Chin. Phys. B Vol. 29, No. 10 (2020) 108901

Modeling and analysis of the ocean dynamic with Gaussian complex network*

Xin Sun(孙鑫)[1], Yongbo Yu(于勇波)[1], Yuting Yang(杨玉婷)[1], Junyu Dong(董军宇)[1,2,†], Christian Böhm[3], and Xueen Chen(陈学恩)[4]

[1] Department of Computer Science and Technology, Ocean University of China, Qingdao 266000, China
[2] Frontiers Science Center for Deep Ocean Multispheres and Earth System, Qingdao 266000, China
[3] Institut für Informatik, Ludwig Maximilian University of Munich, Munich 80331-81929, Germany
[4] College of Physical and Environmental Oceanography, Ocean University of China, Qingdao 266000, China

(Received 17 April 2020; revised manuscript received 19 June 2020; accepted manuscript online 3 July 2020)

The techniques for oceanographic observation have made great progress in both space-time coverage and quality, which make the observation data present some characteristics of big data. We explore the essence of global ocean dynamic via constructing a complex network with regard to sea surface temperature. The global ocean is divided into discrete regions to represent the nodes of the network. To understand the ocean dynamic behavior, we introduce the Gaussian mixture models to describe the nodes as limit-cycle oscillators. The interacting dynamical oscillators form the complex network that simulates the ocean as a stochastic system. Gaussian probability matching is suggested to measure the behavior similarity of regions. Complex network statistical characteristics of the network are analyzed in terms of degree distribution, clustering coefficient and betweenness. Experimental results show a pronounced sensitivity of network characteristics to the climatic anomaly in the oceanic circulation. Particularly, the betweenness reveals the main pathways to transfer thermal energy of El Niño–Southern oscillation. Our works provide new insights into the physical processes of ocean dynamic, as well as climate changes and ocean anomalies.

Keywords: complex networks, ocean dynamic, Gaussian mixture model, physical processes

PACS: 89.75.Fb, 05.45.Tp, 64.60.aq **DOI:** 10.1088/1674-1056/aba27d

1. Introduction

Data analysis has always been one of the most important foundations in marine science research. There are multitudinous ways to collect a variety of marine hydrological observation data in the area of marine and meteorological science,[1] such as sea surface temperature, sea surface wind, waves and currents. It is wonderful to know what we can get from these observation data. Data analysis and mining could provide us some insight information and patterns for discovering and predicting the ocean phenomenon and climate abnormities. Recent years, the techniques for oceanographic observation have made great progress in both spatial/temporal coverage and quality.[2] Nowadays, ocean observation data are acquired and stored at an unprecedented scale and speed,[3] which presents some characteristics of big data and poses a great challenge for data analysis.[4] A pivotal problem of applications of computer science for marine science research is how to effectively extract valuable knowledge from such a huge amount of observation data.[5–7]

One classical research topic in the artificial intelligence, i.e., complex network analysis, recently has attracted more and more attention to analyze the complex big data. It gives new perspective of understanding the massive data by network topology modeling and analysis.[13] The modelled network is a complex system composed of nodes and edges. Currently, there are extensive applications of complex network theory due to its universality. More and more interdisciplinary research between complex network and various traditional fields have been seen with significant success. Can complex network analysis methods be used in the areas of environment data analysis, such as global climate and ocean dynamic? Encouragingly, research in terms of introducing the complex theory into climate areas has already been carried out. Climate plays an important role to the survival and development of human beings, therefore it always draws huge attention of researchers from different areas. As a tool for data analysis, complex network has made great achievements among climatology field in recent years. The relationship between dynamic characteristics and topological structure of the climate network has been widely studied, especially the synchronization of climate phenomena has become a hot research topic.[8,9]

Ocean dynamic is highly related and influences the global climate change. Exploring and understanding the ocean energy and mass exchange are important research issues about ocean and climate. Complex network separates the complex system into a plenty of dynamical subsystems, which can be represented by a limit-cycle oscillator. Meanwhile, the interaction of oscillators is considered as the edge of network.

*Project supported by the National Natural Science Foundation of China (Grant Nos. U1706218, 61971388, and L1824025).

†Corresponding author. E-mail: dongjunyu@ouc.edu.cn

This work focuses on modeling the dynamic ocean system as a complex network to investigate its statistical patterns. We investigate global ocean dynamic via constructing the complex network with regard to sea surface temperature. The global ocean is divided into discrete regions to represent the nodes of the network. In probability theory, the central limit theorem (CLT) establishes that the distribution of most natural data conforms to the Gaussian distribution. To model the random and dynamic of the ocean physical process, we introduce the Gaussian mixture models to describe the nodes as limit-cycle oscillators. What's more, the ocean can be simulated as a stochastic system by forming a complex network with the interacting among these dynamical oscillators. The interacting behavior can be regarded as the edge of the network. Gaussian probability matching is suggested to measure the interacting behavior, i.e., similarity of regions. The statistical characteristics of the ocean network model are investigated in terms of degree distribution, clustering coefficient and betweenness. From the experimental results, we give a preliminary analysis for ocean phenomena with the complex network statistical characteristics. It provides new insights into the physical processes of ocean dynamic, as well as climate changes and ocean anomalies.

2. Related work

Complex networks have been frequently used to represent relationships between entities in many complex systems, and have achieved fruitful results in climatological application. Donges *et al.*[10,11] carried out the pioneering works of introducing the complex network theories and technologies to explore characteristics of spatial and temporal climate data. They constructed a complex climate network model to investigate characteristics of global climate, and pointed out the scale-free properties of the earth's climate system. Charakopoulos *et al.*[12] examined the performance of two different approaches for identification of underlying patterns and understanding complex characteristics of climate dynamics among atmospheric and oceanic variables. Their research findings suggested that the network characteristics reveal significant information and are strongly dependent on the underlying system dynamics. Moreover, their methodologies can also be applied reliably in spatiotemporal pattern identifications and data classifications among global climate observations. Steinhaeuser *et al.*[14] proposed six-dimensional features considering the relationship among temperature, pressure, humidity and precipitation to establish a network model. They further explored different function areas with similar climatic characteristics in 2010. Alexander *et al.*[15] founded that complex networks meet the need for innovation of climate research facing quickly increasing data volumes which were produced by growing observational networks and model inter-comparison exercises. Furthermore, higher-order complex network measures may contain complementary statistical information that is invisible to other methods like EOF analysis. To precisely forecast El Niño events, Meng *et al.*[16] approached a percolation framework based on time-delayed cross correlation as an alarm which forecasts 7 out of 10 El Niño events from 1980 and 2016. Kurths *et al.*[17] found a unique wave structure associated with ocean currents related to the transmission of high energy flow. Boers *et al.*[18,19] identified the temporal and spatial characteristics of extreme precipitation events in the South American monsoon climate region. Ludescher *et al.*[20] forecasted the next El Niña phenomenon ahead of a year's time. Tsonis *et al.*[21] firstly used degree centrality to identify supernodes and associated them with teleconnection patterns in the atmosphere. Furthermore, they found that the climate network shows "small-world" properties due to long-range connections existing in the global climate. The complex network model of global climate provides clues about the collective dynamics of the Earth's climate system.[22] For instance, researchers have found that the change of climate network topology can be used for predicting the El Niño events.[23,24] Iglesias *et al.*[25] developed a multitask deep fully connected neural network trained on historical time series data for heat waves prediction. Climate research is an interdisciplinary science that offers many exciting opportunities and challenges for physicists.[26] From the perspective of complex network theory, the above researches help us get a good illuminating insight into the climate.[27] Wang *et al.*[28] constructed a novel ocean observation complex network (OOCN) with a multilayer structure, based on the continuous observation data, which obtains the hierarchical structure of the mesoscale eddy. Feng *et al.*[29] presented new network based measures of stability of the Pacific climate state. Their studies indicated that climate network based properties can be very useful analysis tools in ENSO dynamics and prediction. Another recent research[30] constructed a network relying on the linear Pearson correlation and evolving the network transitivity as a parameter to distinguish two types of climate, i.e., El Niño and La Niña. The work confirmed the classification of years that all references have in common and provide a discrimination for those years that have been so far ambiguously defined.

However, linear approaches cannot bring out some properties in nonlinear dynamics (i.e., EOF analysis and Pearson correlation). Tsonis *et al.*[31] argued that scale-free phenomena which are associated with nonlinear dynamics cannot be brought out by linear approaches, such as EOF analysis. They also pointed out a limitation that nonlinear correlation measure like mutual information's accurate estimation requires much more data beyond available. Zerenner *et al.*[32] indicated that the climate system does not consist of a structure with identifiable nodes. Hence, they interpreted observation data in ocean fields as the realizations of Gaussian random fields (GRFs) in both grid point and spectral space. Their network showed ordered structure and nodes only connect to first or second order neighbors, which reached a conclusion that multivariate climate variable should be taken into account when constructing the climate network. However, their networks still derived by linear approaches, i.e. Pearson correlation coefficients and partial correlations.

Therefore, it is a tricky task to seek a way that can better discover the characteristics of a nonlinear system. For this purpose, we use the uncertain probability model, i.e., a Gaussian mixture model (GMM), for nonlinear approximation of ocean hydrological dynamics. Meanwhile the oceanographic satellites provide decades of sea surface temperature (SST) data,[33] which gives us opportunity to approximate the nonlinear dynamics by GMMs. This work presents a frontier research on pattern analysis and discovery of ocean dynamics.

3. The data – sea surface temperature

Sea surface temperature (SST) is the water temperature close to the ocean surface. SST is one of the first oceanographic variables to be measured as a most important characteristic of seawater. It plays an important role in the process of interaction between sea surface and atmosphere, and affects the behavior of the Earth's atmosphere and global climate. The formation of some special ocean phenomena is related with the change of sea surface temperature, such as tropical cyclogenesis and ocean front. A variety of techniques can be utilized to measure this vital parameter, of which weather satellites have the ability to provide the SST both in a synoptic view of the ocean and in a high frequency of repeat views.

The Global Ocean Data Assimilation Experiment high-resolution sea surface temperature pilot project[34] provides a new generation of global high resolution SST data to the operational oceanographic, meteorological, climate and general scientific community. The sea surface temperature data used in this work comes from this project, and downloaded from NOAA (National Oceanic and Atmospheric Administration) who implements its data stewardship and reanalysis facility. To construct the complex network model of the ocean temperature dynamic, we investigate the daily global SST data during the years of 2010, 2014 and 2015. El Niño and La Niña are two complex weather patterns resulting from variations in ocean temperatures. As known to all, La Niña phenomena occurred in 2010, which contributed to extreme weather around the globe during the first half of the next year. Meanwhile, the El Niño event which occurred in September 2014 strongly influenced on the 2015 El Niño phenomenon. That El Niño can be regarded as one of the strongest on record and made the year 2015 to be the hottest year since 1998. Therefore, we pay more attention to the SST data of those three year.

4. The complex network model of the ocean

At the beginning, some definitions of the network model concerned in this work are given in the following.

The topology of the complex network can be defined as $N = (V, E, f)$. The set $V = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes (vertices), where n is the number of vertices. $E = \{e_{ij}, \ldots\}$ is the set of edges, where e_{ij} denotes the edge from node v_i to v_j. A weight $w_{ij} \geq 0$ is attached to each edge e_{ij}, and $w_{ij} = 0$ if v_i to v_j are not directly connected. For modeling, we consider only undirected networks where $e_{ij} = e_{ji}$ and $w_{ij} = w_{ji}$, employ a mapping function f to calculate the w_{ij} between the node v_i and v_j.

A well-defined complex network model can show some amazing discoveries from the trivial data. In this section, we will give the node and edge definitions of our ocean network. Specifically, we propose the Gaussian approach to model each node as dynamical oscillators, in order to reflect the stochastic behavior of the ocean.

4.1. The node-subsystem of the global ocean

Generally, a node of an ocean network represents a special region of the ocean. The best way to define the region should be realized by oceanographers according to the marine function and the geological structure. However, the ocean is dynamically changing, which results in hundreds of unpredictable ocean phenomena. The region can hardly be found and located by its function. Climatologists simply arrange the global climate data on a grid with a given resolution.[30,31] For example, Tsonis *et al.*[31] chose a resolution of 5° latitude × 5° longitude, and Wiedermann *et al.*[30] designated the spatial resolution of their study in 2.5° longitudinal and latitudinal. In our work, the global ocean is divided into discrete regions in the form of grids according to the latitude and longitude. The remote sensing data of the sea surface temperature approximately homogeneously cover the earth in 3600 pixels × 7200 pixels. We divide the global data into 90×180 grids where each grid is arranged with a resolution of 2° latitude × 2° longitude. Each grid represents a node of network. The reason of choosing the resolution of 2° latitude × 2° longitude as the size of grid is that ocean mesoscale and sub-mesoscale phenomena could be encircled. By removing the continent and islands, 11769 nodes are created for our ocean network model.

Each grid contains the observation data from 1600 geographic coordinate points, which raises a crucial question how to express the property of a node from the nonlinear voluminous data. Traditionally, researchers can directly calculate the average value of the data in certain geographic region as the feature of the node.[15,30,31] However, simply taking the average of the observations of the region to simulate the dynamical subsystem is not justified statistically. The average would not be an elaborate feature of a region (node) due to dynamical behavior of the sea surface temperature. In probability theory, the central limit theorem states that the distribution of most natural data conforms to the Gaussian distribution. To model the random and dynamic of the natural fluctuations in the ocean temperature, we suggest the Gaussian mixture models (GMMs) to describe the nodes as limit-cycle oscillators. GMMs have outstanding performance in numerous research fields because of the talent of approximating the distribution of many complex probabilities. In the pivotal part of this work, we interpret the node as realizations of Gaussian mixture models to contain the dynamic of the global ocean.

Regardless of knowing which subpopulation a data point belongs to, mixture models allow the model to learn the subpopulations automatically. Generally, a Gaussian mixture model consists of K components, which are parameterized by the mixture component weights, component means and variances/covariances. In a univariate case, there is only one Gaussian distribution in the model denoted by mean μ and variance σ^2. The model also called the single Gaussian model. Its probability density function can be written as

$$p_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{1}$$

The single Gaussian model is also known as the normal distribution. When K is greater than 1, the k-th component is parameterized by mixture component weights φ_k, a mean of μ_k and covariance matrix of $\sum k$. The sum of weights converges to 1 so that the total probability distribution normalizes to 1. In this multivariate case, Gaussian mixture models read

$$p(x) = \sum_{i=1}^{K} \varphi_i N(x|\mu_i, \sigma_i), \tag{2}$$

$$N(x|\mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right), \tag{3}$$

$$\sum_{i=1}^{K} \varphi_i = 1. \tag{4}$$

The distribution of sea surface temperature can be greatly influenced by many ocean and climate behaviors including sea breeze, ocean current and illumination, which cause sea surface temperature to exhibit highly random and dynamical behavior. The distribution of SST may not completely converge to a Gaussian distribution. We thus apply the trivariate model to fit the distribution of SST.

4.2. The edge-interaction of subsystems

The edge of the network is used to connect a pair of nodes and indicates a certain relationship between them. In general, an edge of the ocean network represents the similarity of the temperature dynamic between two ocean regions. In this study, we investigate the ocean network using the undirected network model. The undirected network only has at most one edge between any pair of nodes and self-loops are not allowed. The adjacency matrix can be written as

$$A = (a_{ij}) = \begin{cases} 1, & \text{if} \quad e_{ij} \in E, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where E is the edge set of the network and A_{ij} indicates whether connection between node v_i and v_j exists or not. In this work, we argue that the ocean hydrological dynamics should be regarded as a stochastic system. Instead, the nodes should be modelled as Gaussian distributions. Therefore, we introduce the Gaussian matching probability as the mapping function f to measure the similarity between two nodes.

Matching probability (MP) is a kind of similarity measures with closed-form expressions for GMMs.[35] Single Gaussian matching probability measures the probability of a Gaussian distribution P belonging to the other one named as Q, which can be comprehended as the similarity of two independently distributed Gaussian models. The similarity can be calculated as joint probability of two Gaussian distributions[35]

$$p(p|q) = \int_{-\infty}^{+\infty} N_{\mu_q,\sigma_p}(x) \cdot N_{\mu_q,\sigma_q}(x)\mathrm{d}x, \tag{6}$$

where q and p are the probability density functions of Gaussian models P and Q respectively. Formula (6) can be further simplified to

$$p(p|q) = N_{\mu_q,\sigma_p+\sigma_q}(\mu_p). \tag{7}$$

Mix-Gaussian matching probability measures the similarity between two independent GMMs P and Q with the same number of Gaussian distribution. According to previous description, the similarity between two Gaussian distributions is calculated as the joint probability of them. Inherit the thought of single Gaussian probability matching, the similarity between two independent GMMs can be counted as the product of joint probabilities between every two Gaussian components in disparate GMMs. Therefore, the formula of similarity between two independent GMMs P and Q can be written as

$$P(P|Q) = \prod_{i=1}^{K}\prod_{j=1}^{K}\prod_{n=1}^{D} p(p_i|q_j), \tag{8}$$

where q_i and p_j are i-th and j-th Gaussian distribution in Q and P, respectively. D represents the dimension of time series on each node.

Exhaustively calculating the similarity between two GMMs is intractable since there are multiple components that lie in GMMs. Multiple multiplication will cause floating-point underflow. Thus, the theorem of likelihood estimation can give us help. This theorem allows us to avoid computational floating-point underflow by taking the logarithm of a polynomial product. Consequently, the mix-Gaussian matching probability can be rewritten as

$$\log P(P|Q) = \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{n=1}^{D}\log p(p_i|q_j). \tag{9}$$

After taking the logarithm of a polynomial product, we can obtain the similarity between two GMMs from summating the joint probability of every two Gaussian distributions.

Gaussian distributions provide the opportunity of modelling the ocean as a stochastic system. Different regions are the nodes which display as dynamical oscillators. MP constructs the network edges by measuring the behavior similarity between two nodes.

4.3. Methodology from similarity to networks

Climatologists are committed to finding powerful analytical tool to study the statistical correlation of climatic factors in different geographic areas and simulate the dynamics of the climate system in a large scale (temporally and spatially). Techniques and theories of complex networks can gain a good understanding of the mutation and affiliation of a climate phenomenon in a statistical perspective.

Compared with previous studies, we introduce the Gaussian mixture models to achieve a nonlinear approximation of the SST data, and simulate the nodes as limit-cycle oscillators. Unlike the climate observations, ocean dynamical behaviors show much more stochastic characters due to the sea breeze, ocean current and illumination, etc. As shown in Fig. 1, we proposed to model the dynamical behavior of the ocean node as Gaussian distributions, and simulate the ocean as a stochastic system. After defining the node and edge, we construct the ocean network as a global similarity matrix of size 11769×11769. Then a threshold should be suggested as the basis for retaining or removing edges so as to form the adjacency matrix A. If the similarity between two nodes i and j higher than the threshold, the edge e_{ij} can be constructed as $A(i,j) = 1$, otherwise as $A(i,j) = 0$. Here the threshold is pivotal to network construction. We apply link density ρ to guarantee the statistical significance of the network model. For instance, $\rho = 0.01$ keeps the most pivotal links to emerge the architecture of the networks.

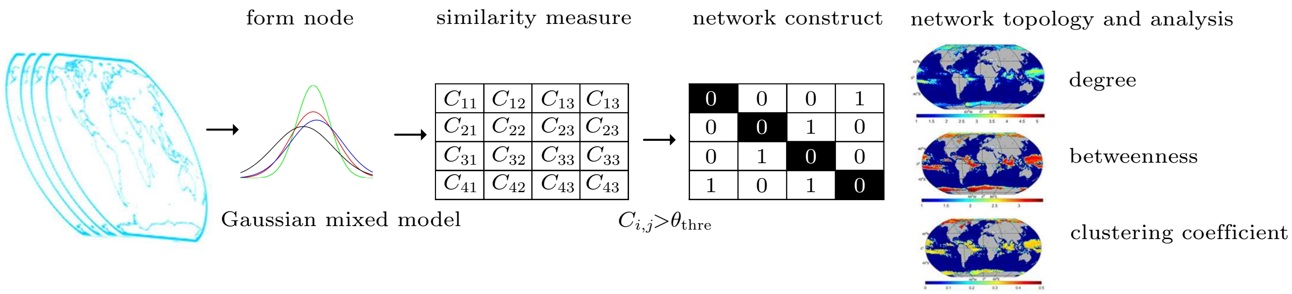


Fig. 1. Process of complex network modeling based on surface sea temperature (SST).

Finally, we can analyze network topology systematically in the complex network perspectives including degree distribution, clustering coefficient and betweenness.

5. Results and analyses

In this section, we analyze and explore the relationship between the topological characteristics of the climate network and the ocean phenomena using the characteristics of the complex network theory, such as degree distribution, aggregation coefficient and betweenness. We model the ocean network as a Gaussian stochastic system with probability matching including single Gaussian probability matching (SG-network) and mix-Gaussian probability matching (MG-network).

5.1. Degree distribution

The degree of node V_i refers to the number of neighbors that are directly connected to node V_i, which is denoted as K_i:

$$K_i = \sum_{j=1}^{N} C_{ij}, \tag{10}$$

where N is the total node number of the network and C_{ij} is the binary cross-correlation matrix. Intuitively, the higher the degree of the node is, the more important the node is in the network system.[36] The degree distribution provides a macro perspective to study the correlation between the individuals in a network and presents the nature of the network. To get an intuitive view of the distribution, we investigate it in a scatter diagram and global map combining geographical location.

Figure 2 shows the degree distribution of ocean networks under different link densities in the scatter diagram. Apparently, although super-nodes have increased with link density, the degree distribution of the network does not change with the edge density. This means that the edge density cannot change the shape of the network topology. Moreover, the power-law properties of SG-network and MG-network are not evident. It is also worth noting that some super-nodes in such networks are very steady with the change of link density. Particularly, the distribution of SG-network is bimodal distribution that the distribution peaks at both ends. According to Galtung's classification system,[37] it can be classified as a U-type bimodal distribution. The bimodal distribution indicates that we can obtain two different groups of nodes in the network which may be resulted from the climate anomaly.

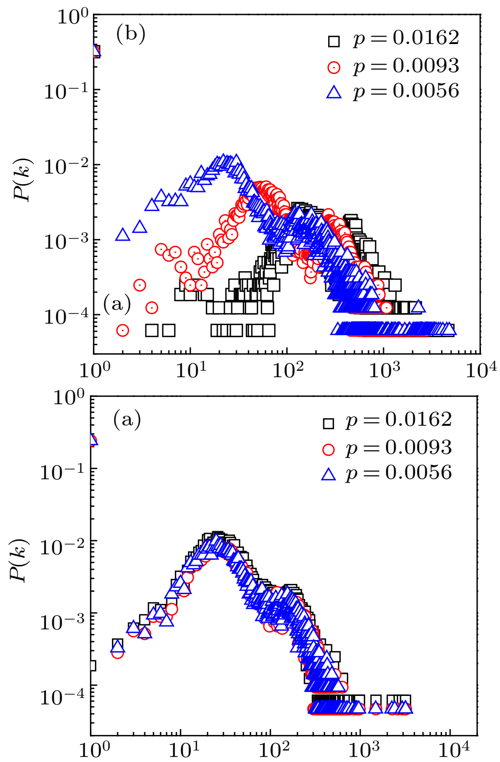


Fig. 2. Degree distribution of nodes in (left) SG-network (right) MG-network.

We investigate the degree distribution combined with geographic location to further understand the network structure. As shown in Fig. 3, the super-nodes are clustered around the tropics areas and high latitudes in southern and northern hemisphere. This means that certain nodes in the tropics and high latitudes in southern and northern hemisphere possess more connections than the rest. An intriguingly observation is that super-nodes may map some multiple-studied teleconnection patterns. The super-nodes in the equatorial Pacific and in the district around the dateline coincide with the well-known La Niña that happened in 2010, which is defined as CP-type La Niña.[30] In the northern hemisphere we also see that super-nodes are located in the north Atlantic, especially surrounded by Canada and close to the eastern United States and Iceland, which is consistent with the salient characters of the North Atlantic oscillation (NAO). It has been proved that there is evident association between two major inter-annual variabilities in the atmospheric circulation, which are the El Niño-Southern oscillation (ENSO) and the NAO.[38]

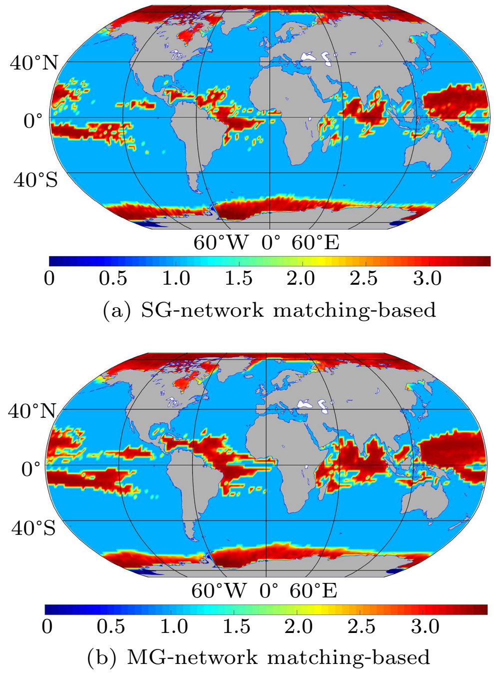


Fig. 3. Degree log$D$ distributions with geographic locations in the global ocean: (a) SG-network, (b) MG-network.

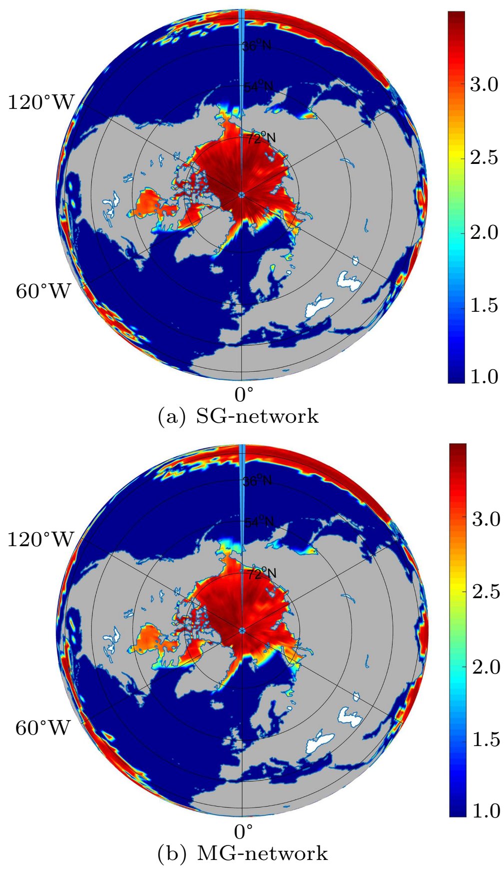


Fig. 4. Degree distributions in the northern hemisphere.

Atmospheric teleconnection is a subject in climatology field. The mathematical terminology of networks possesses compact connection with the physics of the complex system, though it is abstract. Thus, the complex network science is superior to delineate teleconnection patterns without ignoring long-range links and clustering information. In thermodynamics field, the changes of the degree of each node reflects the conjugated flows and the Onsager relations,[21] which give a magnificent answer to why even small perturbations of a parameter can induce fluctuations of other parameters. The limit-cycle oscillators which are represented by nodes in the network may synchronize or experience bifurcations when the amplitude or frequency varies. Therefore, the network statistical characteristic named degree can signify the climate synchronization of two long-distance regions. Figures 3(a) and 3(b) bring out two major teleconnection patterns according to the high-degree property of nodes. To reveal the feature of NAO meticulously, we further consider the degree distribution of network with only nodes in the northern hemisphere. From the viewing angle of 90°N as illustrated in Fig. 4, we intuitively find that locations of super-nodes in the network are consistent with the characters of NAO. Super-nodes are mostly clustered in the Arctic Ocean. Moreover, those super-nodes overlaid in the north Atlantic ocean should worth more attentions as it maps the affected area of NAO. The teleconnection between ENSO and NAO has been widely studied to protect ecological environment, e.g., water quality management by predicting total organic carbon load.[39]

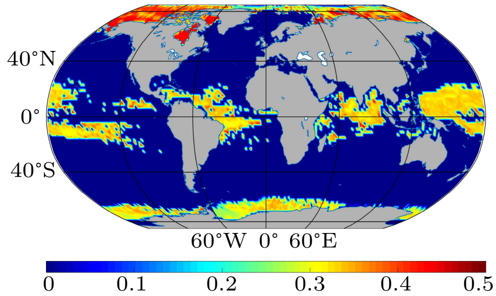
5.2. Clustering coefficient

Clustering coefficient is a measure of the degree which nodes in a graph tend to cluster together. The local clustering coefficient C_i of node V_i is the radio of the numbers of subsistent edges between V_i's neighboring nodes to the numbers of possible edges between V_i's neighbor nodes, which is given by

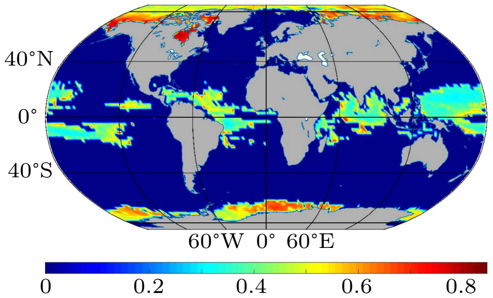
$$C_i = 2E_i/K_i(K_i - 1), \tag{11}$$

where K_i is the degree of V_i, and E_i represents the numbers of subsistent connections of V_i's neighbor nodes. It is evident that the range of clustering coefficients is [0,1], only in the situation that the network is connectionless, the clustering coefficients achieve the minimum value. It is worthy emphasized that in a random network, the clustering coefficient would naturally take a very low value but not come to 0. In the dynamic process of information transmission, clustering coefficients characterize the macro-control of network. The larger average aggregation coefficient of the network, the more robust of the network structure.

In Fig. 5, nodes of strong clustering coefficient are concentrated in the equatorial region, especially the Arctic ocean and the oceans around the Antarctica. The maximum clustering coefficients of two networks are orientated in the north Atlantic ocean, precisely in Hudson Bay beset by Canada. This observation means that the neighbors of nodes in Hudson Bay tend to be a clique, and testifies the synchronicity and relevance of climate in Hudson Bay. Though the maximum clustering coefficient of SG-network is 0.3 larger than the MG-network, the clustering coefficients of the two networks in the tropics are basically the same.



(a) SG-network matching-based



(b) MG-network matching-based

Fig. 5. Clustering coefficient distribution of the network model at geographic location.

5.3. Betweenness

Nodes in the network can play considerable roles in the process of energy transmission for the dynamic system. Assume that the energy is always propagated along the shortest path from the node V_i to the non-neighbor node V_t in the process of propagating. The betweenness of node can quantify the importance of the node to determine the node's role in energy transmission. The betweenness of node V_i is defined as

$$BC(i) = \sum_{i \neq s, i \neq t, s \neq t} \frac{l^i{}_{st}}{l_{st}}, \tag{12}$$

where l_{st} denotes the total number of the shortest paths between V_s and V_t, and $l^i{}_{st}$ the number of the paths which pass through V_i among all the shortest paths between V_s and V_t. High betweenness of the node indicates that the node can reach others on shortest paths, i.e., node of high betweenness lies on many shortest paths. If a node with a large betweenness value is removed, it will lengthen the paths between many pairs of nodes and reduce the energy transfer efficiency in network. As a high order characteristic of the network, betweenness is considered to be the most appropriate structural properties of networks to approximate the flow of energy in the network dynamic. Especially, betweenness is suitable to identify exceptional climate events that are susceptible to perturbations such as volcanic eruptions or anthropogenic influences. Ordinarily, it is used to describe the influence of the consequence of network dynamic changing.

From Fig. 6, it is evident that high betweenness nodes of SG-network and MG-network are concentrated in the western Pacific and the Atlantic ocean. Active regions in equatorial pacific between 120°W and 180°W longitude match the energy flow inspired by a La Niña event. During the La Niña event, the dry sinking of the Walker circulation moves westward to the mid-equatorial Pacific. In addition, there are two wet updrafts near 120°W in the western Pacific and near 90°E in the eastern Pacific respectively. The teleconnection between the positive phase NAO and La Niña at the lower and upper troposphere causes a significant increase of jet stream in the subtropical Atlantic. Plenty of warm and humid air transport from the Atlantic resulted in an unusually warm winter temperature in Western Europe and an increase of precipitation in most parts of Europe. Abnormal climate events are usually accompanied by massive transfers of heat and energy. Although the process of energy transfer is invisible to us, we can use the betweenness to illustrate the transmission accurately. In other words, betweenness has the potential ability to unfold the invisible ocean phenomena in a statistics way.

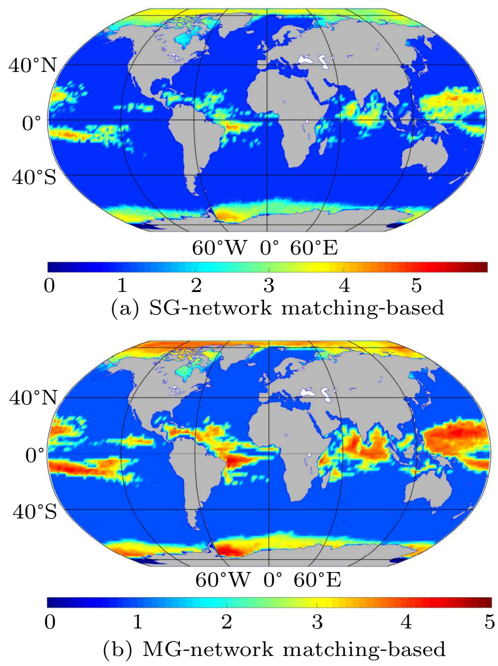


Fig. 6. Betweenness distribution of the network model at geographic location.

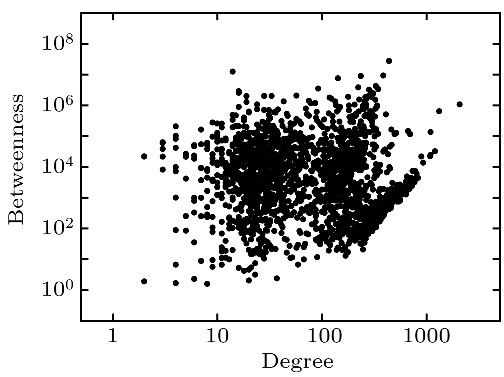


Fig. 7. Scatter plots of betweenness against degree of SG-network

As a statistics property, betweenness may possess some correlations with the basic network character, such as degree. Figure 7 shows the correlation between degree and betweenness of the SG-network. Although there are no obvious patterns between the distributions of betweenness and degree, we can still see that high degree nodes tend to have high betweenness. This phenomenon suggests that super-nodes act as bottlenecks of energy flow in the network. Nodes with low degree but high betweenness in Fig. 7 locate in special geographical locations where the exchange of energy does not depend on other regions.

Table 1. Average clustering coefficient, distance and diameter for the networks constructed by Gaussian approach in three periods.

		Clustering coefficient	Distance	Diameter
La Niña period	MG-network	0.136	2.083	4
	SG-network	0.102	2.905	8
Normal period	MG-network	0.302	1.98	3
	SG-network	0.132	2.07	4
El Niño period	MG-network	0.1585	8189	6
	SG-network	0.261	1.809	4

We have discussed the distribution of network characteristics on the earth intuitively. However, the macroscopic properties of network are still unknown to us. Table 1 compares the average clustering coefficient, average distance and network diameter of different networks constructed by Gaussian approach in three periods. It is realized that the diameter measures the rate of information travels over the network. In the case of the climate network, the information transferring from a node to another can be regard as "fluctuations" induced by a region diffuse to other locations. The transmission of information diffuses local fluctuations so then reduces the chance of prolonged local anomaly and makes the global climate system more stable. The smaller the diameter, the easier the information transfers. Nevertheless, the higher average clustering coefficient means the lower chance of breaking the network into non-communicating parts by accidental removing links from network. In other words, the higher the average clustering coefficient, the greater the robust of the network. The average clustering coefficient of MG-network is higher than SG-network in the three special periods, especially in El Niño period. However, the distance and diameter of MG-network are slightly smaller than SG-network in 2010 and 2014, while in 2015 they are equal. The result shows that MG-network is not only more stable than SG-network, but also more effective for information transmission than SG-network. The distance and diameter of MG-network and SG-network are virtually unchanged over time, while the average clustering coefficient increased a lot in El Niño period. We wonder which network is better to describe the real world ocean dynamic. As we mentioned above, the 2015 El Niño event was the second strongest in history. During El Niño period, the increase of long-range links results in the closest spatial neighborhood of the nodes associated with rise of super-nodes.[17] Therefore, the clustering coefficient will increase. It follows the graph theory that removal of super-nodes makes the network less connected and less stable. The results in Table 1 indicate that Gaussian approach is appropriate to estimate the spatially climate systems.

6. Discussion and conclusion

The techniques for oceanographic observation produce huge amount of oceanic data, of which sea surface temperature is a critical element of the global climatological observations. In this work, we have explored the essence of global ocean dynamic via constructing a complex network with regard to the sea surface temperature. The purpose of this research is providing new insights into dynamic behavior and functional structure of the ocean for oceanographers. Traditional cases commonly average the sea surface temperature of each region as the feature of the node, then apply mutual information and Pearson correlation coefficient to measure the similarity between different nodes. To further understand the ocean dynamic behavior, we introduce the Gaussian mixture models to achieve a nonlinear approximation of the SST data, and simulate the nodes as limit-cycle oscillators. It does not expect to immediately discover new ocean phenomena, but tries to illustrate some well-known phenomena in the complex network viewpoint. The complex network theory based statistical patterns could help researchers further comprehend the causes of ocean phenomena. The interacting dynamical oscillators form the complex network that simulate the ocean as a stochastic system. According to the number of Gaussian distribution in GMMs, we construct different networks, i.e., SG-network and MG-network. The distributions of the two network are shown as bimodal distributions, which means that nodes are grouped by degree. Supernodes are clustered around the tropics areas, the Arctic ocean and high latitudes in southern. In particular, supernodes in the equatorial Pacific denote the origin of ENSO. Nontheless, another climatic anomaly called NAO was highlighted by hypernodes in the area of north Atlantic, to be more precisely, which is surrounded by Canada and close to the eastern United States and Iceland. Furthermore, the supernodes in the most densely distributed range correspond to areas affected by La Niña and El Niño. The high order characteristics including clustering coefficient and betweenness of the network illustrate the transformation of energy and heat between different ocean areas that incur by influential climate phenomenon such as La Niña. To realize the macroscopic properties of the networks, we have compared the degree, clustering coefficient and betweenness distributions in 2010, 2014 and 2015. All the methods can recognize the climatic anomalies from normal period. The GMMs not only helps us simulate the dynamical subsystem from the statistical perspective, but also provides sufficient information to describe the physical processes of ocean dynamic.

Acknowledgment

We would like to express our sincere appreciation to the anonymous reviewers for their insightful comments, which have greatly aided us in improving the quality of this paper. We would like to thank the Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology, Qingdao, China, for the computation support.

References

[1] Zhang L, Huang S X, Shen C and Shi W L 2011 *Chin. Phys. B* **20** 129201
[2] Shen X Y, Qing T and Li X F 2013 *Chin. Phys. B* **22** 94213
[3] Chave A D, Arrott M, Farcas C, Farcas E, Krueger I, Meisinger M, Orcutt J A, Vernon F L, Peach C and S Oscar *Oceans* May 11–14, 2009 Bremen, Germany, p. 1
[4] Liu Y j, Qiu M, Liu C and Guo Z W 2017 *Pers. Ubiquitous Comput.* **21** 55
[5] Owens D, Best M, Guillemot E, Jenkyns R and *Oceans B* 2010 May 24–27 2010 Sydney, Australia, p. 1
[6] Sun X, Chen C, Dong J, Liu D and Hu G 2020 *Knowl.-Based Syst.* **196** 105824
[7] Pradhan R, Aygun R S, Maskey M, Ramachandran R and Cecil D J 2017 *IEEE Trans. Image Process.* **27** 692
[8] Havlin S, Kenett D Y, Jacob E B, Bunde A and Cohen R 2012 *Eur. Phys. J. Spec. Top.* **214** 273
[9] Wu Y, Shang Y, Chen M *et al.* 2008 *IEEE Trans. Circuits Syst.* **214** 1335
[10] Tominski C, Donges J F and Nocke T 2011 *15th International Conference on Information Visualisation* July 13–15, 2011 London, UK pp. 298–305
[11] Donges J F, Zou Y, Marwan N and Kurths J 2009 *Eur. Phys. J. Spec. Top.* **174** 157
[12] Charakopoulos A K, Katsouli G A and Karakasidis T E 2009 *Physica A* **495** 436
[13] Sun X, Song Z, Dong J, Yu Y, Plant C and Böhm C 2019 *Thirty-Third AAAI Conference on Artificial Intelligence* January 29–31, 2019 Hawaii, USA pp. 5041–5048
[14] Steinhaeuser K, Chawla N V and Ganguly A R 2009 *Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data* June 28, 2009 Paris, France pp. 23–31
[15] Donges J F, Petrova I, Loew A, Marwan N and Kurths J 2015 *Clim. Dynamics* **45** 2407
[16] Meng J, Fan J F, Ashkenazy Y and Havlin S 2017 *Chaos: An Interdisciplinary J. Nonlinear Sci.* **27** 035807
[17] Donges J F, Zou Y, Marwan N and Kurths J 2009 *Eur. Phys. J. Special Topics* **174** 157
[18] Boers N, Bookhagen B, Marwan N, Kurths J and Marengo J 2013 *Geophys. Res. Lett.* **40** 4386
[19] Boers N, Bookhagen B, Barbosa H M J, Marwan N, Kurths J and Marengo J A 2014 *Nat. Commun.* **5** 5199
[20] Josef L, Avi G, Mikhail I B, Armin B, Shlomo H and Hans J S 2014 *Proc. Natl. Acad. Sci. USA* **111** 2064
[21] Tsonis A A, Swanson K L and Wang G 2008 *J. Clim.* **21** 2990
[22] Tsonis A A and Roebber P J 2004 *Phys. A: Stat. Mech. Its Appl.* **333** 497
[23] Yamasaki K, Gozolchiani A and Havlin S 2008 *Phys. Review Letters* **100** 228501
[24] Tsonis A A and Swanson K L 2008 *Phys. Rev. Lett.* **100** 228502
[25] Iglesias G, Kale D C and Liu Y 2015 *The 5th International Workshop on Climate Informatics* September 24–25, 2015 Boulder, USA
[26] Fan J F, J M, Chen X S, Ashkenazy Y and Havlin S 2016 *Sci. Chin. Phys. Mech. Astron.* **60** 010531
[27] Ford J D, Tilleard S E, Lea B F, Araos M, Biesbroek R, Lesnikowski A C, MacDonald G K, Hsu A, Chen C and Bizikova L 2016 *Proc. Natl. Acad. Sci. USA* **113** 10729
[28] Wang Y H, Shen X R, Yang S Q and Gao Z K 2020 *Europhys. Lett.* **128** 60005

[29] Feng Q Y and Henk A D 2017 *Chaos: An Interdisciplinary J. Nonlinear Sci.* **27** 035801
[30] Wiedermann M, Radebach A, Donges J F, Kurths J and Donner R V 2016 *Geophys. Res. Lett.* **43** 7176
[31] Tsonis A A, Swanson K L and Roebber P J 2006 *Bull. Am. Meteorological Soc.* **87** 585
[32] Zerenner T, Friederichs P, Lehnertz K and Hense A 2014 *Chaos: An Interdisciplinary J. Nonlinear Sci.* **24** 023103
[33] Zhe J, Li X F, Zhou Y S and Gao S T 2012 *Chin. Phys. B* **21** 054215
[34] Smith N R 2000 *Adv. Space Res.* **25** 1089
[35] Bohm C, Pryakhin A and Schubert M 2006 *Proceedings of the 22nd International Conference on Data Engineering* April 3–7, 2006 Atlanta, GA, USA p. 9
[36] Freeman L C 1978 *Soc. Netw.* **1** 215
[37] Duncan O D 1968 *Am. Sociological Rev.* **33** 457
[38] Huang J P, Kaz H and Amir S 1998 *Geophys. Res. Lett.* **25** 2707
[39] Fronczak A, Fronczak P and Hołyst J A 2007 *Phys. Rev. E* **76** 061106