SENSITIVITY ANALYSIS ABOUT THE METHODS OF UTILIZING THE HIGH RESOLUTION CLIMATE MODEL SIMULATION FOR KOREAN WATER RESOURCES PLANNING (I) : THEORETICAL METHODS AND FORMULATIONS

CHANGSAM JEONG[1], SANGJIN LEE[1], ICKHWAN KO[1], JUNHAENG HEO[2] and DEGHYO BAE[3]

[1] Hydrosystems Engineering Center, Korea Institute of Water and Environment, Korea Water Resources Cooperation, 462-1 Jeonmin-Dong, Uusung-Gu, Daejeon, 305-730, Korea
(Tel: 82-42-860-0343, Fax: 82-42-860-0349, email: chs1229@kowaco.or.kr)
[2] Department of Civil Engineering, Yonsei University, Seoul, Korea
(Tel: 82-2-2123-2805, 82-2-3408-3332, email: jhheo@yonsei.ac.kr)
[3] Department of Civil and Environment Engineering, Sejong University, Seoul, Korea
(Tel: 82-2-3408-3814, Fax: 82-2-364-5300, email: dhbae@sejong.ac.kr)

Abstract: Nowadays Climate disasters are frequently happening due to occasional occurrences of El Nino and La Nina events and among them, water shortage is one of the serious problems. To cope with this problem, climate model simulations can give very helpful information. To utilize the climate model for enhancing the water resources planning techniques, probabilistic measures of the effectiveness of global climate model (GCM) simulations of an indicator variable for discriminating high versus low regional observations of a target variable are proposed in this study. The objective of this study is to present the various analysis methods to find the suitable application methods of GCM information for Korean water resources planning. The basic formulation uses the significance probability of the Kolmogorov-Smirnov test for detecting differences between two variables. The various methods for adopting correct association, changing the window size, discrimination condition, and the use of temporally downscaled data were proposed to find out the suitable way for Korean water resources planning.

Keywords: GCM, Kolmogorov-Smirnov test, correct association, changing the window size, discrimination condition

1. INTRODUCTION

Due to industrial development throughout the world, climates are changing rapidly and this affects us in many ways sometimes causing disasters. Among the various climate disasters, water problems occur to unexpected precipitation variations. Theses water problems can be classified into two categories; water shortage or water exceedance. Korea also suffers from those climatological disasters. For example, there was a nation-wide drought in the spring of 2001, and a flood in the summer of 2002. The damages were so heavy, that the Korean government started to prepare several measures to cope with these problems. There are two measures to overcome water shortage problems in Korea; one is structural method, the other is unstructural method. Structural method implies the construction or extension of every water resources related structure (dam, reservoir, phreatic development, etc.). Unstructural method means enhancing the management techniques of water management or developing the optimal rules using the existing hydro-structures. Recently the policies of the Korean government regarding water resource s management were focused on the unstructural methods instead of structural methods because of serious opposite of NGO and environmental problems caused by the structural methods. With regard to unstructural methods, developing the control rule for water resources management is necessary. The most important factor of developing the control rule is long range forecast of hydrometeorological components like precipitation, temperature, and soil moisture. There are no doubts that GCMs can be the good solution about this problem. However, KOWACO (Korea Water Resources Cooperation), which is the main institution for Korean water resources management, are not using the GCM results. Uncertainty of climate model information and no basin scale forecasts but nation wide scale forecast prohibited its practical use to the water resources planning. They has used the frequency analysis method of the past inflow data to the reservoir for developing the operation rule instead of using the climate model information. If we can get the uncertainties of analyzed basin scale climate forecast, more efficient reservoir control methods will be developed and it will enhance the water supply efficiency. Recently some research about the climate impact assessments for regional weather and water resources have done under the climate changes (e.g., *Changnon, 1985; Gleick, 2000*). In addition, the first reported uses of climate model information for improved water resources management under climatic variability and change (*Georgakakos et al., 1998; Carpenter and Georgakakos, 2001; Yao and Georgakakos, 2001*) have indicated benefits of these uses in some regions, when the considerable uncertainty in climate model simulations is quantified and downscaled to regional spatial scales.

The main purpose of this research is to analyze the uncertainties of climate model information and do sensitivity analysis about the effective factors through the proposed methods to find out the optimal application methodologies for Korean water resources planning. Recently Georgakakos (2003) proposed a methodology for measuring the effectiveness of ensemble climate-model simulations for regional impact assessment. Jeong et al. (2004) showed the effectiveness of GCM information for Korean basin using the same concept.

Sensitivity analyses using probabilistic methodology in various methods were conducted to measure the sensitivities. The next section provides brief reviews of the theory and the concept of sensitivity analysis methods. A case study of the designed test for Korea is given in section 3 and concluding remarks follows in section 5. A companion paper (Jeong et al. this issue) develops the numerical experiment results and analysis.

2. METHODOLOGIES

Recently Georgakakos (2003) proposed a methodology for measuring the effectiveness of ensemble climate-model simulations for regional impact assessment. ECHAM3 ensemble simulations were used to discriminate the observed values of 334 climate divisions in the U.S.A. with proposed probabilistic diagnostic measure. The main focus of this study is to do sensitivity analysis of various climate models in order to determine useful methods for GCM application in Korean finding out the recommendable methods for GCM application in Korea water resources management using the previous concepts. At first, we briefly review the basic concept of analysis method and then set the various analysis methods to detect the sensitivity factor of each test.

2.1 Theory overview

The probabilistic utility index method proposed by Georgakakos (2003) was adopted. This theory is based on the premise that for climate model simulations/predictions of a certain indicator variable (e.g., nodal precipitation) to be useful for water resources impact assessments that involve regional surface target variables (e.g., watershed mean areal precipitation), the extreme values of the climate model nodal values should correspond to extreme values of the target variable. If there is no such "signal" in the climate nodal values, it is unlikely that they will be useful for water resources assessments that involve the mentioned target variable. The theory accounts for potential shifts in climate model output in space and time. It may also be generalized for vector indicator and target variables given adequate record lengths.

We present the mathematical formulation as a means of establishing notation and to fix ideas. The discussion follows Georgakakos (2003) closely, and the interested reader is referred to that reference for more detailed information. Mathematical configuration can be expressed as in Fig. 1. Notation of Fig. 1 can be found in the APPENDIX.

2.2 Analysis methods to detect the sensitivity for water resources management

There are two kinds of basic variables for the analysis. One is indicator variable, which is the nodal values of applied GCM simulation, and the other is target variable, which is the mean averaged observation. Considering the availabilities of each variable, we applied three test methods. Surface precipitation and surface temperature values of each node are used from GCM simulation and values of the basin scale mean averaged observed precipitation, temperature, and observed gauge discharge are also used for these tests. Fig. 2 presents the concept of 3 kinds of analysis method. The first one is precipitation-precipitation test (P-P test) which uses nodal precipitation values for indicator and basin scale mean averaged observed precipitation values for target variables. The second test is precipitation-discharge test (P-Q test) which uses nodal precipitation values for indicator and observed discharge values for target variables. The last test is temperature-temperature test (T-T test) which uses nodal temperature values for indicator and basin scale mean averaged observed temperature values for target variables. The selection of nodal values for each analysis is defined by setting of analysis window.

Let's review the general procedures of the test considering the proposed theory to set the more detailed design of the test for detecting

Discrimination of Target Variable using Indicator

⇩

2 CDF : Divide in two Distribution (F_L, F_U)

$$F_L(u_i^{(j)}/n) = \{j/N_L; u_i \in \Omega_u : (P_n^m \le q_l)/n\}$$

$$F_U(u_i^{(j)}/n) = \{j/N_U; u_i \in \Omega_u : (q_u \le P_n^m)/n\}$$

⇩

Decision of D_n

$$D_n = \max[\,|\,F_L(u_i^{(j)}/n) - F_U(u_i^{(j)}/n)\,|\,]$$

⇩ **K-S Test**

Decision of $P_{KS}(D > D_n)$

$$P_{KS}(D > D_n) = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2\wedge^2}\ ;\ \wedge = [\sqrt{N_e} + 0.12 + 0.11/\sqrt{N_e}]\,D$$

$$with\quad N_e = \frac{N_L N_U}{N_L + N_U}\qquad : N_e \ge 4\ (It\ means\ N_u \ge 5)$$

⇩

Estimator

$$E_P = \frac{1}{N_\lambda}\sum_{n=1}^{N_\lambda} P_{KS}(D > D_n)$$

Fig. 1 Diagram of mathematical formulation for the study

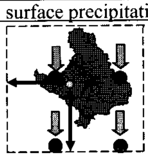
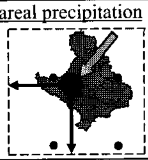
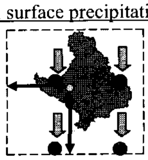
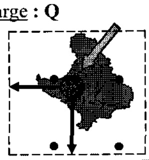
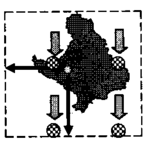
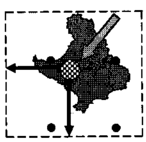
Indicator variable (Nodal values of GCMs)	Target variable (Mean averaged observation)	Test name
Nodal surface precipitation : P 	Mean areal precipitation : P 	**P-P**
Nodal surface precipitation : P 	Discharge : Q 	**P-Q**
Nodal surface temperature : T 	Mean areal temperature : T 	**T-T**

Fig. 2 Conceptual diagrams of analysis method

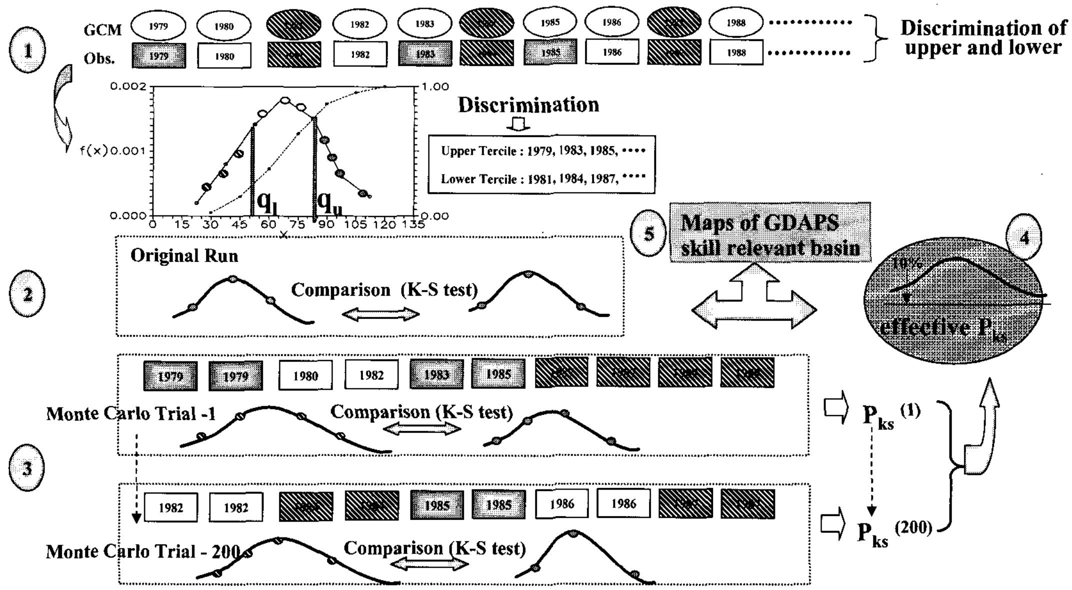


Fig. 3 General procedure to check the applicability of GCM information

sensitivities of the test results with the above three tests. Fig. 3 presents general procedures for the test. The first step is discrimination of the observations using GCM node data (①), the second step is calculation of statistic value of the original run using Kolmogorov-Smirnov test to detect the differences of discriminated values (②), and the third step is Monte Carlo trials (we will call it as 'MCS' bellows) using random sampling to establish the significant threshold (③). In all cases with the available data, 200 times MCS were conducted to estimate the threshold value of the utility index that was significant at the 10% significance level presented in ④. We set this value as significant threshold. Finally in ⑤, we can define whether GCM is significant or not to discriminate the observations.

Considering the conceptual diagram shown in Fig. 2 and general procedures shown in Fig. 3, various analysis methods could be found. Setting of various windows sizes, various discrimination tests to divide the GCM and target values, and whether adopting or not of considering the correct association of discriminated observation variables can be the good methods for the sensitivity analyses. In the next section, more detailed application methods to find out the optimal use of GCM information for water resources managements through the sensitivity analyses of proposed conditions will be described.

2.2.1 Correct association

It is noted that the development of the high and low distributions may be further constrained by ensuring correct association between the seasonal target variables. That is, only including in the distribution values which have greater for high and lower for low conditioning. This will be referred to as *"constrained"* and the opposite case will be referred to as *"unconstrained."*

2.2.2 Window size

Consider a certain watershed of interest and a region centered on it that contains several nodes of a climate model. In Fig. 5, (x_n, y_n) are the

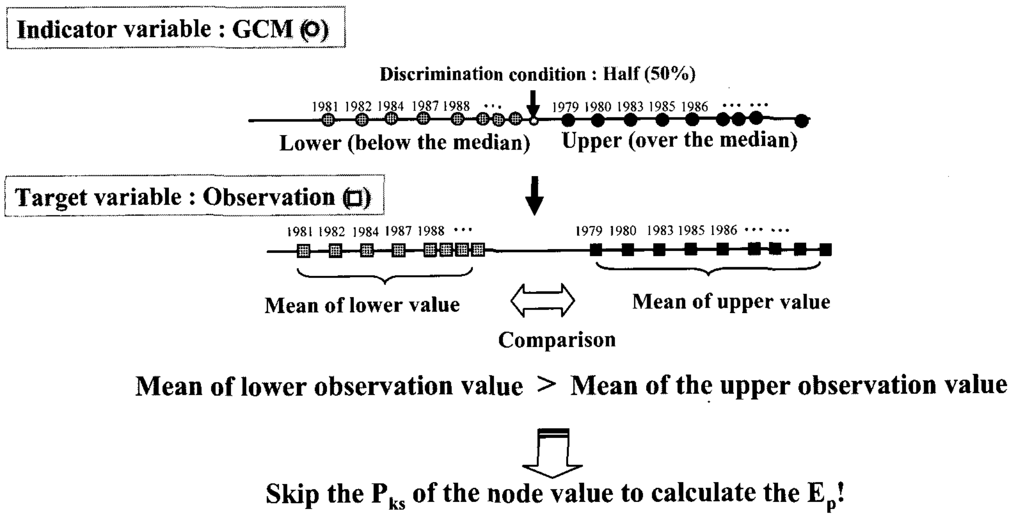


Fig. 4 Concept of correct association ("*Constrained*")

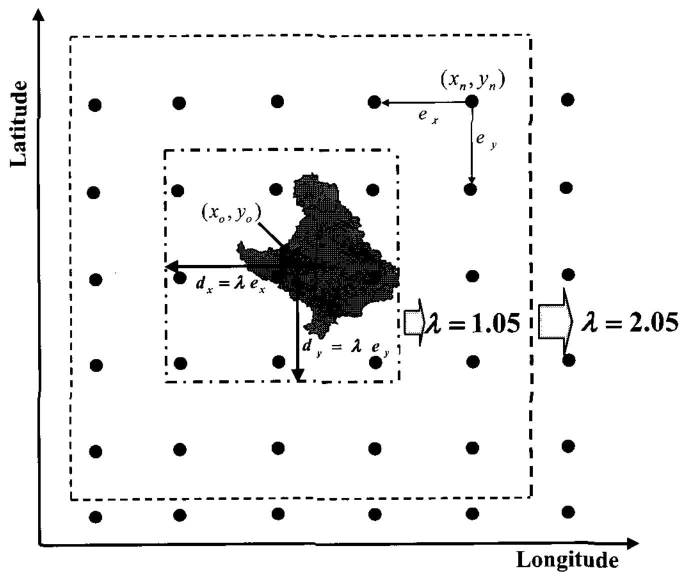


Fig. 5 Setting of analysis window considering the spatial condition

longitude and latitude coordinates of the n^th node of the climate model under consideration, e_x and e_y are the inter-nodal distances in the longitude and latitude directions, (x_o, y_o) are the coordinates of the centroids of the study watershed, and d_x and d_y are the longitude and latitude distances of the watershed centroid from the boundaries of the window. The distances d_x and d_y are multiples of the inter-nodal distances e_x and e_y, respectively, with factor λ (window size) specified by the user. The larger the value of λ, the larger the number of climate model nodes used in the analysis.

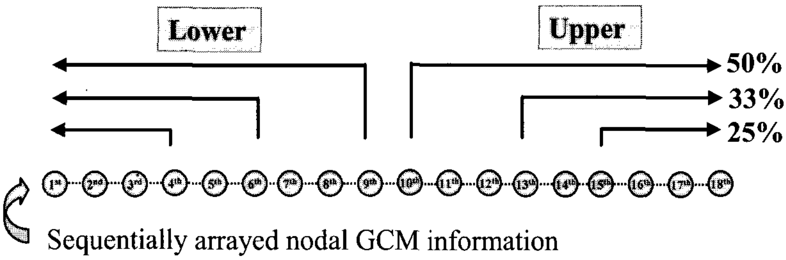


Fig. 6 Discrimination conditions of target variables using climate model information

2.2.3 Discrimination condition

We can set the various kinds of discrimination conditions for the test shown in Fig. 6. The only limit for the discrimination condition is data availability. As shown in Fig. 1, N_e has to be a bigger value than 4 to get the stable P_{ks} values in the K-S test. To satisfy this limitation, the minimum numbers of each part should be over 8. Fig. 6 describes the median (50%), tercile (33%), and quartile (25%) discrimination conditions. To divide the indicator variables with a decided condition, sequential arrays should be preceded. We should then discriminate the data using the decided condition. As shown in Fig. 6, division point of half discrimination of 18 data is 9^{th} value; data are divided with below the median and the upper the median. For the tercile discrimination, 6^{th} and 15^{th} values become division points. But as mentioned before, this case has just 6 members only for each part, and P_{ks} values can not be calculated. Finally discrimination of the target variables is done using these corresponding indicator variable discriminations.

2.2.4 Use of temporally downscaled data

Considering the condition of $N_e \geq 4$ which was described in Fig. 1, there are some limitations for the test. For instance, we need at least 16 years of data collection for the median discrimination analysis and 24 years data for the tercile discrimination analysis. But high resolution climate model simulations take a very long time and a large computational effort for their simulations, so it is difficult to get the desired simulation period data for various analyses. To overcome this limitation, use of temporally downscaled data is recommended. For example, we need at least 24 years of data for the seasonal analysis with tercile discrimination condition. But if we use the temporally downscaled data, each season has 3 individual values instead of one value, so 8 years data is enough for the seasonal analysis with tercile discrimination condition. In this study, we compared both cases for the analysis.

3. CONCLUSION

In this paper, some methods of utilizing the climate model simulation for water resources planning were presented using the probabilistic utility index E_P values under significant model and downscaling uncertainty.

To utilize the climate model for enhancing the water resources planning techniques, probabilistic measures of the effectiveness of global climate model (GCM) simulations of an indicator variable for discriminating high versus low regional observations of a target variable are applied in this study. The objective of this study is to check the sensitivity in accordance with various ways and to find suitable methods for Korean water resources planning. The formulation uses the significance probability of the Kolmogorov-Smirnov test for detecting differences between two variables.

The various methods for adopting correct association, changing the window size, discrimination condition, and the use of temporally downscaled data were proposed to find out the suitable way for Korean water resources planning. In case of correct association, two kinds of analysis (*Constrained, Unconstrained*) can be considered. In case of setting the analysis window sizes, several kinds of sizes can be applicable, but usually 1.05 or 2.05 times of grid area of adopted GCM is recommended. Discrimination condition is constrained by the data period of indicator and target variable. If there is not enough data period for planned time step, use of temporally downscaled data can be the good alternative. Originally use of temporally downscaled data was presented for coping with the shortage of simulation periods, but it can also be good comparison data with original run (averaged ways)

In the companion paper (Jeong et al., this issue) practical sensitivity analysis will be done using the proposed methods.

REMARK

The research work reported herein was sponsored by the Korean 21st Century Frontier Project.

REFERENCES

Carpenter, T.M., and Georgakakos, K.P., 2001. Assessment of Folsom Lake response to historical and
potential future climate scenarios, 1, Forecasting. *J. Hydrol.* 249, 148-175.

Changnon, S.A., 1985. Climate Fluctuations and Impacts: The Illinois Case. *Bull. Amer. Meteor. Soc.* 66, 142-151.

Georgakakos, K.P., 2003. Probabilistic climate-model diagnostics for hydrologic and water resources impact studies. J. Hydrometeorol. 4, pp. 92-105.

Georgakakos, K.P., A.P. Georgakakos and N.E. Graham, 1998: Assessment of Benefits of Climate Forecasts
for Reservoir Management in the GCIP Region. *GEWEX News* 8(3), 5-7.

Jeong, C. S., Heo, J.H., and Bae, D.H., 2004. Uncertainty analysis of GCM information in Korea using probabilistic diagnostics. *KWRA 37(3)*, (in Korean)

MOCT, 2001. Water Vision 2020. Document of Ministry of Construction and Technology, Korea, pp. 5-11. (in Korean).

Press, W. H., Teukolsky, S.A., Vetterling, W. T., and Flannery, B. P., 1992. Numerical Recipes in FORTRAN, The Art of Scientific Computing. Cambridge University Press, New York, 963pp.

Yao, H., and Georgakakos, A.P., 2001. Assessment of Folsom Lake response to historical and potential future climate scenarios, 2, Reservoir management. *J. Hydrol.* 249, 176-196.

APPENDIX

D_n : Test statistic for a given node and climates model realization

$F_L(u_i^{(j)}/n)$: Cumulative distribution of target variables which are divided in low by indicator variable

$F_U(u_i^{(j)}/n)$: Cumulative distribution of target variables which are divided in upper by indicator variable

n : Node

N_L : Total number of sample values of historical target variable for month m for which the global climate model seasonal indicator variable P_n^m is in a low q_l

N_U : Total number of sample values of historical target variable for month m for which the global climate model seasonal indicator variable P_n^m is in an upper q_u

P_n^m : Global climate model seasonal indicator variable for node n, month m

q_l : Lower limit value of indicator variable

q_u : Upper limit value of indicator variable

$u_i \ (i = 1,\Lambda, N_D)$: Observed target variables

$u_i^{(j)}$: j[th] largest value of the observed target variable sample

Ω_u : The set of historical observed values of target variable

Ω_λ : All global climate model nodes

Hydrosystems Engineering Center, Korea Water Resources Cooperation, Daejeon, Korea

(E-mail: chs1229@kowaco.or.kr)

Department of Civil Engineering, Yonsei University, Seoul, Korea

Department of Civil and Environment Engineering, Sejong University, Seoul, Korea