Evaluation of the Simulation Performance of the CMIP5 Climate Model in East Asia Using Self-Organizing Mappings Combined with Multiple Metrics

Yongdi Wang

School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China.

ydwang2003@163.com

Abstract

The fifth Coupled Model Comparison Program (CMIP5) includes 60 coupled models in up to 19 model groups. How to objectively and quantitatively assess and compare the performance of different climate models is becoming increasingly important. Most of the current model evaluation methods are based on a single variable. However, in model comparison and assessment, we often need to assess the ability of a model to simulate a particular climate phenomenon. To address the shortcomings of previous model evaluation methods, this paper applies a self-organizing mapping neural network-based weather-climate modal model evaluation method based on the weather-type classification. Using the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis Data (ERA-40) and National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data as a reference field, the simulation performance of the seven CMIP5 climate models for East Asia over the period 1980-1999 is evaluated in a number of ways: first, the annual mean cycle and interannual variability simulations are evaluated using climate indicators, and then the simulation of weather modal occurrence using the Self-Organizing Mapping (SOM) technique is highlighted. assessment. It is concluded that the choice of the reference field is particularly important and is decisive for the results of the climate model assessment, which is better when ERA-40 is used as the reference field (as opposed to NCEP).

Keywords

Climate Models, Model Evaluation, Self-Organizing Maps, Performance Metrics, East Asia.

1. Introduction

Climate models are powerful tools for studying the climate system and climate change, and their simulation results are an important data base for climate prediction and climate change risk assessment. As the rate of global warming accelerates, surface ecology, dynamic hydrological cycle processes, and socio-economic development are affected, thus affecting human production and livelihoods. The impact of extreme weather events is significant, and the ability of models to simulate extreme weather and climate events is directly related to the accuracy and reliability of future predictions. At present, however, climate models are difficult to simulate extreme weather and climate events in detail, and require significant research.

As a new approach to model evaluation, RADIC′ AND CLARKE (2011) [1] used SOM to evaluate the simulation results of 22 IPCC climate models in North America and its western region. The main method is to evaluate the model simulation capability by calculating the correlation coefficients for the frequency of occurrence at each node of the observation and simulation fields.

In this paper, the SOM is used in combination with other climate indicators to evaluate the simulation results of seven models in East Asia.

2. Model output and validation data (Model and Reference Data Sets)

2.1 The CMIP5 Simulations.

Our evaluation is based on the late-twentieth-century simulations by 7 GCMs from the fifth Coupled Model Comparison Program (CMIP5) [2]. The CMIP5 climate models used in this study are listed in Table 1. We evaluate model performance using various metrics by seven climate variables. These seven variables are monthly grids of relative humidity at 500 hPa (HUR500), precipitation (PR), sea level pressure (SLP), air temperature at 500 hPa (TA500), the eastward wind and northward wind at 500 hPa (UA500, VA500), geopotential height at 500 hPa (ZG500). The data have been downloaded in the form of monthly means from the ESG Data Gateways (http://pcmdi3.llnl.gov/esgcet/home.htm).

Table 1 Model identification, originating center, and atmospheric resolution

Model	Center and location	Atmosphere resolution
MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M)	192×96
MPI-ESM-MR	Max Planck Institute for Meteorology (MPI-M)	192×96
MPI-ESM-P	Max Planck Institute for Meteorology (MPI-M)	192×96
MIROC5	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology (MIROC)	256×128
inmcm4	Institute for Numerical Mathematics (INM)	180×120
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory (NOAA-GFDL)	144×90
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory (NOAA-GFDL)	144×90

The evaluation is performed by comparing GCMs historical simulations with the National Centers for Environmental Prediction / National Center for Atmospheric Research (NCEP/NCAR, Kalnay et al., 1996) [3] and European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis Data (ERA-40, Simmons and Gibson, 2000) [4], which we take as the reference dataset (1970-1999). And all of the data were interpolated to a common grid with a resolution of 2.5°×2.5°.

2.2 Reference Data.

The best available reference data come from two 20-year reanalysis efforts, namely the National Centers for Environmental Prediction / National Center for Atmospheric Research (NCEP/NCAR, Kalnay et al., 1996) [3] and European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis Data (ERA-40, Simmons and Gibson, 2000) [4].

In this study, we use the two sets of reference data for the Self-Organizing Maps [5] procedure and the performance metrics procedure: NCEP reanalysis of daily gridded atmospheric data from 1980 to 1999 with a resolution of 2.5°×2.5° .

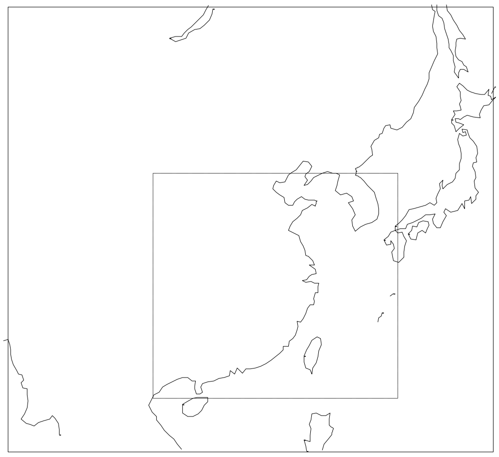

Fig. 1 Analysis domains: large domain (large rectangular) and small domain (rectangle inside the large one)

The area used for the analysis is shown in Fig. 1. The large domain covers the region between 90° and 140° E of longitude and 15° and 55° N of latitude, and the small domain covers the region between 105° and 130° E of longitude and 20° and 40° N of latitude.

3. Validation methods

3.1 Statistical metrics.

3.1.1 Relative Error (RE)

RE [1] is defined as follows.

$$RE = \frac{RMSE - \overline{RMSE}}{\overline{RMSE}} \tag{1}$$

where $\overline{RMSE}$ is the median, not the mean, of the RMSE.

RMSE (Root-mean Square Error) [1] can be calculated according to the following formula (F is the simulated field, R is the reference field).

$$RMSE^2 = \frac{1}{W}\sum_i\sum_j\sum_t w_{ijt}(F_{ijt} - R_{ijt})^2 \tag{2}$$

where *i* is longitude, *j* is latitude, and *t* is time. *W* is the sum of weights w_{ijt}. The calculated *RE* values are both positive and negative, the larger the negative value, the closer it is to the observation field and the better the simulation performance.

3.1.2 Model Variability Index (MVI)

MVI [6] can be calculated by the following formula.

$$MVI = \sum_{n=1}^{N}(\beta_n - \frac{1}{\beta_n})^2 \tag{3}$$

where β^2 is the ratio of variance of the simulated field to the observed field and *N* is the number of variables. The calculated *MVI* is a positive value, and the smaller the value the better.

3.2 Self-organizing maps technique.

3.2.1 SOM Profile

Self-organizing feature map (SOM) [5], proposed in 1981 by the Finnish scholar Professor Teuvo Kohonen of Helsinki University, is an unsupervised artificial neural network primarily used for pattern recognition and classification, and in many respects resembles the traditional form of cluster analysis. The main features of SOM are self-organizing learning, high clustering quality, more objective results, probability distribution of the samples, and topology preservation properties.

The main difference between SOM and other clustering methods is that there is no a priori experience with SOM, the development and fundamentals of which were described in detail by Kohonen in 2000. Here, we use SOM to classify patterns in climate data, with a brief algorithmic procedure as follows.

The output layer of the SOM represents the prototype pattern of the input data with a specified number of nodes. The SOM is first initialized and consists of a specified number of random nodes, each of which is assigned a weight vector and information about its position in the 2D space. The process of mapping the input vector to the SOM nodes is to find the nearest (in European space) weight vector and specify the geographic coordinates of the node. The node with the nearest weight vector is called the "best matching unit". The next step is to adjust and update the neighborhood of the best matching unit, i.e. the distance from the surrounding nodes to the input vector. The formula for calculating the weight vector is as follows.

$$W(t+1) = W(t) + \Theta(t)\alpha(t)[V(t) - W(t)] \tag{4}$$

where *t* is the current iteration, *W* is the weight vector, *V* is the input vector, Θ is the neighborhood function, and α is the learning rate.

The above process is repeated until all input data are trained by the SOM network a predetermined number of times. The number of original model features retained by the output layer depends on the size of the SOM, with smaller input data sets producing broad generalizations and larger SOMs capturing increasingly finer details; the basic feature of the SOM is that neighboring nodes represent similar patterns, while those that are farther apart represent completely different patterns.

SOM is used for pattern evaluation to extract the dominant circulation modes corresponding to different weather processes and to quantitatively assess the ability of the patterns to simulate day-by-day circulation patterns. The goal of our research is to find the best GCM model that can reproduce the weather scale pattern covering the whole region.

3.2.2 Data Preprocessing

a) SLP temporal SLP [1] : Daily SLP value for each grid point minus the daily SLP average over a 20-year base period.

b) SLP spatial SLP [1]: daily SLP value at each grid point minus the average daily SLP value for the entire region (allows the study to focus on the barometric gradient).

4. Results

4.1 Simulation of mean annual cycle

For the seven selected variables (HUR500, PR, SLP, TA500, UA500, VA500 and ZG500), the annual average cycle simulation of the indicator statistics (Relative error) yielded the following results.

a) The value of Relative error can be positive or negative, with a larger negative value indicating that the simulation results are closer to the observational field and the better the simulation performance of the model's annual mean cycle (as shown in Fig. 2).

b) No model has all above-average or all below-average Relative error statistics for all variables.

c) Some models had narrow interval ranges (e.g., in the large domain: MPI-ESM-LR, MPI-ESM-MR, MPI-ESM-P; in the small domain: MPI-ESM-MR, inmcm4).

d) In the small domain, some models had Relative error > 0.5, indicating that the simulation of the annual mean cycle is more difficult in the small domain than in the large domain.

e) Calculated Relative error correlation coefficients for the large and small domains: HUR500 (0.32), PR (0.25), SLP (0.11), TA500 (-0.10), UA500 (0.18), VA500 (-0.41), and ZG500 (0.96). This means that of the seven variables selected, only ZG500 is comparable between the two domains of size (an unsatisfactory result).

f) The time base period is shifted forward by 10 years (switching from 1980-1999 to 1970-1989), and the average annual cycle change of the seven variables is insensitive to the switch in the 10-year base period.

g) Multiple model means and medians are close to the observational field and better than any single model.

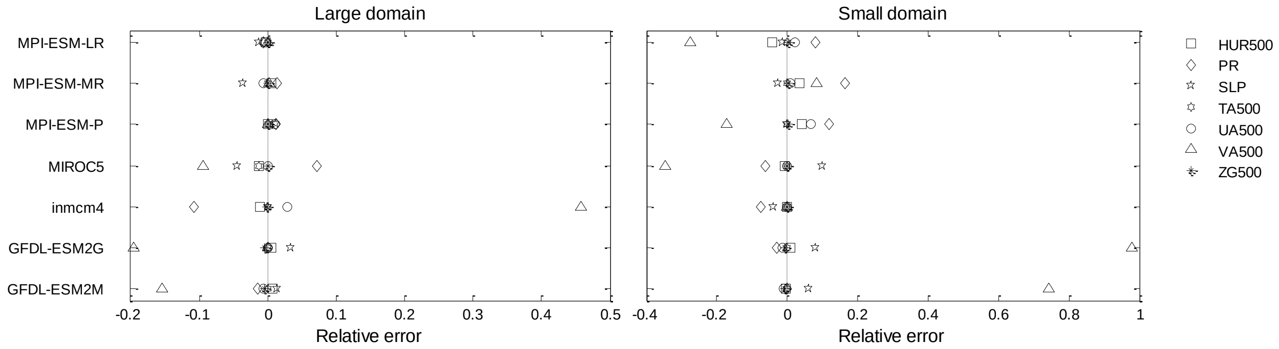

Fig. 2. Relative errors over (left) the large and (right) the small domain for the seven climate variables: HUR500, PR, SLP, TA500, UA500, VA500 and ZG500

h) Some climate variables have severe outlier distributions (e.g., VA500 for inmcm4, GFDL-ESM2G, and GFDL-ESM2M).

i) By examining the correlation coefficients of the Relative error values between two different variables, significant positive correlations were found between some of the different variables, such as in the large domain: HUR500 and TA500 (r = 0.59), SLP and TA500 (r = 0.64), UA500 and VA500 (r = 0.85); and in the small domain: PR and UA500 (r = 0.63), UA500 and ZG500 (r = 0.62).

4.2 Simulation of interannual variability.

For the seven selected variables (HUR500, PR, SLP, TA500, UA500, VA500 and ZG500), simulations of interannual variation yielded the following indicator statistics.

a) has both positive and negative values, with larger negative values indicating that the simulation results are closer to the observational field and the better the model's simulation performance for interannual variation (as shown in Fig. 3).

b) Similar to the previous point, some models produced better results than others, but none of the models had all above-average or all below-average statistics for all variables.

c) Some models had narrow interval ranges (e.g., in the large domain: MPI-ESM-P, MIROC5, GFDL-ESM2G, GFDL-ESM2M; in the small domain: MPI-ESM-MR).

d) In the small domain, some of the models are more diffuse, suggesting that simulating the average annual cycle is more difficult in the small domain than in the large domain.

e) Calculated correlation coefficients for large and small domains: HUR500 (-0.10), PR (0.37), SLP (-0.79), TA500 (0.58), UA500 (-0.17), VA500 (-0.41), and ZG500 (0.86). This means that of the seven variables selected, only two variables, TA500 and ZG500, are comparable between the two domains of size (this result is not ideal).

f) The time base period is shifted forward by 10 years (switching from 1980-1999 to 1970-1989), and the average annual cycle changes of the seven variables are not sensitive to the switching of the 10-year base period.

g) The mean and median of multiple models are very close to the observational field and better than any single model.

h) Some climate variables have severe outlier distributions (e.g., VA500 and SLP for inmcm4, GFDL-ESM2G, and GFDL-ESM2M).

i) By examining the correlation coefficients of the Relative error values between two different variables, a number of different variables were found to be significantly positively correlated, such as in the large domains: HUR500 and SLP (r = 0.85), HUR500 and ZG500 (r = 0.60), PR and UA500 (r = 0.81), SLP and VA500 (r = 0.78), VA500 and ZG500 (r = 0.61), especially notably TA500 and ZG500 (r = 0.98), HUR500 and VA500 (r = 0.90); in the small domains: PR and SLP (r = 0.60), TA500 and ZG500 (r = 0.88).

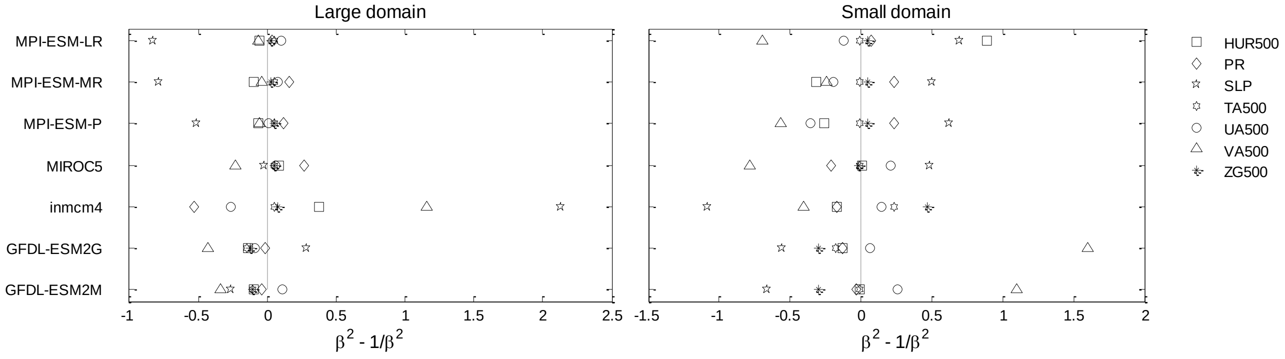

Fig. 3. Values of $\beta^2 - 1/\beta^2$, where β^2 represents the ratios of simulated (GCM) to reference (NCEP) variances, over the large and small domain, for the seven climate variables in Fig. 1

4.3 Simulation of occurrences of synoptic patterns.

Seven models of daily SLP (1980-1999) were used to simulate the occurrence of weather modalities. Data preprocessing was performed separately for the large and small domains.

SLP temporal SLP: Daily SLP value for each grid point minus the daily SLP average over a 20-year base period.

SLP spatial SLP: daily SLP value at each grid point, minus the average daily SLP value for the entire region (allowing the study to focus on the barometric gradient).

Experimental results: (shown in Fig. 4) Temporal SLP calculations were performed on the large and small domains (Fig. 4a, b, e, f), and spatial SLP calculations were performed only on the small domain (Fig. 4c, d).

a) First, the results of temporal SLP are analyzed: in winter, the East Asian continent is under the control of the Mongolian-Siberian high pressure. Most of the SOM nodes represent intensity features of the Mongolian-Siberian high-pressure anomaly.

In summer, the East Asian continent is under the control of the Indian low pressure, while the ocean is under the Pacific subtropical high pressure. Most of the SOM nodes also exhibit the anomalous intensity signature of the Pacific subtropical high pressure.

Among them, the amplitude range is smaller in summer than in winter (large domain: winter -25-20, summer -15-10; small domain: winter -15-20; summer -15-10).

The spring and fall SOM nodes also depict similar modes. temporal SLP represents the daily anomalies relative to the 20-year annual average, and more information is given by spatial SLP.

b) The results of spatial SLP are then analyzed: Fig. 4c depicts the dominant winter weather patterns that allow us to track the development of winter cyclones. Most of the patterns are characterized by the Mongolian-Siberian high pressure located in the East Asian continent, such as nodes (1, 1), (2, 1), (3, 1), (4, 1). And there is a north-south movement of the high-pressure system: (1, 1), (2, 1), (3, 1), (4, 1) are more to the south, while (1, 4), (2, 4), (3, 4), (4, 4) are more to the north.

In summer, the circulation type is mainly influenced by the Pacific subtropical high pressure, and the strength of the anticyclone and pressure is generally lower than in winter (-20 to 25 in winter and -15 to 10 in summer).

c) During the transition seasons of spring and fall, the resulting weather patterns combine the characteristics of the winter patterns with those of the summer patterns (not shown in Fig. 4).

After creating the SOM representing the SLP anomaly characteristics of each model, the next task is to evaluate the simulation performance of each model. Our evaluation is based on the premise that a good model will reproduce the actual atmospheric weather patterns at the same location and at the same frequency. Accordingly, the success of a model's simulation depends on the magnitude of the correlation coefficient between its occurrence frequency and the corresponding occurrence frequency in the reference field.

Next, we can calculate the frequency of each SOM node. In Fig. 5, the frequency of occurrence of each node of ERA-40, NCEP, and GCM (here again, MIROC5 is used as an example) is plotted. Fig. 5a, c correspond to Fig. 4a, and Fig. 5b, d correspond to Fig. 4c.

In Fig. 5a-c (temporal SLP), the frequency correlation coefficients are (ERA-40) 0.74 (significant positive correlation) and (NCEP) -0.51 (significant negative correlation), respectively, indicating that the frequency of modal occurrence of GCMs in a given season during the reference period is often the same with the reference field (ERA-40), but is often the opposite with the reference field (NCEP). In other words, the temporal SLP modalities that frequently occur in the reference field (ERA-40) may also occur in MIROC5, but the temporal SLP modalities that frequently occur in the reference field (NCEP) may not necessarily occur in MIROC5.

In Fig. 5b-d (spatial SLP), the frequency correlation coefficients are 0.44 and 0.30 (weak correlation), respectively, indicating that the case of spatial SLP is also the case where the frequency of modal occurrence in GCM is almost the same as that in the reference field. The results of the correlation coefficients for the small domains are shown in Table 2 (correlation coefficients greater than 0.5 are bolded) for the changing seasons, and the values of the correlation coefficients vary considerably with mode and season.

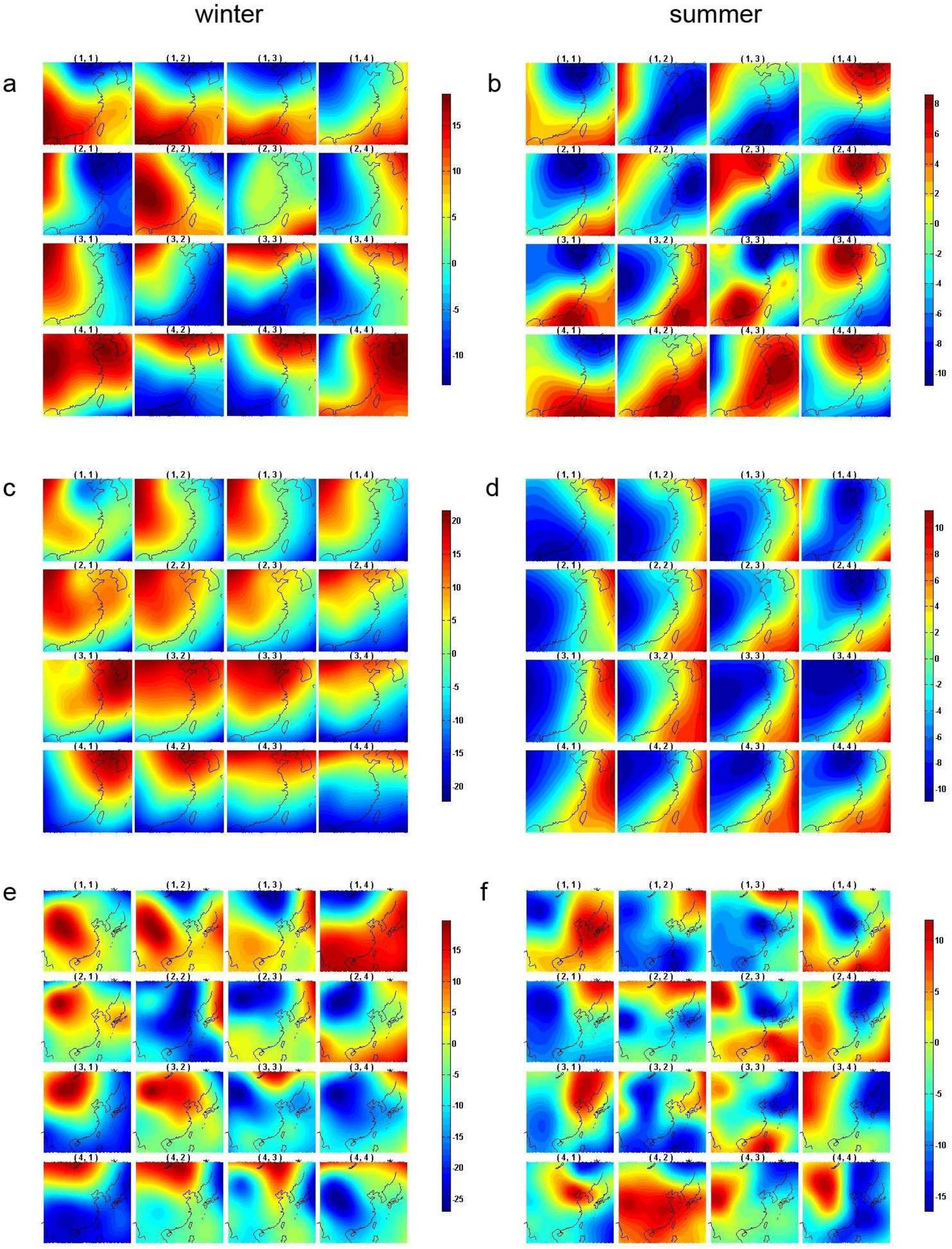

Fig. 4. The 4 ×4 SOMs of SLP anomalies (hPa) trained from Reference (ERA-40, NCEP) and one GCM (MIROC5) over the baseline period 1980–99. Patterns of temporal SLP anomalies over the small domain (a) in winter (DJF) and (b) in summer (JJA). Patterns of spatial SLP anomalies over the small domain (c) in winter and (d) in summer. Patterns of temporal SLP anomalies over the large domain (e) in winter and (f) in summer

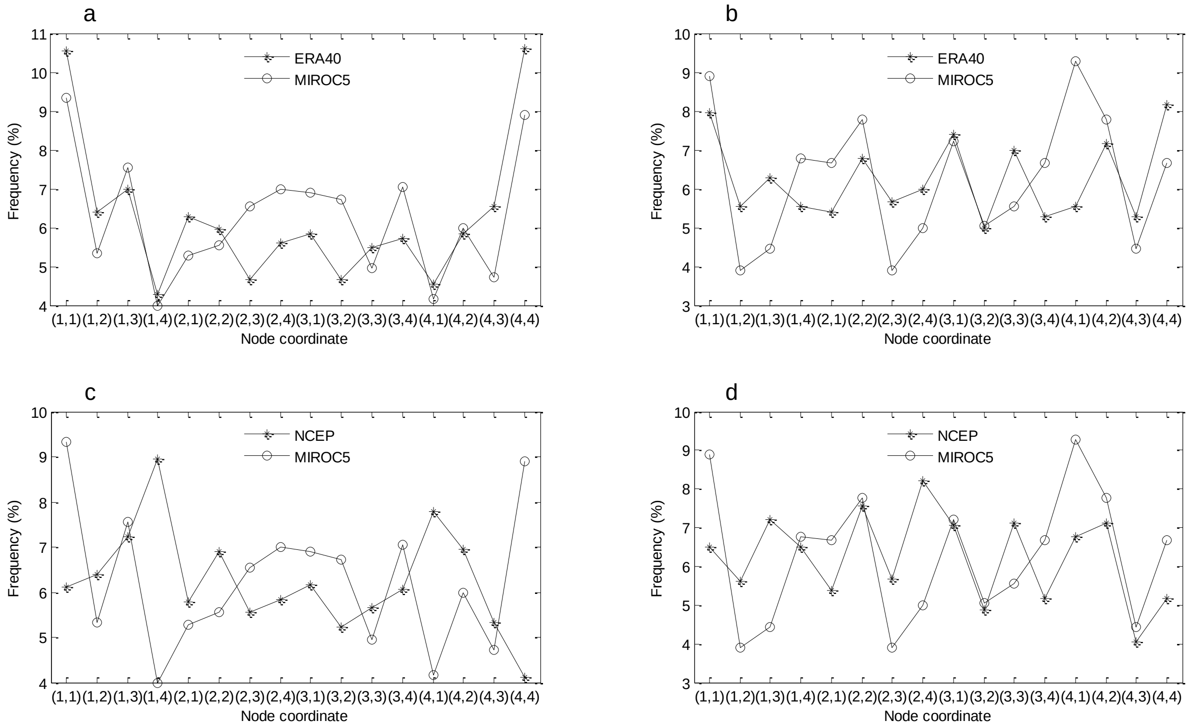

Fig. 5. Comparison of model performance for temporal and spatial SLP anomalies. Reference and GCM (MIROC5) frequency of occurrences (%) for each node on the 4×4 SOM of winter SLP anomalies over the small domain. (a) Node coordinates correspond to the node coordinates of SOM in Fig. 4a; (b) Node coordinates are from SOM in Fig. 4c

Table 2. Correlation coefficients r between node frequencies of 4×4 SOM in the Reference and each GCM, on seasonal basis (DJF, MAM, JJA, and SON). SOMs are given for the temporal and spatial SLP anomalies over the small domain. Bold font marks the correlations significantly>0 (at the 95% confidence level)

Reference Data	Model	Temporal SLP patterns				Spatial SLP patterns			
		MAM	JJA	MAM	JJA	MAMJJA	MAM	JJA	
ERA-40	MPI-ESM-LR	0.76	0.74	0.47	0.51	0.31	0.05	0.65	0.68
	MPI-ESM-MR	0.56	0.68	0.40	0.54	0.60	0.50	0.52	0.50
	MPI-ESM-P	-0.55	0.74	-0.15	0.35	0.25	0.37	0.53	0.67
	MIROC5	0.32	0.70	0.58	0.74	0.51	-0.23	0.45	0.44
	inmcm4	0.76	0.64	0.55	0.18	0.10	-0.13	0.34	0.45
	GFDL-ESM2G	0.67	0.70	0.67	0.76	0.50	0.42	0.39	0.37
	GFDL-ESM2M	0.72	0.62	0.71	0.65	0.50	0.64	0.55	0.49
NCEP	MPI-ESM-LR	0.54	0.40	0.45	-0.17	0.03	0.15	0.73	0.42
	MPI-ESM-MR0.45	0.43	0.48	0.01	0.48	0.46	0.23	0.38	
	MPI-ESM-P	-0.37	0.47	-0.40	0.30	0.35	0.54	0.33	0.56
	MIROC5	0.29	0.43	0.63	-0.51	0.61	0.05	0.32	0.30
	inmcm4	0.74	0.56	0.36	0.44	0.05	-0.21	0.54	0.24
	GFDL-ESM2G	0.43	0.66	0.71	-0.22	0.45	0.58	0.16	0.41
	GFDL-ESM2M	0.59	0.60	0.67	-0.23	0.61	0.52	0.08	0.57

Change the size of the SOM (4×4 → 3×4, 5×4) and perform the same calculation (Table 3).

a) For the temporal SLP, there were a large number of significant positive correlations for all seasons and for all SOM sizes calculated. Specifically, in the small domain, the proportions of significant positive correlations (r > 0.50) for the four seasons were: (ERA-40) spring (47.62%), summer (61.90%), autumn (38.10%), and winter (47.62%); (NCEP) spring (23.81%), summer (19.05%), autumn (23.81%), and winter (28.57%). In the large domain: (ERA-40) spring (52.38%), summer (47.62%), fall (52.38%), and winter (28.57%); (NCEP) spring (28.57%), summer (57.14%), fall (33.33%), and winter (47.62%).

In the small domain, the four seasons were significantly positively correlated (r > 0.20): (ERA-40) spring (80.95%), summer (80.95%), fall (95.23%), and winter (80.95%); (NCEP) spring (66.67%), summer (57.14%), fall (57.14%), and winter (57.14%). In the large domain: (ERA-40) spring (95.24%), summer (100.00%), fall (76.19%), winter (71.43%), (NCEP) spring (95.24%), summer (100.00%), fall (80.95%), winter (76.19%).

b) For spatial SLP, the four seasons were significantly positively correlated (r > 0.50) in the following proportions: (ERA-40) spring (47.62%), summer (23.81%), fall (33.33%), and winter (42.86%). (NCEP) Spring (19.05%), Summer (23.81%), Autumn (38.10%), and Winter (28.57%). (r > 0.20) The proportions were as follows: (ERA-40) spring (85.71%), summer (57.14%), fall (80.95%), and winter (95.24%). (NCEP) Spring (61.90%), Summer (61.90%), Fall (71.43%), and Winter (71.43%).

These large number of significant positive correlations indicated that, on a seasonal basis during the baseline period, almost every GCM could reproduce both temporal SLP abnormalities and spatial SLP abnormalities. Further analysis shows that the success of reproduction depends on the GCM model's seasonal simulation of the 20-year average, with the larger the differences in the simulations, the larger the differences in the frequency characteristics.

Table 3 After varying the size of the SOM (from 4 × 4 to 3 × 4 and 5 × 4) and performing the same calculations, there were a large number of significant positive correlations for all seasons and for all SOM sizes

r	Domain	ERA-40 (%)				NCEP (%)			
		MAM	JJA	SON	DJF	MAM	JJA	SON	DJF
r>0.5	Temporal SLP(Large domain)	52.38	47.62	52.38	28.57	28.57	57.14	33.33	47.62
	Temporal SLP(Small domain)	47.62	61.90	38.10	47.62	23.81	19.05	23.81	28.57
	Spatial SLP(Small domain)	47.62	23.81	33.33	42.86	19.05	23.81	38.10	28.57
r>0.2	Temporal SLP(Large domain)	95.24	100.00	76.19	71.43	95.24	100.00	80.95	76.19
	Temporal SLP(Small domain)	80.95	80.95	95.23	80.95	66.67	57.14	57.14	57.14
	Spatial SLP(Small domain)	85.71	57.14	80.95	95.24	61.90	61.90	71.43	71.43

A sensitivity study of the calculated results: each GCM was moved forward 10 years, while the reference field was kept constant. The difference between the old and new correlation coefficients for the temporal SLP was calculated to be.

Large domain: the average difference in correlation coefficients over the four seasons from small to large. (ERA-40) Spring (-0.01), Fall (0.05), Summer (0.08), Winter (-0.06). (NCEP) Summer (-0.03), Spring (-0.09), Winter (-0.08), and Autumn (-0.10).

Small domains: the four-season average differences in correlation coefficients were from small to large. (ERA-40) Winter (0.20), Summer (0.11), Spring (0.12), Autumn (0.31). (NCEP) fall (-0.10), summer (-0.10), spring (-0.19), and winter (-0.19).

Meanwhile, the old and new differences in correlation coefficients for spatial SLP were.

Small domain: the average difference in correlation coefficients over the four seasons from small to large. (ERA-40) Spring (0.05), Fall (0.09), Summer (0.06), Winter (0.22). (NCEP) winter (0.03), spring (-0.04), summer (0.08), and fall (0.18).

Most of the differences between old and new correlation coefficients are very small, and the above results show that the simulation of the four seasons is insensitive to the switching of the 10-year base period. The results of the model evaluation were basically unchanged.

In order to rate the correlation between the frequency of occurrence of the two model nodes, the following set of evaluation measures is presented (RADIC´ AND CLARKE, 2011) [1].

a) Average correlation coefficient: Average correlation coefficient for all seasons and all SOM sizes.

$$M_C = \frac{1}{m}\sum_{i=1}^{m} r_i \tag{5}$$

where m is the product of the number of seasons and the type of size (4 seasons, 3 sizes, then m=4*3=12). r is the correlation coefficient. The higher the value, the better.

b) Define the cumulative sum of all significantly correlated numbers for the purpose of significant positive correlation of statistical totals.

$$M_S = \sum_{i=1}^{m} \delta_i \begin{cases} \delta_i = 1 & if \quad r_i \geq r_0 \\ \delta_i = 0 & if \quad r_i < r_0 \end{cases} \tag{6}$$

r_0 is the threshold for (t-test at 95% confidence level) significant correlation, the larger the M_S value the better.

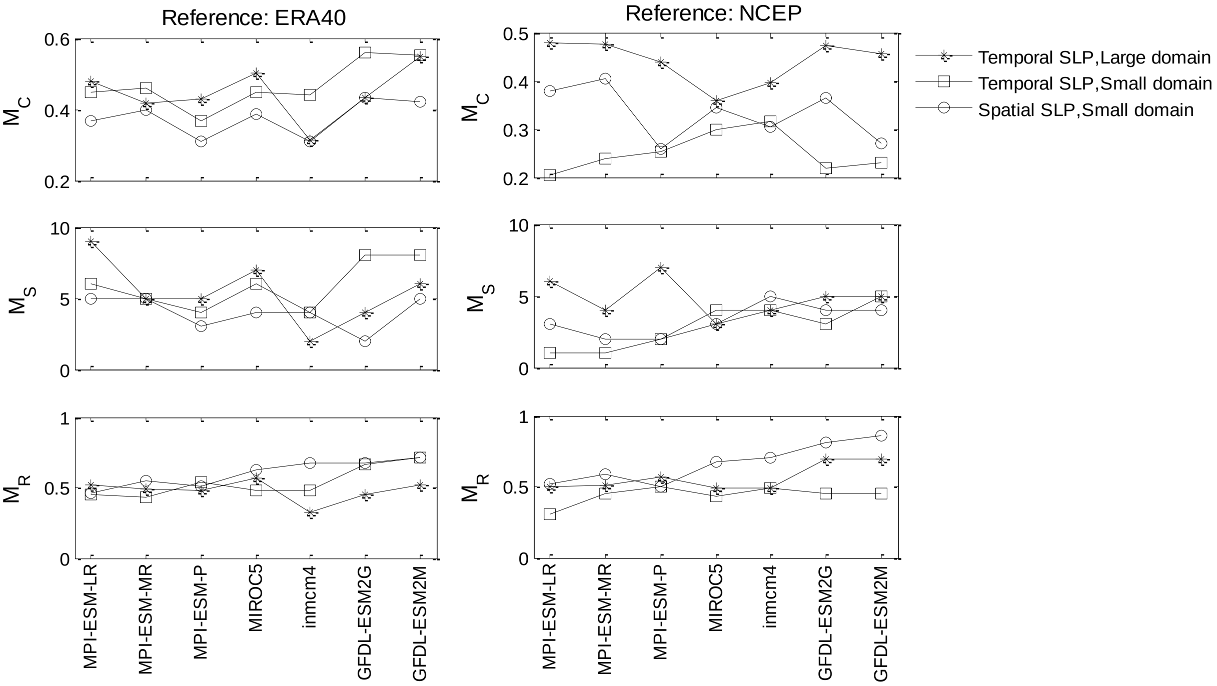

Fig. 6 (top to bottom) Correlation measure MC, significance measure MS, and rank measure MR for each model. The measures are derived for three different cases: SOMs of temporal SLP anomalies over the small domain, SOMs of temporal SLP anomalies over the large domain, and SOMs of spatial SLP anomalies over the small domain

c) Rating Indicators (Schuenemann and Cassano, 2009) [7]

$$M_R = 1 - \frac{1}{252}\sum_{i=1}^{m} rank_i \tag{7}$$

Meaning of rank: The best pattern has a rank value of 1 and the worst is 7. The higher the M_R, the better.

It is possible to evaluate time and space anomalies using and, while only the relative performance of the modes is given.

When ERA-40 is used as the reference field, the correlation between the three evaluation measures is high (r-means 0.84 and 0.80, respectively), i.e., the shape of each polyline is very close in the three plots, and very poor (r-means 0.07) from the Spatial SLP results.

When NCEP was used as the reference field, the correlation between the three evaluation measures became worse, with a mean r value of 0.44 and 0.46 for the Temporal SLP for the large and small domains, respectively, i.e., the shape of each polyline differed greatly among the three plots.

The above calculations show that the results of the seven-mode Temporal SLP simulations are in good agreement with the ERA-40 reanalysis data, while the results differ greatly from the NCEP reanalysis data. In addition, the differences between the Spatial SLP simulation results and the ERA-40 and NCEP reference fields are large.

In addition to the above, we can conclude that Temporal SLP has low sensitivity to the choice of the three different measures when ranking the model evaluations with ERA-40 as the reference field, and that the results of the Temporal SLP model rankings are insensitive to the choice of the domain size. The Spatial SLP results are sensitive to both ERA-40 and NCEP as the reference field.

Table 4 Correlation coefficients for the three indicators

Reference Data	Domain	r (M_C, M_S)	r (M_S, M_R)	r (M_R, M_C)
ERA-40	Temporal SLP(Large domain)	0.76	0.85	0.90
	Temporal SLP(Small domain)	0.91	0.75	0.72
	Spatial SLP(Small domain)	0.05	-0.21	0.37
NCEP	Temporal SLP(Large domain)	0.55	0.31	0.46
	Temporal SLP(Small domain)	0.48	0.40	0.50
	Spatial SLP(Small domain)	-0.23	0.72	-0.18

5. Conclusion

How to objectively and quantitatively assess and compare the simulation capabilities of different climate models for a particular climate phenomenon is becoming increasingly important. To address the shortcomings of previous model evaluation methods, this paper applies a self-organized mapping neural network-based weather-climate model evaluation method based on the weather-type classification. Using the reanalysis data of the ERA-40 and the NCEP/NCAR as the reference field, the simulation performance of seven CMIP5 climate models in East Asia during 1980-1999 is evaluated from several aspects. First, the annual mean cycle and inter-annual variability simulations are evaluated using climate indicators, and then the simulation of weather pattern occurrence using SOM technique is highlighted. The evaluation is then focused on the simulation of weather pattern occurrence using SOM techniques.

It is concluded that the choice of reference field is particularly important and decisive for the results of the climate model assessment, and that the results are better when ERA-40 is used as the reference field (as opposed to NCEP). Although the simulation results vary among different models, most of the models are able to reproduce the evolution of extreme climate events and predict the future extreme climate change effectively. The results of this study are useful for the study of the response to future extreme climate change in China under the greenhouse gas increase scenario.

References

[1] Radi´ c, V., and G. K. C. Clarke: Evaluation of IPCC models' performance in simulating late-twentieth-century climatologies and weather patterns over North America. J. Climate, (2011) No.24, p.5257–5274.

[2] Taylor, K. E., R. J. Stouffer, and G. A. Meehl: An overview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc., (2012) No.93, p485–498.

[3] Kalnay, E., M. Kanamitsu, R. Kistler, et al.: The NCEP/NCAR 40-year reanalysis project. Bull. Amer. Meteor. Soc., (1996) No.77, p437–471.

[4] Simmons, A. J., and J. K. Gibson: The ERA-40 project plan, ERA-40 Project Report Series (2000) No. 1, p62.

[5] Kohonen, T.: Self-organized formation of topologically correct feature maps. Biological Cybernetics, (1982) No.43, p59–69.

[6] Gleckler, P. J., K. E. Taylor, and C. Doutriaux: Performance metrics for climate models, J. Geophys. Res., (1996) No.113, D06104.

[7] Schuenemann, K. C., and J. J. Cassano: Changes in synoptic weather patterns and Greenland precipitation in the 20th and 21st centuries. 1: Evaluation of late 20th century simulations from IPCC models. J. Geophys. Res., (2009), p114.