 are not labelled.

Response: Thanks for the reviewer's insightful comments. We have made a correction. Please refer in the revised version. As follows:

Figure 2 (in previous versions of Figure 1)

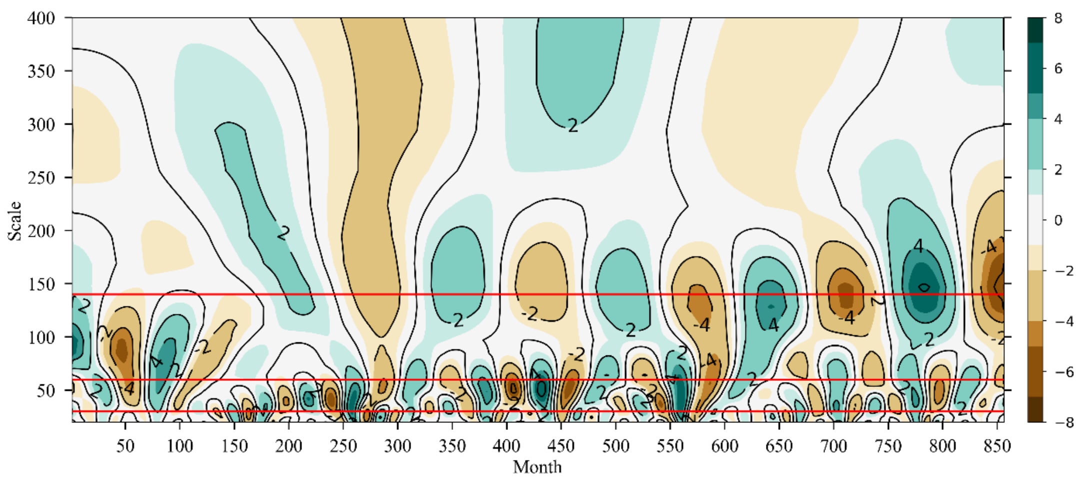

Figure 2: Time-frequency plot of wavelet transform for SST anomaly sequence. The red lines in the figure indicate scales of 140, 60, and 30 months.

4. why do Figure 2, 3 and 6 have the curve in the test period go from red to purple? Is there some significance to this?

Response: We are thankful for the reviewer's thoughtful critique and recommendations. The red curve in the Figure represents the model selection process, while the purple curve shows the predicted Niño 3.4 SSTA based on the optimal model. We have supplemented this section with a detailed description. Please refer in the revised version. As follows:

Line 225: "...The red curve in the Figure represents the model selection process, while the purple curve shows the predicted Niño 3.4 SSTA based on the optimal model."

Line 256: "Figure 4: Same as Figure 3c, but for the integral curve of ODE-MKF for the long sequence."

Line 365: "Figure 7: Same as Figure 3, but for the (a) Integral curve of inertial MKF, (b) integral curve of high-frequency disturbance component, and (c) integral curve of ODE-MKF based on inertial MKF, with initial values considered when constructing MKF is 7 (including the current initial value)."

5. In Figure 2, what is the difference between panel a and c?

Response: Panel (a) shows the integral curve of inertial MKF of equation 4, which includes the memory information of the past 3 months SST and the 60-month cycle of memory signals. Panel (a) more captures the disturbance and memorability of Niño 3.4 SSTA, reflecting more random fluctuations as a whole, and mainly gives the characteristics of SST periodic memory. For the panel (c), it shows the organic combination of periodicity (a) and trend (b), which is ODE-MKF and shows better predictability for Niño 3.4 SSTA.

6. Why do the "high-frequency" components in Figures 2 and 6 actually look low-frequency?

Response: From the higher frequency part f (t, X) differential, the higher frequency component is described by information from past SST observations for the

5th and 6th months. The MKF derived from evolutionary modeling describes the memory signal of the period of 60 months, and includes the memory information of the past 3 months SST. The training results from the figure are not ideal for capturing Niño 3.4 SSTA, but after combining the two into ODE-MKF, the effect is greatly improved.

7. Through the characterization of memory using the tanh function: this was never described. Similarly, the authors say that "the high-frequency component primarily comprised retrospective initial values that included certain nonlinear terms" which is also not described.

Response: When creating MKF of inertial memory, considering the uncertainty of kernel function mathematical form and the possibility of multiple expressions, we use tanh function similar to sigmoid activation function as memory switch function. The Tanh activation function compresses the output in the interval (-1,1), as follows:

$$\tanh(x) = T(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}.$$

f(X) is the differential of multiple initial values, representing the higher frequency component, and X is the initial value vector. The kernel function is embedded in the evolution model as gene code, and the evolution algorithm determines the specific differential expression by itself.

$$f(sst) = \frac{40.2}{sst[-5] + sst[-6] - sst[0] + 135.45} - 1.156,$$

which can be seen that the higher frequency component is described by information from past observations of the 5th and 6th months of the SST. We have revised and supplemented this part of the description in the main paper.

8. “These equations were designed to capture the dynamic behavior of climate systems, where the future state depends not only on the present conditions but also on a series of past states.” This is not true; in fact, the climate is a Markovian system. However, a partially observed Markovian system (e.g., just the ENSO index) will act in a non-Markovian way. This can be understood through the Mori–Zwanzig formalism or Takens’ embedding theorem; see, e.g., Levine & Stuart (2022).

Response: Thanks for the reviewer's insightful comments. Our focus is on capturing the memory of the system, not dynamic behavior. Your insights are very insightful and we have made changes. Please refer in the revised version. As follows:

Lines 74-75: "These equations were designed to capture the memory of climate systems, where the future state depends not only on the present conditions but also on a series of past states."

9. In the introduction, "short-term forecasts" is used, as far as I can tell, to refer to forecasts on timescales shorter than those involved in climate projections (i.e., decadal and longer). This is not standard terminology. I would suggest to be more specific, e.g., refer to seasonal or interannual forecasts.

Response: We appreciate the reviewer's diligence and insightful feedback. In the introduction, we mentioned that "statistical models rely on historical data, are generally less computationally intensive, and can be useful for short-term forecasts". This is an error in our expression, it should actually be an interannual forecast. Please refer in the revised version. As follows:

Lines 45-47: "Additionally, statistical models rely on historical data, are generally less computationally intensive, and can be useful for interannual forecast."

10. The reference for "Predictability of El Niño over the past 148 years" has the wrong year: it should be 2004, not 2024.

Response: We are grateful for the constructive suggestions provided. This was an error in our writing and has now been corrected in the references section. Please refer in the revised version. As follows:

Line 484: "Chen, D., Cane, M. A., Kaplan, A., Zebiak, S. E., and Huang, D.: Predictability of El Niño over the past 148 years, Nature, 428, 733–736, https://doi.org/10.1038/nature02439, 2004."

11. "combining the strengths of dynamic and statistical approaches has emerged as a promising strategy": there is a great deal of work in this area. See for example Bach et al. (2024) and the references therein.

Response: Thanks for the reviewer's insightful comments. This is a very thoughtful and constructive article. We have references therein.

12. There has also been a great deal of work on time-series forecasting methods that make use of memory, such as recurrent neural networks and reservoir computers. Some of this previous work should be cited. See for example Pathak et al. (2018).

Response: Thanks for the reviewer's insightful comments. This is a very thoughtful and constructive article. We have references therein. Please refer in the revised version. As follows:

Line 68-69: Many works on time-series forecasting methods utilize memory, such as recurrent neural networks and reservoir computers (Lindemann et al., 2021; Pathak et al., 2018).

13. The authors claim that numerical models are "powerful for long-term projections but often less effective for short-term predictions involving rapid changes and smaller scales". I don't understand this claim. In fact, numerical models have, until very recently, outperformed data-driven methods for medium-range weather forecasting (3 to 10 days lead time). Data-driven methods have generally been more competitive for subseasonal-to-seasonal prediction (two weeks to several months).

Response: Thank you for your comprehensive review and helpful comments. We made a change in the introduction to correct our misunderstanding. Please refer in the revised version. As follows:

Lines 40-45: "These models require substantial computational resources and detailed initial conditions. While they are powerful for medium- to long-term projections, they often face challenges in short-term predictions involving rapid changes and smaller scales (Alizadeh, 2022). It is important to note that numerical models have traditionally outperformed data-driven methods for medium-range weather forecasting (3 to 10 days lead time). In contrast, data-driven methods have shown more competitiveness in subseasonal-to-seasonal predictions (two weeks to several months)."

14. Why in Eq. 1 do the authors use F and for Eq. 2 they use Fa?

Response: Eq.1 stands for dynamic differential equation of the ENSO system. On assuming that F possesses a specific structure or form that includes the undetermined parameter p, the function F_a can be defined as follows:

$$F_a(x,p,t) = m(x,p,t) + f(q,t).$$

Notably, functions m and f are not fixed functions, but mathematical functions containing self-memorizing information (MKF or basis function), which are also dynamically determined by evolutionary algorithms. In general, F and F_a are different expressions of the same equation, the latter involving the memorability and random perturbations of the system.

15. On line 178, the authors write MKF (sin t, X). I assume MKF (t, X) was meant?

Response: Thanks for the reviewer's insightful comments. We have made changes in the text, MKF (t, X) is correct. Please refer in the revised version. As follows:

Line 229: "$MKF(t,X)$".

16. ENSO should be defined and explained in the introduction.

Response: We have added the corresponding explanation and definition in the introduction, as well as the corresponding impact. And in the discussion, the profound significance related to ENSO is deeply discussed. Please refer in the revised version. As follows:

Lines 90-100: For instance, the El Niño-Southern Oscillation (ENSO) is the most influential driver of global climate variabilities (Philander, 1983) and exhibits complex composite signals. El Niño events lead the central and eastern equatorial Pacific warming, which reduces the west-to-east zonal sea surface temperature (SST) gradient (Trenberth, 1997). The decreased SST gradient causes the western Pacific, the Inter-

Tropical Convergence Zone, and the South Pacific Convergence Zone migrate toward the eastern equatorial Pacific (Cai et al., 2014). As a result, atmospheric convection becomes established in patterns that signal the likelihood of extreme precipitation or droughts globally (Stephens et al., 2015). What's worse, under transient greenhouse warming, ENSO variability is projected to increase pre-2100 (Geng et al., 2024). However, given the nonlinear relationship between ENSO and the magnitude and frequency of extreme events, enhancing ENSO prediction is critical.

Reference:

Alexander, M. A., Shin, S.-I., and Battisti, D. S.: The Influence of the Trend, Basin Interactions, and Ocean Dynamics on Tropical Ocean Prediction, Geophys. Res. Lett., 49, e2021GL096120, https://doi.org/10.1029/2021GL096120, 2022.

Bach, E., Krishnamurthy, V., Mote, S., Shukla, J., Sharma, A. S., Kalnay, E., and Ghil, M.: Improved subseasonal prediction of South Asian monsoon rainfall using data-driven forecasts of oscillatory modes, Proc. Natl. Acad. Sci., 121, e2312573121, https://doi.org/10.1073/pnas.2312573121, 2024.

Bjerknes J.: Atmospheric teleconnections from the equatorial Pacific. Mon. Weather Rev., 97, 163–172, https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2, 1969.

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M. J., Wu, L., England, M. H., Wang, G., Guilyardi, E., and Jin, F.-F.: Increasing frequency of extreme El Niño events due to greenhouse warming, Nat. Clim. Change, 4, 111–116, https://doi.org/10.1038/nclimate2100, 2014.

Geng, T., Cai, W., Jia, F., and Wu, L.: Decreased ENSO post-2100 in response to formation of a permanent El Niño-like state under greenhouse warming, Nat. Commun., 15, 5810, https://doi.org/10.1038/s41467-024-50156-9, 2024.

Ham, Y.-G., Kim, J.-H., and Luo, J.-J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568–572, https://doi.org/10.1038/s41586-019-1559-7, 2019.

Hanlon, H., Hegerl, G., Tett, S., and Smith, D.: Can a Decadal Forecasting System Predict Temperature Extreme Indices?, J. Clim., 26, 3728–3744, https://doi.org/10.1175/JCLI-D-12-00512.1, 2013.

Hu, J., Weng, B., and Huang, T.: Deep residual convolutional neural network combining dropout and transfer learning for ENSO forecasting, Geophys. Res. Lett., 48, e2021GL093531, https://doi.org/10.1029/2021GL093531, 2021.

Liu, Y., Hu, Z.-Z., Wu, R., Ding, Y., and Jha, B.: Evolution of ENSO-related seasonal precipitation anomalies in East Asia: A robustness revisit, Int. J. Climatol., 44, 269–285, https://doi.org/10.1002/joc.8327, 2024.

Mazzola, L., Laine, E., Breuer, H., Maniscalco, S., and Piilo, J.: Phenomenological memory-kernel master equations and time-dependent Markovian processes. Phys. Rev. A., 81, 062120, https://doi.org/10.1103/PhysRevA.81.062120, 2010.

Oueslati, B., Pohl, B., Moron, V., Rome, S., & Janicot, S.: Characterization of heat waves in the Sahel and associated physical mechanisms. Journal of Climate, 30(9), 3095–3115. https://doi.org/10.1175/JCLI-D-16-0432.1, 2017.

Philander, S. G. H.: El Niño Southern Oscillation phenomena, Nature, 302, 295–301, https://doi.org/10.1038/302295a0, 1983.

Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L. V., Hegerl, G., Seneviratne, S. I., Vautard, R., Zhang, X., and Zwiers, F. W.: Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities, Weather Clim. Extremes, 18, 65–74, https://doi.org/10.1016/j.wace.2017.10.003, 2017.

Stephens, E., Day, J. J., Pappenberger, F., and Cloke, H.: Precipitation and floodiness, Geophys. Res. Lett., 42, 316–323, https://doi.org/10.1002/2015GL066779, 2015.

Sun, J., Liu, S., Cohen, J., Yu, S.: Influence and prediction value of Arctic sea ice for spring Eurasian extreme heat events, Commun. Earth Environ., 3, 172. https://doi.org/10.1038/s43247-022-00503-9, 2022.

Trenberth, K. E.: The Definition of El Niño, Bull. Am. Meteorol. Soc., 78, 2771–2778, https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2, 1997.

Wang, C., Deser, C., Yu, J. Y., DiNezio, P., and Clement, A.: El Niño and Southern Oscillation (ENSO): A Review. In: Glynn, P., Manzello, D., Enochs, I. (eds) Coral Reefs of the Eastern Tropical Pacific, Coral Reefs of the World, 8, https://doi.org/10.1007/978-94-017-7499-4_4, 2017.

Wei, J., Wang, W., Shao, Q., Yu, Z., Chen, Z., Huang, Y., and Xing, W.: Heat wave variations across China tied to global SST modes, J. Geophys. Res.: Atmos., 125, e2019JD031612, https://doi.org/10.1029/2019JD031612, 2020.

Zhao, S., Jin, F.-F., Stuecker, M. F., Thompson, P. R., Kug, J.-S., McPhaden, M. J., Cane, M. A., Wittenberg, A. T., and Cai, W.: Explainable El Niño predictability from climate mode interactions, Nature, 630, 891–898, https://doi.org/10.1038/s41586-024-07534-6, 2024.

Zhou, L. and Zhang, R.-H.: A self-attention–based neural network for three-dimensional multivariate modeling and its skillful ENSO predictions, Sci. Adv., 9, eadf2827, https://doi.org/10.1126/sciadv.adf2827, 2023.