

The Stochastic Climate Model helps reveal the role of memory in internal variability in the Bohai and Yellow Sea

Lin Lin ^{1,2}, Hans von Storch ²✉ & Xueen Chen¹

Hasselmann's theory elucidates how short-term random noise leads to longer-term unprovoked variations, i.e., red spectra. Here, we study ensembles of numerical model simulations of the hydrodynamics of the Bohai and Yellow Sea concerning internal variability formation. Short(/long) term variations are associated with small(/large) spatial scales, and the internal variability of long-term temporal and large-scale variations is markedly enhanced, even without external forcing on these scales, when the tides are turned off. This pattern is well explained by Hasselmann's theory. A critical element in this theory is the concept of memory, which in our ensembles exhibits a scale dependence that aligns with the scale-dependent nature of redness. Additionally, this framework clarifies why there is a significant reduction of long-term fluctuations during winter and when tides are active: the system's memory is notably diminished under these conditions.

¹Frontier Science Center for Deep Ocean Multispheres and Earth System (FDOMES) and Physical Oceanography Laboratory, Ocean University of China, Qingdao, China. ²Institute of Coastal Systems, Helmholtz Zentrum Hereon, Geesthacht, Germany. ✉email: hvonstorch@web.de

The behavior of oceans and atmospheres is characterized by the turbulent flow of fluids that exhibit chaotic properties across a wide range of scales, ranging from short scales of a few centimeters to large scales of hundreds of kilometers¹. This turbulent nature results in the emergence of unpredictable, internally generated variability, commonly referred to as noise.

It is common for terms to have varying interpretations across different fields, and in this context, the terms noise, internal variability, and unprovoked variability are used interchangeably. This definition aligns with the terminology used by Penduff et al.^{2,3}, Hasselmann's terminology⁴, and it is used in predictability research of climate and oceanic systems.

Previous studies have shown that internal variability can be analyzed in ensemble simulations with slightly different but consistent initial conditions, the introduction of small random values (well below the level of the analyzed variable)^{5–8}.

While the existence of internally generated variability was acknowledged in models of atmospheric dynamics as early as the 1970s, it was not until a few decades later that the ocean science community began to take notice of this phenomenon^{9–13}. Global ocean studies, for instance, have identified mesoscale eddies and internal waves as potential sources of internal variability^{14–16}. Internal variability in the Gulf of Lyon was demonstrated¹⁷, and the results suggested that baroclinic instability was the cause. Scholars researched the South China Sea's internal variability climatology^{7,18}.

The hypothesis that the lateral boundary may suppress the generation of internal variability, may explain that the internal variability of marginal seas has drawn less attention over a long period. Until recently, researchers understood that shallow area internal variability is also important^{3,5}. The signal-to-noise ratio of sea level trends (ensemble mean divided by the ensemble standard deviation) in the Northwest Pacific, which includes the Bohai and Yellow Sea, was calculated (Fig. 3 in Penduff et al.'s publication³). The variability is mostly caused by external forcing where the signal-to-noise ratio is >2 ; while if the signal-to-noise ratio is <2 , then the variability is mostly attributed by internal forcing. The previous work found out that³ in the Bohai and Yellow Sea the sea level longer timescale variations may be blurred by internal variability, because the whole Bohai and Yellow Sea signal-to-noise ratio is <2 . Based on the above research, the internal variability analysis in the Bohai and Yellow Sea is an issue of scientific interest.

The tidal forcing is a key factor in the Bohai and Yellow Sea, which influences the temperature, salinity, and current distribution^{19–22}. Meanwhile, our previous work found that the tides have an important impact on the formation of internal variability at large and medium scales when simulating the dynamics of the Bohai and Yellow Sea²³. To the best of our knowledge, no previous research has explored why tides influence the internal variability in the Bohai and Yellow Sea. Thus, we reckon the active tides, especially their effect on internal variability is an important scientific problem.

To understand the changes in system properties rather than the chain of mechanisms through which the tides affect internal variability, we use Hasselmann's ansatz, namely that of the Stochastic Climate Model (SCM)⁴. We anticipate observing natural variability with greater internal variability on longer timescales as compared to short term variations.

The simplest, and purest form of the Stochastic Climate Model is an autoregressive process of first-order. An introduction to the SCM with more background information is given in the Methods section. Here we briefly express the concept of SCM from the physical and numerical aspects. Physically, the SCM is based on the assumption that the two distinct times scales exist and can be distinguished in a system, one for long-term trends and another

for short-term fluctuations. The long-term variations are constrained by memory, and forced by short-term variations, permitting transformations from short-scale (white noise) forcing into long-term variation (a red spectrum) with enhanced variance. The memory of the system decides the strength of such transformation. Mathematically, the system is well approximated by an autoregressive process of the first order:

$$y_{t+1} = \alpha y_t + \varepsilon_t \quad (1)$$

with y_{t+1} and y_t representing the value of the slow y at the time step $t+1$ and t , a memory parameter α , and white noise ε_t , representing the impact of the short-term fluctuations. The spectrum of y is red, even if no long-term forcing is present. We will come back to the SCM-associated concepts, such as memory, in the Methods section.

A recent application of the SCM is that of Shi et al.²⁴, who suggested that a decline in ocean memory in recent years is related to the decrease of the upper-ocean mixed layer depth related to global surface warming, based on a simple stochastic model of sea surface temperature variability. A major difference between our work and Shi et al.²⁴, is that we separate the changes into different spatial, and thus temporal, scales by EOF-decomposition and discuss the tidally and seasonally induced changes in ocean memory at different scales.

This paper examines numerical experiment results not by screening a variety of processes, but by examining the output on the system level to explore why internal variability decrease when tidal forcing is considered and in wintertime. An analysis on the process level would also be useful, meanwhile, we believe that our concept of studying the experiment in terms of system properties may need to attract attention. An important insight is the need for significant statistical hypothesis-testing, or similar methods, when evaluating numerical experiments on, say, the effect of using different parameterizations.

Specifically, first, this paper demonstrates the potential of using the SCM for evaluating numerical experiments, namely the change of the memory of the system as a key property, to examine the impact of tides on the Bohai and Yellow Sea, based on scale separation method using the EOF-decomposition. Second, except for the decrease of the internal variability resulting from an activated tidal forcing, which is associated with reduced memory, we demonstrate another case to validate the memory of the system: the seasonal changes of the internal variability in the Bohai and Yellow Sea, with larger memory and internal variability in summer, and less so in winter.

In this paper, we first construct a kind of spectra, to determine if the internal variability is red. This is done by using spatial patterns, namely EOFs, which go with both spatial and temporal scales (see Result sections), in consistency with the well-known general correlation of such scales in atmospheric and oceanic dynamics²⁵, and determine that the variance increases with increasing spatial (and thus temporal) scale. As shown in our previous work²³ this expectation is fulfilled in an ensemble of simulations of the hydrodynamics of the Yellow Sea without tides.

We interpret this scale-dependent internal variability intensity as a trace of a low-dimensional dynamical core within which the internal variability is formed. We find evidence of an underlying low-dimensional dynamical core which is accumulating short-term, small-scale unprovoked variability, and which dominates the whole system in the spirit of the SCM. We test the validity of this assumption by checking if the memory of the system, which is the only relevant parameter in this context, is smaller when less low-frequency noise is formed in the same simulation with active tides compared to the simulations without tides, and in summer and winter. Here, we find that the memory, represented by the decay of the autocorrelation function, is strongly reduced in the

presence of tides, and in winter. The issue of identifying the specific mechanism of dynamic processes within the equations of motion that result in the loss of memory remains unresolved and will be addressed separately. The outcome supports our original hypothesis that the spectrum level of internal variability may be understood conceptually in the framework of the Stochastic Climate Model, with memory being the key parameter.

Results

The intensity of the internal variability is mathematically defined as the time average of the daily standard deviation across all days of the ensemble. When we refer to spectra in the following discussion, we refer to a distribution of variance across time or spatial scales, rather than Fourier spectra.

The Stochastic Climate Model is a model of minimum complexity, intended to describe the fundamental factors involved in generating internal variability within a system. Conversely, the numerical model of the hydrodynamics of the Yellow Sea is a model of maximum complexity, enabling experimental analysis but not providing direct insight into the dynamics of generating internal variability. Our task is to merge these two models in the current analysis, to arrive at a simple understanding.

Scales. Note that the EOFs are derived from the deviations of the ensemble mean, they allow for a decomposition of the internal variability part (see the scale separation based on the EOF analysis portion in the Methods section).

Our previous findings²³ indicated that spatial scales decrease almost monotonously with the EOF rank, even though not in a strict decline, which may be associated with the fact that both, the EOFs themselves and the scales are estimated and subject to random variations.

To establish a connection between spatial and temporal scales, we analyzed each EOF separately and calculated the autocorrelation function for each corresponding coefficient time series. Based on these results, we determined a time scale, represented by the characteristic time. (The definition of the characteristic time is in the Stochastic Climate Model portion of the Methods section). As illustrated in Fig. 1, without tides simulation, the larger the spatial scales, the longer the characteristic time. Such a relationship is evident in the ensemble without tides, but in the ensemble with

tides, such a correlation between spatial and temporal scales exists but is weaker. Consequently, in consistence with the conclusion of previous work^{7,23}, we find the spatial scales decrease with the EOF index increase. Both spatial and temporal scales decreased in parallel and mostly monotonically with the rank for the EOFs.

Change of intensity of internal variations. We establish a connection between the intensity and temporal scales of internal variability when tides are active and passive, by examining the slope of the time spectra, or the degree of redness. In terms of the Stochastic Climate Model and its associated AR(1)-process, a smaller negative slope (with a larger absolute value) indicates a greater accumulation of variance from short-term variability, leading to greater generation of internal variability on larger scales. We calculate the spectra of each EOF time coefficient.

Figure 2 illustrates the variance frequency spectra of scale 81 km (based on the time coefficient of EOF 1), scale 66 (EOF 30), and scale 20 (EOF 2000) to better characterize the internal variability and the slope. The green lines shown in Fig. 2a–c are the fitted slopes of scale 81, 66, and 20 km, from which we can see that the slopes vary from a steeper one to an almost horizontal line. Here, we take scales 81, 66, and 20 km for examples, if we chose scales close to the chosen 81, 66, and 20 km, the spectra are mostly the same and the slopes are mostly the same. In Fig. 2a, the variance of scale 81 km shows a larger value over the period >10 days; the same pattern exists in scale 66 km. But the slope is deeper in scale 81 km, compared with scale 66 km, which means the spectra are redder in larger scales. From Fig. 2a, b, we find that the spectra of large and medium scales are red. By contrast, the spectrum of scale 20 km is almost white, with almost the same variance at different frequencies (Fig. 2c). The scales equal to or <30 km are considered to small spatial scales; they hardly have any memory, thus, the spectra are white (the white noise spectrum is a horizontal line, as introduced in the Stochastic Climate Model of the Methods section).

To establish a connection between the internal variability intensity and the slope of the time spectra, we present the slopes for both ensembles for spatial scales >30 km (Fig. 3). In Fig. 3, the blue cloud representing the slopes in with-tides simulations is above the red cloud representing the slopes in without-tide simulations. For small scales, say 30 km and less, the spectra are almost white (Fig. 2c), so scales <30 km are not considered in the Figs. 3 and 4. The time spectra of the large-scale EOF coefficients are considerably redder in the no tide case compared to the simulations with the activated tide. In general, larger absolute values of slopes appear in the without-tides simulations than with-tides simulations.

Memory changes. In order to relate qualitatively the slopes with the memories of the EOF coefficients, we plot Fig. 4, scatter diagrams of the slopes or redness, and of the characteristic time scales (memory), as derived from the autocorrelation functions.

Figure 4 illustrates a co-variation of the growth of the intensity of the internal variability and of times scales, i.e., the slope of the spectra, and the memory in the no tides case. A more negative slope corresponds to a steeper spectrum, which means that more high-frequency disturbances are accumulated into low-frequency variability. However, when tides are active, this correlation becomes much weaker. This observation provides evidence that the hydrodynamical system of the Yellow Sea may possess a Stochastic-Climate-Model-like dynamical core.

The above results explain the tidal forcing effects from the redness of the spectra. We further explore the time scales and the scale-dependent memory of anomalies in the system by autocorrelation functions for each EOF time series, averaged

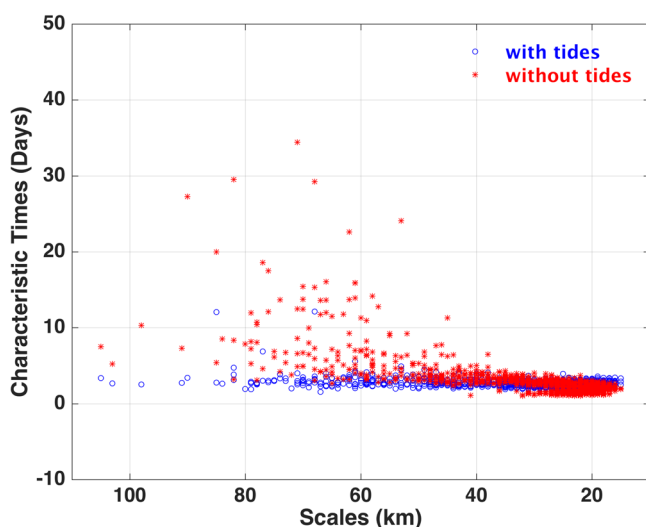


Fig. 1 Scatter diagram of the spatial scales of joint EOFs and their time scales τ in the with tides and without tides ensembles. The blue circles and the red stars represent the characteristic times for the with-tide and no-tide ensembles, respectively.

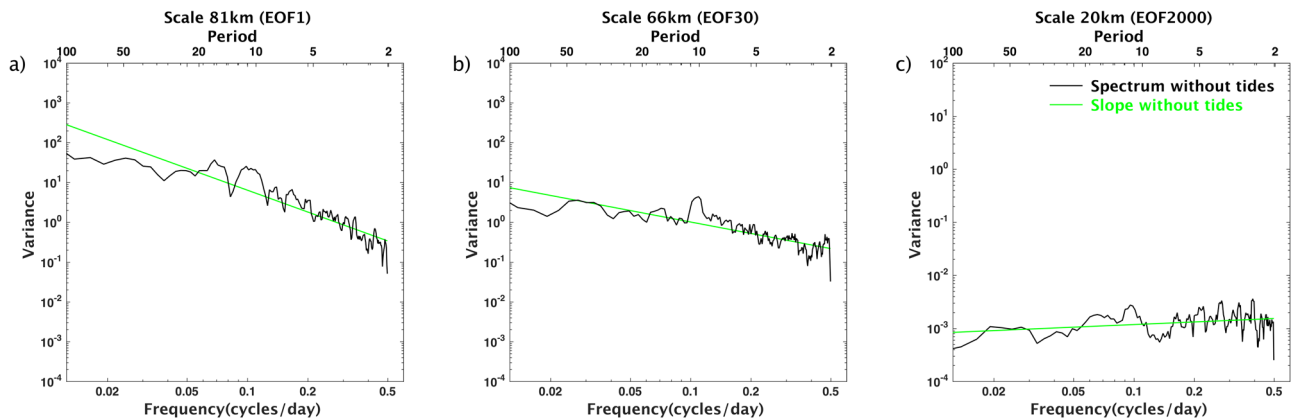


Fig. 2 Variance frequency spectra of the principal components in without tides ensemble. **a** Variance frequency spectra of the principle component for EOF1, which is corresponding to scales 81 km in spatial scale. **b** For EOF30, which is corresponding to scale 66 km in spatial scales. **c** For EOF2000, which is corresponding to scale 20 km in spatial scale. The principle components are derived from the EOF results of the depth average velocity anomalies. The black solid lines show the spectra, and the green lines represent the slope of the spectra.

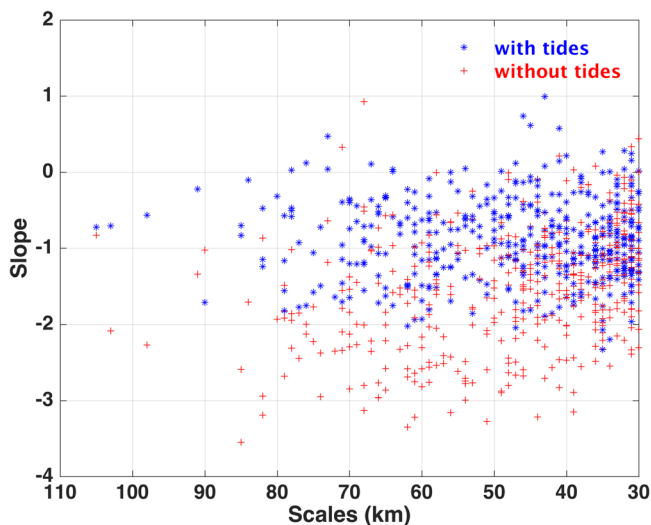


Fig. 3 Slope of the spectrum (steepness of linear fit of the time spectrum in log-log presentation) of spatial scales larger or equal to 30 km. The blue stars and the red crosses represent the slope of the spectrum from with-tide and no-tide ensembles, respectively.

across all members of the ensembles (Fig. 5) for the first 50 large and medium scales EOFs²³. The differences in large-scale (long-term) variance are associated with a difference in temporal autocorrelation. When tides are turned off, then the autocorrelation function decays slower compared to the case with tides. Note that EOF analysis was conducted of depth-averaged velocity anomalies, which represent the internal variability part. Thus, we conclude that larger anomalous patterns prevail longer than smaller anomalous patterns—and that this tendency is much stronger for the case without tide.

To sum up, memory is described by the redness of the spectrum given by the slope of the spectrum. The steeper the slope, the redder the spectrum. From both the autocorrelation and the steepness of the spectrum, we find that for large scales (low-frequency part), the autocorrelation is larger in no-tide and the steepness of the spectrum is steep in the no-tide ensemble, which corresponds to a redder spectrum compared with the ensemble with tidal forcing. Hence, we conclude that the memory of low-frequency noise is stronger in no-tide simulation. Thus, in the no-tide simulation, if an anomaly generates, it will last for a

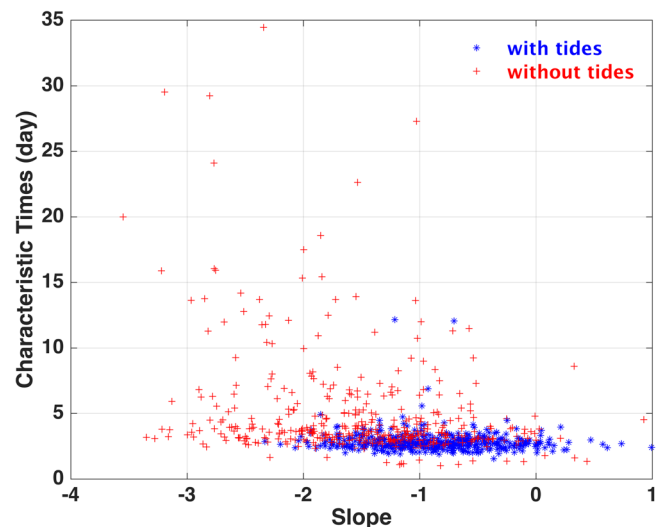


Fig. 4 Scatter diagrams of the slopes of the coefficient times series of EOFs with a spatial scale of 30 km and more and of the memories of the time series. In red: the ensemble without tides, in blue, with tides. The coefficient time series are derived for the internal variations only, i.e., after subtracting the daily ensemble means.

longer period of time. The internal variability level increase with the memory of anomalies.

In addition, the internal variability is shown as the spatial mean standard deviation of depth-averaged velocity anomalies for both configurations, with and without tides in Fig. 6. In both cases, the internal variability is weak in winter, starting to grow in spring, being strongest in summer, and receding gradually in autumn. We further analyze the memory of the system for February (when internal variability is at minimum) and August (when internal variability is at maximum) separately (Fig. 7). Consistently, we find that the memory of the system is smaller in winter, and greater in summer. This is another piece of evidence pointing the co-variation of memory of the system low-frequency internal variability is formed.

Conclusions

In this paper, we apply the concept of the Stochastic Climate Model to evaluate numerical experiments in the Bohai and Yellow Sea with and without tidal forcing by looking at it on the system

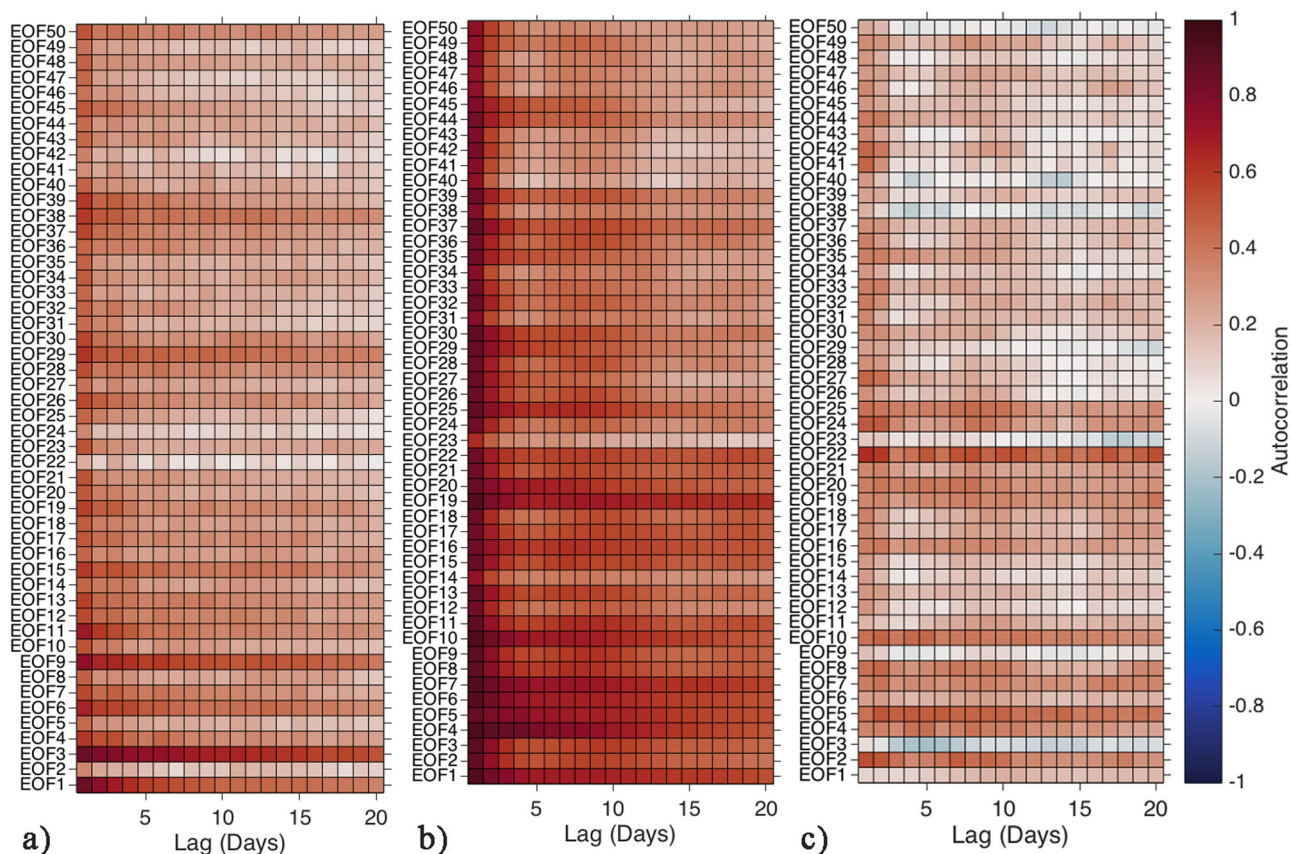


Fig. 5 Autocorrelation functions of with- and without- tidal forcing for lags 1-20 and PCs of EOFs 1-50. **a** With tidal forcing; **b** without tidal forcing (the middle panel); and **c** differences of autocorrelation functions of with tidal forcing ensemble minus without tidal forcing ensemble.

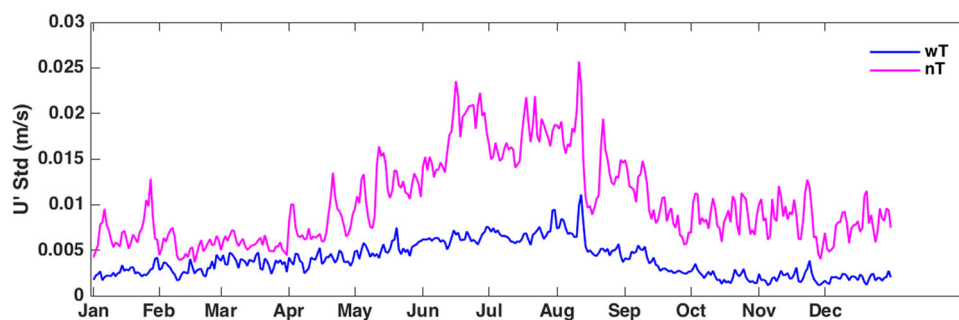


Fig. 6 The time variation of the spatial-mean internal variability intensities. The pink and blue lines are for with-tide and no-tide ensembles, respectively.

level, and to explain why tidal forcing depresses the generation of internal variability. Up to this point, the SCM⁴ has been employed to comprehend recorded spectra^{26, 27}. Over time, the idea has been broadened to encompass additional degrees of freedom to distinguish and define patterns of variability.

Here, we are not examining the specific processes involved in the hydrodynamics of a marginal sea. Instead, we are studying the system properties, specifically the scale-dependent memory, to understand a change in the system's statistics.

Our study yields two primary outcomes. Firstly, we find indications of an underlying low-dimensional dynamical core with relatively sluggish fluctuations, which facilitates the transfer of short-term internal variability to longer timescales, in the Bohai and Yellow Sea. This aligns with Hasselmann's concepts of both the Stochastic Climate model (SCM) and PIPS-and-POPs (see the Methods section). Second, as predicted by the SCM, a modification in memory is associated with a corresponding adjustment in

the long-term variance, particularly on a large-scale basis. We find that active tides, as well as winter conditions, limit the generation of low-frequency internal variability by decreasing the memory of the system.

While these are general results, on the dynamics of internal variability of the relatively shallow marginal seas of the Bohai and Yellow Sea, we are also able to explain the result of the numerical experiment. Specifically, we find that the incorporation of tides reduces the system's memory, which could be due to more efficient damping or a decrease in baroclinic instabilities. While previous work suggested that the latter is the dominant or at least an important factor¹⁷, it is not to say that changes in damping are irrelevant. We plan to explore this further in a separate paper.

The primary finding of the Stochastic Climate Model is the recognition that all hydrodynamic systems, including the atmosphere, the ocean, or the climate system, generate variations on all temporal and spatial scales that are not a response to external

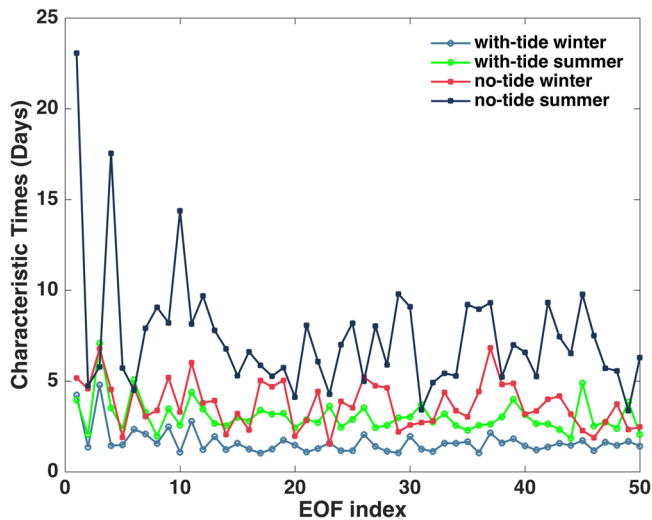


Fig. 7 The characteristic time scales for the first 50 EOF in August (summer) and February (winter). The solid lines with squares are for no-tide ensembles; solid lines with circles are for with-tide ensembles.

drivers but are internally generated. Our analysis affirms this general observation and demonstrates that tides limit this generation by decreasing the system's memory in the hydrodynamics of a marginal sea. This inclination to produce spontaneous variability implies that numerical experiments evaluating scenarios of climate change require a statistical hypothesis test before detection of the effect of an altered configuration can be claimed²⁸.

Methods

Ocean model setup. The Finite-Volume Community Ocean Model (FVCOM) was adopted in this paper^{29,30}. The simulations had a horizontal resolution of about 4 km and 8 km for Bohai and the Yellow Sea, respectively, with 30 sigma layers vertically. The open boundary ranged from Qidong along the Chinese coast to the southern end of the Korean Peninsula, and the simulations were forced by surface forcing, consisting of wind, sea-level pressure, air temperature, precipitation, evaporation, relative humidity, and heat fluxes from the NCEP's Climate Forecast System Version 2. The tidal forcings, which included M_2 , S_2 , N_2 , K_2 , K_1 , O_1 , P_1 , Q_1 , were obtained from TPXO 8³¹. At the open boundary, the tidal forcing is considered by using a time-dependent elevation. The Huanghe, Huaihe, and Haihe are considered and the river discharges are from the "China Sediment Bulletin (2019)". To have a slightly different but generally consistent initial condition for each simulation within the ensemble, an independent 9-year climatological simulation was carried out. The model start time will be shown in the following ensemble design section.

Ensemble design. To explore the impact of tides on hydrodynamics, and the seasonal contrast in the Bohai and Yellow Sea, we considered the same set of two ensembles of numerical simulations: one set included tidal forcing, while the other did not, but both sets had the same other forcing inputs. Each ensemble consisted of 4 simulations of the year 2019, initialized at different months earlier (The model time setup is sketched in Fig. 8). The initial conditions of the four simulations in each

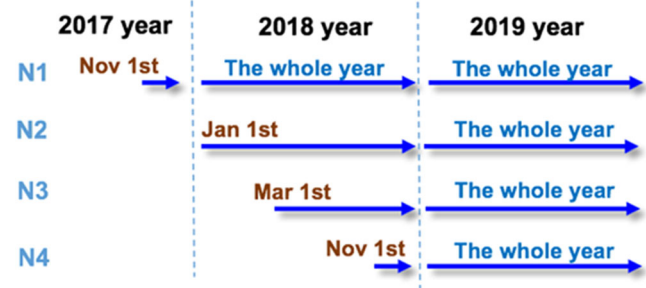


Fig. 8 Schematic diagram of the numerical simulation setup. N1, N2, N3, and N4 are the first, second, third, and fourth members in the ensemble, respectively.

ensemble were taken from Nov. 1st of the 7th year, Jan 1st of the 8th year, Mar. 1st of the 8th year, and Nov. 1st of the 8th year of the separate 9-year climatological simulation.

In the no-tide simulations, the tidal elevations at the open boundary were turned off. The results showed that the activation of tidal forcing led to a substantial reduction in internal variability at large and medium scales but not at small scales²³.

Scale separation based on EOF analysis. We utilized an alternative orthogonal linear basis—the EOFs basis—to describe the fields of depth-averaged velocity. This method, introduced by Tang et al.⁷ and further developed by Lin et al.²³, allowed us to decompose the fields according to both spatial and temporal scales. As we focused on the component of internal variability, which was determined by the difference between ensemble members and the ensemble means, we retrieved EOFs from the daily mean depth-averaged velocity anomalies data (after having subtracted daily ensemble means) of all members of the tide ensemble and no-tide ensemble in 2019. Note that we combined the tide ensemble and no-tide ensemble for EOF analysis, to make sure that the projection is based on identical basic functions and the scales of the same EOF index are identical, representing the same variance. If we do EOF separately for with and without tidal forcing ensembles, there will be some bias because the result will be an optimal decomposition for only the with tidal forcing or only the without tidal forcing ensemble.

To gain the scales involved for the whole 8 simulations from ensembles with and without tidal forcing, the EOF analysis was conducted as follows:

$$B(x, y, t) = \sum_{j=1}^p Q_j(t) e_j(x, y) \quad (2)$$

where B represents the depth-averaged velocity anomalies across all 8 simulations, t is the days in 8 simulations, and x and y represent the regular grid points, after interpolation from the unstructured grid points from FVCOM original model results, in zonal and meridional directions; $e_j(x, y)$ is the j th EOF, and $Q_j(t)$ are the j th principle components, and j represents the EOF index. For each simulation, there are 365 days, so these resulted in 2920 non-zero orthogonal eigenvectors, of which the last one was discarded due to lack of accuracy. Thus, we have 2919 non-zero orthogonal eigenvectors in the end ($p = 2919$).

Using a geostatistical method⁷, we determined a spatial scale for each EOF pattern. Such geostatistical method, which is based on spatial autocorrelation, is computed as

$$c_j(k) = \frac{\sum_{(x,y) \in G^{kz}} [e_j(x + k\Delta, y) - \bar{e}_j] [e_j(x, y) - \bar{e}_j] + \sum_{(x,y) \in G^{km}} [e_j(x, y + k\Delta) - \bar{e}_j] [e_j(x, y) - \bar{e}_j]}{\sum_{(x,y) \in G} [e_j(x, y) - \bar{e}_j]^2 / |G|} \quad (3)$$

where G represents all the grid points within the model domain; $e_j(x + k\Delta, y)$ and $e_j(x, y + k\Delta)$ are $e_j(x, y)$ at lag k grid points in spatial in zonal and meridional direction on the condition that both grid points (x, y) and $(x + k\Delta, y)$ (or $(x, y + k\Delta)$) are in the model domain G . $G^{kz} = \{(x, y) \in G | (x + k\Delta, y) \in G\}$ includes all grid points with k grid points to the right in zonal direction, G^{km} is the same as G^{kz} , but in the meridional direction. $\bar{e}_j$ denotes the spatial average of the j th EOF. $|G|$, $|G^{kz}|$, $|G^{km}|$ are the numbers of grid points in G , G^{kz} , and G^{km} .

We calculate a spatial autocorrelation function $c_j(k)$ varied with spatial lag k grid points for each EOF. To determine the scale of each EOF, we chose a threshold of 0.14. When the spatial autocorrelation function $c_j(k)$ is < 0.14 , the corresponding scale length is set to be that length for the j th EOF.

Fitting method. For Figs. 2, 3, and 4, the slopes of the spectra are estimated by a linear fit in the log-log presentation of the spectra.

The Stochastic Climate Model. Hasselmann's concept, of which he gave a summary in his Nobel speech (Klaus Hasselmann—Nobel Prize lecture. NobelPrize.org. Nobel Prize Outreach AB 2023. Tue. 4 Jul 2023. <https://www.nobelprize.org/prizes/physics/2021/hasselmann/lecture/>). He was confronted with two challenges, if low-frequency variability in the climate system is only reflecting low-frequency forcing, and if part of the very large phase space of the climate system may be considered as a significant core of the dynamics, while the rest may be parameterized stochastically. The answer to the first challenge is the Stochastic Climate Model⁴, and to the second is the concept of PIPs and POPs (Principal Interaction Patterns and Principal Oscillation Patterns)^{32,33}. In the following, we sketch these ideas, without going into mathematical detail. However, it may be good to visit the original publication, which gave Klaus Hasselmann his Nobel Prize in Physics in 2021.

Low-frequency variability in a system could be the integrated response of a linear (or nonlinear) system forced by short-term variations⁴. For doing so, he referred to the Brownian motion as a good analogy: the climate variables Y and weather variables X may be interpreted in the analogous particle picture as the (position and momentum) coordinates of a few large and slow, and many small and fast particles, respectively.

The short-term variations are treated as random. In short: the white noise of the short-term variation transforms into red noise variations. The naming noise refers to the understanding that these variations cannot be traced back to specific events, and are unprovoked by anything in the external forcing or details of the initial state. It is based on the separation of time scales and on linearization not violating the nonlinearities of the system too strongly.

To formulate, a system can be divided into two sub-systems $Z = X + Y$ with strongly different characteristic time scales. X goes with a short timescale τ_x and Y with a long timescale τ_y . Thus, X goes with short-term variations, while Y with long-term variations. For Y we have a differential equation

$$\frac{dY}{dt} = \langle V(X, Y) \rangle + f \quad (4)$$

where $\langle \dots \rangle$ represents the time mean across τ_y , when Y varies little, or an average of trajectories; V represents the dynamical link between X and Y ; f is the external forcing. Since we consider the dynamics of internal variability, we set $f = 0$. If a linearization is permitted, then $\langle V(X, Y) \rangle = -\beta Y + \varepsilon_t$, is a suitable approximation, because of the time-separation of X and Y : X varies on short timescales $\tau_x \ll \tau_y$ and has the effect of a white noise

random variable ε_t after averaging across τ_y . After discretization, we arrive at

$$y_{t+1} = \alpha y_t + \varepsilon_t \text{ with } \alpha = 1 - \beta \quad (5)$$

assuming a time step of 1 for simplicity. With y representing a state variable, y_{t+1} and y_t mean the value of y at the time step $t + 1$ and t , respectively, and ε_t represents external short-term random fluctuations (white noise)²⁵. This is an autoregressive process of first-order (usually called by statisticians). Since the system is assumed to behave stationary, $\alpha < 1$ is required.

The (discretized) Ornstein-Uhlenbeck format of Eq. (5) (usually called by physicists) is

$$\frac{dy}{dt} = \lambda y + \varepsilon_t \quad (6)$$

While ε_t represents external short-term random fluctuations, which has a white spectrum, with equal variance across all scales, y is a stationary variable, which has a red spectrum, with the largest variance for long time scales and the smallest variance with short time scales, λy is a feedback term.

In this simplest form, the spectrum of X is white, i.e., uniform across all scales; in practical situations, it may also be red, or a mix of red processes. In all cases, the spectrum of Y is redder than that of X , i.e., the process of accumulating short-term variations leads to enhanced variability of Y on longer time scales.

The increase of the intensity of the spectrum at the longest time scales of Y is given by α , which may be understood as a memory. An α close to 1 means that once an anomaly has formed, this will persist for a long time, while an α close to zero will be damped out quickly.

The relation between λ in Ornstein-Uhlenbeck process expression and α in the auto-regressive process of the first-order is $1 - |\lambda| = \alpha$. If $\lambda = -0.1$, then $\alpha = 0.9$; if $\lambda = -0.3$, then $\alpha = 0.7$, if we, for simplicity, use a time step of unitless 1 in the discretization.

In the following and Result section, we keep the discussion in the framework of the first-order auto-regressive process, according to which the value of a state variable y_{t+1} at time step $t + 1$ is determined by a memory term (αy_t) and short-term external random fluctuations (ε_t). Then, $(1 - \alpha)y_t$ is a typical feedback term. Stationarity requires $|\alpha| < 1$, so that αy represents negative feedback. The value of α determines the memory of the system, with a large α indicating an extended memory (slowly decaying autocorrelation), and a small α a short memory. With $\alpha = 1$, y is a random walk, and with $\alpha = 0$, y is white noise. As the memory of the system increases, the spectrum of the system becomes redder meaning that the relative increase in variance at long time scales compared to short time scales increase. This increase in memory is a key factor in ocean dynamics and can impact predictability at time scales beyond weather forecasts²⁴.

White noise is an infinite sequence of zero mean independent normal random. It means, for each t , element x_t does not depend on the state at any other time. Thus, the autocovariance function is zero for all lags.

The spectra of the 1st-order auto-regressive process with α_1 (lag-1 correlation coefficient) is

$$\Gamma(\omega) \approx \frac{\sigma_z^2}{1 + \alpha_1^2 - 2\alpha_1 + \alpha_1(2\pi\omega)^2} = \frac{c_1\sigma_z^2}{c_2 + \omega^2} \quad (7)$$

Thus, the spectra of 1st auto-regressive processes follow the linear gradient of -2 in log-log format for $\omega > 0$. When $\alpha_1 > 0$, the 'spectral peak' is located at the frequency $\omega = 0$. Such processes are usually called red noise processes.

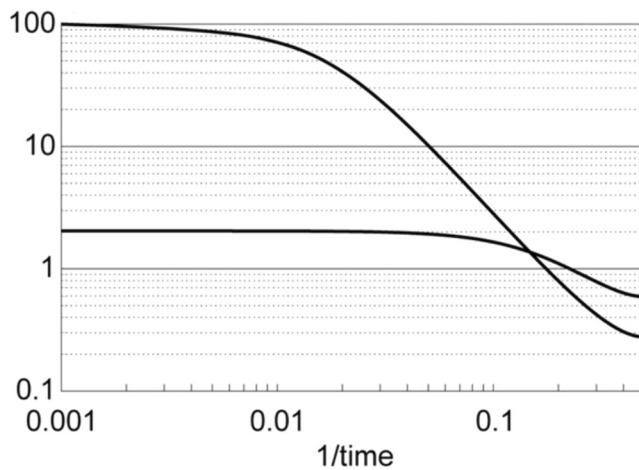


Fig. 9 Spectra of an autoregressive process of the first-order²⁵. The upper curve goes with a large $\alpha = 0.9$, the lower curve with $\alpha = 0.3$.

The spectrum of the white noise is a horizontal line, suggesting that no time scale of variation is preferred, therefore it is an allusion to white light.

Figure 9 shows two 1st auto-regressive process spectra in log-log format with two different α , $\alpha = 0.3, 0.9$. As mentioned before, the relative increase of variance at long time scales compared to that at short time scales depends on α : an α close to 1 has a large increase, and an α close to 0 has a small increase.

There are several methods available to quantify the memory of a system. A common approach is to use autocorrelation^{27, 34,35}. To measure the memory of a stochastic process, we can use the characteristic time, τ , which is defined as the limit of the ratio of sample size and its equivalent sample size²⁵.

For a 1st order auto-regressive process with parameter α , the time scale is $\tau = (1 + \alpha)/(1 - \alpha)$. If $\alpha = 0$, representing white noise, the process has a memory of 1 time step of the time discretization and is considered memoryless. As α approaches 1, the memory becomes stronger with larger τ , while negative values of α do not make sense for this definition.

PIPs and POPs. Originally, the SCM was in most cases used in univariate applications, i.e., Z , as well as X and Y were given by one variable. Of course, one could use several variables, so that Z , as well as X and Y , become vectors, and α a matrix. Then Y may be transformed into a linear basis of eigenvectors of the matrix α , and sorted according to their eigenvalues, which are all real or come in conjugate complex pairs, at least when α is estimated from data. Real eigenvectors go with a red spectrum; when the eigenvectors are conjugate complex, their spectra may be peaked spectra. When α is given by theory, then they are named normal modes; when they are estimated, they are named Principal Oscillation Patterns^{32,33}. The concept has been implemented in many applications³⁶, but because of the enormous size of the phase space of the atmospheric/climate dynamics, before the analysis a reduction of degrees of freedom is needed. A particularly impressive example is in the publication³⁷, which made the Southern Ocean wobble on time scales of 300 years, when disturbances, which were white in time and space, were added to their stationary freshwater flux forcings.

The significant dynamics of the Z -system for the problem considered take place within that POP-part of the phase space, which is spanned by the most inert eigenvectors. Compared to the original SCM analysis, this concept has the advantage that it may

be applied to a subspace of Z which is believed to contain the relevant long-term dynamics.

The POPs concept may be seen as a generalization of the SCM, but also as a simplification of the rather general concept of Principal Interaction Patterns (PIPs)³³, which asks for the determination of a few patterns, whose coefficients allow an optimal prediction of the slow variations of the considered system. Unfortunately, the PIP ansatz has not (yet) been successfully implemented, and thus serves more as a thought-guiding concept, namely of separation of the full system into a slowly varying core (Y), within a dynamical environment (X) with very many short-term fluctuations.

Data availability

The code and data used in the paper are available at: <https://doi.org/10.5281/zenodo.8301348>. The surface forcing data from the database NCEP's Climate Forecast System Version2 is downloaded from: <https://rda.ucar.edu/lookfordata/>. The TPXO 8 database is downloaded from: http://g.hyyb.org/archive/Tide/TPXO/TPXO_WEB/tpxo8_atlas.html.

Code availability

The source code of the used ocean model FVCOM is available at <https://github.com/FVCOM-GitHub/fvcom>. The Matlab code used in the paper is available at <https://doi.org/10.5281/zenodo.8301348>.

Received: 18 April 2023; Accepted: 20 September 2023;

Published online: 02 October 2023

References

- Ilıcak, M., Özgökmen, T. M., Peters, H., Baumert, H. Z. & Iskandarani, M. Very large eddy simulation of the red sea overflow. *Ocean Model.* **20**, 183–206 (2008).
- Penduff, T. et al. Chaotic variability of ocean heat content: climate-relevant features and observational implications. *Oceanogr.* **31**, 63–71 (2018).
- Penduff, T., Llovel, W., Close, S., Garcia-Gomez, I. & Leroux, S. Trends of coastal sea level between 1993 and 2015: imprints of atmospheric forcing and oceanic Chaos. *Surv. Geophys.* **40**, 1543–1562 (2019).
- Hasselmann, K. Stochastic climate models Part I. *Theory. Tellus* **28**, 473–485 (1976).
- Büchmann, B. & Sönderkvist, J. Internal variability of a 3-D ocean model. *Tellus* **68**, 30417 (2016).
- Geyer, B., Ludwig, T. & von Storch, H. Limits of reproducibility and hydrodynamic noise in atmospheric regional modelling. *Commun. Earth Environ.* **2**, 17 (2021).
- Tang, S., von Storch, H. & Chen, X. Atmospherically forced regional ocean simulations of the south China Sea: scale dependency of the signal-to-noise ratio. *J. Phys. Oceanogr.* **50**, 133–144 (2020).
- Weisse, R., Mikolajewicz, U., Sterl, A. & Drijfhout, S. S. Stochastically forced variability in the antarctic circumpolar current. *J. Geophys. Res.* **104**, 11049–11064 (1999).
- Chervin, R. M., Gates, W. L. & Schneider, S. H. The effect of time averaging on the noise level of climatological statistics generated by atmospheric general circulation models. *J. Atmos. Sci.* **31**, 2216–2219 (1974).
- Jochum, M. Internal variability of the tropical Pacific ocean. *Geophys. Res. Lett.* **31**, L14309 (2004).
- Jochum, M., Malanotte-Rizzoli, P. & Busalacchi, A. Tropical instability waves in the atlantic ocean. *Ocean Model.* **7**, 145–163 (2004).
- Mahajan, S., Zhang, R. & Delworth, T. L. Impact of the Atlantic Meridional Overturning Circulation (AMOC) on arctic surface air temperature and sea ice variability. *J. Clim.* **24**, 6573–6581 (2011).
- Kushnir, Y. Interdecadal variations in north atlantic sea surface temperature and associated atmospheric conditions. *J. Clim.* **7**, 141–157 (1994).
- Grégorio, S. et al. Intrinsic variability of the atlantic meridional overturning circulation at interannual-to-multidecadal time scales. *J. Phys. Oceanogr.* **45**, 1929–1946 (2015).
- Hirschi, J. J.-M. et al. Chaotic variability of the meridional overturning circulation on subannual to interannual timescales. *Ocean Sci.* **9**, 805–823 (2013).
- Penduff, T. et al. Sea level expression of intrinsic and forced ocean variabilities at interannual time scales. *J. Clim.* **24**, 5652–5670 (2011).

17. Waldman, R., Somot, S., Herrmann, M., Sevault, F. & Isachsen, P. E. On the chaotic variability of deep convection in the mediterranean sea. *Geophys. Res. Lett.* **45**, 2433–2443 (2018).
18. Zhang, M., von Storch, H., Chen, X., Wang, D. & Li, D. Temporal and spatial statistics of travelling eddy variability in the south China sea. *Ocean Dyn.* **69**, 879–898 (2019).
19. Choi, B. H. *A Tidal Model of the Yellow Sea and the Eastern China Sea* (Korea Ocean Research and Development Institute, 1980).
20. Choi, B. H., Eum, H. M. & Woo, S. B. A synchronously coupled tide–wave–surge model of the yellow sea. *Coast. Eng.* **47**, 381–398 (2003).
21. Fang, G. Empirical cotidal charts of the bohai, yellow, and east China seas from 10 years of Topex/poseidon altimetry. *J. Geophys. Res.* **109**, C11006 (2004).
22. Larsen, L. H., Cannon, G. A. & Choi, B. H. East China sea tide currents. *Cont. Shelf Res.* **4**, 77–103 (1985).
23. Lin, L. et al. The effect of tides on internal variability in the bohai and yellow Sea. *Dyn. Atmos. Oceans* **98**, 101301 (2022).
24. Shi, H. et al. Global decline in ocean memory over the 21st century. *Sci. Adv.* **8**, eabm3468 (2022).
25. von Storch, H. & Zwiers, F. W. *Statistical analysis in climate research*. (Cambridge University Press, 1999).
26. Lemke, P., Trinkl, E. W. & Hasselmann, K. Stochastic dynamic analysis of polar sea ice variability. *J. Phys. Oceanogr.* **10**, 2100–2120 (1980).
27. Frankignoul, C. & Hasselmann, K. Stochastic climate models, part II application to sea-surface temperature anomalies and thermocline variability. *Tellus A Dyn. Meteorol. Oceanogr.* **29**, 289 (1977).
28. Yuan, Q., Thorarindottir, T. L., Beldring, S., Wong, W. K. & Xu, C.-Y. Assessing uncertainty in hydrological projections arising from local-scale internal variability of climate. *J. Hydrol.* **620**, 129415 (2023).
29. Ding, Y. et al. Observational and model studies of synoptic current fluctuations in the bohai strait on the chinese continental shelf. *Ocean Dyn.* **69**, 323–351 (2019).
30. Ding, Y. et al. Modeling the westward transversal current in the southern yellow sea entrance: a case study in winter 2007. *Ocean Dyn.* **70**, 803–825 (2020).
31. Egbert, G. D. & Erofeeva, S. Y. Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Oceanic Technol.* **19**, 183–204 (2002).
32. Von Storch, H., Bruns, T., Fischer-Bruns, I. & Hasselmann, K. Principal oscillation pattern analysis of the 30- to 60-day oscillation in general circulation model equatorial troposphere. *J. Geophys. Res.* **93**, 11022 (1988).
33. Hasselmann, K. PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.* **93**, 11015 (1988).
34. Namias, J. & Born, R. M. Temporal coherence in north Pacific sea-surface temperature patterns. *J. Geophys. Res.* **75**, 5952–5955 (1970).
35. Srivastava, A. & DelSole, T. Decadal predictability without ocean dynamics. *Proc. Natl Acad. Sci. USA* **114**, 2177–2182 (2017).
36. Von Storch, H., Bürger, G., Schnur, R. & Von Storch, J.-S. Principal Oscillation patterns: a review. *J. Clim.* **8**, 377–400 (1995).
37. Mikolajewicz, U. & Maier-Reimer, E. Internal secular variability in an ocean general circulation model. *Clim. Dyn.* **4**, 145–156 (1990).

Acknowledgements

The authors extend their appreciation to the China Scholarship Council for enabling collaboration and communication. Additionally, they give recognition to the German

Climate Center (DKRZ) for granting access to computer resources. The authors are grateful for the fruitful discussions held with Florian Noethen from the University of Hamburg, Valerio Lucarini from the University of Reading, Thierry Penduff from Université Grenoble Alpes, Tang shengquan from ocean university of China and other individuals on diverse scientific topics such as the influence of tides on noise and the definition of signal and noise. Lastly, Hans von Storch expresses his gratitude to Klaus Hasselmann for his unwavering and patient guidance in statistical analysis of systems.

Author contributions

L.L. performed the model simulations, did the stochastic analysis, and wrote the manuscript. Hans von Storch wrote the manuscript and designed the statistical analysis of the model output. Xueen Chen supervised the FVCOM settings, established the cooperation and gave suggestions on the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43247-023-01018-7>.

Correspondence and requests for materials should be addressed to Hans von Storch.

Peer review information *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Kyung-Sook Yun, Clare Davis, Heike Langenberg. A peer review file is available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023