

A score based method for assessing the performance of GCMs in the Yellow-Huai-Hai region

Fangxin Shi¹, Zhihui Wang¹, Liang Qi², Rongxu Chen¹, Shuxia Li¹, Zhizong Tian¹

¹ Yellow River Institute of Hydraulic Research, Zhengzhou 450003, China

2. The Office of Yellow River Flood Control and Drought Relief Headquarters, Zhengzhou 450003, China:

Keywords: multi-criteria rank score, assessment, GCMs, the Yellow-Huai-Hai region

Abstract: A multi-criteria score-based method has been developed to assess General Circulation Models (GCMs) skill in the Yellow-Huai-Hai region for 1970-2005. The rank score of these criteria are applying to comprehensively assess the temporal and spatial performance of precipitation and temperature of 18 GCMs in the study region. The results indicate that: GCMs usually simulate temperature better than rainfall. The temporal and spatial distribution of simulated temperature performed well when compared with the observed data. Compared with temperature, the spatial distribution of simulated precipitation performed poor. All the GCMs underestimate the temperature and overestimate the precipitation in the study region. The method developed in this study could easily be extended to different study regions and results can be used for better informed regional climate change impact analysis.

Introduction

General circulation models (GCMs) are the most common tools for projecting the future climate changes. Errors and uncertainties in GCMs metadata afflict the entire intensity spectrum, causing especially the inability to regenerate the observed meteorological events. GCMs simulation is often characterized by biases and uncertainties that limit their direct application^[1]. Uncertainty is the major shortfall of GCMs for assessing regional impacts of change. The different forcing scenarios, different GCMs and different sub-grid scale forcing and processes caused the uncertainties which limit the application of GCMs^[2]. Despite continuous works to improve the GCMs' simulation capacity, the application of assessment methods is essential for the impact studies of climate change^[3].

To reduce the uncertainties in the GCMs application, the GCMs have been assessed in many researches [4-6]. The assessment places emphasis on various aspects of GCMs according to its different application. For example, the long-term climate change analysis is the main work in a study, the assessment of GCMs performance before its application is focused on the temporal and spatial distribution simulations. However, the drawback of the assessment which use a single criterion is that the assessment could not general describe the skill of GCMs^[4]. More criteria in the assessment process, we could have a more comprehensive understanding of the advantage and disadvantage for the GCMs.

In existing studies, there is still no an assessment method of GCMs has been accepted by all researchers. How to assess the skill of GCMs before using them is becoming an interesting question. In this paper, a multi-criteria score-based method has been analyzed and all the performance of GCMs has been quantitatively calculated and examined. We will study this method with the aim to understand and capture the possible characteristics of future climate change.

Study Region and Data Set

The Yellow-Huai-Hai region is a prime agricultural and industrial region in China. It therefore plays an important role in social and economic development for the country. All these consequences of climate change will seriously restrain economic growth^[7, 8].

All the GCMs data come from the Fifth Phase of the Coupled Model Intercomparison Project



(CMIP5). This data set is the most important tool for analyzing future climate change ^[9]. To simulate future climate change, 18 global general circulation models (GCMs) of CMIP5 are considered in this paper (Table 2-4). Since GCM horizontal resolution varies, the GCM outputs were interpolated to a uniform resolution of 2.5°×2.5°. The grid cells distribution over the study region shows in Figure 1. The precipitation data with good quality and continuous records for the period 1970–2005 are obtained from 128 meteorological stations (Figure 1). To assessment the GCMs skill well, the daily data which observed by the meteorological stations were accumulated to monthly data and interpolated by the inverse distance weighted method to the 2.5°×2.5°cells.

Methods

In this study, a multi- criteria score-based method has been developed to assess the skill of GCMs data in the regional scale. The criteria including mean annual data, standard deviation, annual cycle of the climate, normalized root mean square error, spatial distribution, climate change trend, Empirical Orthogonal Function and Probability Density Function are listed in Table 1.

In the assessment, a rank score value (RS) of 0-9, which has been used to assess each individual assessment criterion, is written in the following form:

$$RS_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \times 9 \tag{1}$$

where x_i is the relative error between the *i*th GCM result and observation or the related statistical value for the ith GCM. As the relative error, the larger x_i means the larger rank score in the GCMs performance assessment. And then total rank score of each GCM's is summated by the rank score of all criteria. Some different criteria which is belong to same statistic character, such as Mann-Kendall test Z and Trend magnitude β for trend and its magnitude analysis, EOF1 and EOF2 for Empirical orthogonal function (EOF) analysis, and the BS and Sscore of PDF (will described later), are weighted 0.5 each in this summation (Table 1) while the other criteria have a 1.0 weight. Relative error (RE) was used to quantify the similarity between simulated and observed values for long-term monthly mean and standard deviation. The skills of bias-corrected time series of water vapor budget in GCMs are evaluated by the NRMSE (normalized root-mean square error)[10, 11]. The correlation coefficient of annual cycle was calculated between observed and modeled long-term monthly mean values. Mann-Kendall test and trend magnitude method were applied to detect the long-term monotonic annual trends and quantify their magnitudes [12]. The trend magnitude b of sen's slope, is the median over all possible combinations of pairs for the whole data set $[7, \frac{12}{2}, \frac{13}{2}]$. The relative error was used to assess how close the Z statistic and b magnitude of each GCM are to the observed values.

Empirical orthogonal function (EOF) analysis in this study was used to compare the spatial distribution difference of modeled climate variable and the observations [14, 15]. Because the first two leading modes of each EOF is accounting for the majority of the total variance, they have been used in this study to compare the observation and modeling data.

The Brier score (BS) and Significance score (Sscore) are used for assess the monthly climate variables probability density functions of GCMs.

$$BS = \frac{1}{n} \sum_{i=1}^{n} (P_{mi} - P_{oi})^2$$
 (2)

$$Sscore = \sum_{i=1}^{n} Minimum(P_{mi}, P_{oi})$$
 (3)

where P_{mi} and P_{oi} are the simulated and observed *i*th probability values of each bin and *n* (set as 16) is the number of bins. In this study, BS is the mean squared error measure for probability forecasts^[16, 17], and the Sscore is the calculated cumulative minimum value of observed and simulated distribution for each bin which could quantifying the overlap between the observed and simulated data^[18, 19]. Therefore, the BS of GCMs is smaller and the Sscore is bigger, the performance of GCMs is better.



Results

1) The assessments of temperature

Table 2 is the assessment results of temperature performance in the Yellow-Huai-Hai region. It can be clearly seen that the mean temperature in the historic period in the study region is 8.49 □, while it is 3.62-8.09 □ of GCMs in the same period. Most GCMs underestimate the mean temperature and the standard deviation. The smaller NRMSE results indicate the better simulation for the set of data. For the monthly mean temperature, the best NRMSE for the GCMs is miroc-esm (0.22), while the NRMSE result of ipsl-cm5b-lr has the largest value in these GCMs. The simulated monthly distribution in annual cycle for each GCMs is quite similar with the observed data, which could be seen from the correlation index (all of them is large than 0.995). As a consequence, the correlation results of monthly distribution in annual cycle have been rounded to 1. The spatial temperature distribution correlation coefficients between each GCMs and observations are also larger than 0.9, which means that all GCMs could well describe the spatial distribution of mean temperature in the Yellow-Huai-Hai region.

The temperature in the Yellow-Huai-Hai region is grown in the last 40 years. The Z value in Mann-Kendall test of observed temperature is 4.81 which means that the observed mean temperature is significant increased in the 0.05 significance level. However, most GCMs underestimate the trend of temperature change in this region. The trend magnitude b in sen's slope shows the similar results.

The results of analyzing the spatial temperature by using the EOF show that the first and the second vector monthly temperature EOFs of observations account for 98.9% and 0.51% of the total variance (Table 2), respectively. Actually, the result provides that the GCMs well perform the physical process of temperature variability. Empirical cumulative probability distribution (Figure 2) shows that the empirical cumulative probability distributions of monthly mean temperature which simulated by most GCMs are quite close to the observation (except the inm-cm4 and ipsl-cm5b-lr). The results of BS and Sscore are consistent with the Figure 2 that the relatively poorer performances of the inm-cm4 and ipsl-cm5b-lr models are confirmed by the larger BS and Sscore. The skill of GCMs has been evaluated and the final score of each model has been calculated. The ccsm4 model gets the highest score while the inm-cm4 gets the lowest score. Figure 3 describes the difference of annual temperature change between the observation and the best and worst skill models in the Yello-Huai-Hai region. We can see clearly that even though the ccsm4 model has underestimates the mean temperature, the model simulates a similar change trend with observations. In contrast, inm-cm4 model has vastly underestimates and simulates a wrong temperature change with the observations.

2) The assessments of precipitation

Table 3 is the assessment results of the skill of GCMs for precipitation. Compare with the temperature, the GCMs have a poor performance for the precipitation. Most GCMs overestimate the values of precipitation and the standard deviations. Especially the precipitation of bun-esm model is 1256mm, which is twice more than the observations. The value of NRMSE for precipitation is much larger than it for temperature. The correlation coefficient of monthly precipitation in annual cycle of GCMs is less than it for temperature but most value of the correlation coefficients is still more than 0.9. The spatial correlation coefficient of GCMs only between 0.45-0.82, which means the simulation of GCMs in spatial precipitation distribution is much worse than it in temperature. By using Mann-Kendall test, the precipitation of GCMs shows different change trends which indicate that the simulated precipitation of GCMs is much uncertainty than temperature. Because the physical mechanism of precipitation occurrence is mainly influenced by the large-scale circulation factors, the uniform spatial distribution of simulated precipitation indicates that it could not explain the influence of circulation factors well.

The empirical cumulative probability distribution of GCMs monthly precipitation has compared with the observation in Figure 4. Most GCMs overestimate the high precipitation in the probability distribution, which is consistent with the result of annual precipitation analysis. The result of higher



BS and lower Sscore also indicate that the temperature simulation of GCMs is better than precipitation.

After we assess the performance of all GCMs by using the Ranking Score, the results have been shown in the Table 3. The csiro-mk3.6 has a best performance while the bnu-esm is the worst. Figure 5 describes the annual precipitation change in the study region. The bnu-esm model vastly overestimates the annual rainfall in the study region and the model seems have a reverse change when compared with the observation. Even the csiro-mk3.6 also appears that it slightly overestimate the annual precipitation in the study area and has a different fluctuant change with observation before 70s, the csrio-mk3.6 model has a quite similar fluctuant change with the observed data after 1975.

According to the Table 4, the simulation of GCMs for temperature is better than precipitation. For the same GCM model, it will provide a different result for the GCM model performance if it had been assessed with the different climate factor, such as the bnu-esm model, it is the 4th better model for simulating the temperature but it is the worst model in the 18 GCMs for simulating the precipitation.

Conclusion

In this paper, a multi-criteria score-based method has been developed to assess GCMs performance in the Yellow-Huai-Hai region. The rank score of these criteria are applying to comprehensively assess the temporal and spatial performance of precipitation and temperature of 18 GCMs in the study region.

All GCMs have a good performance in simulating temperature. Although all the models have underestimated the mean temperature, the result of temporal and spatial distributions is quite close to the observation. The skill of GCMs in simulating precipitation is worse than it in simulating temperature. Most GCMs overestimate the mean precipitation in the study area.

Acknowledgements:

We are very grateful to three anonymous reviewers for their very constructive comments which significantly improved the manuscript. We are grateful to the climate modeling groups for producing and making available their model output to the public. This work was financially supported by the Technology Development Foundation of the Yellow River Institute of Hydraulic Research (HKF201604), the National Natural Science Foundation of China (51409116), the Special Research Fund of Yellow River Institute of Hydraulic Research (HKY-JBYW-2016-05) and the Special Research Fund of Yellow River Institute of Hydraulic Research (HKY-JBYW-2016-08).

Reference

- [1] Sharma D, Das Gupta A, Babel MS. Spatial disaggregation of bias-corrected GCM precipitation for improved hydrologic simulation: Ping River Basin, Thailand[J]. Hydrology and Earth System Sciences, 2007, 11(4): 1373-1390.
- [2] Giorgi Filippo, Francisco Raquel. Evaluating uncertainties in the prediction of regional climate change[J]. Geophysical Research Letters, 2000, 27(9): 1295-1298.
- [3] Seneviratne SI, Nicholls N, Easterling D, et al. Changes in climate extremes and their impacts on the natural physical environment: An overview of the IPCC SREX report[A]. EGU General Assembly Conference Abstracts[C], Vol. 14, 2012, p. 12566.
- [4] Fu Guobin, Liu Zhaofei, Charles Stephen P, et al. A score-based method for assessing the performance of GCMs: A case study of southeastern Australia[J]. Journal of Geophysical Research: Atmospheres, 2013, 118(10): 4154-4167.
- [5] Wang Wanqiu, Kumar Arun. A GCM assessment of atmospheric seasonal predictability associated with soil moisture anomalies over North America[J]. Journal of Geophysical Research, 1998, 103(103): 28637–28646.
- [6] Marengo J. A., Cavalcanti I. F. A., Satyamurty P., et al. Assessment of regional seasonal rainfall



- predictability using the CPTEC/COLA atmospheric GCM[J]. Climate Dynamics, 2003, 21(5-6): 459-475.
- [7] Fu Guobin, Chen Shulin, Liu Changming, et al. Hydro-climatic trends of the Yellow River basin for the last 50 years[J]. Climatic Change, 2004, 65(1-2): 149-178.
- [8] Zhang Y., Arthington A. H., Bunn S. E., et al. Classification of flow regimes for environmental flow assessment in regulated rivers: The Huai River Basin, China[J]. River Research & Applications, 2012, 28(7): 989-1005.
- [9] Taylor Karl E, Stouffer Ronald J, Meehl Gerald A. An overview of CMIP5 and the experiment design[J]. Bulletin of the American Meteorological society, 2012, 93(4): 485-498.
- [10] Hanna Steven R, Heinold David W. Development and application of a simple method for evaluating air quality models[M]. American Petroleum Institute, 1985.
- [11] Randall David A, Wood Richard A, Bony Sandrine, et al. Climate models and their evaluation[M]. Climate Change 2007: The physical science basis Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR). Cambridge University Press. 2007: 589-662.
- [12] Hirsch Robert M, Slack James R, Smith Richard A. Techniques of trend analysis for monthly water quality data[J]. Water Resources Research, 1982, 18(1): 107-121.
- [13] Sen Pranab Kumar. Estimates of the regression coefficient based on Kendall's tau[J]. Journal of the American Statistical Association, 1968, 63(324): 1379-1389.
- [14] Harvey LDD, Wigley TML. Characterizing and comparing control-run variability of eight coupled AOGCMs and of observations. Part 1: temperature[J]. Climate Dynamics, 2003, 21(7-8): 619-646.
- [15] Mu Qiaozhen, Jackson Charles S, Stoffa Paul L. A multivariate empirical-orthogonal-function-based measure of climate model performance[J]. Journal of Geophysical Research: Atmospheres 2004, 109: D15101, doi:15110.11029/12004JD004584.
- [16] Brier Glenn W. Verification of forecasts expressed in terms of probability[J]. Monthly Weather Review, 1950, 78(1): 1-3.
- [17] Fraedrich K, Leslie LM. Evaluation of techniques for the operational, single station, short-term forecasting of rainfall at a midlatitude station (Melbourne)[J]. Monthly Weather Review, 1987, 115(8): 1645-1654.
- [18] Perkins SE, Pitman AJ, Holbrook NJ, et al. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions[J]. Journal of Climate, 2007, 20(17): 4356-4376.
- [19] Watterson Ian G. Calculation of probability density functions for temperature and precipitation change under global warming[J]. Journal of Geophysical Research: Atmospheres (1984–2012), 2008, 113: D12106, doi:12110.11029/12007JD009254.



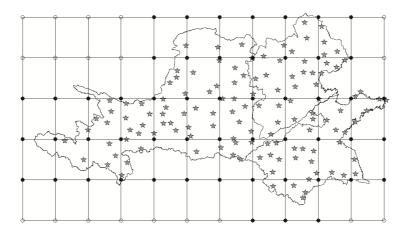


Figure 1. Location of meteorological stations and the selected GCMs grids (solid circle) in the study region

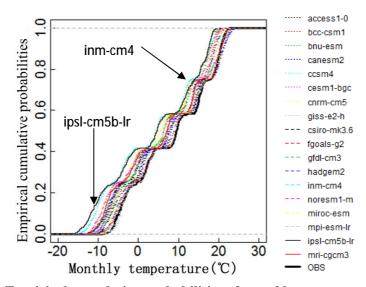


Figure 2. Empirical cumulative probabilities of monthly mean temperature

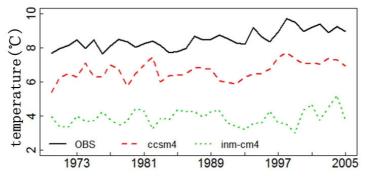


Figure 3. The annual temperature change in the Yello-Huai-Hai region



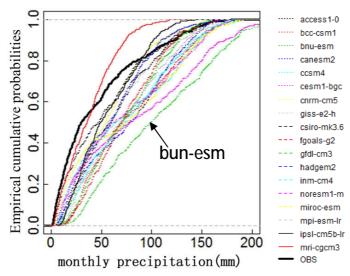


Figure 4. Empirical cumulative probabilities of monthly mean precipitation

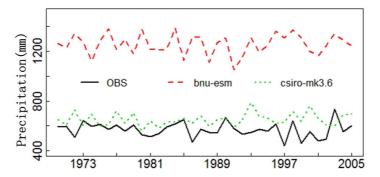


Figure 5. The annual precipitation change in the Yello-Huai-Hai region

Table 1. Statistics of Climate Variables and Their Weights

Statistics of Climate Variables	Criteria	Weights
Mean	Relative error (%)	1
Standard deviation	Relative error (%)	1
Temporal change	NRMSE	1
Monthly distribution	Correlation coefficient	1
Spatial distribution	Correlation coefficient	1
Trend and its magnitude	Mann-Kendall test Z	0.5
Trend and its magnitude	Sen's magnitudeβ	0.5
Space time verichility	EOF1(first vector)	0.5
Space-time variability	EOF2(second vector)	0.5
Probability density	BS (Brier score)	0.5
functions (PDFs)	Sscore (Significance score)	0.5



Table 2. Model Performance for Monthly Mean Temperature

	GCMs	Mean			Monthly	Spatial -	Ma	nn-Kendall	E	OF	P	DF	Total
GCMs	number	Value S (°C)	StDev	NRMSE	Corr	Corr	Zc	Slope(℃/10a)	EOF1	EOF2	BS	Sscore	RS
obs		8.49	0.53				4.81	0.37	98.90	0.51			
access1-0	1	6.81	0.47	0.23	1.00	0.95	3.64	0.28	97.46	0.74	0.05	85.19	16.17
bcc-csm1	2	5.67	0.38	0.33	1.00	0.95	4.40	0.23	97.66	1.19	0.08	82.64	29.98
bnu-esm	3	6.63	0.62	0.27	1.00	0.94	3.69	0.36	98.26	0.64	0.05	85.65	17.56
canesm2	4	7.77	0.51	0.18	1.00	0.93	5.38	0.35	97.66	1.15	0.09	82.41	19.57
ccsm4	5	6.65	0.54	0.27	1.00	0.96	2.98	0.27	98.40	0.63	0.03	90.05	11.91
cesm1-bgc	6	6.79	0.45	0.25	1.00	0.96	2.55	0.20	98.43	0.72	0.03	88.66	15.17
cnrm-cm5	7	5.07	0.46	0.41	1.00	0.95	1.38	0.10	98.03	0.76	0.14	75.00	30.23
giss-e2-h	8	7.12	0.42	0.25	1.00	0.94	-0.10	-0.01	97.95	0.99	0.07	81.94	31.04
csiro-mk3.6	9	6.76	0.42	0.26	1.00	0.93	3.69	0.25	98.52	0.68	0.05	87.04	21.89
fgoals-g2	10	7.49	0.58	0.19	1.00	0.95	4.59	0.44	97.73	1.19	0.35	57.41	24.71
gfdl-cm3	11	5.33	0.61	0.36	1.00	0.97	2.96	0.27	97.74	0.94	0.07	82.87	22.03
hadgem2	12	7.20	0.55	0.26	1.00	0.96	3.17	0.23	97.84	0.64	0.22	67.13	18.19
inm-cm4	13	3.90	0.47	0.53	1.00	0.93	1.13	0.09	96.99	1.08	0.08	85.42	43.78
noresm1-m	14	5.58	0.47	0.35	1.00	0.95	4.24	0.31	98.35	0.75	0.10	81.71	20.93
miroc-esm	15	7.33	0.29	0.22	1.00	0.95	2.60	0.13	97.80	1.23	0.17	72.69	32.88
mpi-esm-lr	16	8.09	0.36	0.16	1.00	0.96	2.90	0.19	98.06	0.93	0.09	78.47	17.92
ipsl-cm5b-lr	17	3.62	0.47	0.55	1.00	0.93	2.17	0.15	98.47	0.59	0.10	80.09	36.20
mri-cgcm3	18	6.26	0.41	0.32	1.00	0.96	2.03	0.13	98.63	0.55	0.13	77.08	21.30

^{*} EOF 1 and 2 is the percentage of the explained variance for the first two leading modes of each EOF; PDF is the probability density function.

Table 3. Model Performance for precipitation

	GCMs	Mean			Monthly	Spatial	M	ann-Kendall	E	OF	PI)F	Total
	number		StDev	NRMSE	Corr	Corr	Zc	Slope(mm/10a)	EOF1	EOF2	BS	Sscore	RS
obs	0	568.2	61.5				-0.86	-11.76	71.72	8.69			
access1-0	1	775.0	62.3	0.72	0.95	0.82	-0.34	-2.63	53.21	17.70	0.31	75.00	15.91
bcc-csm1	2	955.3	67.5	1.00	0.92	0.61	0.12	2.65	61.31	14.08	0.51	65.51	32.25
bnu-esm	3	1255.9	83.4	1.50	0.94	0.45	0.12	1.78	60.56	16.00	0.53	63.89	50.75
canesm2	4	914.3	75.5	1.04	0.85	0.72	1.46	21.48	49.79	21.16	0.36	68.06	41.91
ccsm4	5	891.0	65.6	0.92	0.93	0.70	-1.10	-13.43	60.89	13.03	0.20	78.01	18.57
cesm1-bgc	6	888.2	50.7	0.90	0.93	0.70	-2.30	-19.45	58.82	13.27	0.21	78.01	23.74
cnrm-cm5	7	703.6	71.8	0.65	0.96	0.80	1.21	14.66	55.57	16.66	0.40	69.68	23.16
giss-e2-h	8	802.2	77.5	0.80	0.91	0.57	0.23	2.04	64.49	15.18	0.47	70.83	31.15
csiro-mk3.6	9	650.6	51.7	0.54	0.98	0.75	1.08	10.45	62.30	10.64	0.08	84.26	11.73
fgoals-g2	10	881.0	60.3	1.01	0.78	0.81	-0.59	-6.11	46.99	25.29	0.56	62.50	35.99
gfd1-cm3	11	879.7	65.72	0.88	0.92	0.52	-1.13	-16.57	57.97	19.90	0.53	63.66	32.17
hadgem2	12	725.4	64.31	0.70	0.91	0.76	1.16	9.07	49.51	18.78	0.29	73.38	23.70
inm-cm4	13	923.5	62.61	0.90	0.95	0.68	1.30	13.30	62.19	13.17	0.36	72.22	25.68
noresm1-m	14	1035.7	65.13	1.12	0.94	0.54	-0.80	-5.93	70.20	10.14	0.18	77.78	23.93
miroc-esm	15	898.3	68.81	1.04	0.83	0.57	-0.18	-1.78	65.68	13.81	0.17	74.07	30.06
mpi-esm-lr	16	887.8	63.13	0.95	0.89	0.66	0.45	3.98	55.38	18.05	0.35	71.99	29.70
ipsl-cm5b-lr	17	674.1	66.00	0.71	0.90	0.71	2.39	22.50	49.97	14.25	0.45	69.44	29.99
mri-cgcm3	18	456.1	33.58	0.72	0.93	0.76	-0.15	-1.38	48.21	19.47	0.19	76.85	26.22

^{*}EOF 1 and 2 is the percentage of the explained variance for the first two leading modes of each EOF; PDF is the probability density function.

Table 4. The scores of GCMs for monthly temperature and precipitation

	ID	temperature	precipitation	Total
access1-0	1	16.17	15.91	32.08
bcc-csm1	2	29.98	32.25	62.23
bnu-esm	3	17.56	50.75	68.31
canesm2	4	19.57	41.91	61.48
ccsm4	5	11.91	18.57	30.48
cesm1-bgc	6	15.17	23.74	38.91
cnrm-cm5	7	30.23	23.16	53.39
giss-e2-h	8	31.04	31.15	62.19
csiro-mk3.6	9	21.89	11.73	33.62
fgoals-g2	10	24.71	35.99	60.70
gfdl-cm3	11	22.03	32.17	54.20
hadgem2	12	18.19	23.7	41.89
inm-cm4	13	43.78	25.68	69.46
noresm1-m	14	20.93	23.93	44.86
miroc-esm	15	32.88	30.06	62.94
mpi-esm-lr	16	17.92	29.7	47.62
ipsl-cm5b-lr	17	36.20	29.99	66.19
mri-cgcm3	18	21.30	26.22	47.52