Geoscientific Model Development (gmd)

An improved method of the Globally Resolved Energy Balance Model by the Bayesian Networks

Zhenxia Liu[1], Zengjie Wang[1], Jian Wang[1], Zhengfang Zhang[1], Dongshuang Li[4,5], Zhaoyuan Yu[1,2,3], Linwang Yuan[1,2,3], and Wen Luo[1,2,3]

[1]School of Geography, Nanjing Normal University, Nanjing, 210023, China
[2]Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing, 210023, China
[3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China
[4]Jiangsu Key Laboratory of Crop Genetics and Physiology/Jiangsu Key Laboratory of Crop Cultivation and Physiology, Agricultural College of Yangzhou University, Yangzhou, China
[5]Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Yangzhou University, Yangzhou, China,

Correspondence: Wen Luo (09415@njnu.edu.cn)

Abstract. The accurate simulation of climate is always critically important and also a challenge. This study introduces an improved method of the Globally Resolved Energy Balance Model (GREB) by the Bayesian Networks based on the concept of coarse-fine model. The improved method constructs a coarse-fine structure that combines a dynamical model with a statistical model based on employing the GREB model as the global framework, and utilizing a Bayesian Networks constructed on the interrelationships between internal climate variables of the GREB model to achieve local optimization. To objectively validate the performance and generalization of the improved method, the method is applied to the simulation of surface temperature and temperature of the atmosphere based on the $3.75° \times 3.75°$ global data sets by Environmental Prediction (NCEP)/ National Center for Atmospheric Research(NCAR) from 1985 to 2014. The results demonstrate that the improved model exhibits higher average accuracy and lower spatial differentiation than the original GREB model, and is robustness in long-term simulations. This approach addresses issues with the accuracy of the GREB model in local areas, which can be attributed to an over-reliance on boundary and initial conditions, and a lack of fully using observed data. Additionally, it overcomes the challenge of poor robustness in statistical models due to ambiguous climate inclusions. Thus, the improved method provides a promising way to give reliable and stable simulation of climate.

1 Introduction

As the global warming progresses, extreme events and meteorological disasters occur frequently(Grant, 2017). Thus, the simulation and prediction of climate have become an important topic in current scientific research for the conceptual understanding and development of hypotheses for climate change studies(Dommenget and Flöter, 2011; Huang et al., 2019). Climate models are mathematical models that describe the temporal evolution of climate, oceans, atmosphere, ice, and land-use processes, across a spatial domain via systems of partial differential equations(Berrocal et al., 2012), which can be solved by supercomputer and is an important tool for simulating and predicting future climate change(Kay, 2020).

Generally, climate models mainly include two categories, dynamic model and statistical model. Dynamic model can well understand and express the dynamic process of climate by modeling various complex climate processes or interactions, but it still faces two major problems: i) The simulation process overly relies on initial conditions and boundary conditions(Alley et al., 2019; Zhang et al., 2019; Ludescher et al., 2021); ii) The climate model is too complicated, and its internal characteristics cannot be fully expressed(Fan et al., 2021; Zou et al., 2019; Feng et al., 2020). The Globally Resolved Energy Balance Model (GREB) is a simple but representative dynamic model, which is based on energy balance theory(Dommenget and Flöter, 2011). Compared with other dynamic models, the GREB model is a relatively fast tool for the conceptual understanding and development of hypotheses for climate change studies, because it computes about one model year per second on a standard personal computer, which allows conducting sensitivity studies to external forcing within minutes to hours(Dommenget and Flöter, 2011; Dommenget, 2016; Stassen et al., 2019). However, in addition to the two main problems of dynamic models, the GREB model also faces the problem that the model does not respond well to anomalous climate change because the parameters of the GREB model are predetermined and the observed data can hardly be used to dynamically correct the model parameters(Dommenget and Flöter, 2011; Dommenget, 2016). How to solve these problems is an important research topic to improve the GREB model and further extend it to other dynamic models.

On the contrary, statistical model, as another type of climate models, can make good use of historical observation data to dynamically modify the models from data(Feng et al., 2020), and solve the problem that dynamic climate models rely too much on initial and boundary conditions and underutilize full observation data. Therefore, it provides a possible way to solve those defects of the dynamical model by combining that with the statistical model. Bayesian Networks is a statistical method which combines graph theory and probability (Cai et al., 2013, 2019; Jansen et al., 2003). The method uses graph to express the structure relation of the variables related to the model and has the characteristics of structuring and quantifying the object relation through the causal relation among the parts of the probability computing system(Pearl, 1986), variable logic reasoning and predictive simulation can be realized, and it can use a large amount of historical observation data. As described, it is a possible way to improve the GREB model by the Bayesian Networks.

The concept of coarse-fine model provides a joint modeling approach of dynamical-statistical hybrid model that is different from the traditional use of statistical model to optimize the empirical parameters of the dynamical model. It starts from different coarse and fine granularity of the model(Akgul and Kambhamettu, 2003; Pal and Bhattacharya, 2010; Yibo et al., 2009), uses the dynamical model as a global framework and uses the statistical model to do local optimization, and realizes the unified modeling of both. Based on this idea, this paper introduces a method for improving the GREB model by the Bayesian Networks. The aim of method is to solve the problem of low model accuracy due to over-reliance on boundary conditions and initial conditions and inability to fully utilize historical observation data. The following section presents the improved method. Section 3 presents the study case and data sets to test the new improved model. Finally, we give a discussion and conclusion of the results.

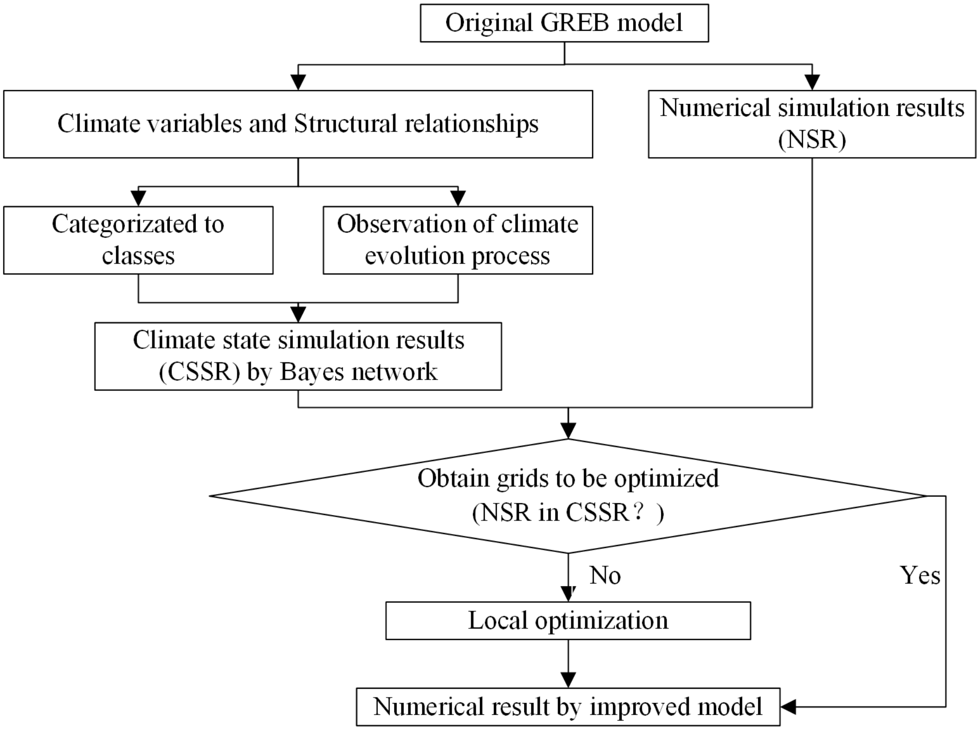

Figure 1. Overall framework of the improved method.

2 Methods

The improved method is developed according to the following procedure. Firstly, climate variables representing different climate processes are chosen as nodes in the Bayesian Networks constructed by the GREB model. And the structural relationships among different nodes are determined to establish an abstract model of the components and structural relationships of climate processes. Secondly, the selected climate variables are categorized into variable ranges based on their numerical values to form different classifications that are used to indicate different climate state. Thirdly, the climate state simulation method is reconstructed based on the Bayesian Networks and climate evolution process to achieve the simulation of the target variables climate state. Finally, the climate state simulation results obtained from the Bayesian Networks are compared with the climate modal simulation results from the original GREB model to get the local optimization grids, and the numerical results of the original GREB model simulation are optimized based on the comparison results. Based on the above considerations, improved method is developed according to the following procedures (Figure 1).

2.1 Structural relationship among climate variables

Based on the energy balance, the GREB model can simulate the main characteristics and climate mean states of global warming, including seven climate processes (solar radiation, thermal radiation, hydrological cycle, sensible heat and atmospheric temperature, atmospheric circulation, sea ice and deep ocean) and four main climatic variables (surface temperature, temperature of the atmosphere, temperature of subsurface the ocean, and humidity of the surface). Each of these processes is represented with strongly simplified equations. Therefore, we can abstract the structural relationship among different climate variables from the simplified equation, i.e., which climate variables control a given climate variable, and which climate variables are influenced by it. This structural relationship provides the possibility to construct Bayesian Networks.

2.2 Categorization of climate state

According to the theory of climate sensitivity(Annan and Hargreaves, 2006; Dommenget, 2016), Climate state, indicated by a range of numerical values, can be used to replace the specific numeric to simulate climate change and characterize the long-term trend of climate change and extreme weather conditions. And it is better suited to capture the similarity of a given climate variable across different spatial and temporal locations compared to the numerical values of specific climate variables. Therefore, it can be used to assess the similarity between the simulated results of a model and the actual results, indicating the accuracy of the simulation. This provides a simple and practical approach to evaluating the accuracy of revealing local abrupt changes in simulation results. Moreover, by simulating state rather than specific numeric, it is possible to significantly reduce computational effort and simulation response time. This is consistent with the primary objective of the GREB model, which is to provides a fast tool for the conceptual understanding and development of hypotheses for climate change studies(Dommenget and Flöter, 2011)

The natural breaks classification (Jenks) method is a commonly used classification method that aims to minimize intra-class variation and maximize inter-class variation. By categorizing the numeric of climate variables into different classifications to indicate climate state using the natural breaks classification method, it can be considered as that numeric within the same classification have less variation, representing that the results of this classification of numeric have similar climate state.

2.3 Climate state simulation

According to the characteristics of Bayesian Networks, the climate state simulation of climate variables is realized by climate evolution process based on Bayesian Networks, i.e., the climate state of unknown climate variables is inferred from the climate state of known climate variables at the same spatial locations.

2.3.1 Bayesian Networks

Bayesian Networks is a probabilistic model that simulates the human reasoning process, which is a combination of graph theory and probability theory, and its network topology is a directed acyclic graph. Where variables are nodes and correlations or causal relationships between variables are directed edges.The dynamic evolution of Bayesian Networks node probabilities is controlled by conditional probabilities, and each node covers a probability distribution table under the joint distribution of the parent nodes, indicating the strength of the relationship between the nodes.(Sahin et al., 2019). When the Bayesian Networks is constructed, given the state of any node, the probability distribution of the states of the remaining nodes can be calculated.

In the Bayesian Networks, the probability of a node can be calculated in the form of probability using prior knowledge and statistical data, namely the Bayes probability (Maher, 2010). Observed sample are defined as: $G = \{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}$, where X is event, x is event value or state. When   is the prior probability of event $X = x$,   is prior knowledge, $P(\ |\ )$ is probability density function, then the probability $P(X_{n+1} = x_{n+1}|\ ,\ )$ of the $n+1$ event $X_{n+1} = x_{n+1}$ can be obtained from the prior probability density $P(\ |\ )$ and the sample G through the Bayes probability. It can be calculated by total probability formula:

$$\begin{aligned} & P(X_{n+1} = x_{n+1}|\ ,\ ) \\ = & \int P(X_{n+1} = x_{n+1}|\ ,G,\ )P(\ |,G,\ )d \\ = & \int \ P(\ |,G,\ )d \end{aligned} \tag{1}$$

Based on Bayes equation, The posterior probability $P(\ |,G,\ )$ is denated as:

$$P(\ |,G,\ ) = \frac{P(\ |\ )P(G|\ ,\ )}{P(G|\ )} \tag{2}$$

Where G is given sample,   is priori probability of G,

2.3.2 Climate evolution process based on Bayesian Networks

In a climatic process composed of several climatic variables, there is an association relationship between climatic variables. These climatic variables are regarded as network nodes, and the association relations between climatic variables are taken as directed edges. The association relationship between nodes is represented by graph model, and the action intensity of association relationship is described quantitatively by conditional probability table. Using the characteristics of Bayesian Networks, the attribute feature state of nodes is inferred by probability. To realize the expression and simulation of the attribute feature state of geographical variables.

A climate process $M_t = \{X(m_1, m_2, \ldots, m_i)|m_{1t}, m_{2t}, \ldots, m_{it}\}$ is composed with i climate variables, $m_1, m_2, \ldots, m_i$, and $X(m_1, m_2, \ldots, m_i)$ is the structural relationship among the variables. Suppose that the climate variable m_i has j states, then the sataes set of m_i is $W_{m_{i1}}, W_{m_{i2}}, \ldots, W_{m_{ij}}$. The climate process is described by a Bayesian Networks $B = (S, X)$. where S is a directed acyclic graph composed of nodes; X is the nodes set of graph, that is climate variables $m_1, m_2, \ldots, m_i$. nodes are connected by directed edges to represent the relationship between climate variables. Each node has an independent conditional probability table, which represents the probability distribution under the joint distribution of its parent nodes. Assume that a climate m_i has one or more parent nodes $m_1, m_2, \ldots, m_e$ $(e \leq i-1)$ and states $d_1, d_2, \ldots d_e$, it can be denoted as :$m_1, m_2, \ldots, m_e \rightarrow m_i$. Under the parent node of all possible states, the conditional probability table composed of the set of state probabilities of mdeteci is follow:

$$\begin{gathered} B_{m_i}^{W_{m_{1r1}}, W_{m_{2r2}}, \ldots, W_{m_{ere}}} = \\ \left\{ (W_{m_{i1}}, P_{W_{m_{i1}}}^{W_{m_{1r1}}, W_{m_{2r2}}, \ldots, W_{m_{ere}}}), (W_{m_{i2}}, P_{W_{m_{i2}}}^{W_{m_{1r1}}, W_{m_{2r2}}, \ldots, W_{m_{ere}}}), \ldots, (W_{m_{ij}}, P_{W_{m_{ij}}}^{W_{m_{1r1}}, W_{m_{2r2}}, \ldots, W_{m_{ere}}}) \right\} \\ (r1 = 1, 2, \ldots, d_1), (r2 = 1, 2, \ldots, d_2), \ldots, (re = 1, 2, \ldots, d_e) \end{gathered} \tag{3}$$

Where $B_{m_i}^{W_{m_{1r1}};W_{m_{2r2}};\dots;W_{m_{ere}}}$ is a conditional probability table of climate variables m_i; $W_{m_{ij}}$ is j th characteristic state of climate variables m_i; $P_{W_{m_{ij}}}^{W_{m_{1r1}};W_{m_{2r2}};\dots;W_{m_{ere}}}$ is the probability of climate variable m_i corresponds to the jth state under the $r1; r2; \dots; re$ characteristic state corresponding to the parent node $m_1; m_2; \dots; m_e$ expression set. The probability set of climate variable m_i at t moment can be denoted as $C_{m_{it}}$:

$$C_{m_{it}} = \left\{ (W_{m_{i1}}; P_{W_{m_{i1}}}^{W_{m_{1r1}};W_{m_{2r2}};\dots;W_{m_{ere}}}); (W_{m_{i2}}; P_{W_{m_{i2}}}^{W_{m_{1r1}};W_{m_{2r2}};\dots;W_{m_{ere}}}); \dots; (W_{m_{ij}}; P_{W_{m_{ij}}}^{W_{m_{1r1}};W_{m_{2r2}};\dots;W_{m_{ere}}}) \right\} \qquad (4)$$

The conditional probability table of each node can be calculated by Eq.2 using training data.

2.4 Local optimization

The numerical results simulated by the original GREB model are compared with the climate state results simulated by the Bayesian Networks, and the grids where the numerical result simulated by the original GREB model are not in the range of the climate state simulated by the Bayesian Networks are used as grids to be optimized.

According to the Third Law of Geography(Zhu et al., 2018), the more similar the geographic environment, the more similar the geographic target characteristics are. Therefore, for an unknown climate variable at a certain spatial and temporal location, the numeric of other known climate variables at that spatial and temporal location can be used to infer. Accordingly, we propose that for an unknown climate variable, the position of its specific value in the range of its classification is related to the position of the specific value of the known climate variable in the range of its classification at the same spatial and temporal location. For a climate variable containing n relevant control variables, the numerical results are calculated as follows:

$$E^{x}_{value} = S^{E}_{lowerlimit} + \frac{1}{n}\left(S^{E}_{upperlimit} - S^{E}_{lowerlimit}\right) \sum \frac{E^{i}_{value} - S^{i}_{lowerlimit}}{S^{i}_{upperlimit} - S^{i}_{lowerlimit}} \qquad (5)$$

Where E^{x}_{value} represents an unknown climate variable; $S^{E}_{lowerlimit}$ represents the lower limit of the range of classification in which the unknown climate variables are simulated by Bayesian Networks; $S^{E}_{upperlimit}$ represents the lower limit of the range of classification in which the unknown climate variables are simulated by Bayesian Networks; n represents the number of known climate variables associated with the unknown variables in the Bayesian Networks; E^{i}_{value} the actual value of the ith known climate variable; $S^{i}_{lowerlimit}$ represents the lower limit of the range of classification in which the ith known climate variables; $S^{i}_{upperlimit}$ represents the upper limit of the range of classification in which the ith known climate variables.

According to the above method, we can improve the accuracy of the model by comparing the climate state, identifying the grid to be optimized, and recalculate the values simulated by the original GREB model within the grid. In this way, the improved model with coarse-fine structure constructed with the GREB model as the global framework and the Bayesian Networks as the local optimization can better reflect the localized abrupt changes in the climate process and achieve the purpose of improving the GREB.

3 Case study

In order to demonstrate the accuracy of the improved model in simulating climate variables and to verify its reliability, surface temperatures and temperature of the atmosphere from the GREB model were selected for simulation objects. The simulation o these two climate variables includes most of the climate processes of the GREB and can reflect the complex coupling process and climate change characteristics of the GREB model.

3.1 Data description

In this paper, data produced by National Centers for Environmental Prediction (NCEP) / National Center for Atmospheric Research (NCAR) is used as the experimental data to evaluate the improved model. The data sets include surface temperature(T_{surf}), temperature of the atmosphere(T_{atmos}), solar radiation(F_{solar}), total cloud cover(CLD), water vapor(q_{air}), temperature of the subsurface ocean(T_{ocean}), and wind speed($\vec{u}$) stored as a 3.75*3.75 (latitude * longitude) grid NC data from 1985 to 2014. In order to facilitate calculation and comparative analysis, all climate data is preprocessed. Firstly, the downloaded climate data is removed from the outliers so that the data are calculated to avoid too large or too small results; secondly, the grid data is resampled and the resampling method is bilinear interpolation. The bilinear interpolation method is used to interpolate the climate data, which not only fills the null values, but also unifies the scale size of the data. Finally, considering that changes in climate variables are usually seasonally related, climate data from 1985 to 2014 were processed as quarterly averages, wher January, February, and March comprised first quarter, April, May, and June formed second quarter, July, August, and September constituted a third quarter, and October, November, and December comprised the fourth quarter.

3.2 Structural relationship among climate variables and climate state

The process of simulating the surface temperature includes solar radiation, thermal radiation, sensible heat and atmospheri temperature, and deep oceanDommenget and Flöter (2011). The main heat source of the surface temperature is solar radiatio some of which is absorbed by the surface temperature, the other part is reflected by the surface temperature, and part of the heat on the surface temperature is transferred in the atmosphere, and some of it is transferred to the ocean below the surfac Each climate variables in this scene can be expressed by a highly simplified equation, which follows the surface temperature tendency equation as follows:

$$_{surf}\frac{dT_{surf}}{dt} = F_{solor} + F_{thermal} + F_{latent} + F_{sense} + F_{ocean} \tag{6}$$

Where T_{surf} is surface temperature; $_{surf}$ is surface heat capacity; F_{solar} is the incoming solar radiation; $F_{thermal}$ is the net thermal radiation; F_{latent} the cooling by latent heat from surface evaporation of water; F_{sense} is the turbulent heat exchange with the atmosphere; F_{ocean} is the heat exchange with the deeper subsurface ocean. The subprocesses of surface temperature are modeled as follows:

$$\begin{cases} F_{solar} = (1 \quad {}_{clouds})(1 \quad {}_{surf})\, S_0 \;\; r(\;; t_{julian}) \\ F_{thermal} = \quad T^4_{surf} + "_{atmos} \;\; T^4_{atmos \;\; rad} \\ "_{atmos} = \frac{pe_8 \;\; CLD}{pe_9}("_0 \quad pe_{10}) + pe_{10} \\ F_{latent} = L \;\; {}_{air} \;\; C_w \;\; \vec{u} \;\; {}_{soil} \; (q_{air} \quad q_{sat}) \\ F_{sense} = c_{atmos} \; (T_{atmos} \quad T_{surf}) \\ F_{ocean} = Fo_{sense} + \;\; {}_{surf} \quad T_{entrain} \\ Fo_{sense} = c_{ocean} \; (T_{ocean} \quad T_{surf}) \end{cases} \tag{7}$$

Where F_{solar} is the incoming solar radiation; ${}_{clouds}$ is the fraction of the incoming solar radiation is reflected by clouds; ${}_{surf}$ is the fraction of the incoming solar radiation is reflected by the surface; S_0 is the solar constant; ${}_{surf}$ is the fraction of the incoming solar radiation is reflected by the surface; r is the 24 h mean fraction reaching a normal surface area on top of the atmosphere; is the function of latitude; t_{julian} the Julian day of the calendar year; $F_{thermal}$ is the net thermal radiation; T_{surf} is surface temperature; $"_{atmos}$ is the effective emissivity; $T_{atmos \;\; rad}$ is the temperature defined in the context of the atmospheric temperature; CLD is the total cloud cover; $"_0$ is the emissivity without considering clouds; pe_i is the parameters; F_{latent} the cooling by latent heat from surface evaporation of water; L is the constant parameters of the latent heat of evaporation and condensation of water; ${}_{air}$ is the density of air; C_w is the transfer coefficient; $\vec{u}$ is the wind speed; ${}_{soil}$ is the Bulk formula is extended by a surface wetness fraction; q_{air} is the actual surface air layer humidity; q_{sat} is the saturation surface air layer specific humidity; F_{sense} is the turbulent heat exchange with the atmosphere; c_{atmos} is the coupling constant; T_{atmos} is temperature of the atmosphere; F_{ocean} is the heat exchange with the deeper subsurface ocean; Fo_{sense} is the turbulent mixing between the two ocean layers; $T_{entrain}$ is the heat exchange with the surface ocean layer due to decreasing of the mixed layer depth; c_{ocean} is the coupling constant; T_{ocean} is the temperature of the subsurface ocean.

In the process of simulating the temperature of the atmosphere by the GREB model Dommenget and Flöter (2011), temperature of the atmosphere is not only related to the thermal radiation reflected from the surface, but also related to the sensible heat exchange with the surface and latent heat release by condensation of atmospheric water vapor. Each climate variables in thi process can be expressed by a highly simplified equation, which follows the temperature of the atmosphere tendency equation as follows:

$${}_{atmos} \frac{dT_{atmos}}{dt} + F_{sense} = F_{thermal} + Q_{latent} + {}_{atmos} \;\; k \;\; r^2 T_{atmos} \quad \vec{u} \;\; r \;\; T_{atmos} \tag{8}$$

Where T_{atmos} is temperature of the atmosphere; ${}_{atmos}$ is atmospheric heat capacity; F_{sense} is the sensible heat exchange with the surface; $Fa_{thermal}$ is net thermal radiation of the atmosphere; Q_{latent} is the latent heat release by condensation of atmospheric water vapor; $\vec{u}$ is the wind speed. The subprocesses of temperature of the atmosphere are modeled as follows:

$$\begin{cases} F_{sense} = c_{atmos} \; (T_{atmos} \quad T_{surf}) \\ Fa_{thermal} = "_{atmos} \quad T^4_{surf} \quad 2"_{atmos} \quad T^4_{atmos \;\; rad} \\ Q_{latent} = \quad 2{:}6736 \;\, 10^3 \;\; kg{=}m^2 \quad q_{precip} \;\; L \end{cases} \tag{9}$$

Where F_{sense} is the turbulent heat exchange with the atmosphere; c_{atmos} is the coupling constant; T_{atmos} is temperature of the atmosphere; $F_{a;thermal}$ is net thermal radiation of the atmosphere; T_{surf} is surface temperature; ε_{atmos} is the effective emissivity; $T_{atmos;rad}$ is the temperature defined in the context of the atmospheric temperature; Q_{latent} is the latent heat release by condensation of atmospheric water vapor; q_{precip} is the condensation or precipitation; L is the constant parameters of the latent heat of evaporation and condensation of water.

For different climate process, the climate subprocesses and relationship structures are different. Therefore, the selection of nodes in each climate process will also be different. Not only the selection of appropriate variables as nodes is very important, but also the number of nodes will directly affect the simulation of the final climate average state. In order to simplify the complex climate evolution process and facilitate calculation, 4-6 climate variables are selected as key nodes in each climate process, and the variables climate state in each processes are simulated by these nodes.

Through the trend equations (Eq. 6 and 7) in the processes of surface temperature, the relation equation of climate variables can be simplified:

$$\begin{cases} T_{surf} = f(F_{solar}; T_{ocean}; q_{air}; CLD) \\ T_{ocean} = f(F_{solar}) \\ q_{air} = f(F_{solar}) \end{cases} \tag{10}$$

Where T_{surf} is surface temperature; F_{solar} is solar radiation; T_{ocean} is temperature of the subsurface ocean; q_{air} is the actual surface air layer humidity, i.e., water vapor content; CLD is total cloud cover. That is, surface temperature, solar radiation, temperature of the subsurface ocean, total cloud cover and water vapor content can be selected as the key nodes of the surfac temperature process.

Through the trend equation (Eq. 8 and 9) in the processes of temperature of the atmosphere, the relation equation of climate variables can be simplified:

$$\begin{cases} T_{atmos} = f(\vec{u}; q_{air}; CLD) \\ q_{air} = f(\vec{u}) \end{cases} \tag{11}$$

Where T_{atmos} is temperature of the atmosphere; $\vec{u}$ is wind speed; CLD is total cloud cover; q_{air} is water vapor content. That is, temperature of the atmosphere, wind speed, total cloud cover and water vapor content can be selected as the key node of the temperature of the atmosphere process.

According to Eq. 10 , in the surface temperature process, surface temperature (T_{surf}) is controlled by solar radiation (F_{solar}), cloud cover (CLD), water vapor (q_{air}) and temperature of the subsurface ocean (T_{ocean}). Temperature of the subsurface ocean (T_{ocean}) and water vapor (q_{air}) are controlled by solar radiation (F_{solar}). For the above relationship, the Bayesian Networks structure in the surface temperature process can be constructed (Figure 2a). According to Eq. 11, in the temperature of the atmosphere process, temperature of the atmosphere (T_{atmos}) is controlled by cloud cover (CLD), water vapor (q_{air}) and wind speed ($\vec{u}$). And water vapor (q_{air}) is controlled by wind speed ($\vec{u}$). For the above relationship, the Bayesian Networks structure in the temperature of the atmosphere process can be constructed (Figure 2b).

Figure 2. Structural relationship among climate variables in the different simulation processes. (a) Surface temperature process; (b) Temperature of the atmosphere process.

The climate state of the variables in the above climate processes was performed using the natural breaks classification method. The climate variables data are categorized into five, seven and nine different classifications to indicated different climate state to test the improved model and verify the effect of the classification number of climate variables data on the simulation results. Detailed schemes are shown in Appendix Table A1, A2, A3.

3.3 Climate state simulation

Surface temperature and temperature of the atmosphere are considered as a simulation object, and other climate variables a known objects, and uses the historical data to calculate the conditional probability tables of each nodes through the Bayesian Networks structure with Eq 4. Among them, the training data is the 10-year historical data from 1985 to 1994.

In each simulation process, there are two training methods for the simulated object. The first is to train a conditional probability table using the data in all the grids, and then use the conditional probability table to simulate the states of all grids. The conditional probability table obtained by this training method can reflect the numerical characteristic relationship between climate variables in the whole region. However, it can not show the distribution pattern of the characteristics of the simulated state in space. The second is to train the data in each grid separately. Because the state grading data in each grids is differen the conditional probability table of the simulated object trained in each grids is also different, and a total of 4896 conditional probability tables are obtained. The conditional probability tables obtained by this training method can accurately reflect the different numerical characteristic relationship between the simulated object and the known object in different regions. However, due to the training of more conditional probability tables, the running time of this training method will be a little longer. Considering the great differences in the pattern of climate evolution in different regions, this paper uses the second data training

Figure 3. Climate state simulation results of the surface temperature by Bayesian Networks for 80 seasons (the period 1995-2014). (a) Categorized into 5 classifications; (b) Categorized into 7 classifications; (c) Categorized into 9 classifications.

Figure 4. Climate state simulation results of the temperature of the atmosphere by Bayesian Networks for 80 seasons (the period 1995-2014). (a) Categorized into 5 classifications; (b) Categorized into 7 classifications; (c) Categorized into 9 classifications.

method in state simulation, which first divides the whole world into 9648 grids, and then uses the data in each grids to train the conditional probability tables of the grid. After the training is completed, the data of the known climate variables will be used to simulate the unknown climate variables from 1995 to 2014. The simulation results are shown in Figure 3 and 4.

Figure 3 shows the climate state of the quarterly average of surface temperature from 1995-2014. The simulation results under different classifications all clearly show the global quarterly average surface temperature distribution with latitudinal variations. The surface temperature starts from the equator and decreases with the increase of latitude, so the temperature in the North and South Pole is the lowest. The climate state distribution of surface temperature is basically in line with the real world. Different from the simulation result of surface temperature, the quarterly average temperature of the atmosphere rises