

# 입자군집최적화 알고리즘을 이용한 효율적인 TOPMODEL의 불확실도 분석

## Efficient Uncertainty Analysis of TOPMODEL Using Particle Swarm Optimization

조 희 대\* / 김 동 균\*\* / 이 강 희\*\*\*

Cho, Huidae / Kim, Dongkyun / Lee, Kanghee

---

### Abstract

We applied the ISPSO-GLUE method, which integrates the Isolated-Speciation-based Particle Swarm Optimization (ISPSO) with the Generalized Likelihood Uncertainty Estimation (GLUE) method, to the uncertainty analysis of the Topography Model (TOPMODEL) and compared its performance with that of the GLUE method. When we performed the same number of model runs for the both methods, we were able to identify the point where the performance of ISPSO-GLUE exceeded that of GLUE, after which ISPSO-GLUE kept improving its performance steadily while GLUE did not. When we compared the 95% uncertainty bounds of the two methods, their general shapes and trends were very similar, but those of ISPSO-GLUE enclosed about 5.4 times more observed values than those of GLUE did. What it means is that ISPSO-GLUE requires much less number of parameter samples to generate better performing uncertainty bounds. When compared to ISPSO-GLUE, GLUE overestimated uncertainty in the recession limb following the maximum peak streamflow. For this recession period, GLUE requires to find more behavioral models to reduce the uncertainty. ISPSO-GLUE can be a promising alternative to GLUE because the uncertainty bounds of the method were quantitatively superior to those of GLUE and, especially, computationally expensive hydrologic models are expected to greatly take advantage of the feature.

**Keywords** : hydrology, optimization, uncertainty analysis, TOPMODEL, particle swarm optimization, GLUE

---

### 요 지

멀티모달 최적화알고리즘의 일종인 ISPSO와 불확실도분석기법인 GLUE를 결합한 ISPSO-GLUE 기법을 TOPMODEL의 불확실도분석에 적용하였으며, 그 결과를 GLUE 기법과 비교하였다. 두 기법 모두 같은 횟수만큼 모형을 실행하였을 때 ISPSO-GLUE 기법의 누적성능이 더 좋아지는 시점을 발견할 수 있었으며, 그 이후로도 ISPSO-GLUE 기법은 GLUE 기법과는 달리 점진적인 성능의 향상을 보여 주었다. 두 기법이 비슷한 모양과 양상의 95% 불확실도구간을 생성하였다. 하지만 ISPSO-GLUE 기법이 약 5.4배 더 많은 관측치를 포함하는 것으로 나타났으며 GLUE 기법에 비해 훨씬 적은 횟수의 모형실행으로도 좋은 성능의 불확실도구간을 얻을 수 있는 것으로 나타났다. ISPSO-GLUE 기법과 비교했을 때 GLUE 기법이 최대 침투유량의 감쇠곡선 부분에서 불확실도를 과대평가하였다. 이 시간대에 대해서는 GLUE의 경우 불확실도

---

\* 美 듀베리社 수자원부 연구원 Atlanta, GA (e-mail: hcho.eng@gmail.com)  
Staff Water Resources Engineer, Dewberry, 2835 Brandywine Road Suite 100, Atlanta, GA

\*\* 교신저자, 홍익대학교 공과대학 토목공학과 조교수 (e-mail: deokaykim@gmail.com)  
*Corresponding Author*, Assistant Professor, Department of Civil Engineering, Hongik University

\*\*\* 홍익대학교 공과대학 토목공학과 학부과정 (e-mail: ozxco9050@gmail.com)  
Undergraduate Student, Department of Civil Engineering, Hongik University

를 줄이기 위해 더 많은 행동모형들을 찾을 필요가 있다. ISPSO-GLUE 기법이 정량적인 성능평가에서 훨씬 많은 관측치를 포함할 수 있었다는 것은 이 기법의 가능성을 잘 보여 주었다고 할 수 있으며, 특히 계산적으로 값비싼 수문모형에서는 보다 큰 성능의 차이를 보일 것으로 기대된다.

**핵심용어 :** 수문학, 최적화, 불확실도 분석, TOPMODEL, 입자군집최적화, GLUE

## 1. 서 론

Draper and Box (1987)는 “모든 모형은 틀렸다”고 말했다. 이는 모형을 통해서 시스템을 모의하고자 할 때 모형이 얼마나 틀릴 수 있는지를 잘 고찰해야 한다는 뜻이다. 결정론적인 모델링 기법은 모의된 결과에 불확실도를 전혀 부여하지 않기 때문에 모형이 얼마나 틀릴 수 있는지 파악하기가 어렵다. 불확실도가 생길 수밖에 없는 이유는 우리가 자연이라는 시스템을 완벽하게 이해하지 못하기 때문이다. 모형은 우리가 모의하고자 하는 시스템을 극도로 단순화한 수식에 불과하며 그 수식에 입력으로 들어가는 매개변수 또한 직접적인 측정이 불가능해서 보정을 통해서만 얻을 수 있는 경우도 있다. 뿐만 아니라 관측치에도 측정의 오차가 있을 수밖에 없다. 이처럼 시스템의 단순화로 인한 오류, 매개변수의 추측으로 인한 오류, 측정에 따른 오차 등 다양한 방식으로 불확실도가 모의결과로 전파된다. 이런 다양한 오류들은 직접적인 관측이 어렵기 때문에 확률론적으로 오류의 발생양상을 가정하고 베イズ 추론을 통해서 모의결과에 얼마만큼의 불확실도가 존재하는지 추측하게 된다.

수문학에서도 베イズ 추론을 이용한 불확실도 기법이 꾸준히 개발되어 왔다. Beven and Binley (1992), Vrugt et al. (2009), 그리고 여러 학자들은 베イズ 추론을 통해 모델링 과정에서 필연적으로 발생할 수밖에 없는 불확실도를 예측하는 기법을 연구해 왔다. 특히, Beven and Binley (1992)의 GLUE (Generalized Likelihood Uncertainty Estimation) 기법은 비교적 구현이 간단하기 때문에 다양한 연구에서 널리 쓰여왔다(Muleta and Nicklow, 2005; Zheng and Keller, 2007; Cho and Olivera, 2014). GLUE 기법은 몬테카를로 기법(Monte Carlo method)을 통해 무작위로 많은 개수의 매개변수표본을 생성해서 모형을 수행한 후 미리 정해 둔 우도(likelihood)의 한계치 보다 더 좋은 모형들(behavioral models 또는 행동모형)만 모아서 예측불확실도를 결정하게 된다. 이 때 우도는 반드시 통계적인 의미를 가지는 확률분포함수일 필요가 없으며 모형을 수행하는 자의 주관적인 판단에 의해 임의적인 목적함수를

사용할 수 있다. 이런 우도의 임의적인 주관성은 종종 비판의 대상이 되어왔다(Mantovan and Todini, 2006). 그러나 Beven et al. (2008)은 모형의 오류가 반드시 수학적인 확률론적 분포를 따르는 것은 아니며, 이를 확률적으로 잘못 가정할 경우 모의치에 편차가 생길 수 있음을 보였다. 하지만, 이 기법은 예측불확실도를 결정하기 위해서 많은 수의 표본을 필요로 한다는 단점이 있다. 효율적인 불확실도 분석을 위해 Blasone et al. (2008)은 마르코프체인 몬테카를로 표본추출법(Markov Chain Monte Carlo sampling)을 GLUE 기법에 적용하였다. 이들은 전역 최적화 알고리즘을 이용해서 기존의 GLUE 기법보다 나은 중앙예측값을 얻었으며 GLUE 기법의 효율을 향상시켰다. 이 기법은 전역 최적화 알고리즘을 사용하였기 때문에 멀티모달(Multi-Modal)한 문제에서는 지역해를 효율적으로 찾을 수 없는 단점이 있다. 이를 보완하기 위해 Cho and Olivera (2014)는 멀티모달한 목적함수에 적합한 입자군집최적화 알고리즘의 일종인 고립종기반 입자군집최적화(Isolated-Speciation-based Particle Swarm Optimization, 이후부터 ISPSO; Cho et al., 2011) 알고리즘과 GLUE기법을 결합한 ISPSO-GLUE 기법을 소개하였다. 이 새로운 기법은 전역해 뿐만 아니라 지역해까지 효율적으로 찾을 수 있으며 따라서 지역해 부근에 분포되어 있는 행동모형들도 불확실도 분석에 포함할 수 있게 되었다. 이들은 이 기법을 통해 계산적으로 값비싼 분산형모형인 SWAT (Soil and Water Assessment Tool; Arnold et al., 1998)의 불확실도 분석을 효율적으로 수행하였다.

본 연구에서는 ISPSO-GLUE 기법을 TOPMODEL (Topography Model; Beven and Kirkby, 1979)의 불확실도 분석에 적용하여 그 효율성을 검토하였다. 본 연구에서 사용한 TOPMODEL 버전은 오픈소스 GRASS GIS (GRASS Development Team, 2012)의 모듈로 개발된 프로그램으로서 지형자료를 처리하고 분석하는 데에 효율성을 크게 높여준다. ISPSO-GLUE 기법을 통한 불확실도 분석은 TOPMODEL 모형에만 국한되는 것이 아니며, 특히 계산시간이 긴 다른 분산형 모형에 적용했을 경우 불확실도 구간이 빠르게 수렴함과 동시에 더 많은 관측치

를 포함함으로써 모형의 실행횟수를 크게 줄일 수 있을 것으로 기대된다. 이런 ISPSO-GLUE의 특징은 단순히 무작위로 매개변수표본을 추출하는 기존의 GLUE기법과는 차별되는 방법으로서 계산적으로 값비싼 모형의 불확실도 분석에 기여할 수 있을 것이다.

## 2. 이론소개

### 2.1 GLUE

GLUE 기법은 몬테카를로 기법을 이용해 매개변수공간에서 무작위로 매개변수표본을 추출한다. 이렇게 추출한 매개변수표본들로 모형을 반복적으로 실행한 다음 미리 정의해 놓은 우도를 이용해 각 모형의 성능을 평가해서 행동모형(behavioral model)과 비행동모형(non-behavioral model)으로 분류한다. 여기서 우도는 반드시 통계적인 의미의 우도일 필요는 없으며 모형의 성능을 측정할 수 있는 0부터 1까지의 값을 가지는 목적함수로 정의할 수 있다. 우도는 각 매개변수표본에 할당되며 각 모의시간대의 모의치의 크기에 따라 정렬되어 누적된다. 이 누적우도곡선에서 불확실도구간을 정의하게 된다. Algorithm 1은 GLUE의 의사코드를 보여준다.

### 2.2 ISPSO-GLUE

ISPSO-GLUE 기법은 전역해 뿐만 아니라 지역해까지

찾을 수 있는 멀티모달 최적화 알고리즘의 일종인 ISPSO를 이용해서 표본을 추출한다. ISPSO는 추가적으로 결정론적인 표본추출법을 이용해서 기존의 입자군집최적화(Particle Swarm Optimization)가 선호하지 않았던 검색공간까지 표본을 추출해서 불확실도 분석에 기여하는 표본을 다양화하는데 도움을 준다. 이로 인해 지역해에 근접한 행동모형을 보다 쉽게 찾을 수 있으며 비선형 검색공간에서의 행동모형 발견도 가능하게 된다. Cho and Olivera (2014)는 이 기법을 이용해서 매개변수의 최적화와 결정론적 표본추출을 동시에 수행함으로써 SWAT의 불확실도 분석에 필요한 모형의 실행횟수를 크게 줄일 수 있음을 보여주었다. ISPSO-GLUE 기법은 이외에도 Kim et al. (2013)의 추계학적 강우생성모형의 최적화 및 Cho et al. (2013)의 타원을 이용한 레이다 상의 폭풍우추적 연구에도 성공적으로 적용된 바 있다.

Algorithm 1에서 5번과 6번 줄까지의 무작위 표본추출을 ISPSO의 멀티모달 최적화를 통한 표본추출로 대체함으로써 매개변수공간에서 더 좋은 모형들을 보다 효율적으로 찾을 수 있게 된다. 하지만 최적해에 편중된 모형들 때문에 불확실도를 과소평가할 수 있다. 이를 방지하기 위해서 9번 줄의 정상화방법을 표본의 밀도를 고려한 방법으로 대체한다(Cho and Olivera, 2014). 이처럼 ISPSO-GLUE 기법은 GLUE 기법에 비해 보다 효율적이고 적극적으로 최적의 매개변수공간을 검색함으로써 같은 수의

1. Define the likelihood function  $L(\theta|X)$ , where  $\theta$  is a parameter set and  $X$  is observed data.
2. Select a likelihood threshold value,  $L_B$ , that separates models into behavioral and non-behavioral.
3. Determine the number of samples,  $N$ .
4. Let  $B=\{\}$ .
5. For  $i=1$  to  $N$ ,
6.     Sample a set of parameters,  $\theta_i$ , randomly from the parameter space.
7.     Evaluate  $L(\theta_i|X)$ .
8.     If  $L(\theta_i|X)$  is greater than  $L_B$ , let  $B=\{\theta_i, B\}$  and store the simulated output,  $S_{i,t}$ , where  $t=1$  to  $T$  and  $T$  is the number of simulated time steps.
9. For all  $\theta_i \in B$ , calculate  $P(X|\theta_i) = \frac{L(\theta_i|X)P(X)}{C}$ ,  
     where  $P(X|\theta_i)$  is the posterior likelihood and  $C$  is a normalizing constant,  $\sum_{\theta_i \in B} L(\theta_i|X)P(X)$ .
10. For  $t=1$  to  $T$ ,
11.     Let  $S=\{\}$ .
12.     For each  $\theta_i \in B$ , let  $S=\{S_{i,t}, S\}$ .
13.     Sort  $S_{i,t}$  in  $S$  and accumulate  $P(X|\theta_i) = \frac{L(\theta_i|X)P(X)}{C}$  in the same sorted order as  $S_{i,t}$ ,  
     where  $P(X|\theta_i)$  is the posterior likelihood and  $C$  is a normalizing constant  
     such that  $\sum_{\theta_i \in B} L(\theta_i|X)P(X) = 1$ .
14.     Find the lower and upper percentiles of the cumulative posterior likelihood curve to determine the uncertainty range at time step  $t$ .

Algorithm 1. Pseudo Code for GLUE

모형실행으로도 더 나은 불확실도구간을 구할 수 있거나 더 적은 모형실행으로 GLUE 기법과 비슷한 불확실도구간을 얻을 수 있는 특징이 있다.

### 2.3 ISPSO-GLUE와 GLUE의 비교

ISPSO-GLUE와 GLUE 기법간의 근본적인 차이점은 표본을 추출하는 방법에서 시작된다. GLUE 기법의 경우 각 매개변수에 대해 선행분포함수를 가정했을 때 이 분포함수를 따르는 균일한 밀도의 표본들이 매개변수끼리의 상호관계를 고려하지 않은 채 다차원 매개변수공간에서 추출된다. 이렇게 추출된 표본들은 모형을 평가하고 행동모형들을 찾는데 쓰인다. 미리 정해진 우도값보다 큰 값을 가지는 행동모형들만을 모아서 신뢰구간을 만들게 된다. 이 기법의 단점으로는 고차원 검색공간에서 단순히 많은 수의 표본을 추출하는 것만으로는 최적해에 근접한 행동모형들을 쉽게 찾을 수가 없다는 것이다. 예를 들어 매개변수가 5개인 모형을 각 매개변수의 값을 열개의 구간으로 나누어 표본을 추출할 경우 표본이  $10^5=100,000$ 개나 필요하게 된다. 이런 단점 때문에 많은 표본추출 기법들이 소개되었다(예를 들어, McKay et al., 1979; Evans, 1991). 하지만 이런 다양한 표본추출 기법들은 최적의 해를 찾는 데는 적합하지 않기 때문에 최적해 근처의 행동모형들을 효율적으로 찾을 수는 없다.

반면, ISPSO-GLUE에서도 선행분포를 가정하나 표본추출은 멀티모달 최적화와 결정론적 방법으로 이루어진다. 먼저 멀티모달 최적화에 의한 표본추출은 ISPSO에서 군집을 이루는 각 입자들의 움직임에 의해 결정되며 이 입자들은 최적의 전역해 및 지역해를 찾아가는 경향을 보이므로 이 과정에서 좋은 행동모형들을 찾을 확률이 높아지게 된다. 하지만 특정 영역에 집중된 행동모형들은 그 영역의 중요도를 과장할 수 있으므로 입자의 밀도를 이용해서 정상화하는 과정을 거치게 된다. 다음으로 결정론적 표본추출을 통해서 검색공간 전반에 걸쳐서 입자들이 선호하지 않는 공간을 검색하게 된다. 이 두 가지 표본추출 방법을 통해 최적해에 근접한 행동모형들 뿐만 아니라 고르게 분포되어 있는 행동모형들을 찾을 수 있게 된다.

### 2.4 TOPMODEL

TOPMODEL에서는 하천으로 유출되는 유량을 크게 초과강우로 인한 유량, 하천으로 복귀하는 환원수, 그리고 지표하 유량으로 나누었다. 이 세 가지 기본적인 하천유량 생성기구를 바탕으로 다양한 TOPMODEL들이 개발되었다. Cho (2000)는 Beven et al. (1995)이 공개한 TMOD9502를

GRASS GIS와 연계하기 위하여 FORTRAN 버전을 C언어로 이식하면서 r.topmodel이라는 GRASS GIS의 모듈을 개발하였다. R.topmodel은 이후 SAGA GIS (Olaya, 2004)와 R 언어(R Development Core Team, 2006)의 TOPMODEL 모듈을 개발하는데 영향을 미쳤다(Hornik, 2008; Conrad, 2003). R.topmodel은 문자파일 기반의 공간자료를 사용하는 TMOD9502와는 달리 GIS 환경 내에서 작동하는 방식으로 공간정보를 효율적으로 처리하고 모형을 수행하는데 있어서 뛰어난 환경을 제공한다. 뿐만 아니라, r.topidx 모듈을 이용해서 지형지수도를 GIS 자료에서 직접 계산할 수 있기 때문에 본 연구에서는 r.topmodel을 이용해서 지형공간정보의 전처리, TOPMODEL의 모의 및 모형 결과의 후처리를 실행하였다.

## 3. 대상 및 방법

### 3.1 대상유역 및 관측자료

본 연구에서는 강우자료 뿐 아니라 증발산자료도 필요하기 때문에 NOAA (National Oceanic and Atmospheric Administration)의 증발산계측기가 많이 분포된 미국 텍사스주의 덴턴시(City of Denton, Texas)를 대상유역으로 선정하였다. Fig. 1은 본 연구에 쓰인 덴턴시에 있는 클리어크릭(Clear Creek) 유역을 보여준다. 이 유역의 면적은  $763 \text{ km}^2$ 이며 USGS (2013a)의 NED (National Elevation Data) 1 아크초(arc second) 해상도의 DEM (Digital Elevation Model) 자료를 사용하여 유역경계를 계산하였다. 이 유역의 유출구에는 USGS (U.S. Geological Survey)의 유량관측소 08051500이 위치하고 있으며 USGS (2013b)에서 일일 유출량자료를 구하였다. 일일 강우자료와 증발산자료는 각각 NOAA-NCDC (2013)와 NOAA-CPC (2013)에서 구할 수 있었다.

TOPMODEL은 하나의 강우 시계열자료만 입력으로 받을 수 있기 때문에 유역에 영향을 미치는 총 네 개의 NCDC 강우계측기들(Table 1)의 자료를 티센망(Thiessen polygons; Thiessen and Alter, 1911)의 기여면적으로 가중치를 줘서 Eq. (1)과 같이 하나의 시계열자료를 계산하였다.

$$P_{final} = \frac{\sum_{i=1}^n P_i \times A_i}{\sum_{i=1}^n A_i} \quad (1)$$

여기서,  $P_{final}$ 은 최종강우량(m/day),  $n$ 은 강우계측기 개수,  $P_i$ 는  $i$  계측기의 강우(m/day) 그리고  $A_i$ 는  $i$  계측기의 기여

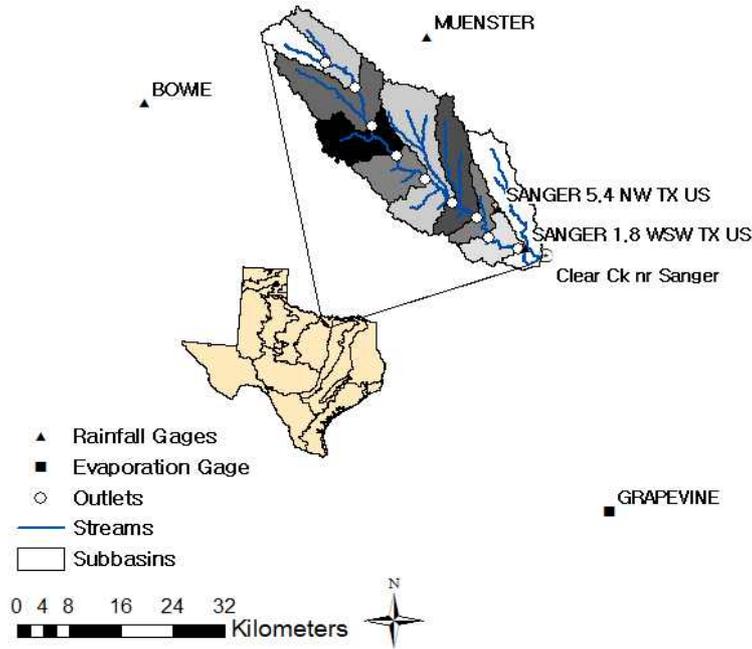


Fig. 1. Clear Creek Watershed in City of Denton, Texas

Table 1. NCDG Precipitation Gages and Contribution Areas

Gage Number	Name	Latitude	Longitude	Contribution area (km <sup>2</sup> )	Contribution rate (%)
GHCND:US1TXDN0015	Sanger 1.8	33.35040	-97.20630	51	6.6
GHCND:US1TXDN0003	Sanger 5.4	33.40700	-97.25220	326	42.7
GHCND:USC00416130	Muenster	33.65361	-97.37528	372	48.8
GHCND:USC0041098	Bowie	33.55110	-97.84720	14	1.9
Total				763	100

면적(km<sup>2</sup>)이다. 증발산계측기의 경우 티센망 분석결과 GPVT2 계측기 하나만 유역에 영향을 미치는 것으로 나타났다. 이 계측기의 시계열자료를 그대로 사용하였다.

모의기간은 자료의 기간이 제일 짧은 증발산자료를 기준으로 2009년 5월 1일부터 2011년 4월 30일까지로 정하였다. 모형의 초기경계조건이 안정화된다고 가정된 최초 1년치의 모의결과는 모형의 성능계산에서 제외하였다. 즉, 모의는 2009년 5월 1일부터 2011년 4월 30일까지 하였으나 모형의 우도계산은 초기 1년을 제외한 2010년 5월 1일부터 2011년 4월 30일까지 1년의 결과를 이용하였다.

### 3.2 모의방법

Table 2는 본 연구에서 최적화한 매개변수와 그 값의 범위를 보여준다. 모형내에 소유역을 하나만 생성해서 소유역의 유출구에서 전체유역의 유출구까지의 거리를 0으로 만들었다. 이렇게 함으로써 주수로 유하속도인 CHV

매개변수를 제거하였으며 최적화하는 검색공간의 차원수를 낮출 수 있었다.

매개변수의 선행분포는 값의 범위 내에서 균등분포로 가정하였으며 우도는 Nash-Sutcliffe 계수(Eq. (2); Nash and Sutcliffe, 1970)를 이용해서 Eq. (3)과 같이 정의하였다.

$$NS = 1 - \frac{\sum_{t=1}^T (S_t - O_t)^2}{\sum_{t=1}^T (\bar{O} - O_t)^2} \quad (2)$$

여기서, NS는 Nash-Sutcliffe 계수, T는 모의날짜수, S<sub>t</sub>는 t 시간대의 모의유량(m<sup>3</sup>/day), O<sub>t</sub>는 t 시간대의 관측유량(m<sup>3</sup>/day), 그리고  $\bar{O}$ 는 관측유량의 평균값(m<sup>3</sup>/day)이다.

$$L = \begin{cases} NS & \text{if } NS \geq 0, \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Table 2. Ranges of the TOPMODEL Parameters Used in this Study

Parameter	Units	Minimum	Maximum	Description
Q0	m/h	0	50	Initial subsurface flow per unit area
T0	ln (m <sup>2</sup> /h)	-7	10	Areal average of ln (T0)
SZM	m	0.001	0.25	Scaling parameter
SR0	m	0	0.01	Initial root zone storage deficit
SRMAX	m	0.005	0.08	Maximum root zone storage deficit
TD	h	0.001	40	Unsaturated zone time delay per unit storage deficit
RV	m/h	50	2000	Internal subcatchment routing velocity
XK0	m/h	0.0001	0.2	Surface hydraulic conductivity
HF	m	0.01	0.5	Wetting front suction
DTH		0.01	0.6	Water content change across the wetting front

여기서, L은 우도이다. 행동모형의 우도입계치는 0으로 정의하여 우도가 0보다 큰 모형들을 행동모형으로 정의하였다. 이 모형들은 “No Model” (Cho and Olivera, 2009)로 알려진 관측치의 평균값보다 더 좋은 성능을 보이는 모형들이다.

ISPSO-GLUE와 GLUE 기법 모두 모형을 10,000번 수행하였으며 모의된 각 시간대마다 같은 개수의 모의유량을 얻었다. 각 모의값에 해당 모형의 정상화된 우도를 할당한 후 모의값의 크기로 정렬하면서 우도를 누적하였다. 이렇게 누적한 0부터 1까지의 누적우도분포에서 하위 및 상위 2.5%의 값을 제거한 모의값의 구간을 95% 불확실도 구간으로 정의하였다. 본 연구에서는 편의상 ISPSO-GLUE와 GLUE 기법으로 구한 95% 불확실도구간을 각각 ISPSO-GLUE 95% 불확실도구간 및 GLUE 95% 불확실도구간으로 정의하였다.

#### 4. 결과 및 분석

Fig. 2는 누적최대 NS값이 모형의 실행횟수에 따라 어떻게 변하는지를 보여준다. 각 기법의 최대 NS값을 비교해 봤을 때, GLUE의 경우 가장 좋은 행동모형의 NS값이 0.47이었으며 ISPSO-GLUE의 경우 0.82로 나타났다. GLUE의 0.47은 아주 좋지도 아주 나쁘지도 않은 성능이라고 할 수 있다. 하지만, ISPSO-GLUE의 0.82는 완벽한 모형의 NS값이 1임을 감안할 때 아주 좋은 성능임을 알 수 있다. 이처럼 약 2배에 가까운 ISPSO-GLUE의 성능은 이후 설명할 불확실도구간의 성능에도 큰 역할을 하게 된다. 모형실행 횟수의 경우, 약 2,300번까지는 무작위로 표본을 추출하는 GLUE 기법이 더 좋은 성능을 보이는 경우

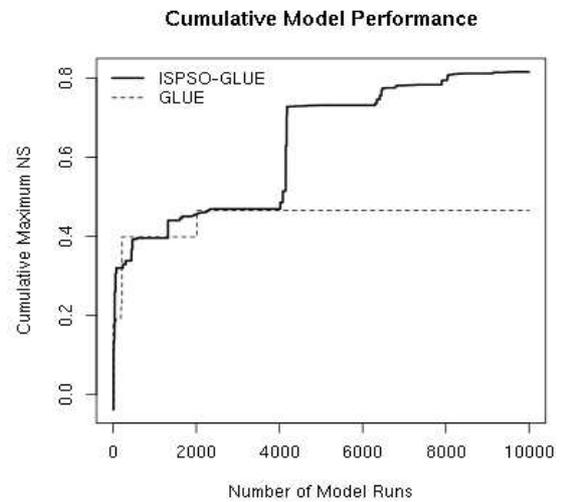


Fig. 2. Cumulative Model Performance. This Figure Shows the Maximum Model Performance Achieved by a Number of Model Runs. ISPSO-GLUE Starts to Perform better than GLUE after 2,310 Model Runs

도 가끔 있었으나 그 이후로는 ISPSO-GLUE가 일관되게 우도가 더 높은 행동모형들을 찾기 시작했다. 누적최대 NS값의 곡선을 비교해 보면 GLUE의 경우 수평으로 진행하다가 수직으로 크게 상승하는 양상만 보인 반면, ISPSO-GLUE의 경우 성능이 급격히 향상되는 NS값의 수평선 사이에 점진적으로 모형이 향상되는 양상을 관측할 수 있다. 이런 ISPSO-GLUE의 특징은 알고리즘내의 입자들이 군집을 이루면서 서서히 좋은 해를 찾아가는 과정에 있음을 뚜렷하게 보여준다. 반면 GLUE의 경우 무작위로 검색공간을 검색하기 때문에 좋은 모형으로 수렴하는 양상을 보이지는 못하며 우연히 좋은 모형을 발견했을 때에만 급격한 성능향상을 보이게 된다. 이처럼 우연한

발견에만 의지하는 GLUE의 경우 ISPSO-GLUE에 비해 좋은 행동모형을 더 많이 찾지는 못하였다.

Fig. 3은 ISPSO-GLUE에서 표본을 추출할수록 최적화 알고리즘내의 입자들이 더 좋은 모형으로 수렴하려는 경향을 잘 보여준다. 반면, 표본을 무작위로 추출하는 GLUE에서는 모형의 성능 또한 향상되는 경향을 보이는 대신 무작위에 가까운 양상을 보인다. Table 3은 각 기법으로 구한 NS값들의 통계를 보여준다. GLUE에 비해 ISPSO-GLUE가 더 큰 최소 및 최대 NS값을 찾았으며 전반적인 분포는 넓은 것으로 나타났다. 이는 ISPSO-GLUE가 성능이 좋은 행동모형들을 보다 더 광역적으로 검색하고 있음을 보여준다. 또한, ISPSO-GLUE가 같은 구간의 NS값에서 GLUE보다 월등히 많은 행동모형들을 찾았다는 것을 알 수 있다. 극단적인 예로 NS값이 0.6보다 큰 구간에서는 GLUE가 단 하나의 행동모형도 찾지 못한 반면 ISPSO-GLUE는 360개의 행동모형들을 찾았다.

앞서 보인 바와 같이 ISPSO-GLUE가 GLUE에 비해 우도가 더 높은 행동모형들을 많이 찾을 수 있었다. 하지만, 불확실도분석에 있어서 표본들의 분포밀도가 좋은 행동모형에만 편중되어 불확실도구간이 좁아지는 경향이 있을 수 있으므로(Cho and Olivera, 2014) ISPSO-GLUE

에서는 검색공간을 여러 구간으로 나누어 각 구간내의 표본들의 개수로 우도를 정상화하는 과정을 거치게 된다. 뿐만 아니라 결정론적인 표본추출이 행동모형에 표본들이 집중되는 경향을 완화시켜준다. Fig. 4는 두 기법의 95% 불확실도구간을 각각 보여준다. 두 불확실도구간의 전체적인 모양과 양상은 비슷하게 나타났으나 GLUE는 작은 유량관측치를 잘 모의하지 못하고 대부분의 시간대에서 불확실도구간이 0으로 수렴하는 현상을 보였다. 모의기간의 약 76% 정도에 걸쳐 GLUE의 불확실도구간이 0으로 나타났으며, 최소관측유량이 48.9m<sup>3</sup>/d임을 감안할 때 GLUE는 이 76% 구간 전체에 걸쳐 어떤 관측치도 포함하지 못했다. 반면, ISPSO-GLUE의 경우 모든 시간대에 걸쳐 불확실도구간의 폭이 0보다 컸다. 전체적으로는 ISPSO-GLUE가 97%에 가까운 관측치를 포함하는 반면 GLUE는 18% 정도의 관측치만 포함하는데 그쳤다. GLUE가 기저유량을 잘 모의하지 못하는 이유를 이해하기 위해서는 본 연구에 쓰인 유량관측치의 특징을 살펴볼 필요가 있다. 유량관측치의 평균값과 표준편차가 각각 140,415m<sup>3</sup>/d 및 719,745m<sup>3</sup>/d로 표준편차가 평균의 500%를 넘을 만큼 관측치의 변동이 아주 크다. 이렇게 큰 유량의 시간적 변동성을 TOPMODEL이 모든 시간대에 대해

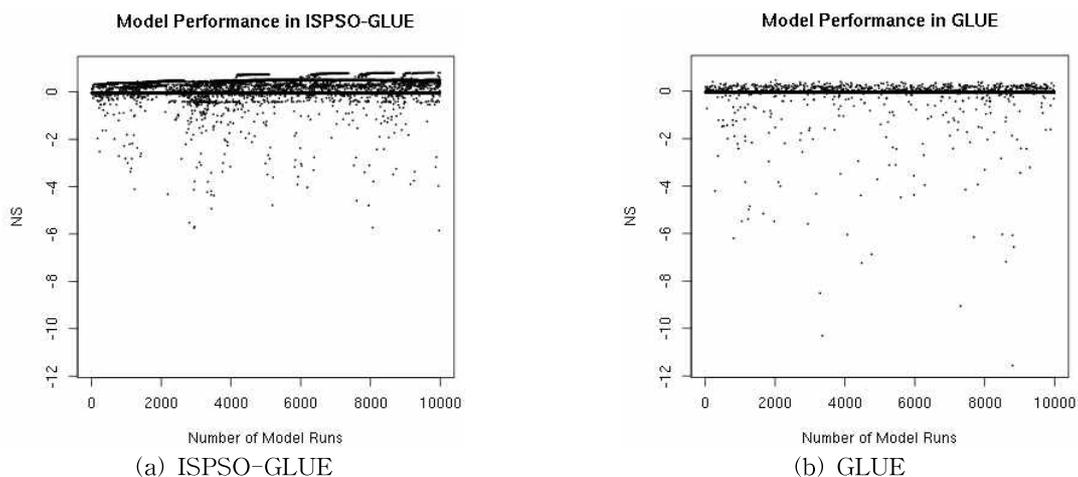


Fig. 3. Model Performance. (a) As More Samples are Taken by ISPSO, It Can Clearly be Seen that Particles are Climbing up the Performance Hills. (b) Taking More Samples Randomly does not Help Increase the Model Performance in General unless a New Sample Happens to be Located at a High Performing Model

Table 3. NS Statistics for the Two Methods

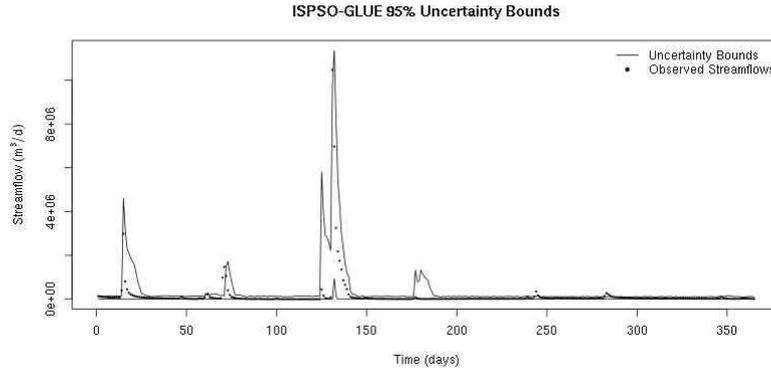
Method	Min NS	Max NS	SD NS	# NS>0	# NS>0.2	# NS>0.4	# NS>0.6	# NS>0.8
ISPSO-GLUE	-5.85	0.82	0.44	3899	3426	2441	360	84
GLUE	-11.56	0.47	0.39	584	174	2	0	0

Min: minimum, Max: maximum, SD: standard deviation, # NS>x: number of NS values greater than x.

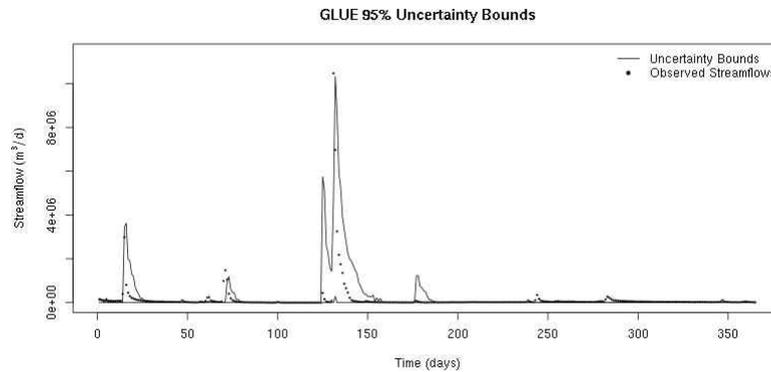
만족스럽게 모의하지는 못하였다. 특히 Fig. 5에서 보인 바와 같이 GLUE의 경우 많은 행동모형들이 기저유량을 0으로 모의함으로써 불확실도구간의 구축에 쓰인 누적우도곡선이 유량값 0인 시점에서 1가까이 수직으로 상승함을 관찰할 수 있다. 이로 인해 누적우도 2.5% (0.025)와 97.5% (0.975)에 해당하는 모의치가 0이 되어 이 시간대의

불확실도구간의 값과 폭이 모두 0이 된다. GLUE의 경우 이런 시간대가 대부분이기 때문에 기저유량을 제대로 모의할 수가 없었다.

ISPSO-GLUE의 경우에도 모의치 0에 해당하는 많은 행동모형들의 영향으로 같은 지점에서 누적우도곡선이 수직상승하는 것을 볼 수 있으나 모의치가 0이 아닌 행동

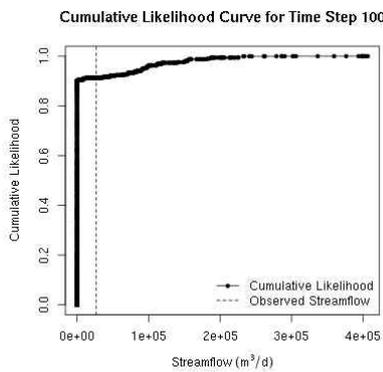


(a) ISPSO-GLUE

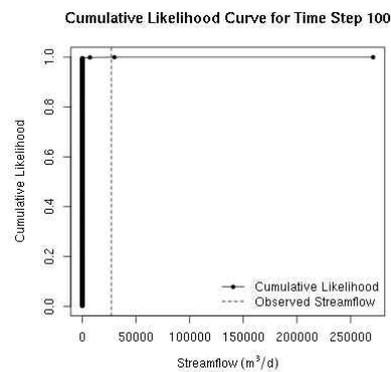


(b) GLUE

Fig. 4. 95% Uncertainty Bounds for the Two Methods. Solid Lines Represent the upper and lower Bounds of the 95% Uncertainty and Dots Represent the Observed Streamflow. Because of a Peak Value at Day 131, it is Difficult to Observe How Well the Uncertainty Bounds Bracket the Observed Data. Out of the 365 Observed Values, ISPSO-GLUE Enclosed 353 (97%) while GLUE Enclosed 67 (18%)



(a) ISPSO-GLUE



(b) GLUE

Fig. 5. Cumulative Likelihood Curves for Time Step 100. In GLUE, the Cumulative Likelihood Jumps to Close to 1 at a Simulated Streamflow 0

모형들 또한 상당수 있기 때문에 이 곡선이 누적값 1에 미치지 못해 오른쪽으로 움직이고 있다. 이로 인해 누적우도 2.5%의 모의치는 0이지만 97.5%의 모의치는 0보다 큰 값이 되어 관측유량을 포함할 수 있게 된다. 이와 같이 관측치를 잘 모의하기 위해서는 단순히 무작위로 표본을 추출하는 것보다 최적화를 통해서 좋은 행동모형들을 적극적으로 찾을 필요가 있다는 것을 알 수 있다. 달리 말하면 GLUE의 경우 ISPSO-GLUE보다 훨씬 더 많은 표본을 추출할 경우에만 ISPSO-GLUE가 찾은 행동모형에 가까운 모형들을 발견할 수 있으며 기저유량의 모의치에서 누적우도곡선이 수직이 되어서 불확실도구간의 폭이 0이 되는 것을 방지할 수 있을 것이다.

ISPSO-GLUE의 경우 131일에 위치한 최대 첨두유량의 감쇠부분에서 불확실도구간이 GLUE에 비해 좁아지는 경향을 보였으나 관측치는 모두 잘 모의하였다. 이 결과는 NS값을 크게하는 행동모형들이 감쇠곡선 부분을 충실히 잘 모의하였기 때문에 ISPSO-GLUE 불확실도구간이 감쇠곡선 부분에 더 밀착되어 모의하는 것으로 해석된다. 반면, 이 보다 더 낮은 NS값을 가진 GLUE 행동모형들의 경우 이 감쇠곡선을 과대평가함으로써 GLUE 불확실도구간 또한 ISPSO-GLUE에 비해 더 넓게 나타나 불확실도를 과대평가하는 것으로 해석된다. 이런 GLUE 기법의 불확실도 과대평가는 더 많은 표본추출을 통해서 좋은 행동모형들을 더 많이 찾을 필요가 있다는 것을 의미한다.

ISPSO-GLUE 95% 불확실도구간이 97%의 관측치를 포함함으로써 GLUE의 18%보다 정량적으로 약 5.4배 정도 좋은 성능을 보여주었다. 같은 수의 표본추출을 통해서 정도의 성능향상 효과를 가져왔으나 이것이 반드시 GLUE 기법이 5.4배 더 많은 표본추출이 필요할 것이라는 의미는 아니다. 하지만, GLUE 기법으로 ISPSO-GLUE 기법과 비슷한 성능을 얻기 위해서는 더 많은 표본추출이 필요한 것은 본 연구에서 잘 보여 주었다. ISPSO-GLUE 기법이 GLUE 기법에 비해 기저유량을 잘 모의하였으며, 최대첨두유량의 감쇠곡선 부분에서 불확실도구간이 좁아졌으나 관측치를 모두 포함했다. 정량적으로 비교해 보았을 때 ISPSO-GLUE 기법이 95% 불확실도구간에서 더 신뢰할 수 있는 결과를 보였다. 이런 점에서 볼 때 GLUE 기법에만 의존할 경우 관측치를 제대로 모의하지 못하는 구간에 대해 모형의 한계라고 해석해서 모형을 평가절하할 수도 있을 것이다. 하지만, 많은 수의 매개변수를 고려했을 때 고차원의 문제를 단순히 무작위 표본추출에만 의존해서는 모형의 성능을 제대로 파악하는데 한계가 있음

을 본 연구를 통해 확인하였다. ISPSO-GLUE 기법을 통한 이런 성능의 향상은 특히 계산적으로 값비싼 분산형 수문모형에서 그 차이가 훨씬 두드러질 것으로 기대된다. 향후 GLUE 기법으로 ISPSO-GLUE 불확실도구간과 비슷한 불확실도구간을 얻기 위해서는 얼마나 더 많은 모형 실행이 필요한지를 연구할 필요가 있을 것이다.

## 5. 결 론

본 연구에서는 ISPSO-GLUE 기법과 GLUE 기법을 GRASS GIS의 TOPMODEL 모듈인 r.topmodel에 적용하여 그 결과를 비교분석하였다. ISPSO-GLUE 기법에서는 ISPSO 최적화 알고리즘을 이용해서 표본을 추출하였고, GLUE 기법에서는 무작위로 표본을 추출하였다. ISPSO-GLUE 기법은 최적의 해를 추적하는 동시에 결정론적인 표본도 추출하게 된다. 반면 GLUE 기법은 무작위로 표본을 추출하기 때문에 좋은 행동모형의 검색을 우연한 발견에 의지하게 된다. 두 기법의 불확실도구간의 모양이나 양상이 비슷하게 나타났음에도 불구하고 이런 표본추출의 차이점으로 인해 ISPSO-GLUE 기법이 누적 최대 NS값에서 더 좋은 성능을 보였으며, 95% 불확실도구간 또한 GLUE 기법에 비해 약 5.4배 많은 관측치를 포함하는 것으로 나타났다. 이는 ISPSO-GLUE 기법을 통해서 훨씬 적은 수의 모형실행으로도 더 좋은 성능의 불확실도구간을 구할 수 있음을 보여준다. 이처럼 매개변수의 수가 증가할수록 GLUE 기법만으로는 모형의 성능을 제대로 파악하기에는 무리가 있기 때문에 최적화를 통해서 모형의 성능을 최대화할 필요가 있을 것이다. 이런 특징은 계산적으로 값비싼 모형에서 보다 큰 차이를 보일 것으로 기대된다. 두 기법이 비슷한 성능의 불확실도구간을 구하는 시점에서 두 불확실도구간이 얼마나 비슷할 것인지는 향후 더 연구가 필요한 과제이다.

## 감사의 글

본 논문은 2013년 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(한국연구재단-2013R1A1A1011676).

## References

Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R. (1998). "Large Area Hydrologic Modelling and

- Assessment, Part I: Model Development.” *Journal of the American Water Resources Association*, Vol. 34, No. 1, pp. 73–89.
- Beven, K.J., and Kirkby, M.J. (1979). “A Physically Based, Variable Contributing Area Model of Basin Hydrology.” *Hydrological Sciences Bulletin*, Vol. 24, pp. 43–69.
- Beven, K., and Binley, A. (1992). “The Future of Distributed Models: Model Calibration and Uncertainty Prediction.” *Hydrological Processes*, Vol. 6, pp. 279–298.
- Beven, K.J., Quinn, P., Romanowicz, R., Freer, J., Fisher, J., and Lamb, R. (1995). TOPMODEL and GRIDATB: A User’s Guide to the Distribution Versions, Lancaster University, p. 31.
- Beven, K., Smith, P., and Freer, J. (2008). “So Just Why Would A Modeller Choose To Be Incoherent?” *Journal of Hydrology*, Vol. 354, pp. 15–32.
- Blasone, R.-S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Robinson, B.A., and Zyvoloski, G.A. (2008). Generalized Likelihood Uncertainty Estimation (GLUE) Using Adaptive Markov Chain Monte Carlo Sampling. *Advances in Water Resources*, Vol. 31, pp. 630–648.
- Cho, H. (2000). Development of a GIS Hydrologic Modeling System by Using the Programming Interface of GRASS GIS, Master’s Thesis. Department of Civil Engineering, Kyungpook National University.
- Cho, H., Lee, D., Lee, K., Lee, J., and Kim, D. (2013). “Development and Application of a Storm Identification Algorithm that Conceptualizes Storms by Elliptical Shape.” *Journal of the Korean Society of Hazard Mitigation*, Vol. 13, No. 5, pp. 325–335.
- Cho, H., and Olivera, F. (2009). “Effect of the Spatial Variability of Land Use, Soil Type, and Precipitation on Streamflows in Small Watersheds.” *Journal of the American Water Resources Association*, Vol. 45, No. 3, pp. 673–686.
- Cho, H., Kim, D., Olivera, F., and Guikema, S.D. (2011). “Enhanced Speciation in Particle Swarm Optimization for Multi-Modal Problems.” *European Journal of Operational Research*, Vol. 213, No. 1, pp. 15–23.
- Cho, H., and Olivera, F. (2014). “Application of Multi-modal Optimization for Uncertainty Estimation of Computationally Expensive Hydrologic Models.” *Journal of Water Resources Planning and Management*, Vol. 140, No. 3, pp. 313–321.
- Conrad, O. (2003). System for Automated Geoscientific Analyses Module Library: sim\_hydrology, topmodel.cpp. [http://sourceforge.net/apps/trac/saga-gis/browser/trunk/saga-gis/src/modules\\_simulation/hydrology/sim\\_hydrology/topmodel.cpp](http://sourceforge.net/apps/trac/saga-gis/browser/trunk/saga-gis/src/modules_simulation/hydrology/sim_hydrology/topmodel.cpp), accessed in October 2013.
- Draper, N.R., and Box, G.E. (1987). *Empirical Model-Building and Response Surfaces*. John Wiley and Sons Inc.
- Evans, M. (1991). Adaptive Importance Sampling and Chaining.” *Contemporary Mathematics*, Vol. 115, pp. 137–143.
- GRASS Development Team. (2012). Geographic Resources Analysis Support System (GRASS GIS) Software. Open Source Geospatial Foundation Project. <http://grass.osgeo.org>.
- Hornik, K. (2008). Changes on CRAN. R News 8(2), 60–68. [http://www.r-project.org/doc/Rnews/Rnews\\_2008-2.pdf](http://www.r-project.org/doc/Rnews/Rnews_2008-2.pdf), accessed in October 2013.
- Kim, D., Olivera, F., Cho, H., and Socolofsky, S. (2013). “Regionalization of the Modified Bartlett–Lewis Rectangular Pulse Stochastic Rainfall Model.” *Terrestrial, Atmospheric and Oceanic Sciences*, Vol. 24, No. 3, pp. 421–436.
- Mantovan, P., and Todini, E. (2006). Hydrological Forecasting Uncertainty Assessment: Incoherence of the GLUE Methodology. *Journal of Hydrology*, Vol. 330, pp. 368–381.
- McKay, M.D., Beckman, R.J., and Conover, W.J. (1979). “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code.” *Technometrics* Vol. 21, No. 2, pp. 239–245.
- Muleta, M.K., and Nicklow, J.W. (2005). “Sensitivity and Uncertainty Analysis Coupled with Automatic Calibration for a Distributed Watershed Model.” *Journal of Hydrology* Vol. 306, pp. 127–145.
- Nash, J.E., and Sutcliffe, J.V. (1970). River Flow Forecasting Through Conceptual Models, Part I—A Discussion of Principles. *Journal of Hydrology*, Vol. 10, pp. 282–290.

- NOAA-CPC. (2013). National Oceanic & Atmospheric Administration-Climate Prediction Center. Joint Agricultural Weather Facility, U.S. Evaporation Data. [http://www.cpc.ncep.noaa.gov/products/GIS/GIS\\_DATA/JAWF](http://www.cpc.ncep.noaa.gov/products/GIS/GIS_DATA/JAWF), accessed in September 2013.
- NOAA-NCDC. (2013). National Oceanic & Atmospheric Administration-National Climatic Data Center. Surface Data, Daily US. <http://www.ncdc.noaa.gov>, accessed in September 2013.
- Olaya, V. (2004). A Gentle Introduction to SAGA GIS. <http://downloads.sourceforge.net/saga-gis/SagaManual.pdf>, Accessed on September 13, 2013.
- R Development Core Team. (2006). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Thiessen, A.H., and Alter, J.C. (1911). Climatological Data for July, 1911: District No. 10, Great Basin. Monthly Weather Review July 1911:1082-1089.
- USGS. (2013a). U.S. Geological Survey. National Elevation Dataset (NED). <http://ned.usgs.gov>, accessed in September 2013
- USGS. (2013b). U.S. Geological Survey. Surface-Water Daily Data for the Nation. <http://waterdata.usgs.gov/nwis>, accessed in September 2013.
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., and Higdon, D. (2009). "Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling." *International Journal of Nonlinear Sciences & Numerical Simulation*, Vol. 10, No. 3, pp. 271-288.
- Zheng, Y., and Keller, A.A. (2007). *Uncertainty Assessment in Watershed-Scale Water Quality Modeling and Management: 1. Framework and Application of Generalized Likelihood Uncertainty Estimation (GLUE) Approach*. Water Resources Research 43, W08407.

논문번호: 13-095	접수: 2013.11.28
수정일자: 2014.01.28/02.07	심사완료: 2014.02.07