

A New Algorithm of Bayesian Model Averaging Based on SCE - UA Collection Averaging

Liu Jun-Hua, Zhang Hong-Qin, Zhang Cheng-Ming, Zhao Tianyu, Ma Jing

► **To cite this version:**

Liu Jun-Hua, Zhang Hong-Qin, Zhang Cheng-Ming, Zhao Tianyu, Ma Jing. A New Algorithm of Bayesian Model Averaging Based on SCE - UA Collection Averaging. Daoliang Li; Yingyi Chen. 8th International Conference on Computer and Computing Technologies in Agriculture (CCTA), Sep 2014, Beijing, China. IFIP Advances in Information and Communication Technology, AICT-452, pp.351-358, 2015, Computer and Computing Technologies in Agriculture VIII. <10.1007/978-3-319-19620-6_40>. <hal-01420251>

HAL Id: hal-01420251

<https://hal.inria.fr/hal-01420251>

Submitted on 20 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A New Algorithm of Bayesian Model Averaging Based on SCE - UA Collection Averaging

LIU Jun-hua^{1,2}, ZHANG Hong-qin¹, ZHANG Cheng-ming^{1,*}, Zhao Tianyu³, Ma Jing¹

1. Shandong Agricultural University, College of Information Science and Engineering, Taian, Shandong

2. Institute of Agricultural standards and Detecting Technology SAAS, Jinan, China

3. Shandong South-North Water Diversion Corporation Limited, Jinan, China

Abstract: Bayesian Model Averaging (BMA) is a statistical method used for multi-model ensemble forecast system. Firstly, the likelihood function of BMA is improved by eliminating the explicit constraint, that the sum of weights is 1, and use SCE-UA for the minimization of its, which presents a new method for solving the Bayesian model averaging, that is the BMA-SCE-UA method. With three land surface models of soil moisture simulation test of multiple numerical model. By comparing the common Expectation Maximization (EM) method with the SCE-UA method, the results show that: SCE-UA method can improve the simulation performance of soil moisture in a large extent, and the soil moisture obtained by the BMA collection simulation and observation matches well, no matter from the amplitude variation and seasonal variability, which makes it possible that generating high accuracy data set of soil moisture with the method of BMA-SCE-UA and using multiple land surface models.

Key Words: Bayesian Model Averaging SCE-UA soil moisture optimization algorithm

Introduction

The model prediction is an import method that has been used in many areas. It contains the physical mechanism between the variables that can be more deeply reveal the changes and regulations of variables. But the prediction performance is not satisfied because there are many uncertainties in the models. Bayesian Model Averaging, a statistical probability forecast post-processing method that put forward by Raferly et al. This method obtains more reliable prediction results by giving the model forecast probability distribution some weights which is after the deviation correction. The weights are posterior probability, which represent the contribution of each model for the ensemble forecast. There are many research results have been already existed by using the BMA method show that the prediction performances are accurate, reliable and with incomparable advantages over other approaches^[1-3].

The key of BMA is accurately to estimate the weights of each model. Now the Expectation Maximization (EM) method is widely used. EM algorithm is easy, low calculation cost and requires that the weights are nonnegative and the sum of weights is 1. EM algorithm is used effectively and can generate relatively stable weights and variances^[1,4]. Unfortunately, this method cannot guarantee to get the global optimal solution, especially in high-dimensional case. In addition, EM algorithm assumes the variables are normal distribution. However, it is difficult to satisfy the assumption in many cases.

SCE-UA algorithm is not only an efficient optimization method to solve the nonlinear constraint problems but also can obtain the global optimal solution. Calculating the weights and variances are aimed at finding a set of optimal values to meet the constraints of the objective function. At this time, the weights can be as parameters, and rating by SCE-UA method. Structuring the objective function used for estimating weights is the key of BMA method, therefore proposes the new method for solving the Bayesian model averaging (BMA-SCE-UA)^[2,5,6]. The BMA-SCE-UA method is different

* correspondent author

from the EM algorithm without the assumption that forecast variables are normal distribution. This article selects three models (CLM, VIC and Noah) as the prediction models of soil moisture to test the performance. Comparing the results obtained by the EM, BMA-SCE-UA, the observation and the prediction of three models shows that in the calculation precision aspect, the BMA-SCE-UA method is better than EM algorithm.

1 Experiments and Methods

SCE-UA method is an efficient optimization method to solve the nonlinear minimization constraint problems, which is combining the deterministic complex search technology and the principle of biological evolution in nature. It combines with deterministic methods, stochastic methods, competition evolution methods and complex mixture methods to ensure the flexibility, global, consistency, and validity of the SCE-UA algorithm. The core of SCE-UA algorithm is the Competition Complex Evolutionary algorithm (CCE) [7-9]. In the CCE method, each vertex is the potential father and is likely to produce the next generation.

There are multiple parameters in SCE-UA algorithm, mainly including: the vertices of each complex polygon m ($m \geq n+1$), the vertices of each child complex number q ($2 \leq q \leq m$), the complex number of participating in the evolution p ($p \geq 1$), the sample size s ($s = p*m$), consecutive reproduce complex y ($y = 1$) of each child, each child complex evolution steps z ($z = 2n + 1$), the number of the waiting for optimization parameters n . The m value should not be too large, otherwise the calculation time will be too long and efficiency also will not be high yet [10-12].

Bayesian Model Averaging needs to estimate the weights and variances of each member of the mode of conditional probability function. Firstly, this paper improves the likelihood function of BMA by eliminating the explicit constraint, that the sum of weights is 1, and minimizes the likelihood function by the SCE-UA method. To solve the Bayesian model averaging, we put forward a new method that is the BMA-SCE-UA method.

Actually, estimating the weights and variances is to maximum the objective function as follows:

$$\max\left(\sum_{s=1}^S \sum_{t=1}^T \log\left(\sum_{k=1}^K \omega_k g_k\right)\right) \quad (1)$$

Here $0 \leq \omega_k \leq 1$, and $\sum_{k=1}^K \omega_k = 1$. In order to get rid of the restrain $\sum_{k=1}^K \omega_k = 1$, this paper introduces

$$\omega_k = \frac{\alpha_k}{\sum_{k=1}^K \alpha_k}$$

an intermediate variable and normalizes. Make to eliminate the constraint

$\sum_{k=1}^K \omega_k = 1$, and at the same time, the above maximum problem (1) is converted into minimum

problem by inversing, namely:

$$\min\left(\frac{1}{\left(\sum_{s=1}^S \sum_{t=1}^T \log\left(\sum_{k=1}^K \frac{\alpha_k}{\sum_{k=1}^K \alpha_k} g_k\right)\right)}\right) \quad (2)$$

$$(0 \leq \alpha_k \leq 1) \quad (k=1, \dots, k) \quad (3)$$

The flowchart of solving the minimization problem by using SCE-UA algorithm is as following figure 1:

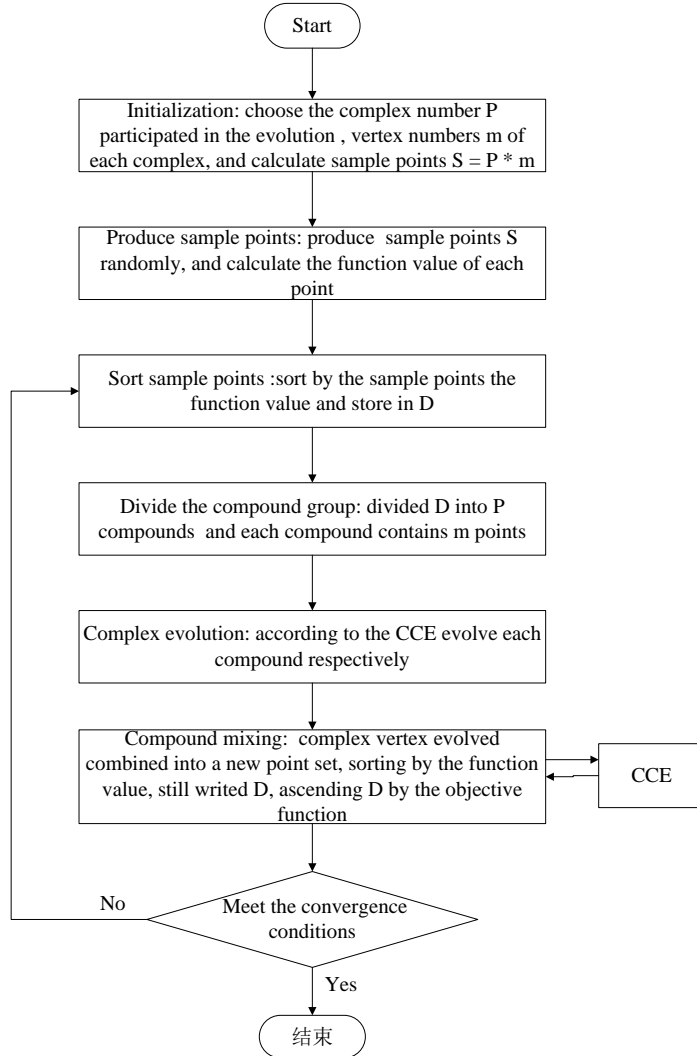


Fig. 1 The flowchart of solving the minimization problem by using SCE-UA algorithm

Calculating to obtain the optimal value of $\alpha_k, \sigma_k (k=1, \dots, k)$, then we can obtain the weight and variance of the competition model.

2 Experiment Data

This paper uses three land surface models (CLM3.5, Clm-Vic, Noah) with the atmospheric forcing data drive models to generate the soil moisture data to test the effectiveness of BMA-SCE-UA method. We choose the north region of China (110 °~ 120 °E, 40 °to 45 °N) containing the observation of 14 sites

from 1993 to 2006 to train the BMA. Each site is observed at 8, 18, 28 monthly. Observation level is divided into three layers: 0~10 cm, 40~50 cm and 90~100 cm. Due to observation at 40~50 cm and 90~100 cm is relatively scarce, the experiment is considered only 0~10 cm level. We can analysis the calculation accuracy by comparing the EM with SCE-UA method,.

3 Experiment Process and Analysis

Figure 2 shows the soil moisture data results of three models prediction (CLM3.5, CLM - VIC, Noah), average prediction and observation from 1993 to 2006. As can be seen from figure 2, the soil moisture is changing with season regulation and the trend of change each year. At the same time, three model forecast precisions still need to be improved.

Figure 2 shows the forecast results of three models ensemble mean have no better than a single mode. On one hand, each model prediction results can have equal coefficient by using simple arithmetic average, and without considering the performance of each model. On the other hand, it could be that the number of used models (only three) is less, which makes the simple average simulation results to be uncertainty.

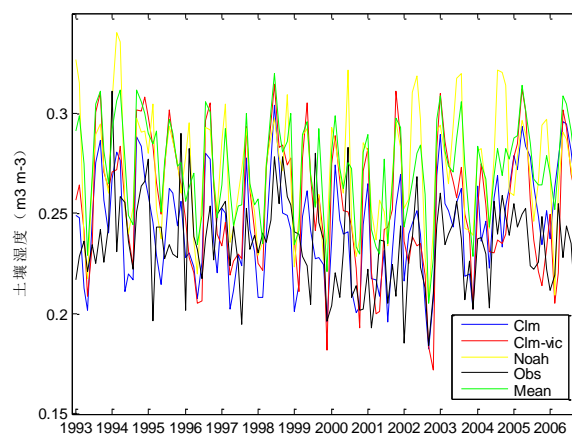


Fig. 2 Soil moisture time series

Figure 3 depicts the soil moisture time series results simulated by two BMA methods and observations. As can be seen from figure 2, the simulation results of two BMA methods have obvious improvement in comparison with the results of each model and model averaging. This is mainly because the BMA method can give full consideration to the prediction ability of each model and endow the different coefficient according to their respective contributions. Their weight coefficients are the posteriori probability representing the prediction techniques in training model, which will improve the simulation precision in a large extent. Firstly a simple linear regression is used for the model forecast results in BMA method, which is actually error correction and bound to improve the simulation precision in certain extent.

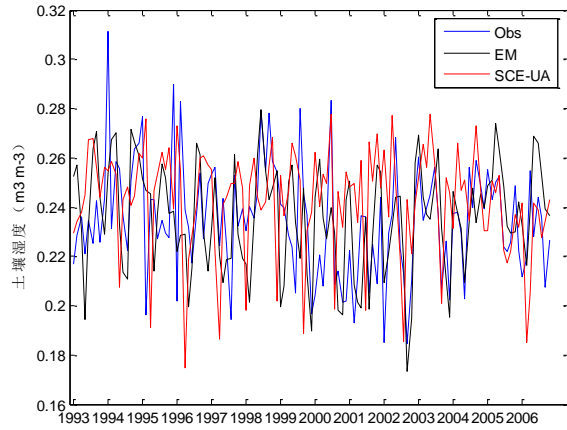


Fig. 3 Simulated time series of two methods of BMA

Figure 4 shows the correlation coefficient and the root mean square error of soil moisture respectively representing the two BMA simulations, three land surface models directly simulation, their collection average simulation and observation. Generally the correlation coefficient of two BMA methods is above 0.5. Compared with two BMA algorithms, the EM algorithm is poorer than SCE-UA algorithm. The correlation coefficient of BMA-SCE-UA method is above 0.6, while the direct simulation of each model and their simple average simulation is less than 0.5, and the effect of Noah is the best, its correlation coefficient reaches above 0.4, the other is less than 0.4, the ensemble average simulation results is better than CLM-VIC model. From the root mean square error results, it is significant that the BMA-SCE-UA method is superior to the EM method.

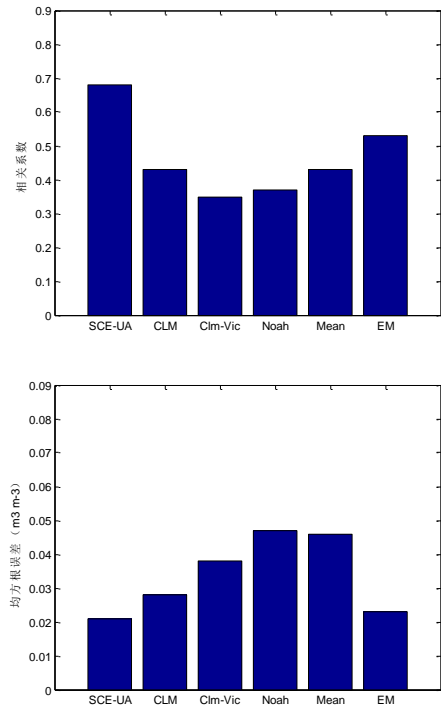


Fig. 4 The correlation coefficient and the root mean square error of soil moisture time series of different methods of simulation results and the observation

4 Discussions

Firstly, the BMA likelihood function is improved in this paper, getting rid of the restraint that sum of weights is 1 and transforming into minimization problem, then solving by the SCE-UA. In this thing called BMA-SCE-UA method. The soil moisture multi-model numerical simulation test using three land surface models shows that the BMA-SCE-UA method is better than the EM algorithm in calculation precision. In addition, the BMA-SCE-UA method is rid of the assumption that forecast variables are normal distribution, and broadening the application. Numerical experiments show that the BMA-SCE-UA method can largely improve the simulation effect, the BMA collection simulation of soil moisture and the observation of soil moisture match well, no matter from the amplitude variation and seasonal variability, which makes it possible that makes use of multiple land surface process models combining the BMA-SCE-UA method to generate higher accuracy of soil moisture data integration. It should be pointed out that, due to the nature of the BMA multi-model prediction method is a model training method, so we must use observed data in the collection each forecast model for training and testing to determine the coefficient of the BMA. But as for some regions with no observation data, this method is invalid.

Acknowledgments: National High Technology Research and Development Program of China (863 Program) (2013AA122003), Foundation for Outstanding Young Scientist in Shandong Province (BS2 011DX031), and Science and Technology Develop Project in Shandong Province (2012GSF11713), Funded by Key Laboratory of Geo-informatics of State Bureau of Surveying and Mapping(201319)

References

1. Tian X J, Xie Z H, Wang A H, et al. A new approach for Bayesian model averaging. *Sci China Earth Sci*, 2011,doi:10.1007/s11430-011-4307-x.
2. Ciccone, A. and Jarociński, M. (2010). Determinants of Economic Growth: Will Data Tell? *American Economic Journal: Macroeconomics*, forthcoming. Stefan Zeugner (2011) . Bayesian Model Averaging with BMS.
3. Vrugt JA, ter Braak CJF, Diks CGH, Robinson BA, Hyman JM, Higdon D (2008a) Accelerating Markov chain Monte Carlo simulation by self-adaptive differential evolution with randomized subspace sampling. *Int J Nonlinear Sci Numer Simul* (in press)
4. Vrugt JA, ter Braak CJF, Clark MP, Hyman JM, Robinson BA (2008b) Treatment of input uncertainty in hydrologic modeling: doing hydrology backwards with Markov Chain Monte Carlo simulation. *Water Resour Res* doi:10.1029/2007WR006720
5. Wöhling T, Vrugt JA (2008) Combining multi-objective optimization and Bayesian model averaging to calibrate forecast ensembles of soil hydraulic models. *Water Resour Res* doi:10.1029/2008WR007154
6. Ajami NK, Duan Q, Sorooshian S. An integrated hydrologic Bayesian multimodel Combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resour Res*, 2007, 43: W01403, doi: 10.1029/2005WR004745
7. Raftery A E, Gneiting T, Balabdaoui F, et al. Using Bayesian model averaging to calibrate forecast ensembles. *Mon Weather Rev*, 2005, 133: 1155–1174

8. N.K. Ajami, Q. Duan, X. Gao, S. Sorooshian Multi-model combination techniques for hydrological forecasting: application to distributed model intercomparison project results *J Hydrometeorol*, 8 (2006), pp. 755 - 768
9. Ramkumar Rajagopal & Enrique del Castillo (2005) Model-Robust Process Optimization Using Bayesian Model Averaging, *Technometrics*, 47:2, 152-163.
10. Xie, H., Eheart J W, Chen Y, et al., An approach for improving the sampling efficiency in the Bayesian calibration of computationally expensive simulation models, *Water Resour. Res.*, 2009, 45:W06419, doi:10.1029/2007WR006773
11. Tian, X. J, Xie Z H , Dai A G, et al, A dual-pass variational data assimilation framework for estimating soil moisture profiles from AMSR-E microwave brightness temperature, *J. Geophys. Res.*, 2009, 114, D16102, doi:10.1029/2008JD011600.
12. Tian, X. J, Xie Z H , Dai A G, et al, A microwave land data assimilation system: Scheme and preliminary evaluation over China, *J. Geophys. Res.*, 2010, 115, D21113, doi:10.1029/2010JD014370.