이학석사 학위논문

Analysis of stepwise uncertainty of water resource forecast

수자원 전망의 단계별 불확실성 분석

2019 년 2 월

서울대학교 대학원

통계학과

이 지 수

Analysis of stepwise uncertainty of water resource

지도교수 김 용 대

이 논문을 이학석사 학위논문으로 제출함

2018 년 12 월

서울대학교 대학원

통 계 학 과

이 지 수

이지수의 이학석사 학위논문을 인준함

2018 년 12 월

위 원 장 　 박 병 욱 　 (인)

부위원장 　 김 용 대 　 (인)

위 　 원 　 원 중 호 　 (인)

Abstract

Analysis of stepwise uncertainty of water resource forecast

Ji su Lee
The Department of Statistics
The Graduate School
Seoul National University

The prospect of various resources considering climate change is academically meaningful but also politically important. Among them, the importance of water resources can not be overstressed at all, because there are not so many important resources like water to human beings. As recent droughts and floods become frequent and severe, there is a growing need to predict the outlook for water resources due to climate change. Water resources forecasts due to climate change are largely the emission scenarios, global circulation models, downscaling technique, and hydrological models. It is also important to quantify the uncertainty of each outlook, as well as forecasts interaction with each outlook phase. In this paper, we propose a method to quantify the uncertainty of each forecasting stage, with the sum of each forecasting step being equal to the total uncertainty. It also suggests ways in which uncertainty at each stage of the forecast can be resolved once more at various stages of the forecasting phase to enable comparisons within the forecasting stage.

keywords : Uncertainty analysis, water resource forecasting due to climate change, disaggregation of each forecasting stage
Student Number : 2017-25184

Contents

List of Tables

Chapter 1

Introduction

While policy prospects for various resources are crucial, water is the most important resource for humans. And the scarcity of water is becoming increasingly important, the need for water resources is increasing. The characteristics of water resources are difficult to solve short-term problems such as floods and droughts, and long-term anticipation and defense measures are needed. Therefore, considering climate change in advance and accurately estimating its effect is also important for establishing policies.

In Chapter 2 of this study, we will describe in detail the data to be applied to the modeling and the forecasting stage of the water outlook. And I will explain the research model of the data theoretically. The goal to pursue in this study is divided into two stages. The first step is to quantify the total uncertainty and distribute each outlook phase. The second step is to break down each of the forecasting steps one more time, so that it is possible to clarify in detail which uncertainties are large. In this way, uncertainty can be accurately measured, not vaguely predicted by detailed steps.

In Chapter 3, we analyze the results of applying the model presented in chapter 2. This stage is also divided into two stages. The first step is to divide the total uncertainty. And the second step is to check the result of the uncertainty of each forecast step.

In chapter 4, we will present the conclusion about the analysis of the result value and significance of the research model.

In this paper, we propose a method to quantify the uncertainty of each forecasting stage. It also suggests ways in which uncertainty at each stage of the forecast can be resolved once more at various stages of the forecasting phase to enable comparisons within the forecasting stage.

Chapter 2

Preliminaries

Two categories are considered in this chapter. The first step introduces the data and background to be applied in detail. In the second phase, we will describe a model that quantifies and assigns the uncertainty of water forecasting measurements in this paper.

2.1 Detailed description of the data

Water forecasts are usually made in four steps. The first step is to develop a different scenario depending on how much greenhouse gas emissions are reduced or increased by the emission scenario. The second step is the global circulation model. In this model, simulations are made using parameters that mainly change the future climate, such as temperature, precipitation, and wind. As the spacing for each variable decreases, the prediction becomes more accurate and the uncertainty decreases. However, uncertainty can not but exist because it is almost impossible to include all characteristics of the atmosphere in the model. The third step is the downscaling technique. This is a step that uses the global circulation model to change the climate forecast for a specific region. The fourth step is the hydrological model. This is the stage of forecasting the flow through the three steps of forecast climate.

The emission scenarios used in this study were the A2 and B1 scenarios presented in IPCC SRES AR4. The global circulation model used four techniques, CSIRO-Mk30, MPI-M-ECHAM5, MIUB-ECHO-G and MRI-M-CGCM232. The downscaling technique used bilinear regression and two artificial neural networks. And the hydrological model used two models abcd and GR2M.

The number of cases is two for the emission scenarios and four for the global circulation model. and There are two downscaling techniques and two hydrological models. Therefore, the water resources forecast calculation can be expressed as the number of cases of 32 (2 × 4 × 2 × 2). This is summarized in Table 2.1.

Table 2.1 Models used in the data

Model	Explanation
Emission scenarios	A2, B1
GCMs	CSIRO-Mk30, MPI-M-ECHAM5, MIUB-ECHO-G, MRI-M-CGCM232
Downscaling techniques	BR, ANN
Hydrological model	abcd, GR2M

These forecasts are based on the flow data of Chungju Dam in July, August, December, February, and February. July, August, and September are bound to summer, and December, February and February are bound to winter. The period was from January 2030 to December 2059. The average flow of summer and winter was used for uncertainty analysis.

2.2 Uncertainty decomposition modeling

Let $x=(x_1,\ldots,x_4)\in\prod_{k=1}^{4}M_k$ denote the combination of emission scenarios, global circulation model, downscaling technique, and hydrological model. A2 and B1 of emission scenarios are the elements M_1, CSIRO-Mk30, MPI-M-ECHAM5, MIUB-ECHO-G and MRI-M-CGCM232 of global circulation model are the elements of the M_2, BR, ANN of downscaling techniques are the elements of the M_3 and abcd, and GR2M of the hydrological model are the elements of M_4. That is, k is a step estimated at the M_k stage, and the number of all cases can be considered in the same manner as above.

The prospect value Y_x through $x=(x_1,\ldots,x_4)\in\prod_{k=1}^{4}M_k$ can be expressed as $Y_x=\mu_x+\epsilon_x$ by the ANOVA model. ϵ_x is a random variable whose mean is zero and variance is σ^2 which is independent of other random variables. ϵ_x is an error in the variance analysis model, which is assumed to be caused by internal variability.

It is one of the core processes of this analysis to express the value μ_x which does not take into account the errors related to internal variability, as the sum of the main effects of each of the forecasting steps and the interactions of each step.

$$\mu_x=\mu_0+\sum_{k=1}^{4}\beta_{x_k}^{(k)}+\sum_{a=1}^{3}\sum_{b=a+1}^{4}\beta_{x_a,x_b}^{(a,b)} \tag{2.1}$$

μ_x can be expressed as the following equation, where $\beta_{x_a,x_b}^{(a,b)}$ is the intersection effect of the ath observation phase and the bth observation phase, and $\beta_{x_k}^{(k)}$ is the main effect of the kth observation phase.

$$\sum_{x\in\prod_{k=1}^{K}M_k}\left(Y_x-\mu_0+\sum_{k=1}^{4}\beta_{x_k}^{(k)}+\sum_{a=1}^{3}\sum_{b=a+1}^{4}\beta_{x_a,x_b}^{(a,b)}\right)^2 \tag{2.2}$$

An analytical model that minimizes the above equation is used. Using the parameter estimation results in the ANOVA, the uncertainty due to the main effect is quantified as follows.

$$U_k^{MAIN} = \frac{1}{n_k} \sum_{x_k \in M_k} \left(\overline{Y_{x_k}^{(k)}} - \overline{Y} \right)^2 \tag{2.3}$$

$\overline{Y_{x_k}^{(k)}}$ is the average of all forecasts using scenarios at the x_kth forecast stage, and $\overline{Y}$ is the overall average of the forecasts. We quantify $\frac{1}{n_k} \sum_{x_k \in M_k} \left(\beta_{x_k}^{(k)} \right)^2$ as uncertainty due to the k-th main effect. The reason for dividing by n_k is that the total number of scenarios in the step M_k is n_k. The uncertainty due to the interaction effect between the l-th forecasting stage and the m-th forecasting stage is quantified as follows.

$$U_{a,b}^{I} = \frac{1}{n_a n_b} \sum_{x_a \in M_a} \sum_{x_b \in M_b} \left(\overline{Y_{x_a, x_b}^{(a,b)}} - \overline{Y_{x_a}^{(a)}} - \overline{Y_{x_b}^{(b)}} + \overline{Y} \right)^2 \tag{2.4}$$

Where $\overline{Y_{x_l, x_m}^{(l,m)}}$ means the averages of all forecasting values using scenario x_l at the l-th forecasting stage and the scenario x_m at the m-th forecasting stage. And $\frac{1}{n_a n_b} \sum_{x_a \in M_a} \sum_{x_b \in M_b} \left(\beta_{x_a, x_b}^{(a,b)} \right)^2$ is quantified by the uncertainty of the interaction effect between the a-th stage and the b-th stage. The reason for dividing by $n_a \times n_b$ is that the number of total cases in which the alternation effect of the a-th stage and the b-th stage can be represented is $n_a \times n_b$.

And the total uncertainty minus the main effect and the interactions effect is the partial reliability by internal variability. Based on this, we can define U_k, the uncertainty of each forecast step as follows.

$$U_k = U_k^{MAIN} + \frac{1}{2}\sum_{a \neq b} U_{a,b}^{I} \tag{2.5}$$

The above definition is proposed by Ohn, I. et al. (2018). The uncertainty in the outlook phase is defined as the uncertainty due to the main effect plus the half of the uncertainty sum due to all interactions associated with that outlook phase. If the sum of the uncertainties due to the interactions is multiplied and then multiplied by half, the sum of the uncertainties for each of the forecasting steps can be estimated to be equal to the uncertainty of the whole step.

$U_{a,b}^{I}$ of equation (2.5) can be expressed as equation (2.6) by decomposing each of the scenarios in each of the forecasting stages through Equation (2.4).

$$U_{a,x_a} = \frac{1}{n_a}\left(\overline{Y_{x_a}^{(a)}} - \overline{Y}\right)^2 + \frac{1}{2n_a}\sum_{a \neq b}\left[\frac{1}{n_a}\sum_{x_a \in M_a}\left(\overline{Y_{x_a,x_b}^{(a,b)}} - \overline{Y_{x_a}^{(a)}} - \overline{Y_{x_b}^{(b)}} + \overline{Y}\right)^2\right] \tag{2.6}$$

Using Equation (2.6), we can accurately measure the uncertainty in the sub-step of model M_a used in the a-th step. Knowing precisely the uncertainties in the detailed steps is important in policy formulation and decision making, and it can be assumed that using this technique can be very helpful.

Chapter 3

Results

Two categories are considered in this chapter. The first step introduces uncertainty decomposition results at each forecasting stage. In the second step, we will explain the analysis by dividing each view step into detailed steps.

3.1 The result of decomposing uncertainty

The results of analysis using the ANOVA are as follows Table 3.1. It consists of uncertainties due to main effects of each forecast, interactions and internal variability. Uncertainties due to the main effects in each of the lines 1 to 4 are quantified using equation (2.3). The results from lines 5 to 10 are then quantified using the equation (2.4), which quantifies the interactions between the two different observation phases. It is divided into summer season and winter season. The uncertainty due to the main effects of GCM was the largest in the summer season followed by the uncertainty due to the main effect of the downscaling technique, and uncertainty due to the interaction effect between the GCM and the downscaling technique.

In the winter season, the uncertainty due to the main effect of GCM was the largest, followed by the uncertainty due to the interaction effect between GCM and downscaling technique, and uncertainty due to the main effect of GCM.

In this process, we can see that the order of uncertainty in the summer and winter seasons is different. Uncertainty due to the main effect of GCM was overwhelmingly high at 68.908% in the summer season, while uncertainty due to the main effect of downscaling was the largest at 52.746% in the winter season. This suggests that uncertainties in the forecast phase will vary significantly as the season changes. In summer, uncertainty due to the main effect of GCM is the biggest cause of uncertainty due to summer drought or typhoon, which suggests a big change in model selection.

Also, except for uncertainties due to GCM main effects in summer, the uncertainty due to the main effect of the downscaling technique was the largest at 13.315%. Main effect of the downscaling technique was the largest in the winter season, and it can be seen that the downscaling technique is the most important step to be taken except for the summer season's climate change like the typhoon or drought.

And the singularities obtained by quantifying the interaction effects are that the uncertainty due to the interaction effect between GCM and the downscaling technique is larger than expected. Researching or supplementing this point in the forecasting phase or in the policy setting phase may result in a further reduction of uncertainty. In addition, it can be seen that the internal variability does not have a large value in both the summer season and the winter season.

Table 3.1 uncertainty of main effects and interaction effects

	Season	
Effects	Summer	Winter
ES	868.41 (3.637%)	2.59 (0.091%)
GCM	16454.18 (68.908%)	496.76 (17.544%)
DS	3179.37 (13.315%)	1493.48 (52.746%)
HM	84.04 (0.352%)	3.16 (0.111%)
ES:GCM	1007.43 (4.219%)	76.15 (2.689%)
ES:DS	33.1 (0.139%)	14.3 (0.505%)
ES:HM	0.03 (0%)	0.08 (0.003%)
GCM:DS	1323.42 (5.542%)	646.91 (22.848%)
GCM:HM	158.17 (0.662%)	10.75 (0.38%)
DS:HM	1.06 (0.004%)	9.07 (0.32%)
Internal variability	769.37 (3.222%)	78.2 (2.762%)
Total	23878.6 (100%)	2831.44 (100%)

3.2 Break down into smaller steps

If the items in table 3.1 are summarized according to equation (2.5), the uncertainty can be quantified and distributed according to each of the four forecasting stages. The results are summarized in the following Table 3.2.

Table 3.2 Uncertainties of four stages

	Season	
Stage	Summer	Winter
Emission scenario	1388.7 (5.816%)	47.85 (1.69%)
GCM	17698.7 (74.119%)	863.66 (30.503%)
Downscaling technique	3858.16 (16.157%)	1828.61 (64.583%)
Hydrological model	163.67 (0.685%)	13.11 (0.463%)
Internal variability	769.37 (3.222%)	78.2 (2.762%)
Total	23878.6 (100%)	2831.44 (100%)

As shown in Table 3.1, the summer season has the largest uncertainty due to GCM and the downscaling technique is the largest in winter. However, it can be seen that the rankings have not changed since the uncertainties due to the interaction effects between the stages are relatively small. In other words, if the weights of uncertainties due to the interactions between the stages were large, the order of the uncertainties due to the main effects and the interactions was different from the order of the uncertainties due to the main effects.

And the metric of uncertainty at each forecasting stage in Table 3.2 is decomposed by equation (2.6). This is shown in Table 3.3 for the summer season and table 3.4 for the winter season.

The third line shows the uncertainty due to the main effect and the fourth line shows the uncertainty due to interactions effect. The uncertainties in the CSIRO-Mk30 method of GCM was the greatest in the summer season, followed by the MIUB-ECHO-G method in GCM, BR in the downscaling method, and ANN method in the downscaling method. In the winter season, the uncertainty due to the BR of the downscaling technique, ANN of the downscaling technique method was the largest, follwed by CSIRO-Mk30, MIUB-ECHO-G and MRI-CGCM232 of GCM method.

What is notable about the results is that the size can be compared a little more accurately by decomposing it into subdivided scenarios within the view stage. For example, in the case of the summer season, it can be concluded that the GCM stage has the greatest uncertainty when analyzed in Table 3.2. This analysis is entirely correct, but you can analyze the results of the slightly different by viewing and analyzing table 3.3. The analysis shows that the ratio of the uncertainties of the CSIRO-Mk30 method and the MIUB-ECHO-G method of the GCM stage are 50.463% and 14.117%, respectively, which are larger than other forecasting steps. However, the MPI-M-ECHAM5 and MRI-CGCM232 methods of the GCM stage account for 2.484% and 7.056% of the total uncertainty, respectively. This is below the uncertainty of 8.08% for the ANN method and the BR method of the downscaling technique. This confirms that a more detailed analysis of the prospecting phase can yield meaningful results. In addition, we can see that the uncertainty values of both methods are the same when there are two methods within each forecasting step. This is due to the fact that the distances to the outlook for the two methods from the overall average of the outlook are same.

Table 3.3 Uncertainties of four stages for summer season

stage	model	Uncertainty from main effect	Uncertainty from interaction effect	Scenario Uncertainty
ES	A2	434.205	260.145	694.35 (2.908%)
	B1	434.205	260.145	694.35 (2.908%)
GCM	CSIRO-Mk30	11744.18	305.821	12050 (50.463%)
	MIUB-ECHO-G	3069.38	301.514	3370.894 (14.117%)
	MPI-M-ECHAM5	242.752	350.364	593.116 (2.484%)
	MRI-CGCM232	1397.869	286.811	1684.68 (7.056%)
DS	ANN	1589.685	339.395	1929.08 (8.078%)
	BR	1589.685	339.395	1929.08 (8.078%)
HM	abcd	42.02	39.815	81.835 (0.343%)
	GR2M	42.02	39.815	81.835 (0.343%)
Internal variability				769.37 (3.222%)
Total				23878.6 (100%)

Table 3.4 Uncertainties of four stages for winter season

stage	model	Uncertainty from main effect	Uncertainty from interaction effect	Scenario Uncertainty
ES	A2	1.295	22.63	23.927 (0.845%)
	B1	1.295	22.63	23.927 (0.845%)
GCM	CSIRO-Mk30	348.865	206.821	555.685 (19.625%)
	MIUB-ECHO-G	93.0486	55.822	148.871 (5.257%)
	MPI-M-ECHAM5	3.476	28.2101	31.686 (1.119%)
	MRI-CGCM232	51.37	76.0512	127.422 (4.5%)
DS	ANN	746.74	167.57	914.305 (32.291%)
	BR	746.74	167.57	914.305 (32.291%)
HM	abcd	1.58	4.975	6.554 (0.231%)
	GR2M	1.58	4.975	6.555 (0.231%)
Internal variability				78.2 (2.762%)
Total				2831.44 (100%)

Chapter 4

Conclusions

This study quantifies the uncertainty of the quantity forecast value for each forecast step, and furthermore, introduces the method of quantifying the uncertainty one more time by detailed method of each forecast step. In this way, it is possible to analyze exactly how much each of the forecasting steps takes up the total uncertainty. In addition, we can clearly see which method has the greatest uncertainty within each of the forecasting stages And it can help predict the water outlook and can be helpful in policy direction. In addition, the ranking of uncertainty for each stage of the forecast may be slightly different from that of the detailed method within the forecasting stage, which is also meaningful when research or policy is implemented.

And this method has advantages not only to predict water resources outlook but also to be universally used in all other fields. It is more useful if it is used in fields related to fields where damage can be large according to uncertainty.

The method also decomposes the view step once more according to the detailed method. If there are detailed steps within the detail step, I think that the method of disassembling several times can also be devised.

Two categories are considered in this chapter. The first step introduces the data and background to be applied in detail. In the second phase, we will describe a model that quantifies and assigns the uncertainty of water forecasting mdevelop a different scenario depending on how much greenhouse gas emissions are reduced or increased by the emission scenario. The second step is the global circulation model. In this model, simulations are made using parameters that mainly change the future climate, such as temperature, precipitation,

Bibliography

Xiu D. & Karniadakis G. 2003. Modeling uncertainty in flow simulations via generalized polynomial chaos. J. 187(1), pp. 158-166.

Ralph Smith. 2013. Uncertainty quantication: theory, implementation, and applications, volume 12. SIAM.

Yang, M. Choi, G. Lin, and G. E. Karniadakis, 2012 Adaptive ANOVA decomposition of stochastic incompressible and compressible flows, Phys., vol. 231, no. 4, pp. 1592 - 1607

Ohn, I., Kim, Y. and Kim, Y.-O. 2018. Uncertainty decompositon in hydrological projection considering interaction effects. Journal of Korea Water Resources Association, submitted.

Eurachem. 2000. Quantifying Uncertainty in Analytical Measurement, Eurachem, pp. 23-48.

Yeo H, Choi M, and Sunwoo Y. 2004. Seasonal variations in atmospheric concentrations of organochlorine pesticides in urban and rural areas of Korea.

국문초록

기후 변화를 고려한 여러 자원의 전망은 학문적으로도 의의가 있지만 정책적으로도 중요한 사안이다. 그 중 수자원의 중요성은 단연 강조해도 지나치지 않은데 인간에게 물처럼 중요한 자원은 그렇게 많지 않기 때문이다. 최근의 가뭄이나 홍수가 빈번해지고 심해짐에 따라 기후 변화에 따른 수자원 전망에 대한 예측 필요성이 점점 대두되고 있다. 기후 변화에 따른 수자원 전망은 크게 배출 시나리오, 전지구적 순환 모형, 상세화 기법, 수문 모형이 있다. 각 전망 단계들을 통한 예측뿐만 아니라 각 전망의 불확실성을 계량화하는 것 또한 중요하다. 각 전망 단계의 불확실성을 계량화하는 방법은 제시되었지만 본 논문에서는 각 전망 단계의 합이 총 불확실성과 같게 계량화된 방법을 제시하고자 한다. 또한 각 전망 단계의 불확실성을 전망 단계의 여러 방법에 따라 한 번 더 분해해 전망 단계 내에서의 비교가 가능할 수 있는 방법을 제시한다.

주요어 : 불확실성 분석 , 기후 변화에 따른 수자원 전망, 각 전망 단계의 분해

학 번 : 2017-25184