ISSN 2040-2244

Journal of
**Water and Climate Change**

IWA PUBLISHING

Volume 5  Number 1  March 2014

# Selection of global climate models for India using cluster analysis

K. Srinivasa Raju and D. Nagesh Kumar

## ABSTRACT

Global climate models (GCMs) are gaining importance due to their capability to ascertain climate variables that will be useful to develop long, medium and short term water resources planning strategies. The applicability of K-Means cluster analysis is explored for grouping 36 GCMs from Coupled Model Intercomparison Project 5 for maximum temperature (MAXT), minimum temperature (MINT) and a combination of maximum and minimum temperature (COMBT) over India. Cluster validation methods, namely the Davies–Bouldin Index (DBI) and F-statistic, are used to obtain an optimal number of clusters of GCMs for India. The indicator chosen for evaluation of GCMs is the probability density function based skill score. It is noticed that the optimal number of clusters for MAXT, MINT and COMBT scenarios are 3, 2 and 2, respectively. Accordingly, suitable ensembles of GCMs are suggested for India for MAXT, MINT and COMBT individually. The suggested methodology can be extended to any number of GCMs and indicators, with minor modifications.

**Key words** | cluster, Davies–Bouldin Index, ensemble, global climate models, India, K-means

**K. Srinivasa Raju**
Department of Civil Engineering,
Birla Institute of Technology and Science-Pilani,
Hyderabad,
India

**D. Nagesh Kumar** (corresponding author)
Department of Civil Engineering,
Indian Institute of Science,
Bangalore,
India
and
Centre for Earth Sciences,
Indian Institute of Science,
Bangalore,
India
E-mail: nagesh@civil.iisc.ernet.in

## INTRODUCTION

Assessment and analysis of maximum and minimum temperatures have become important due to their increasing impact on climate. Surface temperature has a significant effect on most of the parameters of the hydrological cycle. However, climate impact assessment studies based on temperature are few when compared to precipitation over India (Kothawale et al. 2010; Sonali & Nagesh Kumar 2013). Global climate models (GCMs), which are simulation tools for the Earth's climate, are gaining importance due to their capability to predict future climate variables that will be useful to develop long, medium and short term water resources planning strategies. This raises the question of choosing suitable GCMs from the large number available, making the problem more complex. The problem becomes aggravated as most of the GCMs are similar, i.e.: (a) some GCMs developed by various agencies share part of the same code by way of exploring similar numerical schemes, parameterisations, and resolutions, and some are even developed by the same agency; (b) some institutions provide

simulations from more than one variation of the same GCM; (c) employing many GCMs results in similar forecasts due to sharing of similar oceanic and atmospheric components (Pennell & Reichler 2011). In the present study clusters and groups are used synonymously.

In this regard, cluster analysis is becoming prominent due to its capability of reducing the number of GCMs, eliminating those of a similar nature, to a manageable homogeneous subset. Complimentarily, cluster validation methods have proved to be efficient and can be explored as the basis to determine the optimal number of clusters of GCMs (Wang et al. 2009). Thirty-six GCMs are grouped homogeneously by using K-means cluster analysis. Skill score is the indicator employed for evaluation. Cluster validation methods, namely the Davies–Bouldin Index (DBI) and F-statistic, are applied to determine an optimal number of clusters of GCMs. Outputs from the F-statistic and DBI are used for formulating an ensemble of GCMs, or in other words, a group of GCMs. The following sections present a

literature review, objectives of the study, K-means cluster analysis, cluster validation methods, performance indicator, data collection, results, discussion and conclusions.

## LITERATURE REVIEW

A brief but relevant literature review related to the evaluation and clustering of GCMs is presented below.

### Evaluation of GCMs

Cai et al. (2009) assessed the performance of 17 GCMs for simulation of precipitation and temperature using a skill score for the periods of 1931–1960 and 1961–1990. No single GCM was found to be advantageous in predicting precipitation or temperature for the whole world, even though some GCMs performed better for particular regions. Chen et al. (2012) evaluated the ability of nine GCMs of the Coupled Model Intercomparison Project 5 (CMIP5) category to simulate the near surface wind over China. Spatial fields of wind speed at the end of the 21st century are almost similar to those of the last 35 years. Gómez-Navarro et al. (2012) studied the ranking of an ensemble for predicting the regional climate for Spain. They discussed the weighted ensemble average quality with reference to the uncertainties in the observational dataset.

Fu et al. (2013) evaluated the performance of 25 GCMs using the multicriterion method at the regional scale for air temperature, monthly mean sea level pressure and rainfall, and annual rainfall over south eastern Australia for 1960–1961 to 1999–2000. They concluded that simulated monthly rainfall and monthly mean sea level pressures are not satisfactory compared to monthly temperature. Su et al. (2013) assessed the performance of 24 GCMs of the CMIP5 category for the Eastern Tibetan Plateau by comparing the GCM outputs with field data for the period 1961–2005. They concluded that most of the GCMs satisfactorily assessed the spatial variations and climatological patterns of temperature. Tiwari et al. (2014) analysed the skill of five GCMs in predicting precipitation in winter over Northern India for the period 1982–2009. The climatology, correlation coefficients, and inter annual standard deviation were computed and compared with the observed values. It

was noticed that the GCMs were able to replicate the climatology and inter annual standard deviation to varying degrees, whereas prediction skill is too low. Barfus & Bernhofer (2015) assessed the capabilities of 18 GCMs using indices such as Vertical Totals, Total Totals, Cross Totals, Severe WEAther Threat (SWEAT), K, and Showalter to the case study of the Arabian Peninsula. No overall best performing GCM was identified. Grose et al. (2014) evaluated the 27 GCMs' performance for the Western tropical Pacific, and their differences from the Coupled Model Intercomparison Project 3 (CMIP3) were estimated. The CMIP5-based GCMs showed some improvements over CMIP3 in the late 20th century. Perez et al. (2014) evaluated the skill of GCMs from CMIP3 and CMIP5 databases for the North-East Atlantic Ocean region. The skill of GCMs to replicate the synoptic situations, the historical inter-annual variability, and the consistency of GCM experiments for 21st century projections were considered.

Hidalgo & Alfaro (2015) evaluated 107 climate runs from 48 GCMs for Central America. A research basis to ascertain the performance of the GCMs is suggested. Similar studies were reported by Palazzi et al. (2015) for Karakoram-Himalaya for the precipitation rate with 32 GCMs and Kumar et al. (2015) for 22 regions across the world for wind extremes with 15 GCMs.

### Clustering of GCMs

Pennell & Reichler (2011) mentioned the limitations about GCMs as 'the limitations of these models tend to be fairly similar, contributing to the well-known problem of common model biases and possibly to an unrealistically small spread in the outcomes of model predictions'. They evaluated 24 GCMs for their capability to simulate 20th century climate. The effective ensemble size of GCMs was observed to be less than the actual number. They also used hierarchical clustering to assess the similarities among different GCMs and concluded that the present approach of interpreting multi-model ensembles may induce over prediction of climate. Yokoi et al. (2011) applied cluster analysis to 43 model performance metrics, which were evaluated using 22 GCMs. They employed K-means and Ward's method and found that both methods provided a similar dynamical and physical result pattern. Yuan & Wood

(2012) studied issues such as overconfidence of the ensemble of GCMs and the necessity of an effective sub set of GCMs that can maintain skill and predictability. They stressed the necessity for clustering of models, and applied hierarchical cluster analysis for grouping 12 seasonal forecast models at different lead times for surface air temperature and precipitation. It was concluded that cluster analysis was found to be useful for grouping climate models. Similar studies were reported by Knutti (2008), Knutti *et al.* (2013), Reifen & Toumi (2009), Eden *et al.* (2012), and Jurya (2015).

## OBJECTIVES OF THE STUDY

It is observed from the aforementioned literature that no significant work was conducted in the clustering of GCMs in association with cluster validation methods for Indian climate conditions. The formulated objectives are as follows:

1. Identification of a performance indicator for evaluating GCMs.
2. Examining the suitability of K-Means cluster analysis for grouping 36 GCMs for maximum temperature (MAXT), minimum temperature (MINT) and a combination of maximum and minimum temperature (COMBT).
3. Examining the suitability of DBI and F-statistic to achieve optimal clusters of GCMs.
4. Suggesting ensembles of GCMs for MAXT, MINT, and COMBT.

## METHODS AND PERFORMANCE INDICATOR

### K-means cluster analysis

K-means is an iterative algorithm used for clustering GCMs. The algorithmic steps consist of: (a) assignment of each GCM initially to any one group (g) and obtaining the average of the assigned group(s); (b) identification of the objective function; (c) identification of termination criteria for stopping the algorithm, which provides the basis for choosing final groups and corresponding GCMs; and (d) computation of total squared error, as the error in cluster analysis significantly represents the deviation of individual data set $x_i$ from

the group average $\bar{x}$. Computation of total squared error value for cluster K is done using $ER_K = \sum_{k=1}^{K} e_k^2$ where $e_k$ is error value $(x_i - \bar{x})$ (Jain & Dubes 1988; Raju & Nagesh Kumar 2014a). K-means cluster analysis can be tried a good number of times with possible minimum and maximum groups for effective appraisal of the problem.

## CLUSTER VALIDATION METHODS

For clustering algorithms, the number of groups is an important input. However, it is difficult to assess optimal clusters manually for a set of GCMs. In this regard, cluster validation methods inherently indicate the quality of the chosen group. Brief descriptions of the DBI and F-statistic are presented below.

The DBI relates to the ratio of the sum of the intra-cluster scatter to the inter-cluster separation (Davies & Bouldin 1979). The DBI is expressed as:

$$DBI(U) = \frac{1}{K} \sum_{i=1}^{K} \max \left[ \frac{\Delta(X_i) + \Delta(X_j)}{\delta(X_i, X_j)} \right] \; i \neq j \tag{1}$$

where DBI (U) is the Davies–Bouldin Index for cluster U, $K$ is the number of clusters, $\Delta(X_i)$ is the average distance of all GCMs in the group from the group average, $\delta(X_i, X_j)$ is the distance between the group averages of $i$ and $j$ respectively. The ratio will be smaller if the clusters are far away from each other and compact. The DBI lies between 0 and $\infty$. A lower value indicates that clustering is good, while a higher value indicates bad clustering. This can be explained by the fact that a lower value indicates that the numerator is small compared to the denominator. The denominator is indicative of inter-cluster distance, whereas the numerator is indicative of intra-cluster distance. Thus, the smallest value of overall DBI indicates the optimum cluster.

The F-statistic is a statistical test that measures the variance reduction from clusters $K$ to $K + 1$ (Burn 1989) and is expressed as:

$$F_K = \left( \frac{ER_K}{ER_{K+1}} - 1 \right) (N - K + 1) \tag{2}$$

where $F_K$ = F-statistic value for cluster $K$, $N$ = number of GCMs and $ER_{K+1}$ = total squared error value for all clusters

$(K+1)$. An F-statistic value greater than 10 establishes a transformation from $K$ to $K+1$ clusters (Burn 1989).

## Performance indicator

The simulating capacity of GCMs compared to the observed data is evaluated using skill score, which evaluates the similarity between the probability density functions of both of them (Perkins *et al.* 2007; Raju & Nagesh Kumar 2014b). Skill score is expressed as:

$$SS = \sum_{i=1}^{nbin} \min (fr_m, fr_o) \qquad (3)$$

where *nbin* is the number of bins used to calculate the probability density functions, $fr_m$, $fr_o$ are the frequencies of values in the given bin from the chosen GCM and the observed data. Skill score values vary between zero and one.

## DATA COLLECTION

Thirty-six GCMs belonging to CMIP5 presented in Table 1 provide both monthly maximum and minimum temperatures and are chosen for evaluation. It may be noted that these 36 GCMs were chosen based on data availability from among the CMIP5 models. CMIP5 has selected these models based on their performance amongst all the existing GCMs. In addition, more than one GCM from the same institution is explored to observe the similarities/differences between them in the classification perspective, instead of only one model from the same institution (Knutti 2008). Accordingly, outputs from GCMs are compared with the Climate Research Unit 2.1 datasets for the model evaluation study, and the period considered is 1961–1999. Forty grid points of $2.5^0 \times 2.5^0$ covering India are considered.

## RESULTS

A skill score of 40 grid points corresponding to each GCM is computed using MATLAB (Equation (3)). A data matrix of 36 GCMs and skill scores for each grid point are developed for three variables, i.e. MAXT, MINT, COMBT, and are used as the basis for K-means cluster analysis. The Cluster Validity Analysis Platform (CVAP) developed by Wang (2007) is used for K-means cluster analysis and performed for 2–9 clusters and even beyond. CVAP is run for 2–30 clusters (five times for each cluster) to assess the occupancy of GCMs in each cluster, totaling 145 runs. It is observed that as the number of clusters is increased beyond nine, many clusters are not occupied. Therefore, the analysis is restricted to nine clusters. For each chosen group, the DBI and F-statistic is evaluated (Equations (1) and (2)) and used as the basis for determining the optimal cluster. The following sections present the results relating to MAXT, MINT, and COMBT.

## MAXT

It is observed that (results not presented) the maximum skill score is 0.9788 at $92.5^0 \times 22.5^0$ grid point for the FIO-ESM, whereas the minimum skill score is 0.0502 at $75^0 \times 27.5^0$ grid point for HadCM3. The range of skill scores is 0.9286. Figure 1 presents the number of GCMs in clusters 2–9 and a representative GCM in each sub cluster. The notation for cluster 2 in Figure 1 is as follows: 1 represents the sub cluster, GFDL-ESM2M is the representative GCM in that sub cluster; 32 represents the number of GCMs in that sub cluster whereas 2 represents the sub cluster, GISS-E2-H is the representative GCMs in that sub cluster; 4 represents the number of GCMs in that sub cluster. It is observed that as the number of clusters is increased, there is a spread in the number of GCMs. It is noted that as the number of clusters is increased to four and more, only one GCMs (HadCM3) is observed in sub clusters belonging to clusters 4–9, and only two GCMs are observed in some sub clusters belonging to clusters 7–9. This outcome is on the expected lines, as the GCMs are only 36 and the number of clusters are too many. Note that there is no specific trend while classifying GCMs from clusters 2 to 9. Most of the time, GCMs are in the same sub cluster irrespective of increase in the size of cluster. This may be due to the similarities in their structure as well as them having been developed by the same agency. Substantial similarities are noticed between GCMs from the same institution in the cluster analysis outcome. For example, GISS-E2-H, GISS-E2-R, GISS-E2-R-CC, developed by NASA Goddard Institute for Space Studies, are sub clustering always

**Table 1** | Details of CMIP5 models: modeling center acronym, model name and modelling institutions

| S. no. | Modeling center_acronym | Model_name | Institution |
|---|---|---|---|
| 1 | BCC | BCC-CSM1.1 | Beijing Climate Center, China Meteorological Administration |
| 2 | BCC | BCC-CSM1.1-m | Beijing Climate Center, China Meteorological Administration |
| 3 | CCCma | CanESM2 | Canadian Centre for Climate Modelling and Analysis |
| 4 | CNRM-CERFACS | CNRM-CM5 | Centre National de Recherches Meteorologiques/Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique |
| 5 | CSIRO-BOM | ACCESS1.0 | CSIRO (Commonwealth Scientific and Industrial Research Organization, Australia), and BOM (Bureau of Meteorology, Australia) |
| 6 | CSIRO-BOM | ACCESS1.3 | CSIRO (Commonwealth Scientific and Industrial Research Organization, Australia), and BOM (Bureau of Meteorology, Australia) |
| 7 | CSIRO-QCCCE | CSIRO-Mk3.6 | Commonwealth Scientific and Industrial Research Organization in collaboration with the Queensland Climate Change Centre of Excellence |
| 8 | FIO | FIO-ESM | The First Institute of Oceanography, SOA, China |
| 9 | GCESS | BNU-ESM | College of Global Change and Earth System Science, Beijing Normal University |
| 10 | INM | INM-CM4 | Institute for Numerical Mathematics |
| 11 | IPSL | IPSL-CM5A-LR | Institute Pierre-Simon Laplace |
| 12 | IPSL | IPSL-CM5A-MR | Institute Pierre-Simon Laplace |
| 13 | IPSL | IPSL-CM5B-LR | Institute Pierre-Simon Laplace |
| 14 | LASG-IAP | FGOALS-s2 | LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences |
| 15 | MIROC | MIROC4h | Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology |
| 16 | MIROC | MIROC5 | Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology |
| 17 | MIROC | MIROC-ESM | Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies |
| 18 | MIROC | MIROC-ESM-CHEM | Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies |
| 19 | MOHC | HadCM3 | Met Office Hadley Centre |
| 20 | MPI-M` | MPI-ESM-LR | Max Planck Institute for Meteorology (MPI-M) |
| 21 | MPI-M` | MPI-ESM-MR | Max Planck Institute for Meteorology (MPI-M) |
| 22 | MPI-M` | MPI-ESM-P | Max Planck Institute for Meteorology (MPI-M) |
| 23 | MRI | MRI-CGCM3 | Meteorological Research Institute |
| 24 | NASA GISS | GISS-E2-H | NASA Goddard Institute for Space Studies |
| 25 | NASA GISS | GISS-E2-R | NASA Goddard Institute for Space Studies |
| 26 | NASA GISS | GISS-E2-R-CC | NASA Goddard Institute for Space Studies |
| 27 | NCAR | CCSM4 | National Center for Atmospheric Research |
| 28 | NCC | NorESM1-M | Norwegian Climate Centre |
| 29 | NIMR/KMA | HadGEM2-AO | National Institute of Meteorological Research/Korea Meteorological Administration |
| 30 | NOAA GFDL | GFDL-CM3 | Geophysical Fluid Dynamics Laboratory |
| 31 | NOAA GFDL | GFDL-ESM2G | Geophysical Fluid Dynamics Laboratory |

**Table 1** | continued

| S. no. | Modeling center_acronym | Model_name | Institution |
|---|---|---|---|
| 32 | NOAA GFDL | GFDL-ESM2M | Geophysical Fluid Dynamics Laboratory |
| 33 | NSF-DOE-NCAR | CESM1-BGC | National Science Foundation, Department of Energy, National Center for Atmospheric Research |
| 34 | NSF-DOE-NCAR | CESM1-CAM5 | National Science Foundation, Department of Energy, National Center for Atmospheric Research |
| 35 | NSF-DOE-NCAR | CESM1-FAST CHEM | National Science Foundation, Department of Energy, National Center for Atmospheric Research |
| 36 | NSF-DOE-NCAR | CESM1-WACCM | National Science Foundation, Department of Energy, National Center for Atmospheric Research |
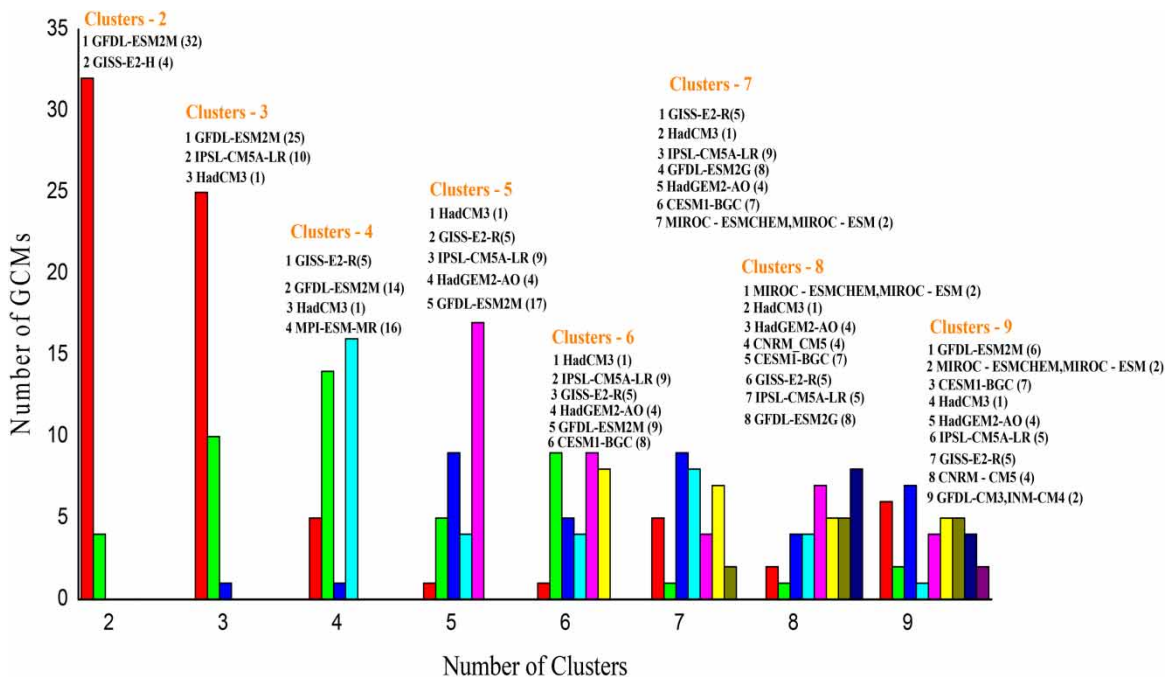


**Figure 1** | Number of GCMs in clusters 2–9 for MAXT.

jointly; similarly IPSL-CM5A-MR and IPSL-CM5A-LR, developed by Institute Pierre-Simon Laplace, were in the same sub-cluster whereas IPSL-CM5B-LR, which is also developed by the same institute, is not part of the same sub cluster.

The variation of total error for clusters 2–9 for three variables, MAXT, MINT and COMBT, is shown in Figure 2. It is noticed from Figure 2 that the total error is reducing from 15.1453 (the error values for sub clusters 1 and 2 are 11.248 and 3.8973, totalling 15.1453) to 4.1211 from clusters

2–9 with a range of 11.0242. The average error per cluster is found to be 1.378. There is a steep fall of error from cluster 2 to 3, i.e. 4.4097. Thereafter the difference is approximately 1 unit each from clusters 3–8, whereas it is very nominal from 8 to 9. The error values between the group average and the skill score values for each GCM in that sub cluster are computed. The summation of the error values for all 40 grid points gives the total error value corresponding to each GCM in that sub cluster. The GCM that gives the minimum error value in that sub cluster is chosen as the representative
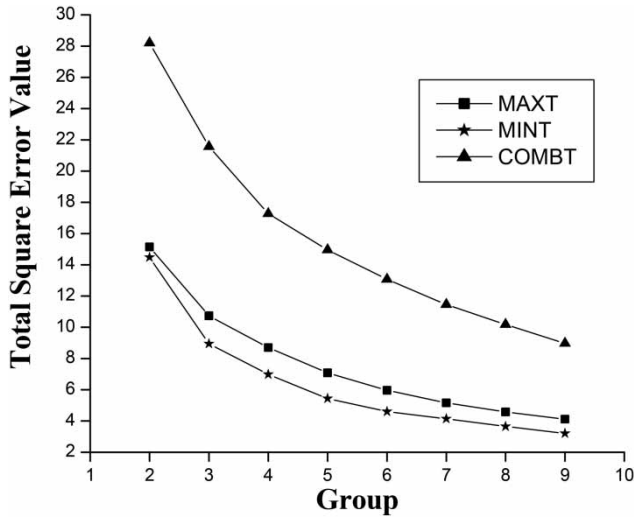
**Figure 2** │ Variation of total squared error value for three variables MAXT, MINT, COMBT.

GCM for that sub cluster. The error value is one of the important input parameters in the computation of the DBI and F-statistic, which provides the basis for optimal clusters.

Figure 3 presents DBI values, and these are varying from cluster 2 (1.2675) to cluster 7 (0.8261), but not in sequential order. According to DBI principles, the optimum cluster size is 7 for MAXT. However, it is also interesting to note that groups 3, 7 and 9 have almost equal DBI values, i.e. 0.8829, 0.8261, and 0.8909 with a nominal difference of 0.0568 (with reference to clusters 7 and 3) and 0.0648 (with reference to clusters 7 and 9). This aspect is making
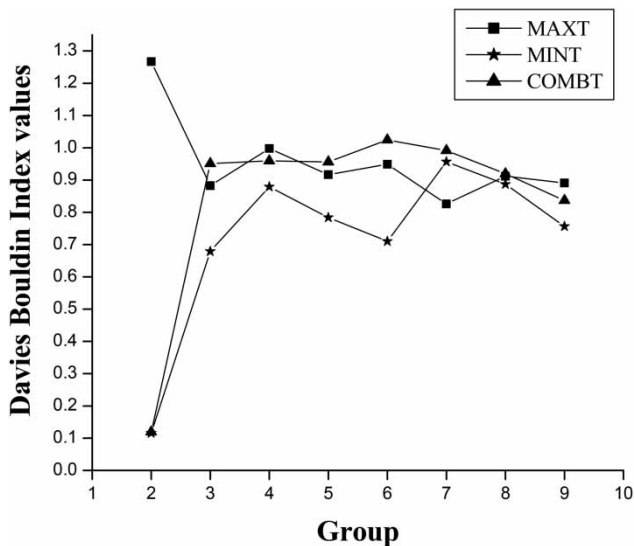


**Figure 3** │ Variation of DBI values for the three variables MAXT, MINT, COMBT.

it difficult to assume even the reasonable margin of error to determine optimum clusters with confidence. This necessitates the use of complimentary F-statistic (Burn 1989) along with DBI to obtain optimality more effectively.

Figure 4 presents the F-statistic values for clusters 2–8. It is observed that the F-statistic value varies from 14.3764 (cluster 2) to 3.2060 (cluster 8) in sequential order. Optimal clusters are 2 (based on the philosophy of preferring a cluster whose F-statistic value is greater than 10). However, the optimum number of clusters for MAXT is fixed as 3 based on: (a) preference of cluster 3 over 7 due to the narrow marginal difference of DBI values between clusters 3 and 7; (b) compatibility with the output of F-statistic; and (c) the steep error reduction from cluster 2 to 3. Accordingly, the ensemble of HadCM3, IPSL-CM5A-LR and GFDL-ESM2M is suggested for MAXT.

## MINT

Maximum skill score is 0.9815 at $70^0 \times 27.5^0$ grid point for the NorESM1-M, whereas the minimum skill score is 0.0194 at $90^0 \times 25^0$ grid point for HadCM3 and the range of skill score is 0.9621. The total error reduces from 14.4766 to 3.2036 from 2 to 9 clusters, with a range of 11.273 (Figure 2) with the average error per cluster as 1.4091. A steep fall in error of 5.532 is observed from cluster 2 to 3. Thereafter the difference is approximately 2 units from cluster 3 to 4, whereas it is 1 unit each from cluster 4 to cluster 5 and very nominal from cluster 5 to 9. Figure 5 presents the number of GCMs in clusters 2 to 9 and the representative GCM in
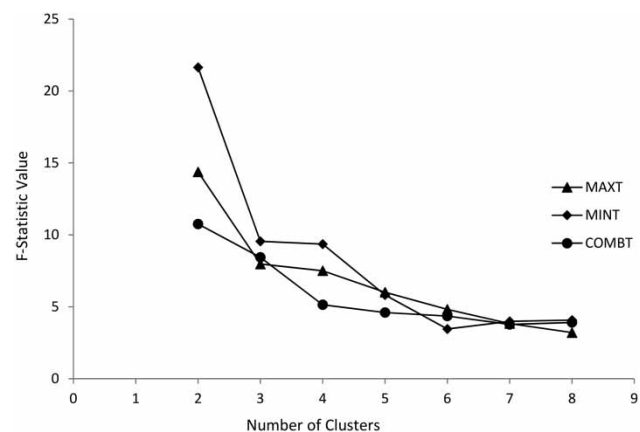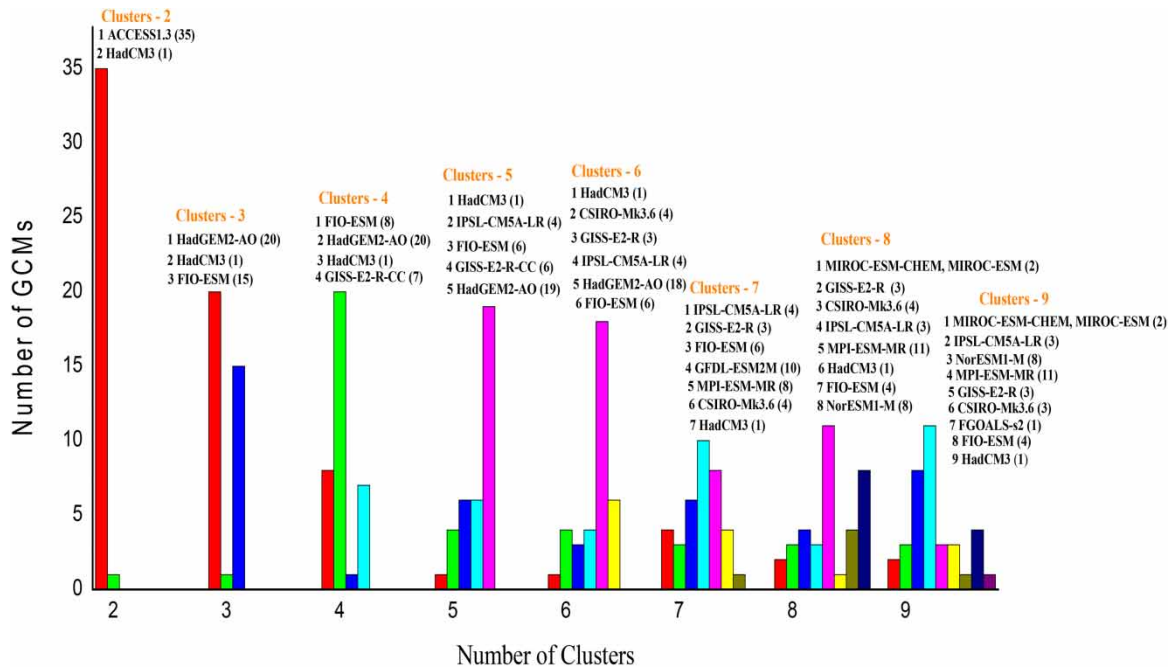


**Figure 4** │ Variation of F-statistic values for the three variables MAXT, MINT, COMBT.

**Figure 5** │ Number of GCMs in clusters 2–9 for MINT.

each sub cluster. No trend is observed while classifying GCMs from cluster 2 to cluster 9. For example, 36 GCMs are sub clustered into 32 and four GCMs in MAXT, whereas these are 35 and 1 in the case of MINT. GISS-E2-R-CC, GISS-E2-H, GISS-E2-R, developed by the NASA Goddard Institute for Space Studies, are always jointly sub clustering, similar to IPSL-CM5A-LR and IPSL-CM5B-LR developed by Institute Pierre-Simon Laplace, whereas IPSL-CM5A-MR is not part of the same sub cluster. This is contrary to that for MAXT. DBI values vary from cluster 2 (0.1156) to cluster 7 (0.9568), but not in sequential order. The range of DBI is 0.8412 with an average DBI of 0.1051 (Figure 3). It is noticed from Figure 4 that the F-statistic value varies from 21.646 (cluster 2) to 4.077 (cluster 8) in sequential order. The optimum number of clusters is 2 based on the F-statistic value of 21.64 and low DBI of 0.1156. An ensemble of ACCESS1.3 and HadCM3 is suggested for MINT, which means that the outputs of both GCMs will be used for further processing (Figure 5).

## COMBT

Total error is reduced from 28.20 to 8.97 from 2 to 9 clusters, with a range of 19.23. It is observed that there is a steep fall of error from cluster 2 to cluster 9 (Figure 2). The highest total error for COMBT is due to the combined coverage of maximum and minimum temperature (covering 40 grid points each) as compared to MAXT and MINT. Figure 6 presents the number of GCMs falling in each sub cluster and the representative GCMs in each sub cluster. DBI values vary from cluster 2 (0.1189) to cluster 6 (1.0245) but not in sequential order (Figure 3). It is noticed from Figure 4 that the F-statistic value varies from 10.758 (cluster 2) to 3.911 (cluster 8) in sequential order. The optimum number of clusters is 2 based on the philosophy of the preferred F-statistic value of more than 10 and the low DBI (0.1189). An ensemble of MPI-ESM-MR and HadCM3 is suggested for COMBT, which means that the outputs of both GCMs will be used for further processing.

## DISCUSSION

Studies on three variables, namely, MAXT, MINT and COMBT, using the skill score as indicator, show the existence of similarity for most of the GCMs (Figures 1, 5 and 6). It is observed from Figure 4 that the F-statistic values decrease with an increase in the number of clusters for MAXT, MINT and COMBT. F-statistic values are almost the same for cluster 7 and very nominal
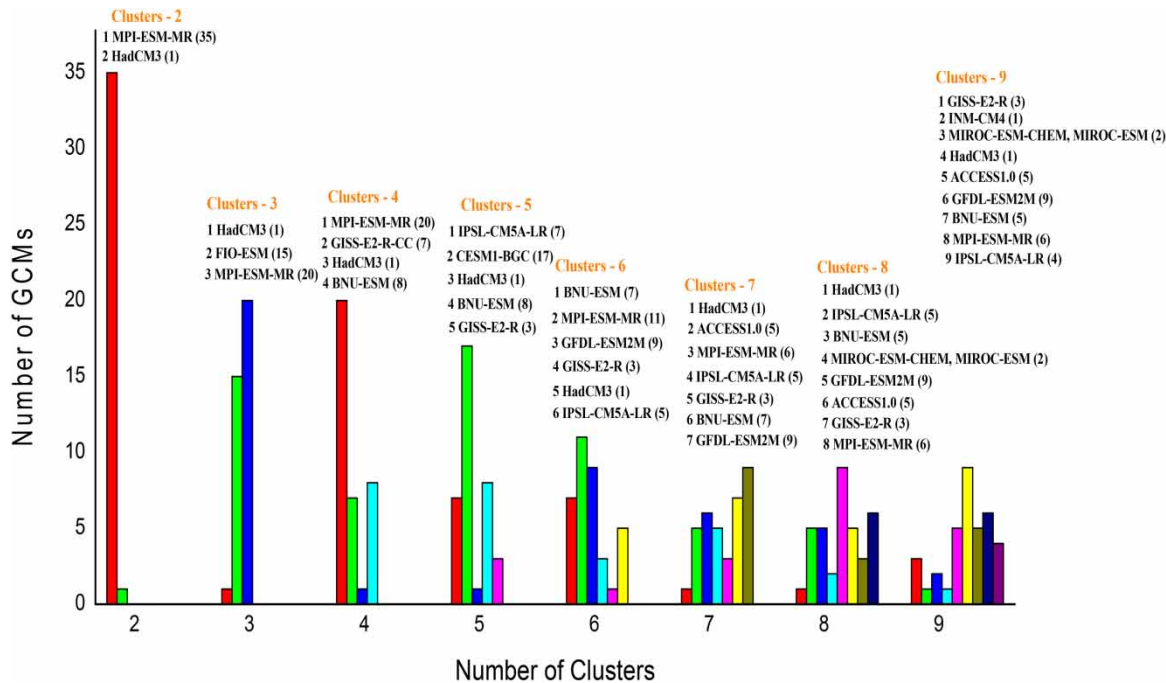
**Figure 6** │ Number of GCMs in clusters 2–9 for COMBT.

difference for clusters 6 and 8 is observed for MAXT, MINT and COMBT, mainly due to the similar ratio values (Equation (2)). It is observed that HadCM3 is the common GCM in the three ensembles of MAXT, MINT, and COMBT. Numerous authors suggested the necessity for an ensemble of GCMs (or more than one GCM) (Mujumdar & Nagesh Kumar 2012; Raju & Nagesh Kumar 2014b; Hidalgo & Alfaro 2015; Palazzi *et al.* 2015; Raju *et al.* 2016). In this context, heterogeneous GCMs obtained through cluster analysis can be used as support for formulating the ensemble, which in our opinion is innovative as compared to an ensemble of GCMs of homogeneous nature. In addition, the F-statistic and DBI values are found to be complimentary to each other and provide an opportunity to find the optimum ensemble size with more confidence, as more GCMs than required burden the process whereas a smaller ensemble size may not serve the intended purpose.

## CONCLUSIONS

K-means cluster analysis along with F-statistic and DBI is applied to a case study of India. It is observed that the optimum cluster is 2 for both MINT and COMBT whereas it is 3 in the case of

MAXT on the basis of the F-statistic test and DBI. Ensembles of {HadCM3, IPSL-CM5A-LR, GFDL-ESM2M}, {ACCESS1.3, HadCM3} and {MPI-ESM-MR, HadCM3} are suggested for MAXT, MINT and COMBT, respectively.

In our opinion, this is the first application of K-means cluster analysis and optimal clustering of GCMs, and it can be used as the basis for choosing an ensemble of GCMs for Indian climate conditions.

It is relevant to note that the inferences emanating from the present paper depend on the chosen grid points as well as on the chosen clustering and validation method. However, the thrust of the present paper is on a methodology that can be applied and replicated with ease.

Future work is targeted to employ fuzzy cluster analysis to handle impreciseness in the skill score values and skewness in the data. There is a possibility that fuzzy cluster analysis may give better results, but it needs to be explored in detail.

## ACKNOWLEDGEMENTS

# REFERENCES

Barfus, K. & Bernhofer, C. 2015 Assessment of GCM capabilities to simulate tropospheric stability on the Arabian Peninsula. *Int. J. Climatol.* **35**, 1682–1696.

Burn, D. H. 1989 Cluster analysis as applied to regional flood frequency. *J. Wat. Reso. Plan. Manage. ASCE* **115**, 567–582.

Cai, X., Wang, D., Zhu, T. & Ringler, C. 2009 Assessing the regional variability of GCM simulations. *Geophys. Res. Lett.* **36**, L02706.

Chen, L., Pryor, S. C. & Li, D. 2012 Assessing the performance of Intergovernmental Panel on Climate Change AR5 climate models in simulating and projecting wind speeds over China. *J. Geophys. Res.* **117**, D24102.

Davies, D. L. & Bouldin, D. W. 1979 A cluster separation measure. *IEEE Trans. Pattern. Anal. Mach. Intell.* **1**, 224–227.

Eden, J. M., Widmann, M., Grawe, D. & Rast, S. 2012 Skill, correction, and downscaling of GCM-simulated precipitation. *J. Clim.* **25**, 3970–3984.

Fu, G., Liu, Z., Charles, S. P., Xu, Z. & Yao, Z. 2013 A score-based method for assessing the performance of GCMs: a case study of southeastern Australia. *J. Geophys. Res. Atmos.* **118**, 4154–4167.

Gómez-Navarro, J. J., Montávez, J. P., Jerez, S., Jiménez-Guerrero, P. & Zorita, E. 2012 What is the role of the observational dataset in the evaluation and scoring of climate models?. *Geophys. Res. Lett.* **39**, L24701.

Grose, M. R., Brown, J. N., Narsey, S., Brown, J. R., Murphy, B. F., Langlais, C., Gupta, A. S., Moisec, A. F. & Irvinga, D. B. 2014 Assessment of the CMIP5 global climate model simulations of the western tropical Pacific climate system and comparison to CMIP3. *Int. J. Climatol.* **34**, 3382–3399.

Hidalgo, H. G. & Alfaro, E. J. 2015 Skill of CMIP5 climate models in reproducing 20th century basic climate features in Central America. *Int. J. Climatol.* **35**, 3397–3421.

Jain, A. K. & Dubes, R. C. 1988 *Algorithms for Clustering Data.* Prentice-Hall, Englewood Cliffs, New Jersey.

Jurya, M. R. 2015 Statistical evaluation of CMIP5 climate change model simulations for the Ethiopian highlands. *Int. J. Climatol.* **35**, 37–44.

Knutti, R. 2008 Should we believe model predictions of future climate change? *Philos. Trans R. Soc.* **366**, 4647–4664.

Knutti, R., Masson, D. & Gettelman, A. 2013 Climate model genealogy: generation CMIP5 and how we got there. *Geophys. Res. Lett.* **40**, 1194–1199.

Kothawale, D. R., Revadekar, J. V. & Rupa Kumar, K. 2010 Recent trends in pre-monsoon daily temperature extremes over India. *J. Earth. Syst. Sci.* **119**, 51–65.

Kumar, D., Mishra, V. & Ganguly, A. R. 2015 Evaluating wind extremes in CMIP5 climate models. *Clim. Dyn.* **45**, 441–453.

Mujumdar, P. P. & Nagesh Kumar, D. 2012 *Floods in a Changing Climate Hydrologic Modeling.* International Hydrology Series, Cambridge University Press, Cambridge, UK.

Palazzi, E., Hardenberg, J. V., Terzago, S. & Provenzale, A. 2015 Precipitation in the Karakoram-Himalaya: a CMIP5 view. *Clim. Dyn.* **45**, 21–45.

Pennell, C. & Reichler, T. 2011 On the effective number of climate models. *J. Clim.* **24**, 2358–2367.

Perez, J., Menendez, M., Mendez, F. J. & Losada, I. J. 2014 Evaluating the performance of CMIP3 and CMIP5 global climate models over the north-east Atlantic region. *Clim. Dyn.* **43**, 2663–2680.

Perkins, S. E., Pitman, A. J., Holbrook, N. J. & McAveney, J. 2007 Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature and precipitation over Australia using probability density functions. *J. Clim.* **20**, 4356–4376.

Raju, K. S. & Nagesh Kumar, D. 2014a *Multicriterion Analysis in Engineering and Management.* Prentice Hall of India, New Delhi.

Raju, K. S. & Nagesh Kumar, D. 2014b Ranking general circulation models for India using TOPSIS. *J. Water Clim. Change* **6**, 288–299.

Raju, K. S., Sonali, P. & Nagesh Kumar, D. 2016 Ranking of CMIP5 based global climate models for India using compromise programming. *Theor. Appl. Climatol.* Doi: 10.1007/s00704-015-1721-6.

Reifen, C. & Toumi, R. 2009 Climate projections: past performance no guarantee of future skill? *Geophys. Res. Lett.* **36**, L13704.

Sonali, P. & Nagesh Kumar, D. 2013 Review of trend detection methods and their application to detect temperature changes in India. *J. Hydrol.* **476**, 212–227.

Su, F., Duan, X., Chen, D., Hao, Z. & Cuo, L. 2013 Evaluation of the global climate models in the CMIP5 over the Tibetan Plateau. *J. Clim.* **26**, 3187–3208.

Tiwari, P. R., Kar, S. C., Mohanty, U. C., Kumari, S., Sinha, P., Nair, A. & Dey, S. 2014 Skill of precipitation prediction with GCMs over north India during winter season. *Int. J. Climatol.* **34**, 3440–3455.

Wang, K. 2007 *Cluster validation toolbox, MATLAB central file exchange*, MathWorks. Available from: www.mathworks.

com/matlabcentral/fileexchange/14620 (accessed 1 May 2015).

Wang, K., Wang, B. & Peng, L. 2009 CVAP: validation for cluster analyses. *Data. Sci. J.* **8**, 88–93.

Yokoi, S., Takayabu, Y. N., Nishii, K., Nakamura, H., Endo, H., Ichikawa, H., Inoue, T., Kimoto, M., Kosaka, Y., Miyasaka, T., Oshima, K., Sato, N., Tsushima, Y. & Watanabe, M. 2011 Application of cluster analysis to climate model performance metrics. *J. Appl. Meteorol. Clim.* **50**, 1666–1675.

Yuan, X. & Wood, E. F. 2012 On the clustering of climate models in ensemble seasonal forecasting. *Geophy. Res. Lett.* **39**, L18701.