

# 확률강우량의 공간분포에 대한 불확실성 해석: CEM과 SGS 기법의 비교

## Uncertainty Analysis of Spatial Distribution of Probability Rainfall: Comparison of CEM and SGS Methods

서 영 민\* / 여 운 기\*\* / 이 승 윤\*\*\* / 지 흥 기\*\*\*\*

Seo, Young Min / Yeo, Woon Ki / Lee, Seung Yoon / Jee, Hong Kee

### Abstract

This study compares the CEM and SGS methods which are geostatistical stochastic simulation methods for assessing the uncertainty by spatial variability in the estimation of the spatial distribution of probability rainfall. In the stochastic simulations using CEM and SGS, two methods show almost similar results for the reproduction of spatial correlation structure, the statistics (standard deviation, coefficient of variation, interquartile range, and range) of realizations as uncertainty measures, and the uncertainty distribution of basin mean rainfall. However, the CEM is superior to SGS in aspect of simulation efficiency.

**Keywords** : probability rainfall, spatial distribution, uncertainty, SGS, CEM

### 요 지

본 논문에서는 확률강우량에 대한 공간분포 추정에 있어서 공간변동성에 따른 불확실성을 평가하기 위하여 지구통계학적 추계모의기법인 CEM과 SGS 기법을 비교하였다. CEM과 SGS를 이용한 추계모의에 있어서 공간상관구조의 재생성, 확률강우량에 대한 불확실성 평가측도로서 실현치에 대한 통계치(표준편차, 변동계수, 사분위수 범위 및 범위의) 공간분포, 유역평균강우량의 불확실성 분포의 경우 두 기법이 대체로 비슷한 결과를 보이는 것으로 분석되었다. 그러나 모의 효율성 측면에서는 CEM이 SGS에 비해 우수한 결과를 나타내는 것으로 분석되었다.

**핵심용어** : 확률강우량, 공간분포, 불확실성, SGS, CEM

### 1. 서 론

우리나라는 강우의 시공간적 변동이 매우 심하다. 특히 산악지역의 경우 산악지형에 의한 국지기상변화가 때

우 심하며, 강우지속시간이 짧고 강우강도가 매우 높기 때문에 돌발홍수 등에 의해 상당한 피해를 입을 수 있다. 따라서 강우의 시공간적 변동성에 대한 정보는 하천유역 관리, 홍수위험방어, 기후변화 및 침식모의와 관련된 연

\* 교신저자, 영남대학교 건설시스템공학과 박사과정 (e-mail: elofy@nate.com)  
*Corresponding Author*, Ph.D. Student, Department of Civil Engineering, Yeungnam Univ., Gyeongsangbuk-do 712-749, Korea  
\*\* 영남대학교 건설시스템공학과 박사과정 (e-mail: adonas@nafree.net)  
Ph.D. Student, Department of Civil Engineering, Yeungnam Univ., Gyeongsangbuk-do 712-749, Korea  
\*\*\* 한국수자원공사, K-water 수자원연구원 선임연구원 (e-mail: leesy@kwater.or.kr)  
Senior Researcher, K-water Institute, 462-1 Jeonmin-dong, Daejeon 305-730, Korea  
\*\*\*\* 영남대학교 건설시스템공학과 교수 (e-mail: hkjee@yu.ac.kr)  
Professor, Department of Civil Engineering, Yeungnam Univ., Gyeongsangbuk-do 712-749, Korea

구에 있어서 중요하며, 특히 최근 활발한 연구가 이루어지고 있는 분포형 수문모델에 있어서 공간분포된 강우자료에 대한 신뢰성 있는 정보의 필요성은 더욱 증가하고 있는 실정이다.

대부분의 강우관측망은 강우패턴에 대한 공간 변동성을 적절하게 파악할 수 있을 정도로 조밀하지 않기 때문에 기존의 강우관측망으로부터의 강우 변동성 추정, 특히 복잡한 지형에서 극치호우사상에 대한 공간 변동성 추정은 한계가 있다. 지상 레이더로부터의 강우 변동성 추정은 조밀하지 않은 레이더 탐지범위와 지형차단(terrain blocking) 효과, 반사도(reflectivity)와 강우량 관계에서의 변동성, 혼합위상(mixed-phase) 및 고체강수(solid precipitation) 추정과 관련하여 어려움이 따른다(Westrick et al., 1999; Young et al., 1999; Dinku et al., 2002). 그리고 대기모델(atmospheric model)로부터 강우의 변동성에 대한 정량적 추정은 모델의 수평해상도, 미세물리적 매개변수화 기법(microphysical parameterization scheme)의 부족 및 강우과정에 대한 불충분한 이해 때문에 제한을 받는다(Colle et al., 2000; Schultz et al., 2002; Stoelinga et al., 2003). 따라서 정량적 강우추정에 대한 이러한 확정론적 기법은 많은 제약을 가진다.

한편, 공간분포 추정에 있어서 크리깅은 최소오차분산(minimum error variance)을 고려한 최적추정치를 제공한다. 그러나 이러한 최적추정치는 평활효과(smoothing effect)로 인하여 평균보다 큰 값은 과소추정되고 평균보다 작은 값은 과대추정되기 때문에 참값에 비해 변동성이 감소하여 현실성이 떨어지는 문제점을 가진다. 즉, 크리깅은 원자료가 가지는 분산을 상실하기 때문에 관심변수에 대한 최적추정치를 제공하기는 하지만 변동성에 대한 정보는 상실하게 된다(Goovaerts, 1997; Webster and Oliver, 2007; Hengl, 2009). 이러한 변동성의 상실은 입력치와 출력치 간의 높은 비선형성과 극치값이 모델반응을 결정하는 홍수 또는 다른 극치수문과정을 모델링할 경우 오차를 발생할 수 있기 때문에 바람직하지 않다(Haberlandt and Gattke, 2004; Haberlandt, 2007). 그리고 크리깅은 관심변수에 대한 단일 점추정치를 제공하며, 이러한 단일 추정치는 관심변수에 대한 불확실성 범위를 설명할 수 없다(Kanevski, 2008).

크리깅과 같은 기존의 공간보간기법들과는 달리 지구통계학적 추계모의(geostatistical stochastic simulation)는 변동성이 큰 현상에 대한 공간분포와 불확실성을 재생성하기 위한 목적으로 널리 적용되고 있다(Casta et al., 2008). 지구통계학적 추계모의는 동일한 발생확률을 가지는 많은 실현치(realization)에 근거하여 자료의 변동성을 설명하기 때문에 관심속성의 공간분포에 대한 더욱 현실

적인 시각화를 제공할 수 있으며, 실현치들의 후처리를 통하여 공간적 불확실성을 요약할 수 있다(Goovaerts, 1997; Hengl, 2009). 지구통계학적 추계모의는 동일한 발생확률의 많은 실현치를 제공하기 때문에 변동성을 보존하며, 이러한 여러 실현치를 이용한 수문모델의 연속실행, 즉 Monte Carlo 기법을 이용하여 입력강우로부터의 불확실성 평가 및 더 작은 편의(bias)를 가지는 모델링 결과를 제공한다. 이러한 Monte Carlo 기법은 수문모델링에서 매개변수의 불확실성을 평가하기 위하여 널리 적용되고 있으나 입력 강우량에 대한 불확실성 해석을 위해서는 드물게 적용되고 있다(Haberlandt and Gattke, 2004).

지구통계학적 추계모의기법은 크게 픽셀기반기법(pixel-based method)과 객체기반기법(object-based method)으로 크게 나눌 수 있다. 픽셀기반기법은 어떤 공간특성을 나타내기 위하여 일반적으로 규칙격자의 형태를 보이는 픽셀 또는 점들의 배열에 대한 정의에 기반한다. 최근 지구통계학적 추계모의에 대한 대부분의 적용, 특히 위험도의 특성화와 관련된 적용은 픽셀기반기법을 채택하고 있다. 픽셀기반기법은 다시 비매개변수적 기법(non-parametric method), 가우시안 기반 기법(Gaussian-based method), 프랙탈기법(fractal method)으로 나누어진다. 비매개변수적 기법의 대표적 기법으로는 SIS(sequential indicator simulation), p-field 기법 등이 있으며, 가우시안 기반 기법으로는 TBM(turning bands method), SGS(sequential Gaussian simulation) 등이 있다. 한편, 객체기반기법으로는 Poisson 또는 Cox 과정과 같은 점과정(point process)을 이용한 방법과 랜덤집합(random set)을 이용한 부울 기법(Boolean method) 등이 있다(Vann et al., 2002).

한편, 강우의 공간변동성에 대한 정량적 추정은 다양한 불확실성 요인에 의해 영향을 받으며, 이러한 불확실성 요인들을 추계모의에 반영하기 위해서는 모의 알고리즘에 있어서 필연적으로 루프구조의 중첩이 요구된다. 불확실성 요인을 모의 알고리즘에 추가할 때 마다 루프의 중첩이 이루어지며, 각 루프구조에서는 다수의 표본발생이 이루어지게 된다. 어떤 유역에 대한 강우의 공간변동성을 분석하기 위해서 일반적으로 격자형태의 자료구조를 채택하게 된다. 이러한 격자구조는 각 격자노드에 대한 연산이 이루어지기 때문에 상당한 계산상의 부담이 요구되며, 이러한 격자구조에 대한 중첩된 루프의 반복계산은 상당한 연산시간과 저장공간을 요구하게 된다. 따라서 강우의 공간변동성에 대한 불확실성 분석에 있어서 다양한 불확실성 요인을 반영하기 위해서는 모의 알고리즘의 정확성뿐만 아니라 특히 알고리즘의 효율성이 중요한 검토 항목이라 할 수 있다.

따라서 본 연구에서는 강우의 공간변동성에 대한 불확

실성 분석에 있어 다양한 불확실성 인자들을 모의 알고리즘에 반영하기 위한 사전연구로서 지구통계학적 추계 모의기법의 비교를 통해 모의 알고리즘의 효율성을 검토하고자 하였다. 이를 위하여 가우시안 랜덤필드모델(Gaussian random field model)에 기반한 지구통계학적 추계모의기법 중 여러 연구분야에서 가장 널리 적용되고 있는 SGS와 2차원 푸리에 변환(two-dimensional Fourier transform)에 기반하여 고속연산이 가능하고 최근 적용이 증가하고 있는 CEM(circulant embedding method)을 비교하였다.

## 2. 가우시안 랜덤필드모델

### 2.1 SGS

SGS는 그 기법이 직관적이고 알고리즘이 단순하기 때문에 공간적으로 분포된 연속변수를 모델링하기 위하여 여러 분야에서 널리 사용되고 있다. SGS 알고리즘의 중요한 가정은 공간변수에 대한 결합정규분포(joint normal distribution)이며, 이것은 모든 평가위치에서의 모든 가변 요소들이 결합정규분포되어 있음을 의미한다. 다중정규성(multi-normality)이라고 불리는 이러한 특성은 모든 국부적 확률분포(local probability distribution)들이 정규분포를 따르게 한다.

SGS는 순차모의원칙(sequential simulation principle)에 기반하며, 이것은 Eq. (1)과 같이  $n$ 개의 관측자료에 의해 조건지워진  $N$ 개의 국부적 확률분포함수의 곱으로서 전체 공간패턴에 대한 결합확률분포함수를 나타낸다.

$$F(x_1, \dots, x_N; z_1, \dots, z_N | (n)) = F(x_N; z_N | (n+N-1)) \cdot F(x_{N-1}; z_{N-1} | (n+N-2)) \dots F(x_1; z_1 | (n)) \quad (1)$$

여기서,  $F(\cdot)$ 는 확률분포함수,  $i=1, \dots, N$ 에 대하여  $x_i$ 는 격자위치,  $z_i$ 는 확률변수  $Z(x_i)$ 에 대한 실현값이다.

알고리즘의 실제 실행에 있어서 SGS는 이전 모의값들과 조건자료(conditioning data)를 이용하여 다음 격자위치에서의 모의값을 순차적으로 발생시키며, 모의계열은 각 실현치에 대하여 발생하는 랜덤경로(random path)에 의해 결정된다. 단일 실현치에 있어서 모의값은 국부적 확률밀도함수로부터 무작위적으로 추출되고 다음 격자위치에서의 모의를 위한 조건자료집합에 추가된다.

SGS는 개념상 초기화 과정(initialization process), 전역확률보행과정(global random walk process), 지역탐색(local search) 및 조건부 추정과정(conditional estimation process)과 같은 세 가지 과정으로 구성되며(Deutsch

and Journel, 1997), SGS 알고리즘을 정리하면 다음과 같다(Kanevski, 2008).

- i) 만약 자료가 정규분포되어 있지 않다면 먼저 원자료를 표준정규분포  $N(0, 1)$ 로 변환한다.
- ii) 랜덤경로를 따라 모의위치를 선택한다.
- iii) 베리오그램(variogram) 모델과 조건자료를 이용하여 단순크리깅 추정치와 분산을 계산한다.
- iv) 모의지점에서 평균과 분산이 단순크리깅 추정치와 분산인 국부적 정규분포를 구축한다.
- v) 구축된 국부적 정규분포로부터 실현치들을 발생시킨다.
- vi) 조건자료에 모의값을 추가하고 모의할 새로운 위치를 선택한다. 이전 모의값은 관측치와 함께 새로운 위치에서의 모의에 사용된다.
- vii) 모의된 실현치에 대한 역변환을 실시한다.
- viii) 단계 ii)~vii)을 반복하여 다수의 실현치를 모의한다.

### 2.2 CEM

#### 2.2.1 블록순환행렬

CEM은 규칙격자(regular grid)에 걸쳐 정상 가우시안 필드(stationary Gaussian field)에 대한 실현치를 발생시키는 기법이다. 이러한 조건하에서 결과되는 벡터  $Z \sim N(0, T)$ 의 공분산 행렬  $T$ 는 테플리츠 블록(Toeplitz block)을 가지는 블록 테플리츠 행렬(block Toeplitz matrix)이 된다(Zimmerman, 1989). 정의에 의해 정칙 테플리츠 행렬(nonsingular Toeplitz matrix)은 모든 대각선을 따라 일정한 값을 가진다. 테플리츠 블록을 가지는 블록 테플리츠 행렬은 테플리츠 구조가 전체 블록 및 각 블록 내에서 적용될 수 있는 블록행렬로서 정의된다.

한편, 순환행렬(circulant matrix)은 테플리츠 행렬의 특별한 경우로서 두 번째 열부터 모든 열의 시작값은 이전 열의 순환에 의해 얻어진다. 즉, 순환행렬에서 열의 마지막 값은 다음 열의 첫 번째 위치로 이동하고 그 열의 다른 모든 값들은 이전 열의 값들이 한 단계씩 아래로 이동하는 순환을 통해 얻어진다. 테플리츠 행렬의 경우와 마찬가지로 순환블록(circulant block)을 가지는 블록순환행렬(block circulant matrix)은 순환구조가 전체 블록 및 각 블록 내에서 적용될 수 있는 블록행렬로서 정의된다. CEM의 기본개념은 표준임베딩(standard embedding) 절차를 이용하여 블록테플리츠 행렬을 순환블록으로서 더 큰 블록순환행렬에 삽입하는 것이다.

#### 2.2.2 2차원 푸리에 변환

$n_1 \times n_2$  직사각형 행렬(rectangular matrix)  $A$ 에 대한

2차원 푸리에 변환은 Eq. (2)와 같은 요소들을 가지는 동일한 크기의 행렬  $B$ 이다.

$$B_{lm} = \frac{1}{\sqrt{n_1 n_2}} \sum_{g=0}^{n_1-1} \sum_{h=0}^{n_2-1} A_{gh} \exp\left(-\frac{2\pi i g l}{n_1}\right) \exp\left(-\frac{2\pi i h m}{n_2}\right) \quad (2)$$

for  $l = 0, \dots, n_1 - 1, m = 0, \dots, n_2 - 1$

2차원 푸리에 변환은 행렬곱 (matrix multiplication)의 형태로 또한 나타낼 수 있다. Eq. (3)과 같은 요소를 가지는  $n_1 n_2 \times n_1 n_2$  행렬  $F$ 를 고려할 경우 행렬  $B$ 는 행렬  $A$ 로부터 다음과 같이 구할 수 있다.

$$F_{lm} = \frac{1}{\sqrt{n_1 n_2}} \exp\left(-\frac{2\pi i}{n_1} \left\lfloor \frac{l}{n_2} \right\rfloor \left\lfloor \frac{m}{n_2} \right\rfloor\right) \exp\left(-\frac{2\pi i}{n_2} (l \bmod n_2)(m \bmod n_2)\right) \quad (3)$$

for  $l, m = 0, 1, \dots, n_1 n_2 - 1$

- 행렬  $A$ 의 요소들을 행방향으로 나열하여  $n_1 n_2$  벡터  $a$ 를 만든다.
- $b = Fa$ 를 계산한다.
- 벡터  $b$ 의 요소들을 행방향으로 나열하여  $n_1 \times n_2$  행렬  $B$ 를 만든다.

### 2.2.3 CEM의 기본이론 및 알고리즘

$n_1 \times n_2$  격자  $S = \{s_{kj}\}$ 에 걸쳐 어떤 주어진 공분산 함수  $r$ 을 가지는 영평균 가우시안 필드  $Z$ 로부터 표본을 발생시킨다고 하자. 격자는 고정간격  $x$  및  $y$ 를 가지는 Eq. (4)와 같은 규칙격자라고 가정한다.

$$s_{kj} = s_{00} + kyv_2 + jxv_1, \quad (4)$$

$$k = 0, \dots, n_1 - 1, j = 0, \dots, n_2 - 1$$

여기서,  $v_1 = (1, 0)$ ,  $v_2 = (0, 1)$ 이다.

위치  $s_{kj}$ 에서  $Z$ 의 값을  $Z_{kj}$ 라고 나타낼 수 있다. 그러나 기법의 목적상 Eq. (5)와 같이  $Z$ 의 값들을  $n_1 n_2$  벡터로 정리하고 행방향으로 표시된 격자의  $i$ 번째 위치에서  $Z$ 의 값을  $Z_i$ 로 나타내는 것이 더욱 편리하다.

$$Z_i \equiv Z_{\lfloor i/n_2 \rfloor, i \bmod n_2}, \quad i = 0, \dots, n_1 n_2 - 1 \quad (5)$$

격자  $S$ 의 규칙성과 필드의 정상성을 가정할 때  $n_1 n_2$  벡터  $Z$ 의  $n_1 n_2 \times n_1 n_2$  공분산 행렬  $T$ 는 테플리츠 블록을 가지는 블록테플리츠행렬이 된다. 즉, 그것은 Eq. (6)과 같이 각각 크기가  $n_2 \times n_2$ 인  $n_1$ 개의 다른 테플리츠 블록들로 구성된다.

$$T = \begin{pmatrix} T^{(1)} & T^{(2)} & \dots & \dots & T^{(n_1)} \\ T^{(2)T} & \ddots & \ddots & \ddots & \ddots \\ T^{(3)T} & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ T^{(n_1)T} & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \quad (6)$$

테플리츠 블록을 가지는 대칭블록테플리츠행렬  $T$ 는 신속한 계산을 위하여 순환블록을 가지는 대칭블록순환행렬  $C$ 에 삽입되어야 한다. 이를 위하여  $C$ 는 비음정치 (non-negative definite)라고 가정되어야 하며, 그렇지 않으면 알고리즘은 성립되지 않는다. 이것은 격자크기가 너무 작거나 공분산 함수  $r$ 이 거리에 따라 급격히 감소되지 않음을 나타내며, 이러한 경우 격자크기를 증가시켜 격자의 반대편 점들과의 공분산이 거의 영이 되도록 해야 한다.  $C$ 가 비음정치가 되는 충분조건에 대한 결과는 Dietrich and Newsam (1997)를 참조할 수 있다.

$C$ 를 양정치 (positive definite)라고 가정하면  $4n_1 n_2$  벡터  $W \sim N(0, C)$ 를 발생시킬 수 있다. 다음으로  $W$ 의 특정요소들을 추출할 수 있으며,  $n_1 n_2$  벡터  $Z$ 로 정리될 경우 결합정규분포를 나타내게 된다.  $W$ 를 발생시키기 위한 알고리즘은 다음과 같다. 먼저  $2n_1 \times 2n_2$  2차원 푸리에 변환행렬  $F$ 에 대하여 Eqs. (7) and (8)을 정의한다.

$$F_1 = Re(F) \quad (7)$$

$$F_2 = Im(F) \quad (8)$$

여기서,  $Re$ 와  $Im$ 은 각각 복소행렬 (complex matrix)  $F$ 에 대한 실수부와 허수부를 리턴한다.

다음으로 대각화 정리 (diagonalization theorem) (Nott and Wilson, 1997)에 의해 Eq. (9)와 같이 나타낼 수 있다.

$$C = F^H A F$$

$$= (F_1 + iF_2)^H A (F_1 + iF_2)$$

$$= (F_1 - iF_2)^T A (F_1 + iF_2) \quad (9)$$

$$= (F_1 - iF_2) A (F_1 + iF_2)$$

$$= (F_1 A F_1 + F_2 A F_2) + i(F_1 A F_2 - F_2 A F_1)$$

여기서,  $F^H$ 는  $F$ 의 공액전치 (conjugate transpose) 또는 에르미트 수반행렬 (Hermitian adjoint),  $A$ 는  $C$ 에 대한 고유치의 대각행렬,  $i$ 는 표준허수단위 (standard imaginary unit)로서  $i^2 = -1$ 이다.

Eq. (9)에서 실수부와 허수부를 같게 놓으면 Eqs. (10) and (11)과 같은 관계를 얻을 수 있다.

$$F_1 A F_1 + F_2 A F_2 = C \quad (10)$$

$$F_1 \Delta F_2 - F_2 \Delta F_1 = 0 \quad (11)$$

다음으로 독립요소를 가지는 두 개의 독립벡터  $e_1, e_2 \sim N(0, \mathbf{A})$  를 발생시킨 후 복소벡터 (complex vector)  $e = e_1 + ie_2$  에 결합시킨다.  $\mathbf{A}$  에 대한 고유치 벡터  $\lambda$  는  $\mathbf{C}$  의 첫 번째 열로부터 얻어진  $2n_1 \times 2n_2$  행렬에 대한 2차원 푸리에 변환에 의해 계산될 수 있다. 모든  $\lambda_j$  가 음수가 아니라고 가정하면 Eq. (12)와 같이 나타낼 수 있다.

$$\begin{aligned} w &= Fe \\ &= (F_1 + iF_2)(e_1 + ie_2) \\ &= (F_1 e_1 - F_2 e_2) + i(F_2 e_1 + F_1 e_2) \\ &\equiv w_1 + iw_2 \end{aligned} \quad (12)$$

비록 여기서  $w = Fe$  라고 나타내지만 사실상 실제 계산은 행렬로 재배치된  $e$  에 대한 2차원 푸리에 변환이다.

다음으로  $W \sim N(0, \mathbf{C})$  의 실현치로부터  $Z \sim N(0, \mathbf{T})$  의 발생은 다음과 같은 절차를 따른다.  $W$  의 첫 번째 절반을 가져오고 그것을 길이  $n_2$  의 그룹으로 나눈 Eq. (13)과 같이 다른 모든 그룹을 가져온다. 그 결과는  $Z \sim N(0, \mathbf{T})$  가 된다.

$$\begin{aligned} Z = & (W_0, W_1, \dots, W_{n_2-1}, \\ & W_{2n_2}, W_{2n_2+1}, \dots, W_{3n_2-1}, \\ & \dots, \dots, \dots, \dots, \\ & W_{2(n_1-1)n_2}, W_{2(n_1-1)n_2+1}, \dots, W_{2(n_1-1)n_2+n_2-1}) \end{aligned} \quad (13)$$

CEM을 이용하여  $Z$  를 발생시키기 위한 알고리즘을 정리하면 다음과 같다 (Kozintsev, 1999).

- i) 주어진 공분산 행렬  $\mathbf{T}$ 로부터 블록순환행렬  $\mathbf{C}$  의 첫 번째 열을 얻은 후 그것을  $2n_1 \times 2n_2$  행렬로 행방향으로 재배열한다.
- ii)  $\mathbf{C}$  의 고유치 벡터  $\lambda$  를 얻기 위하여 i)에서 구한 행렬에 대한 2차원 푸리에 변환을 실시한다.
- iii) 각  $\lambda_j$  가 음수인지 아닌지 확인한다. 만약 음수가 아니면 알고리즘을 계속 진행하고 그렇지 않으면 알고리즘이 실패하기 때문에  $n_1$  과  $n_2$  를 증가시키고 알고리즘을 다시 시작한다.
- iv)  $j = 1, \dots, 4n_1n_2$  에 대하여  $e_1, e_2 \sim N(0, \lambda_j)$  를 발생시킨 후 복소벡터  $e = e_1 + ie_2$  에 결합하고 그것을 행방향으로  $2n_1 \times 2n_2$  행렬로 재배열한다.
- v) 복소벡터  $w$  를 얻기 위하여 iv)에서 구한 행렬에 대한 2차원 푸리에 변환을 실시한다.
- vi)  $w$  의 실수부와 허수부를 분리하여 각각으로부터 필요한 요소를 추출하며, 이것은 독립된  $z_1, z_2 \sim N(0, \mathbf{T})$  가 된다.

### 3. 적용 및 분석

#### 3.1 적용구역 및 분석개요

본 연구는 확률강우량의 공간변동성에 대한 불확실성 분석 시 다양한 불확실성 인자들을 모의 알고리즘에 반영함에 있어서 효율적인 모의 알고리즘을 선택하기 위하여 지구통계학적 추계모의기법인 SGS와 CEM을 비교하였다. 이를 위하여 국제수문개발계획 (IHP)에 의해 선정 및 운영된 바 있는 위천대표유역을 포함하고 있어 다수의 양질의 수문자료를 보유하고 있으며, 유역의 대부분이 산지



Fig. 1. Study Area and Location of Raingauges

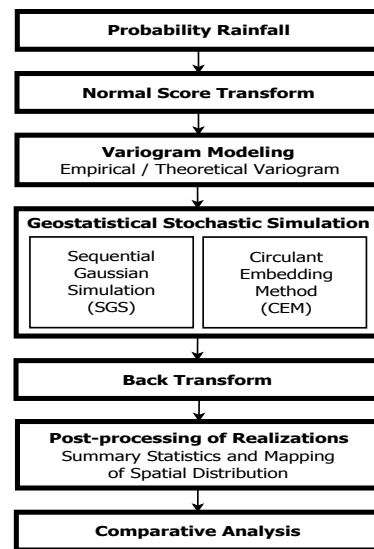


Fig. 2. Flow Chart for Geostatistical Stochastic Simulation

로 이루어져 있고 도시화율이 낮은 위천유역을 분석유역으로 선정하였다. SGS 및 CEM의 적용에 있어서 크리깅 추정치와 분산을 산정하기 위하여 단순크리깅을 적용하였으며, 추계모의를 위한 입력자료인 대상유역 내 지점 확률강우량은 서영민 등 (2010)이 분석한 위천유역의 18개 관측지점에 대한 확률강우량 분석결과 (확률분포형: Gumbel, 100년 빈도, 지속시간 24시간)를 사용하였다. 그리고 베리오그램 모델링, 공간추계모의 및 후처리를 위하여 통계계산 및 그래픽을 목적으로 GNU 프로젝트로 개발된 프로그래밍 언어인 R을 이용하였다. 본 연구의 대상유역 및 모의절차는 Figs. 1 and 2와 같다.

베리오그램 모델링 및 공간추계모의에 앞서 입력자료로 사용한 지점확률강우량값들이 정규분포를 따르지 않기 때문에 먼저 정규점수변환 (normal score transform)을 실시하였다. 입력자료의 분포가 크게 왜곡되어 있을 경우 베리오그램 모델링, 크리깅 분산추정 및 랜덤필드의 모의발생에 있어서 잘못된 결과를 발생할 수 있기 때문에 사전에 입력자료를 정규분포로 변환하는 것이 유리하다. 일반적으로 널리 적용되는 변환기법으로는 대수정규변환, Box-Cox 변환, 정규점수변환 등이 있으나 대수정규변환과 Box-Cox 변환의 경우 입력자료로부터 근사적인 정규분포를 유도하고 분포의 꼬리부분에서 다소 편차가 발생할 수 있기 때문에 대부분의 경우 상대적으로 정규분포로의 정확한 변환이 가능한 정규점수변환이 채택되고 있다 (Goovaerts, 1997).

다음으로 경험베리오그램을 산정하고 가중최소제곱법을 이용하여 베리오그램 모델의 매개변수를 추정하였으며, 베리오그램 모델로서 구형모델 (spherical model)을 선택하였다. 그리고 각 모의지점에서의 추정치와 분산을 산정하기 위하여 단순크리깅을 적용하였으며, SGS 및 CEM을 이용한 추계모의를 통해 500개의 실현치를 생성하였다. 이러한 실현치들로부터 후처리를 통해 확률강우량의 공간분포에 대한 불확실성을 평가하였으며, 불확실성에 대한 정량적 측도로서 표준편차, 변동계수, 사분위 범위 (제3사분위수-제1사분위수) 및 범위 (최대값-최소값)에 대한 공간분포를 고려하였다.

### 3.2 모의 효율성 검토

SGS는 모의과정을 통하여 이전 모의값과 입력자료를 조건자료로서 사용하며, 모의과정 중 크리깅 단계에서 조건자료수가 증가함에 따라 공분산 행렬에 대한 역행렬 산정 및 행렬-벡터곱 연산단계가 CPU 시간에 지배적인 영향을 미치기 때문에 합리적인 CPU 시간을 유지하기 위하여 사전에 미리 설정된 최대 개수 이하의 최근린 조건자료만을 사용하여 모의가 이루어진다. Fig. 3은 SGS를 이

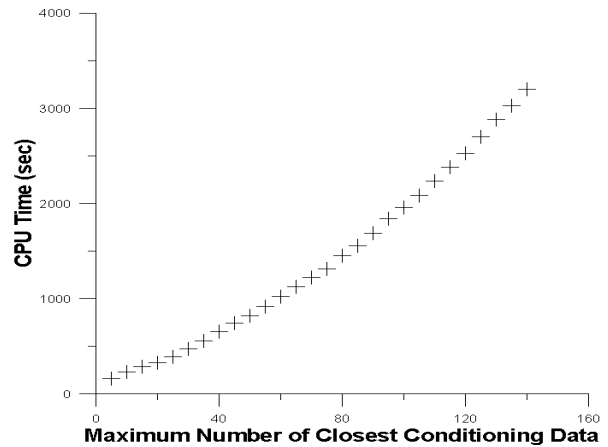


Fig. 3. CPU Time for Maximum Number of Closest Conditioning Data

※ System information - OS, MS Windows 7; CPU, Intel Mobile Core2 Duo T5550, 1.83GHz; RAM, 2.0GB Single-Channel DDR2, 332MHz

용한 추계모의에 있어서 지역탐색 및 조건부 추정시 최근린 조건자료의 최대 개수에 따른 CPU 시간을 나타낸 것으로서 조건자료의 수가 증가함에 따라 모의실행에 소요되는 CPU 시간이 증가하는 추세를 나타내고 있다.

기 설정된 최대 개수 이하로 조건자료를 제한함으로써 모의지점으로부터 원거리에 위치한 자료를 걸러내는 효과를 가져오며, 이러한 자료들이 가지는 추가적인 정보는 무시할 수 있을 정도로 충분히 작다고 간주한다. 만약 작은 수의 조건자료가 사용될 경우 추정치의 정확도가 떨어지고 높은 분산을 가지게 되며, 이것은 가능한 범위를 벗어나는 모의값을 발생시킬 가능성을 증가시키게 된다. 반면에 더 많은 조건자료를 사용함에 따라 추정치가 개선되고 크리깅 분산이 감소하게 되며, 감소된 분산은 불가능한 값이 발생할 가능성을 제한시키지만 CPU 시간이 증가하는 단점을 가지게 된다. 따라서 모의의 정확성 측면에서 조건자료의 수를 증가시키는 것이 유리하지만 Fig. 3과 같이 조건자료의 수가 증가함에 따라 CPU 시간이 증가하여 모의의 효율성 측면에서는 불리하기 때문에 조건자료의 수에 따른 크리깅 추정치 및 분산의 민감도를 검토할 필요가 있다.

Figs. 4 and 5는 조건자료의 수에 따른 크리깅 추정치 및 분산의 변화를 분석한 것으로서 무작위로 선택된 임의의 두 모의위치에 대한 크리깅 추정치 및 분산과 전체 모의위치에 대한 크리깅 추정치 및 분산의 평균을 나타내고 있다. Figs. 4 and 5로부터 조건자료의 수가 20개 이상일 경우 크리깅 추정치 및 분산이 일정하게 나타나는 것으로 분석되었다. 따라서 본 적용유역의 경우 추계모의의 정확도 및 효율성 측면에서 최소 20개 이상의 조건자료를 최근린 조건자료의 최대개수로 설정해야 하며, 이러한 경우

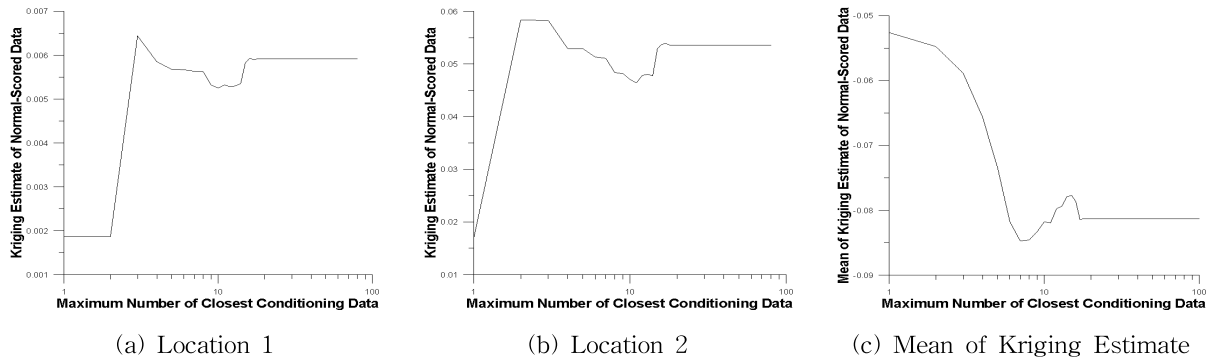


Fig. 4. Kriging Estimates for Maximum Number of Closest Conditioning Data

※ Location 1: (361918.7, 330980.1), Location 2: (353018.7, 329080.1)

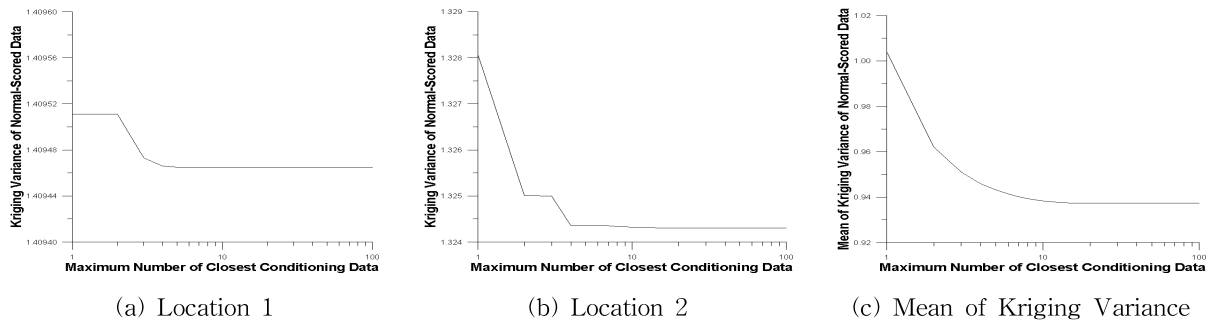


Fig. 5. Kriging Variances for Maximum Number of Closest Conditioning Data

SGS에 대한 CPU 시간은 약 6.7분이 소요된다.

반면에 CEM의 경우 모의과정에서 조건자료 개수를 설정할 필요가 없기 때문에 조건자료의 개수에 따른 민감도를 검토할 필요가 없으며, 블록테플리즈 및 블록순환행렬에 의해 공분산 행렬에 대한 연산을 크게 단순화시킬 수 있고 2차원 푸리에 변환에 의한 대각화를 통하여 신속한 연산이 가능하기 때문에 SGS에 비해 상대적으로 모의 효율성 측면에서 유리하다. 그리고 SGS는 이동근린탐색(moving neighborhood search)을 통한 근사화에 기반하기 때문에 크리깅 단계의 구조가 유동적인 반면 CEM은 이동근린탐색과 같은 단순화가 필요하지 않기 때문에 크리깅 구조가 고정적이며, 조건부 크리깅에 원자료만을 포함한다. 또한 CEM은 모든 실현치에 대해 단 한 번의 조건부 크리깅 단계가 수행되지만 SGS는 동일한 경로를 사용하지 않을 경우 실현치 당 한 번의 조건부 크리깅 단계가 수행되기 때문에 모의 효율성 측면에서 CEM은 SGS보다 유리하다.

본 연구의 경우 CEM에 대한 CPU 시간은 약 2.7분으로 분석되었으며, 이것은 SGS에서 조건자료의 최대 개수 5개에 상응하는 CPU 시간(약 3.7분)보다도 낮기 때문에 모의 시간 측면에서는 CEM이 SGS보다 우수함을 알 수 있다. 특히 이러한 결과는 모의과정에서 추가적인 불확실성을

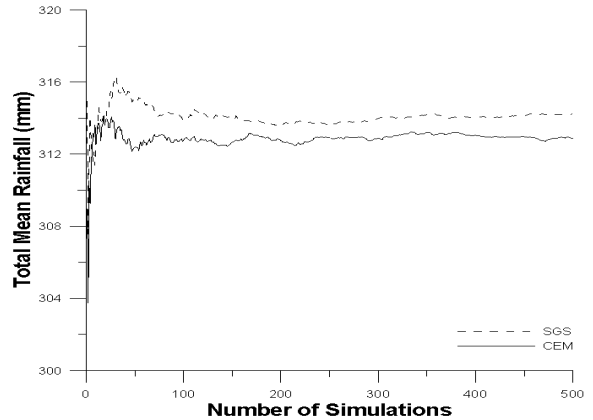


Fig. 6. Total Mean for Number of Simulations

고려할 경우 모의 알고리즘에서의 추가적인 루프의 중첩으로 인하여 CPU 시간이 기하급수적으로 증가할 수 있기 때문에 공간추계모의 및 불확실성 분석에 있어서의 효율적인 모의 알고리즘 선택 측면에서 중요하다고 할 수 있다.

Fig. 6은 모의횟수 증가에 따른 실현치의 공간평균에 대한 평균, 즉 총평균의 변화를 나타낸 것으로서 SGS와 CEM 모두 모의횟수가 약 350회 이상일 때 총평균이 거의 일정하게 유지되는 것으로 분석되었다. 이러한 결과로부터 본 연구에서는 SGS 및 CEM에 대한 모의횟수를 500회

로 하여 확률강우량의 공간분포에 대한 불확실성 측도, 즉 표준편차, 변동계수, 사분위수 범위 및 범위를 추정하였다.

Fig. 7은 본 연구의 비교대상 알고리즘인 SGS와 CEM이 베리오그램으로 표현된 공간상관구조에 대한 재생성 정도를 비교하기 위하여 SGS 및 CEM을 이용하여 발생된 실현치들로부터 경험베리오그램을 산정하여 베리오그램 모델과 도식적으로 비교하였다. 그 결과 SGS와 CEM

을 이용하여 발생된 실현치들에 대한 경험베리오그램은 공간상관구조의 재생성에 있어서 비슷한 패턴을 나타내고 있으며, 평균적으로 베리오그램 모델을 재생성하는 것으로 분석되었다.

### 3.3 확률강우량의 공간분포에 대한 불확실성 분석

Table 1은 확률강우량의 공간분포에 대한 불확실성을

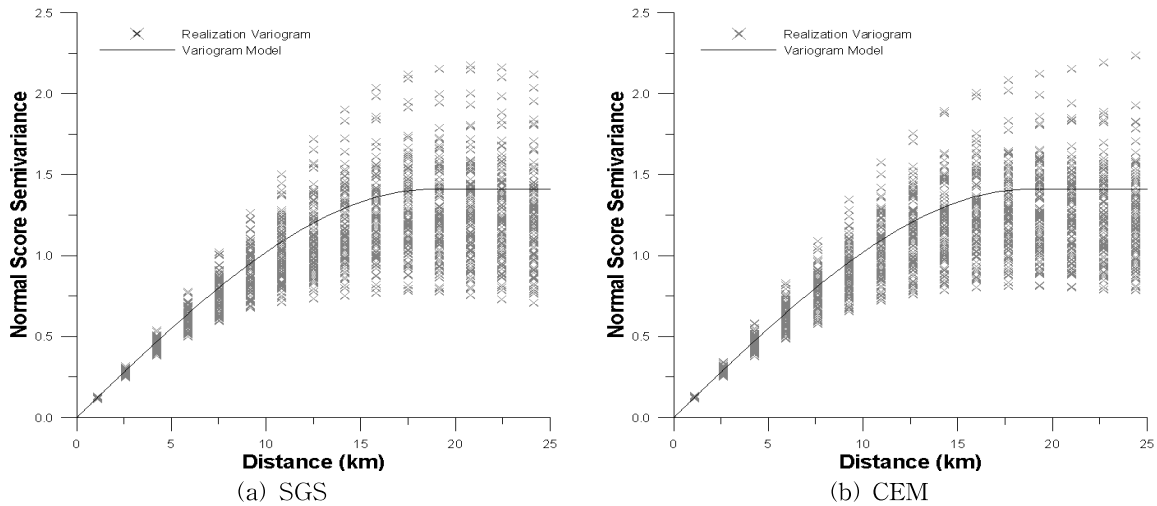


Fig. 7. Realization Variograms and Variogram Model of Normal-Scored Data

Table 1. Comparison of Spatial Distributions of Uncertainties

| Realization Statistics               | Simulation Methods | Spatial Statistics |           |        |        |           |        |
|--------------------------------------|--------------------|--------------------|-----------|--------|--------|-----------|--------|
|                                      |                    | Min.               | 1st Quan. | Median | Mean   | 3rd Quan. | Max.   |
| Realization SD (mm)                  | SGS                | 0.00               | 24.50     | 32.49  | 31.07  | 38.50     | 46.99  |
|                                      | CEM                | 0.00               | 27.54     | 36.38  | 34.55  | 42.41     | 50.08  |
| Realization CV                       | SGS                | 0.000              | 0.079     | 0.103  | 0.099  | 0.121     | 0.151  |
|                                      | CEM                | 0.000              | 0.088     | 0.114  | 0.111  | 0.138     | 0.161  |
| 1st Quantile of Realization (mm)     | SGS                | 258.20             | 279.00    | 293.50 | 295.60 | 308.10    | 420.80 |
|                                      | CEM                | 235.10             | 277.60    | 293.50 | 293.10 | 305.60    | 420.80 |
| 3rd Quantile of Realization (mm)     | SGS                | 279.50             | 320.40    | 323.30 | 329.60 | 334.10    | 424.10 |
|                                      | CEM                | 276.30             | 319.70    | 324.30 | 329.60 | 337.50    | 424.30 |
| Realization Interquartile Range (mm) | SGS                | 0.00               | 26.70     | 34.18  | 34.02  | 40.70     | 91.78  |
|                                      | CEM                | 0.00               | 29.08     | 36.73  | 36.51  | 43.36     | 90.36  |
| Realization Minimum (mm)             | SGS                | 118.20             | 191.10    | 208.30 | 218.40 | 236.70    | 420.80 |
|                                      | CEM                | 71.25              | 183.40    | 199.50 | 209.80 | 227.50    | 420.80 |
| Realization Maximum (mm)             | SGS                | 279.50             | 422.40    | 433.80 | 426.40 | 447.60    | 502.90 |
|                                      | CEM                | 276.30             | 426.70    | 441.20 | 433.80 | 454.30    | 513.70 |
| Realization Range (mm)               | SGS                | 0.00               | 163.70    | 222.20 | 208.10 | 248.60    | 347.30 |
|                                      | CEM                | 0.00               | 177.20    | 238.10 | 224.10 | 268.30    | 359.40 |

SD, standard deviation; CV, coefficient of variation; MIN., Minimum; 1st QUAN., 1st Quantile; 3rd QUAN., 3rd Quantile; MAX., Maximum.



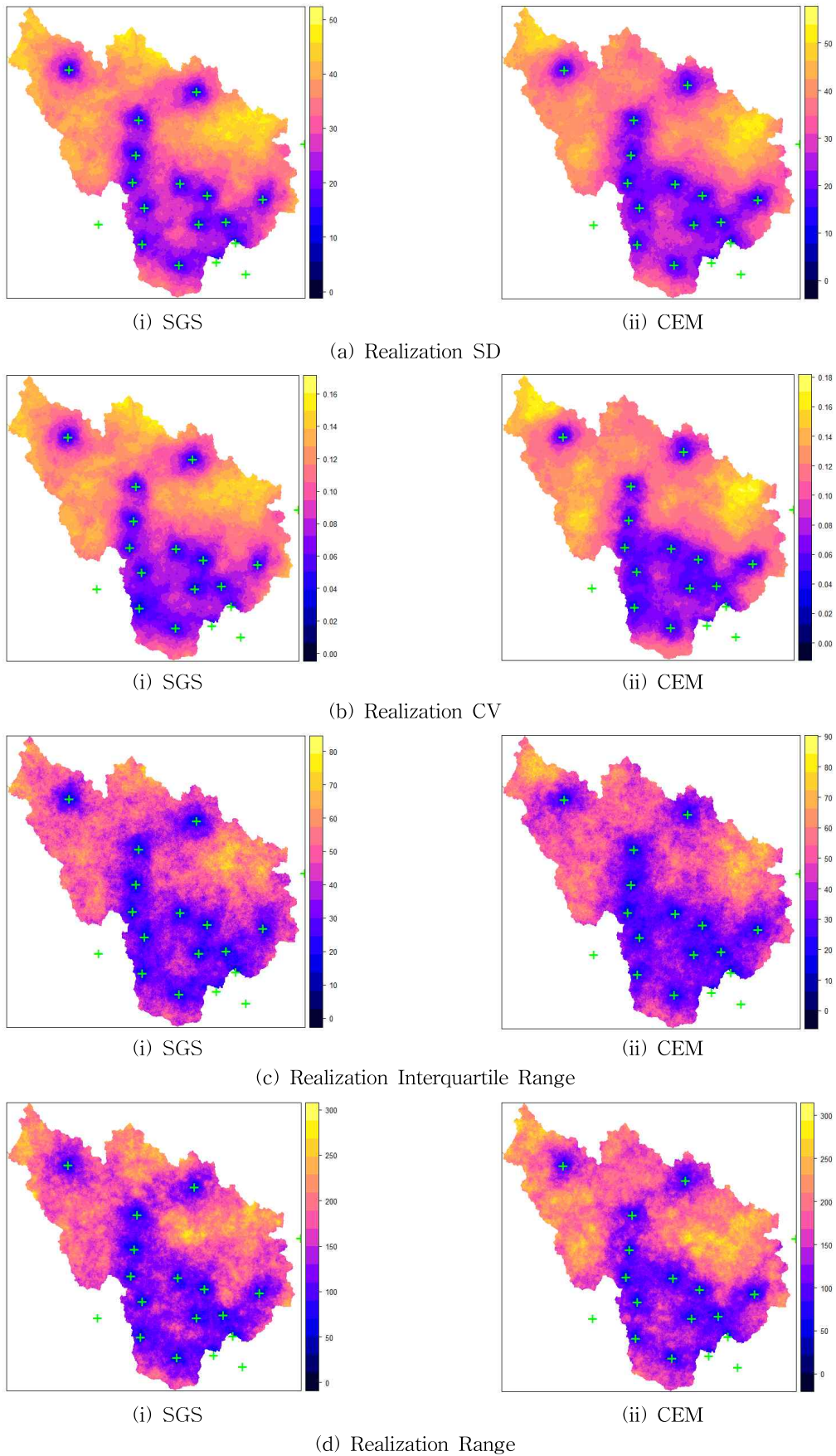
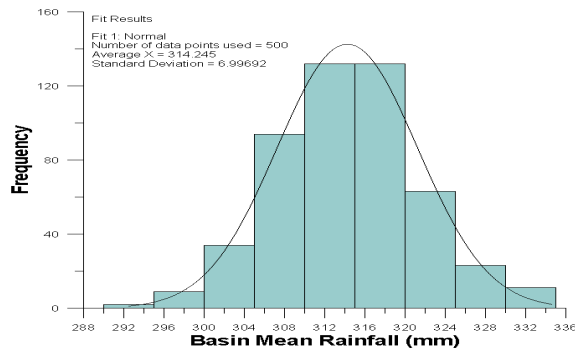
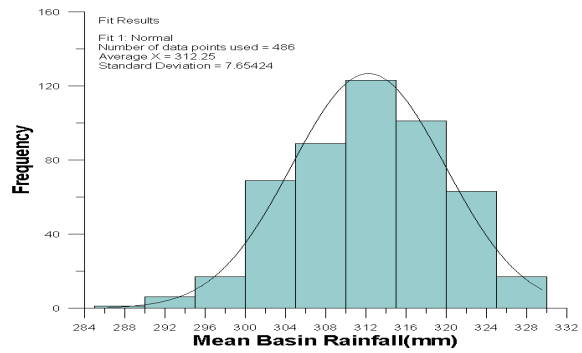


Fig. 8. Spatial Distributions of Uncertainties



(a) SGS



(b) CEM

Fig. 9. Uncertainty Distribution of Basin Mean Rainfall

Table 2. Summary of Uncertainty Distribution of Basin Mean Rainfall

| METHODS | MIN.  | 1st QUAN. | MEDIAN | MEAN  | 3rd QUAN. | MAX.  |
|---------|-------|-----------|--------|-------|-----------|-------|
| SGS     | 292.5 | 309.5     | 314.4  | 314.2 | 318.3     | 334.6 |
| CEM     | 286.3 | 306.9     | 313.2  | 312.9 | 318.6     | 348.0 |

분석한 결과로서 본 연구에서는 불확실성을 평가하기 위한 통계치로서 실현치에 대한 표준편차, 변동계수, 사분위수 범위 및 범위의 공간분포를 분석하여 요약통계치로 정리하였다. SGS 및 CEM 모두 조건부 모의를 실시하였기 때문에 관측소 위치들에서의 확률강우량값은 모의과정 내에서 항상 유지되며, 따라서 실현치의 불확실성 구간에 대한 최소값은 0이 된다. 즉, 실현치에 대한 표준편차, 변동계수, 사분위 범위 및 범위의 공간분포에 대한 최소값은 0이 되며, 실현치에 대한 최소값 및 제1사분위수의 공간분포에 대한 최대값은 관측소 위치들에서의 확률강우량값의 최대값과 같게 된다. 그 밖의 통계치를 살펴보면 실현치에 대한 사분위 범위의 경우 두 기법이 거의 비슷한 결과를 나타내고 있으나 사분위 범위에 대한 공간평균이 각각 36.51 및 34.02mm로서 CEM이 SGS보다 약간 큰 것으로 나타난 반면 실현치에 대한 범위의 경우 공간평균이 각각 224.10 및 208.10mm로서 CEM이 SGS보다 다소 큰 결과를 나타내는 것으로 분석되었다. 실현치에 대한 표준편차 및 변동계수의 경우도 마찬가지로 CEM이 SGS보다 약간 큰 결과를 나타내는 것으로 분석되었다.

한편, Fig. 8은 확률강우량의 공간분포에 대한 불확실성 추정치의 공간분포를 나타낸 것으로서 불확실성 추정치는 관측지점에 가까울수록 작고 관측지점에서 멀어질수록 커지는 양상을 나타내고 있으며, SGS와 CEM의 모의결과가 비슷한 것으로 분석되었다. 따라서 확률강우량의 공간분포에 대한 불확실성에 대하여 SGS와 CEM의 모의결과는 대체로 비슷한 것으로 분석된 반면 모의 효율성 측면에서는 SGS 보다 CEM이 우수하기 때문에 확률

강우량의 공간분포에 대한 불확실성 평가는 CEM을 이용하는 것이 유리함을 알 수 있다.

Fig. 9 및 Table 2는 SGS 및 CEM을 이용한 추계모의로부터 추정된 유역평균강우량의 불확실성 분포를 나타낸 것으로서 각각 평균이 314.2, 312.9 mm, 표준편차가 6.997, 7.654 mm이며, 히스토그램으로부터 두 기법의 결과가 대체로 비슷한 분포를 나타냄을 알 수 있다.

#### 4. 결론 및 향후과제

본 논문에서는 확률강우량에 대한 공간분포 추정에 있어서 공간변동성에 따른 불확실성을 평가하기 위하여 가우시안 랜덤필드모델을 이용한 지구통계학적 추계모의기법인 CEM과 SGS 기법을 적용하였으며, 위천유역을 대상으로 한 적용분석으로부터 그 결과를 비교하였다. CEM과 SGS를 이용한 공간추계모의에 있어서 모의 효율성을 검토하였으며, 확률강우량의 공간분포에 대한 불확실성 평가측도로서 실현치에 대한 통계치(표준편차, 변동계수, 사분위수 범위 및 범위)의 공간분포를 비교하였고 유역평균강우량의 불확실성 분포를 비교하였다. 그 결과 확률강우량의 공간분포 및 유역평균강우량에 대한 불확실성 추정결과는 두 기법이 대체로 비슷한 결과를 보이는 것으로 분석되었으나 모의 효율성 측면, 특히 모의시간에 있어서는 CEM이 SGS에 비해 우수한 결과를 나타내는 것으로 분석되었다.

본 논문에서는 확률강우량에 대한 공간분포 추정에 있어서 공간변동성에 따른 불확실성만을 다루고 있으며,

공간상관특성을 설명하는 베리오그램 모델의 매개변수에 대한 불확실성 추정에는 포함하지 않는다. 따라서 향후 연구에서 베리오그램 모델의 매개변수에 대한 불확실성을 고려하기 위하여 Bootstrap, Jackknife 및 Bayesian MCMC 등과 같은 기법들과 공간추계모의기법의 연계와 관련된 연구가 필요할 것으로 판단된다. 그리고 SGS 및 CEM 이외의 공간추계모의기법들에 대한 비교, 모의과정에서의 효율적인 샘플링을 위하여 Latin Hypercube 기법과 같은 층화표본추출기법의 도입, 연구결과의 일반성을 위한 다수의 유역에 대한 적용분석 등과 같은 추가연구가 제안될 수 있을 것이다.

## 감사의 글

본 연구는 소방방재청에서 주관하고 있는 자연재해저감기술개발사업단 연구개발사업 (NEMA-자연-08-06)의 일환으로 수행된 “지구단위홍수재해 정보시스템 및 침수재해저감기술개발”의 일부로서 본 연구의 재정지원에 감사드립니다.

## 참고문헌

- 서영민, 여운기, 이승윤, 지흥기 (2010). “확률강우량의 공간분포 추정을 위한 KED 기법의 적용.” **한국수자원학회 논문집**, 한국수자원학회, 제43권, 제8호, pp. 757-767.
- Casta, A.C., Durão, R., Pereira, M.J., and Soares, A. (2008). “Using stochastic space-time models to map extreme precipitation in southern Portugal.” *Natural Hazards and Earth System Sciences*, Vol. 8, pp. 763-773.
- Colle, B.A., Mass, C.F., and Westrick, K.J. (2000). “MM5 precipitation verification over the Pacific Northwest during the 1997-99 cool seasons.” *Weather and Forecasting*, Vol. 15, pp. 730-744.
- Deutsch, C.V., and Journel, A.G. (1997). *GSLIB: Geostatistical software library and user's guide*. 2nd Edition, Oxford University Press, New York.
- Dietrich, C., and Newsam, G. (1997). “Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix.” *SIAM Journal of Scientific Computation*, Vol. 18, No. 4, pp. 1088-1107.
- Dinku, T., Anagnostou, M., and Borga, M. (2002). “Improving radar-based estimation of rainfall over complex terrain.” *Journal of Applied Meteorology*, Vol. 41, pp. 1163-1178.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. Oxford University Press, New York, p. 496.
- Haberlandt, U. (2007). “Geostatistical interpolation of hourly precipitation from rain gauge and radar for a large-scale extreme rainfall event.” *Journal of Hydrology*, Vol. 332, pp. 144-157.
- Haberlandt, U., and Gattke, C. (2004). “Spatial interpolation vs. simulation of precipitation for rainfall-runoff modelling—a case study in the Lippe river basin.” *Proceedings of the British Hydrological Society International Conference: Hydrology: Science and Practice for the 21st Century*, London, Vol. 1, pp. 120-127.
- Hengl, T. (2009). *A practical guide to geostatistical mapping*. 2nd Edition, University of Amsterdam.
- Kanevski, M. (2008). *Advanced mapping of environmental data: geostatistics, machine learning and Bayesian maximum entropy*. John Wiley & Sons, Inc.
- Kozintsev, B. (1999). *Computations with Gaussian random fields*. Ph.D. dissertation, University of Maryland, College Park, MD, USA, p. 33.
- Nott, D.J., and Wilson, R.J. (1997). “Parameter estimation for excursion set texture models.” *Signal Processing*, Vol. 63, pp. 199-210.
- Schultz, D.M., Steenburgh, W.J., Trapp, R.J., Horel, J., Kingsmill, D.E., Dunn, L.B., Rust, W.D., Cheng, L., Bansaer, A., Cox, J., Daugherty, J., Jorgensen, D.P., Meitin, J., Showell, L., Smull, B.F., Tarp, K., and Trainor, M. (2002). “Understanding Utah winter storms - The international precipitation experiment.” *Bulletin of the American Meteorological Society*, Vol. 83, pp. 189-199.
- Stoelinga, M.T., Hobbs, P.V., Mass, C.F., Locatelli, J.D., Colle, B.A., Houze, R.A.J., Rangno, A.L., Bond, N.A., Smull, B.F., Rasmussen, R.M., Thompson, G., and Colman, B.R. (2003). “Improvement of microphysical parameterization through observational verification experiment.” *Bulletin of the American Meteorological Society*, Vol. 84, pp. 1807-1813.
- Vann, J., Bertoli, O., and Jackson, S. (2002). “An overview of geostatistical simulation for quantifying risk.” *Proceedings of the Geostatistical Association*

- of Australasia Symposium: Quantifying Risk and Error*, Perth, Western Australia, pp. 13-28.
- Webster, R., and Oliver, M.A. (2007). *Geostatistics for environmental scientists*. 2nd Edition, John Wiley & Sons, Ltd.
- Westrick, K.J., Mass, C.F., and Colle, B.A. (1999). "The limitations of the WSR-88D radar network for quantitative precipitation measurement over the coastal western United States." *Bulletin of the American Meteorological Society*, Vol. 80, pp. 2289-2298.
- Young, C.B., Nelson, B.R., Bradley, A.A., Smith, J.A., Peters-Lidard, C.D., Kruger, A., and Baeck, M.L. (1999). "An evaluation of NEXRAD precipitation estimates in complex terrain." *Journal of Geophysical Research*, Vol. 104, No. D16, pp. 19691-19703.
- Zimmerman, D.L. (1989). "Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models." *Journal of Statistical Computation and Simulation*, Vol. 32, pp. 1-15.

|                  |                  |
|------------------|------------------|
| 논문번호: 10-085     | 접수: 2010.10.07   |
| 수정일자: 2010.10.22 | 심사완료: 2010.10.22 |