Influence of Normalization Techniques in CMIP Model Selection Using an MCDM Method MOORA

Gaurav Patel, Subhasish Das, and Rajib Das

1 Introduction

Researchers have been motivated to develop new techniques in response to the requirement for a variety of decision-making procedures for addressing various model selection difficulties. The use of multi-criteria decision-making (MCDM) methodologies has the potential to produce superior results. Finding a good model is critical for predicting forthcoming environmental difficulties. Global climate models (GCMs) are commonly used to forecast future climate. Various groups have created a significant number of GCMs. Climate datasets from many organizations are being used by the IPCC for climate impact assessments [1]. The Coupled Model Intercomparison Project Phase 6 (CMIP6) is commonly utilized for present and future climate analysis and projections. Despite substantial advancements in CMIP6, large uncertainty remains under a variety of climate circumstances. Many assumptions were made throughout the creation of GCMs due to a lack of accurate information on atmospheric events, leading to exaggerations or underestimations of climate change. This allows us to see where the climate models and observations diverge. Climate projection uncertainty can be decreased by using a proper collection of GCMs. The preliminary goal of any climate change impact research or climate modeling is to pick the best group of GCMs [2]. Typically, the ability of climate models to simulate past climate is utilized as the basis for selecting GCMs. The uncertainty in climate projections has a significant impact on impact estimation. A small adjustment in climate projection can drastically alter the return duration of hydrological disasters such as floods and droughts. As a result, selecting credible GCMs is regarded as one of the most successful methods of lowering uncertainty in climate change estimates. GCMs are typically chosen based on their capacity to recreate historical climates. To evaluate the performance measure of GCMs, time series of monthly or

G. Patel (✉) · S. Das · R. Das
School of Water Resources Engineering, Jadavpur University, Kolkata, West Bengal, India
e-mail: gauravp.wre.rs@jadavpuruniversity.in

B. P. Swain and U. S. Dixit (eds.), *Recent Advances in Civil Engineering*, Lecture Notes in Civil Engineering 431, https://doi.org/10.1007/978-981-99-4665-5_6

annual observed and GCM simulated climate are typically compared. The typical method of selecting GCMs emphasizes their ability to replicate temporal variability in precipitation or temperature.

The biggest drawback of employing other performance indicators is that the timing info for the error positions is lost. As a result, any metric used to evaluate performance has advantages and disadvantages that are unique to it, and there is no single criterion for judging the quality of a model that has been established as standard across a wide range of contexts and purposes. To address these issues, MCDM methodologies were used to integrate the findings of different performance indicators into a single score. Because the aggregated results that are produced by the MCDM techniques are strongly dependent on the weights that are assigned to the numerous individual evaluation criteria, techniques for weightage division play a key role in the classification process. The reference gridded datasets that are employed and the interpolation techniques that correspond to them have a significant impact on the estimated values of the various performance metrics that are driven into the MCDM algorithms. There are various approaches for determining the efficacy of a classification problem, but because there haven't been many studies done on the effectiveness of normalization techniques for MCDM methods, it is still unclear how to select the most appropriate one. The process of normalization involves making adjustments to the values of the criterion so that they are approximately the same size. However, utilizing a variety of normalization approaches can result in a variety of solutions, which in turn can result in variations from the outcomes that were initially proposed.

In relation to the ranking of CMIP6 GCMs, the newly released CMIP6 model has only a limited amount of deployment in India. The selection of GCMs has been done using a variety of different methodologies. However, a good selection of normalizing approaches has not yet been explored in any of the relevant literature. The purpose of this study is to apply Multi-Objective Optimization based on Ratio Analysis (MOORA) MCDM with four different normalization techniques to determine which normalization technique is the most effective for particular this method that can be used to rank the CMIP6 GCMs utilizing the historical dataset to simulate precipitation. The goal of the research is to identify the finest normalization technique for particular this method. Finding the optimal normalization strategy can be of great assistance when selecting models, methods, and criteria for MCDM applications in a variety of fields.

2 Data Used and Methodology

The IMD Pune compiled daily gridded precipitation data from 1901 to 2020. With regard to precipitation, there is historical climatic data available with a grid precision of 0.25 degrees by 0.25 degrees [3]. These gridded datasets are widely used in India for climate-related studies and applications. This research makes use of downscaled datasets from the Coupled Model Inter-comparison Project, Phase 6 (CMIP6) that are made available by the NASA NEX-GDDP Program. The procedure known as

Table 1 Details of the 10 GCMs of the CMIP6

Model	Model name	Country origin	Resolution (km)	References
Mod-1	IITM ESM	India	250	Krishnan et al. [4]
Mod-2	CMCC CM2	Italy	100	Cherchi et al. [5]
Mod-3	CNRM ESM2-1	France	250	Séférian et al. [6]
Mod-4	INM CM4-8	Russia	100	Volodin et al. [7]
Mod-5	IPSL CM6A LR	France	250	Boucher et al. [8]
Mod-6	MIROC-6	Japan	250	Tatebe et al. [9]
Mod-7	MPI ESM1-2 h	Germany	100	Müller et al. [10]
Mod-8	MPI ESM1-2 LR	Germany	250	Mauritsen et al. [11]
Mod-9	NorESM2 LM	Norway	250	Seland et al. [12]
Mod-10	TaiESM 1	Taiwan	100	Wang et al. [13]

Bias Correction Spatial Disaggregation (BCSD) is utilized in its production. Table 1 presents the details along with a list of the 10 GCMs that are chosen based on the available data.

This study takes use of the historical CMIP6 GCMs as well as the IMD gridded precipitation of 64 years, from 1950 to 2014, to analyze various ranking and normalization procedures. The goal of this assessment is to determine the effectiveness of these strategies. Ten different performance criteria have been utilized in the process of assessing GCMs. These standards are determined by a combination of experience, trial and error, and effectiveness. In addition to that, the Shannon entropy approach is used for the weightage of different performance criteria to determine which is more important. In order to rank the concerned models, the MOORA technique is applied, along with four alternative types of normalization. It is proposed to use the normalization technique with the best performance, and using those strategies is how the ranking is established.

2.1 Indicators

Indicators are utilized in the process of comparing the GCMs' simulation of projected data with actual data. In this investigation, ten indicators (Table 2) are chosen after considering both their error rates and their overall productivity.

2.2 Weight Criteria Technique (Shannon Entropy)

There are several different approaches to choose from when figuring out weights. The entropy weight is the criterion that is employed for weighting in this study. It is

Table 2 Ten indicators selected to compare GCMs' simulation of projected data with actual data

Indicators	Abbreviations	Indicators	Abbreviations
Standard deviation	SD	Correlation coefficient	CC
Taylor skill score	TSS	Normalized root mean square error	NRMSE
Index agreement	IA	Kolmogorov–Smirnov test	KST
Skill score	SS	Nash–Sutcliffe efficiency	NSE
Percentage bias	PBIAS	Kling-Gupta efficiency	KGE

the procedure that is utilized most frequently for determining weights. It is necessary to take into account all of the parameters before assigning relative weight.

The entropy technique is established on a theory that takes both the amount of information that is now accessible as well as the significance of a criterion concerning that information. It does not take into account the preferences of the individual making the decision and instead uses the provided payoff matrix to determine the relative importance of the various criteria [14]. The following are the actions that need to be done to calculate the weightage of each indicator.

Step 1. Construct a payoff matrix and normalized it (Eq. 1)

$$N_{ij} = x_{ij} \Big/ \sum\nolimits_{i=1}^{m} x_{ij}^2 \text{ for } i = 1, 2, ..., m \text{ and } j = 1, 2, ..., n \tag{1}$$

where i is the index for GCM, j is the index of the indicator, x_{ij} is the scale of the indicator j for GCM i, m is the total number of models and n represents the total number of indicators.

Step 2. Entropy and degree of diversification are calculated by using Eqs. 2 and 3,

$$E_j = -\sum\nolimits_{i=1}^{m} N_{ij} \times \ln N_{ij} \Big/ \ln m \tag{2}$$

$$D_j = 1 - E_j \tag{3}$$

Step 3. Calculate the normalized weight matrix of indicators by Eq. 4,

$$w_j = D_j \Big/ \sum\nolimits_{k=1}^{n} D_k \tag{4}$$

where Ej is entropy, Dj is the degree of diversification, w_j is the individual indicator's weightage.

2.3 MOORA MCDM

The MOORA technique begins with the creation of a decision criteria matrix, which compares the effectiveness of various options based on several distinct criteria (objectives). After that, a ratio matrix system is constructed, in which each parametric performance of a potential alternative on a particular attribute is equated to a denominator that serves as a representative for all of the potential alternatives concerning that attribute. In the context of multi-objective optimization methodology, the normalizing performances are increased in the event of maximization (for beneficial qualities), and they are decreased if minimization is the goal (for non-beneficial attributes) [15].

$$W_{ij} = w_j N_{ij} \tag{5}$$

$$y_i = \sum\nolimits_{j=1}^{h} W_{ij} - \sum\nolimits_{j=h+1}^{n} W_{ij} \tag{6}$$

where h is the number of parameters that aim to be maximized, $(n - h)$ is the number of parameters to be minimized, and y_i is the normalized value of ith alternative with respect to all the parameters.

2.4 Normalization Techniques

Normalization is a procedure to analyze various datasets on a similar comparable scale. Normalization is typically required when working with qualities that are measured on more than one scale. If normalization is not performed, the influence of a significant and equally important attribute that is measured on a smaller scale may be diminished because other attributes have values that are measured on a greater scale. Within the scope of this investigation, we have implemented a total of four distinct normalization strategies. Table 3 represents the different normalization techniques along with their formula.

2.5 Index Ranking (IR)

The IR describes the extent to which the rankings obtained from one normalization method are consistent with the results obtained from other methods. The consistency index is a metric that indicates how well one normalization procedure produces rankings that are comparable to those produced by other normalization procedures [16]. This comparison indicates how well one normalization procedure produces rankings that are comparable to those produced by other normalization procedures. They determined the IR for each method of normalization by counting the total number of

Table 3 Four different normalization techniques employed to analyze datasets

Normalization technique	Formula	Criteria
Linear: Max (N_1)	$N_{ij} = x_{ij} \big/ x_{\max}$	Beneficial criteria
	$N_{ij} = x_{\min} \big/ x_{ij}$	Non- beneficial criteria
Linear: Max- Min (N_2)	$N_{ij} = \left(x_{ij} - x_{\min}\right) \big/ \left(x_{\max} - x_{\min}\right)$	Beneficial criteria
	$N_{ij} = \left(x_{\max} - x_{ij}\right) \big/ \left(x_{\max} - x_{\min}\right)$	Non- beneficial criteria
Linear: Sum (N_3)	$N_{ij} = x_{ij} \big/ \sum_{i=1}^{m} x_{ij}$	Beneficial criteria
	$N_{ij} = 1 - \left(x_{ij} \big/ \sum_{i=1}^{m} x_{ij}\right)$	Non- beneficial criteria
Vector (N_4)	$N_{ij} = x_{ij} \Big/ \sqrt{\sum_{i=1}^{m} x_{ij}^2}$	Beneficial criteria
	$N_{ij} = 1 - \left(x_{ij} \Big/ \sqrt{\sum_{i=1}^{m} x_{ij}^2}\right)$	Non- beneficial criteria

instances in which these normalizations produce results that are comparable to one another or dissimilar to one another in the problem. This allowed them to determine how each normalization contributed to the overall solution. Because of this, they were able to discover which methods of normalization were more successful.

In the first phase, we determine the IR by counting the number of ways in which each of the normalizations that have been evaluated is the same and different from one another [16]. Because there are four distinct approaches to the process of normalization, we will start by defining the weight (WT) in the following way:

- if a ranking is similar with all approaches, then $WT_1 = 3/3 = 1$;
- if a ranking is similar to two of the three approaches, then $WT_2 = 2/3$;
- if a ranking is similar to one of the three approaches, then $WT_3 = 1/3$;
- if a ranking is not consistent, then $WT_4 = 0/3 = 0$.

and then the Index Ranking, for N_1, is calculated as in Eq. 7 [16].

$$IR_{N_1} = \left[\begin{array}{l}(SR_{1234} \times WT_1) + (SR_{123} \times WT_2) + (SR_{124} \times WT_2) + (SR_{134} \times WT_2) \\ +(SR_{12} \times WT_3) + (SR_{13} \times WT_3) + (SR_{14} \times WT_3) + (DR_{1234} \times WT_4)\end{array}\right] \Big/ CC \tag{7}$$

where, IR_i: IR for normalization procedure ($i = N_1, N_2, N_3$ *and* N_4); *CC:* Number of complete cycles ($CC = 1$); SR_{1234}*:* Total number of similar ranking with N_1, N_2, N_3 and N_4; SR_{123}*:* Total number of similar ranking with N_1, N_2 and N_3; SR_{12}*:* Total number of similar rankings with N_1 and N_2; DR_{1234}*:* Total number of different ranking with N_1, N_2, N_3 and N_4.

Table 4 Ranking of models using MOORA

Models	Normalization techniques				Models	Normalization techniques			
	N_1	N_2	N_3	N_4		N_1	N_2	N_3	N_4
Mod-1	8	1	10	10	Mod-6	2	5	3	2
Mod-2	3	4	4	3	Mod-7	6	8	7	7
Mod-3	7	9	6	6	Mod-8	1	2	1	1
Mod-4	9	7	8	8	Mod-9	10	6	9	9
Mod-5	4	3	5	4	Mod-10	5	10	2	5

3 Results and Discussion

Several metrics have been provided as a result of the research that may be utilized to evaluate the various normalization strategies that are utilized in MCDM procedures. During our investigation, we made use of the IR that was initially suggested in literature [16]. To identify which normalizing method led to the production of the best accurate rankings, we put the MOORA method and four distinct normalization methods through their paces.

3.1 Ranking of GCMs

The MOORA method is executed to calculate the ranking of ten GCMs. This is done for the variable of precipitation. The ranks of GCMs for precipitation, as determined by the MOORA methodology, are presented in Table 4. The findings of the ten GCMs, which are based on the four distinct normalization methods, each yield a different rating for the normalization method. The Mod-8 seems to be the most effective of all the other choices that are on the table, at least according to MOORA.

3.2 Index Ranking (IR)

The final findings for MOORA procedures based on IR suggest that N_4 has the greatest IR values, followed by N_3, N_1, and N_2. Based on the results, it is simple to conclude that the vector normalizing technique is the most effective normalization technique for MOORA. The overall IR value is represented in Fig. 1 for each of the normalization procedures.

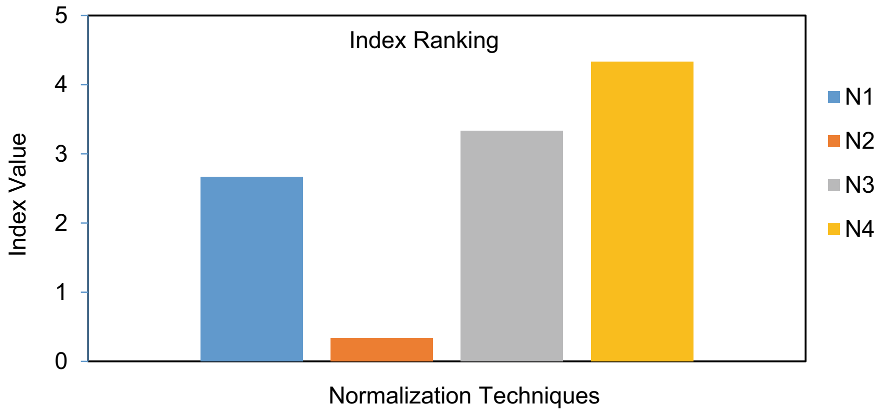

Fig. 1 Index ranking of normalizations

4 Conclusions

Because we need to obtain dimensionless units to calculate the final rating for each alternative, the process of decision-making cannot function without normalization playing a crucial role. Consequently, we cannot conduct one without the other. In this exploratory study, we demonstrated the effects of applying four distinct normalizing approaches that are common and well-known in their respective fields. In order to evaluate which normalizing approach is most suited for the MOORA method, we compared all four normalization strategies by utilizing a simple illustrative case. We are able to show that the vector normalization technique (N_4) is the one that works best for the MOORA. The outcome of various normalization strategies for MCDM approaches is similar to the findings of earlier research [16]. The procedure gave a ranking for the four standardization methods that are commonly used in order to offer decision-makers assistance in making better-informed judgments. When researchers and practitioners normalize their data in an MCDM application using the MCDM approach stated, they should keep these results in mind and consider them. Keeping these results in mind and considering them will help ensure accurate outcomes.

References

1. IPCC (2021) Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Masson-Delmotte V, Zhai P, Pirani A et al. (eds.). Cambridge University Press
2. Samadi SZ, Sagareswar G, Tajiki M (2010) Comparison of general circulation models: methodology for selecting the best GCM in Kermanshah Synoptic Station, Iran. Int J Glob Warm 2(4):347–365

3. Pai DS, Sridhar L, Rajeevan M et al. (2014) Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. Mausam 65(1):1–18
4. Krishnan R, Swapna P, Vellore R et al. (2019) The IITM earth system model (ESM): development and future roadmap. In: D Randall, J Srinivasan, R Nanjundiah, P Mukhopadhyay (eds.) Current trends in the representation of physical processes in weather and climate models. Springer, Singapore, pp 183–195
5. Cherchi A, Fogli PG, Lovato T et al (2019) Global mean climate and main patterns of variability in the CMCC-CM2 coupled model. J Adv Model Earth Syst 11(1):185–209
6. Séférian R, Nabat P, Michou M et al (2019) Evaluation of CNRM earth system model, CNRM-ESM2-1: role of earth system processes in present-day and future climate. J Adv Model Earth Syst 11(12):4182–4227
7. Volodin E, Mortikov EV, Kostrykin SV et al (2018) Simulation of the modern climate using the INM-CM48 climate model. Russ J Numer Anal Math Model 33(6):367–374
8. Boucher O, Servonnat J, Albright AL et al. (2020) Presentation and evaluation of the IPSL-CM6A-LR climate model. J Adv Model Earth Syst 12(7):e2019MS002010
9. Tatebe H, Ogura T, Nitta T et al (2019) Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. Geosci Model Dev 12(7):2727–2765
10. Müller WA, Jungclaus JH, Mauritsen T et al. (2018) A higher-resolution version of the max planck institute earth system model (MPI-ESM1.2-HR). J Adv Model Earth Syst 10:1383–1413
11. Mauritsen T, Bader J, Becker T et al. (2019) Developments in the MPI-M earth system model version 1.2 (MPI-ESM1.2) and Its response to increasing CO_2. J Adv Model Earth Syst 11(4):998–1038
12. Seland Ø, Bentsen M, Olivié D et al (2020) Overview of the Norwegian earth system model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. Geosci Model Dev 13(12):6165–6200
13. Wang YC, Hsu HH, Chen CA et al (2021) Performance of the Taiwan earth system model in simulating climate variability compared with observations and CMIP6 model simulations. J Adv Model Earth Syst 13(7):1–28
14. Pomerol JC, Romero SB (2000) Weighting methods and associated problems. In: Multicriterion Decision in Management. International Series in Operations Research & Management Science, vol 25, pp 75–104. Springer, Boston, MA
15. Brauers WKM (2013) Multi-objective seaport planning by MOORA decision making. Ann Oper Res 206(1):39–58
16. Chakraborty S, Yeh CH (2009) A simulation comparison of normalization procedures for TOPSIS. International conference on computers & industrial engineering. IEEE Press, France, pp 1815–1820