

## 시공간 상관성을 고려한 일기산출기 모형을 이용한 4대강 유역별 미래 일기 변수 산출

이동환<sup>1</sup> · 이재용<sup>2</sup> · 오희석<sup>3</sup> · 이영조<sup>4</sup>

<sup>1</sup>서울대학교 통계학과, <sup>2</sup>서울대학교 통계학과, <sup>3</sup>서울대학교 통계학과, <sup>4</sup>서울대학교 통계학과

(2011년 12월 12일 접수, 2012년 1월 25일 수정, 2012년 1월 31일 채택)

### 요약

일기 산출기 모형은 가상의 일기 자료를 생성하는 통계 모형이다. 본 연구는 시공간 상관성이 고려된 다중지점에서의 일기산출 모형을 제안하고, 온실가스 배출 미래 시나리오에 따라 강수량과 평균 기온 일기산출이 가능한 알고리즘을 개발하였다. 제안된 알고리즘은 다단계 일반화 선형모형 하에서 필요한 모수들을 추정하고, 적합한 모형 하에서 일기변수들을 랜덤하게 산출하는 절차이다. 과거 30년간 관측된 우리나라 4대강 유역의 일 강수량 자료와 평균 기온 자료를 가지고 모형을 적합하고, 미래 일별 일기자료 산출에 적용하였다.

주요어: 일기산출기, 다단계 일반화 선형 모형, 시공간 상관성.

### 1. 서론

확률적 일기산출기(Stochastic weather generator)는 관측된 강수량, 기온, 습도, 태양복사 등의 기상 일기자료를 가지고 가상의 일별 자료를 생산하는 통계 모형으로, 산출된 일기 자료는 관측 자료와 통계적으로 유사한 구조를 가지게 된다 (Richardson, 1981; Wilks와 Wilby, 1999). 현재까지 제안된 대부분의 일기산출기들은 기상요소별 확률분포함수를 이용하여 관련 모수들을 추정하고, 난수 발생을 통해 일련의 기상자료들을 생성한다 (강문성 등, 1998). 대표적으로 가장 널리 쓰이고 있는 일기산출기 모형 중의 하나인 WGEN (Weather Generator; Richardson과 Wright, 1984)은 강수 발생과정을 일차 마코프 모형을 통해 모형화하고, 강수의 유무에 따라 0이 아닌 강수량은 감마분포를 가정하고, 다른 기상변수들(최고기온, 최저기온, 일복사량 등)도 모수적 분포 가정을 통해 기상변수들을 생성한다. 최근 Furrer와 Katz (2007)은 Nelder와 Wedderburn (1972)이 제안한 일반화 선형모형(Generalized linear models; GLMs)을 이용하여 이진형 자료인 강수 발생 자료는 이항분포를 가정하고, 강수량은 감마분포를 따른다고 가정하여 일기산출기 모형을 제안한바 있다.

그러나 WGEN을 비롯하여 기존에 개발된 대부분의 일기산출기는 단일 지점(single-site)에서의 일별 기상자료를 산출하는 모형으로 다중지점(multi-sites)의 일기를 함께 산출할 경우에 시간적, 공간적 상관성을 고려하지 못하는 한계점이 존재한다. 따라서 최근에는 이러한 시·공간적 상관성을 고려한 다중지점의 일기산출방법과 실제 일기변수들의 특징들을 보다 잘 반영할 수 있는 다양한 모수적, 비모수적 통계모형을 이용한 일기산출기의 개발이 활발히 연구되고 있다. Wilks (1998)은 다변량 정규 분포를 이용하여 WGEN을 다중지점간의 공간 상관성을 허용하는 모형을 제안한바 있다. 그러나 이 방법은 관측지

이 연구는 기상청 기후변화 감시·예측 및 국가정책지원 강화사업(RACS\_2010-4010)의 지원으로 수행되었습니다.

<sup>4</sup>교신저자: (151-747) 서울시 관악구 관악로 599, 서울대학교 통계학과, 교수. E-mail: [youngjo@snu.ac.kr](mailto:youngjo@snu.ac.kr)

표 2.1. 4대강 유역별, 월별 일평균 강수량 및 기온 통계표

일기변수	유역	관측지점 수	1월	2월	3월	4월	5월	6월
강수량	한강	14	0.80	1.06	1.41	2.57	2.89	4.77
	낙동강	18	0.91	1.36	2.01	3.30	3.29	6.16
	금강	12	0.93	1.23	1.59	2.72	2.79	5.64
	영산강	10	1.03	1.59	2.20	3.43	3.43	7.14
기온	한강	14	-3.64	-1.02	4.60	11.42	16.64	21.04
	낙동강	18	-0.05	1.93	6.65	12.79	17.36	21.17
	금강	12	-1.98	0.06	5.14	11.77	17.03	21.51
	영산강	10	0.38	1.98	6.40	12.32	17.11	21.19
일기변수	유역	관측지점 수	7월	8월	9월	10월	11월	12월
강수량	한강	14	9.82	9.99	5.25	1.82	1.64	0.86
	낙동강	18	7.75	8.61	5.16	1.62	1.49	0.70
	금강	12	8.74	9.17	4.51	1.73	1.74	1.02
	영산강	10	8.35	9.01	5.03	1.56	1.75	0.89
기온	한강	14	24.11	24.32	19.48	12.98	5.87	-0.56
	낙동강	18	24.47	24.95	20.47	14.64	8.24	2.45
	금강	12	24.80	25.02	20.12	13.61	6.84	0.79
	영산강	10	24.64	25.32	21.06	15.07	8.63	2.89

점의 개수가 많을 때, 계산적 어려움이 있어 Kim 등 (2008)과 Lee 등 (2010)에서 개선된 모형과 계산 방법들을 제시하였다.

Lee와 Nelder (1996)은 관측되지 않는 변량 효과(random effects)를 도입하여 일반화 선형모형을 확장한 다단계 일반화 선형 모형(Hierarchical generalized linear models; HGLMs)을 제안하였다. 이 모형은 변량 효과를 통해 시간적, 공간적 상관성을 설명함으로써, 다변량 관측치들의 복잡한 공분산을 쉽게 표현하고 계산할 수 있는 장점이 있다. Jang 등 (2007)은 다단계 일반화 선형 모형을 이용해 공간 상관성을 가진 질병지도 모형을 제안하여 지역간 상관성이 있는 자료들의 분석에서의 우수성을 보인바 있다. 본 연구는 다단계 일반화 선형모형을 이용하여 다중 지점에서의 시공간 상관성이 고려된 일기산출 모형을 제안한다. 우리나라의 미래 일별 일기자료 산출을 위해 전국 54개 기상관측소에서 30년(1976년~2005년)간 관측된 일별 강수량과 평균기온 자료를 이용하였다. 제안된 일기산출 알고리즘은 다음의 두 단계로 이루어진다. 첫 번째 단계는 54개 기상관측소를 국가 수자원 관리 종합 정보 시스템(Water Management Information System; WAMIS)을 참고하여 4대강(한강, 낙동강, 금강, 영산강 및 섬진강) 유역별로 나누고, 강수 발생자료, 강수량 자료, 평균 기온 자료를 다단계 일반화 선형모형을 이용하여 모형화 하고, 모형에 필요한 모수들을 추정하였다. 두 번째 단계는 국립기상연구소에서 제공하는 전 지구 모델 결과로부터 구한 54개 지점의 온실가스 배출 시나리오의 3가지 유형별 미래 일기 예측 자료를 이용하여, 향후 90년간(2011년~2100년)의 모의 일별 일기자료를 산출하였다.

## 2. 일기 자료

우리나라 다중지점 일기 산출 모형을 적합하기 위해, 기상청으로부터 제공받은 전국 54개 강우, 기상관측소의 지리정보시스템(Geographic information system; GIS) 자료를 사용하였다. 본 연구에서 여러 일기자료 중에 일별 강수량(mm)과 평균 기온(°C) 자료를 사용하였다. 관측기간은 1976년부터 2005년까지 30년(10,958일)이며, 해당기간동안 자료들의 결측값은 없다. 또한, 국가 수자원 관리 종합 정보 시

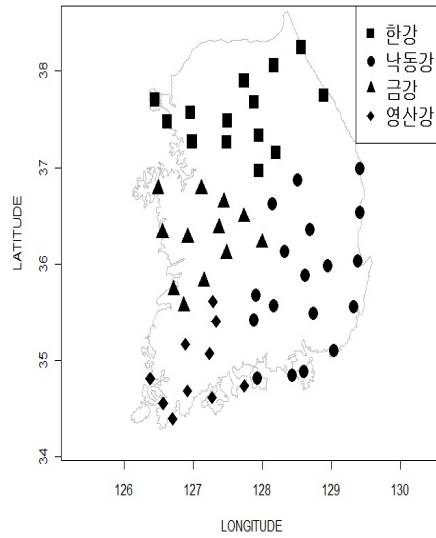


그림 2.1. 우리나라 4대강 유역별 기상 관측소의 위치

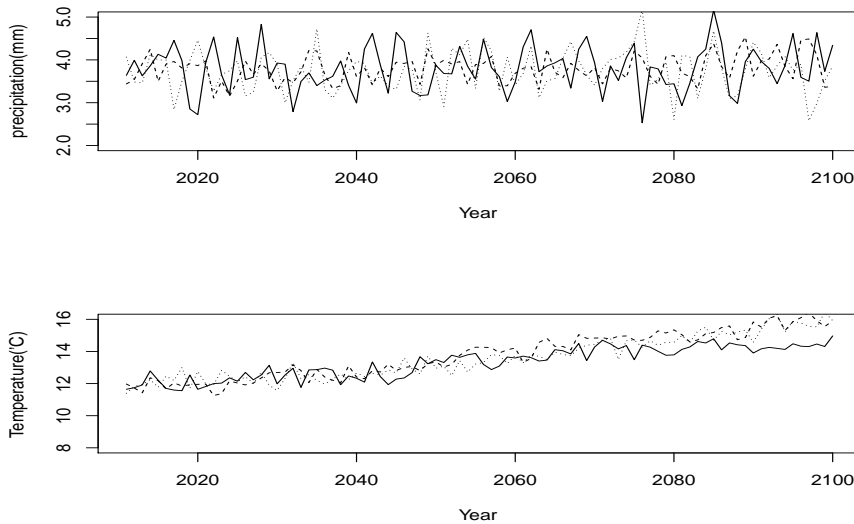


그림 2.2. 온실가스 배출 시나리오에 따른 우리나라 미래의 평균 강수 및 기온 예측치 (B1-실선, A1B-파선, A2-점선)

시스템(WAMIS)을 참고하여 54개 지점을 4대강(한강, 낙동강, 금강, 영산강 (섬진강 포함)) 유역별로 나누었다. 4대강 유역별, 월별 일평균 강수량 및 기온의 평균값은 표 2.1에 정리하였고, 그림 2.1은 우리나라 54개 기상 관측소의 위치와 4대강 유역의 구분을 보여준다.

국립기상연구소는 전 지구 모델로부터 통계적 상세화 기법을 적용하여 기상관측소 54개 지점의 월별 일평균 강수량과 기온의 예측값을 제공하고 있다. 일기 예측값은 온실가스 배출 시나리오에 따라 3가지 유형(B1: 저배출, A1B: 중배출, A2: 고배출)에 따라 다르게 생산된다. 온실가스 배출 시나리오에 따른 향후 90년간(2011년~2100년) 연도별 일평균 강수량과 기온 예측 값의 추세는 그림 2.2와 같다. 그

림 2.2에서 온실가스 배출에 따라 향후 90년간 강수량의 평균은 차이가 거의 없지만, A1B(중배출) 시나리오에 비해 B1(저배출)과 A2(고배출) 시나리오의 경우 연간 변동성이 커짐을 알 수 있다. 평균 기온의 경우 세 가지 시나리오 모두 기온이 상승함을 알 수 있는데, B1(저배출) 시나리오의 경우가 다른 두 시나리오에 비해 기온이 가장 완만하게 상승함을 알 수 있다.

### 3. 방법론

#### 3.1. 다단계 일반화 선형모형과 모수추정

Lee와 Nelder (1996)가 제안한 다단계 일반화 선형모형을 이용하여 다음과 같이 관측된 일기변수들을 모형화 한다. 특정한 연도(year)와 월(month)이 주어졌을 때,  $i$ 번째 관측 지점 ( $i = 1, \dots, K$ )의  $j$ 번째 일(day) ( $j = 1, \dots, T$ )에 관측된 일기 자료(강수 발생 유무, 강수량, 평균 기온 등)를 종속 변수  $y_{ij}$ 라고 하자. 이 때, 변량효과  $u = (u_1, \dots, u_K)^T$ 와  $v = (v_1, \dots, v_T)^T$ 가 주어졌을 때, 종속 변수  $y_{ij}$ 는 조건부 평균  $\mu_{ij}$ 를 갖는 지수족(exponential family)의 분포를 따르고, 다음을 만족한다.

$$g(\mu_{ij}) = x_{ij}^T \beta + u_i + v_j, \quad (3.1)$$

여기서  $g(\cdot)$ 는 연결 함수(link function),  $x_{ij}$ 는 설명 변수(explanatory),  $\beta$ 은 고정효과(fixed parameter)를 나타낸다. 변량효과 벡터  $u = (u_1, \dots, u_K)^T$ 는 평균이 0이고 공분산 행렬이  $\Sigma_u$ 인 다변량 정규 분포를 따른다고 가정한다. 이때, 공간상관성을 반영하기 위해  $\Sigma_u$ 의  $(i_1, i_2)$ 번째 원소  $\sigma_{i_1, i_2}^u$ 를 다음과 같이 모형화 한다.

$$\sigma_{i_1, i_2}^u = \lambda_1 \exp(-\rho_1 d_{i_1 i_2}), \quad (3.2)$$

여기서  $\lambda_1$ 과  $\rho_1$ 은 양수인 분산 모수이고,  $d_{i_1 i_2}$ 는  $i_1$ 번째 지점과  $i_2$ 번째 지점사이의 거리(Euclidean distance)를 뜻한다. 즉, 두 지점간의 거리가 가까울수록, 두 지점의 변량효과  $v_{i_1}$ 와  $v_{i_2}$ 의 공분산 값은 지수적으로 감소하게 됨을 의미한다 (Diggle 등, 1998). 마찬가지로, 변량효과 벡터  $v = (v_1, \dots, v_T)^T$ 는 시간 상관성을 반영하기 위해 1차 자기 상관 회귀 모형 (AR(1))을 따른다고 가정한다. 즉,  $v$ 는 평균이 0이고 공분산 행렬이  $\Sigma_v$ 인 다변량 정규분포를 가정하고,  $\Sigma_v$ 의  $(j_1, j_2)$ 번째 원소  $\sigma_{j_1, j_2}^v$ 를 다음과 같이 모형화 한다.

$$\sigma_{j_1, j_2}^v = \lambda_2 \rho_2^{|j_1 - j_2|}, \quad (3.3)$$

여기서  $\lambda_2$ 는 양수인 분산 모수이고,  $\rho_2$ 은 -1과 1사이의 값을 가지는 분산 모수이다.

모형을 적합하는데 필요한 모수 및 변량효과들을 추정하기 위해, Lee와 Nelder (1996)으로 부터 다음과 같은 다단계 우도(hierarchical likelihood)를 정의한다.

$$h = \sum_{i=1}^K \sum_{j=1}^T \log f_\phi(y_{ij}|u, v) + \log f(u) + \log f(v), \quad (3.4)$$

여기서  $\log f_\phi(y_{ij}|u, v)$ 는 분산모수  $\phi$ 를 가지는  $y_{ij}$ 의 로그 조건부 확률 밀도 함수를 의미하고,

$$\begin{aligned} \log f(u) &= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_u| - \frac{1}{2} u^T \Sigma_u^{-1} u \\ \log f(v) &= -\frac{T}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_v| - \frac{1}{2} v^T \Sigma_v^{-1} v \end{aligned}$$

이다. 위의 모형 (3.1), (3.2), (3.3)에서 추정해야할 대상은 변량효과  $u, v$ , 평균 모수  $\beta$ 와 분산 모수  $\tau = (\phi, \lambda_1, \lambda_2, \rho_1, \rho_2)$ 이다. 또한 Lee와 Nelder (2001)은 다음과 같은 수정된 단면 우도를 정의하였다.

$$p_\alpha(l) = \left[ l - \frac{1}{2} \log \left| \frac{D(l, \alpha)}{2\pi} \right| \right] \Big|_{\alpha=\bar{\alpha}}$$

여기서  $l$ 은 로그 주변 우도 또는 다단계 우도를 의미하고,  $D(l, \alpha) = -\partial^2 l / \partial \alpha \partial \alpha^T$ 이고  $\bar{\alpha}$ 는  $\partial l / \partial \alpha = 0$ 의 해이다. 모형에 필요한 값들을 추정하는데 있어서 Lee와 Nelder (2001)은 변량효과는 식 (3.4)의  $h$ 를 최대화 시키는 값을, 평균 모수인  $\beta$ 는 수정된 단면우도  $p_{u,v}(h)$ 를, 분산 성분인  $\tau$ 는  $p_{u,v,\beta}(h)$ 를 최대화 시키는 값으로 구하였다. 본 연구에서는 필요한 모수들 추정하기 위해 통계패키지 R의 library(dhglm)을 사용하였다.

### 3.2. 강수 발생 모형

많은 모수적 일기산출기들은 Richardson의 WGEN 모형과 같이 강수량 과정을 강수 발생 유무(occurrence)와 강수량(amount) 두 가지의 과정으로 분리해서 모형화 한다. Furrer와 Katz (2007)에서는 강수 발생 과정에서는 logit 연결 함수를 가지는 이항분포 일반화선형모형을 적합시켜서 모형화 하였다. 하지만 시공간 상관성을 고려한 다중지점 일기산출을 위해 Lee와 Nelder (1996)에서 제안한 다음과 같이 다단계 일반화 선형 모형으로 강수 발생을 모형화 하도록 한다.

특정한 연도(year)와 월(month)이 주어졌을 때,  $i$ 번째 관측 지점 ( $i = 1, \dots, K$ )의  $j$ 번째 시점(day) ( $j = 1, \dots, T$ )에 비가 오면 1, 오지 않으면 0의 값을 가지는 강수 유무를 나타내는 이진 관측 자료를  $y_{ij}^O$ 라고 하자. 이 때, 변량효과  $u^O = (u_1^O, \dots, u_K^O)^T$ 와  $v^O = (v_1^O, \dots, v_T^O)^T$ 가 있어서, 종속 변수  $y_{ij}^O$ 는 조건부 확률이  $\mu_{ij}^O$ 인 이항 분포 Bernoulli( $\mu_{ij}^O$ )를 따르고, 식 (3.1)에서의 연결함수로 로짓(logit) 연결 함수를 사용하여 모형화 한다:

$$\text{logit}(\mu_{ij}^O) = \log \frac{\mu_{ij}^O}{1 - \mu_{ij}^O} = \beta^O + u_i^O + v_j^O, \tag{3.5}$$

여기서  $\beta^O$ 은 고정 효과(fixed parameter)를 나타낸다. 공간 상관성을 나타내는 변량효과  $u^O$ 와 시간 상관성을 나타내는  $v^O$ 에 대한 모형은 식 (3.2)와 (3.3)과 같은 공분산을 가지는 다변량 정규 분포를 따른다고 가정한다.

### 3.3. 강수량 모형

앞 절에서와 같이  $i$ 번째 관측 지점의  $j$ 번째 시점에 관측된 강수량 자료(precipitation amounts)를  $y_{ij}^A$ 라고 하자. 또한 양수 값을 가지는 강수량 자료만을 모형화 하기위해,  $\mathcal{A} = \{(i, j); y_{ij}^O = 1\}$ 이라고 정의 하자. 이번 절에서는  $(i, j) \in \mathcal{A}$ 인 강수량 자료  $y_{ij}^A$ 만을 고려한다. 변량효과  $u^A = (u_1^A, \dots, u_K^A)^T$ 와  $v^A = (v_1^A, \dots, v_T^A)^T$ 가 주어졌을 때,  $y_{ij}^A$ 는 조건부 평균이  $\mu_{ij}^A$ 인 감마 분포를 따르고, (3.1)의 연결함수로 로그(log) 연결함수를 사용하였다:

$$\log(\mu_{ij}^A) = \beta^A + u_i^A + v_j^A,$$

여기서  $\beta^A$ 은 고정 효과(fixed parameter)를 나타낸다. 공간 상관성을 나타내는 변량효과  $u^A$ 와 시간 상관성을 나타내는  $v^A$ 에 대한 모형은 식 (3.2)와 (3.3)과 같은 공분산을 가지는 다변량 정규 분포를 따른다고 가정한다.

### 3.4. 평균 기온

앞 절에서와 같이  $i$ 번째 관측 지점의  $j$ 번째 시점에 관측된 일평균 기온 자료를  $z_{ij}$ 라고 하자. 변량효과  $u^z = (u_1^z, \dots, u_K^z)^T$ 와  $v^z = (v_1^z, \dots, v_T^z)^T$ 가 주어졌을 때,  $z_{ij}$ 는 조건부 평균이  $\mu_{ij}^z$ 인 정규 분포를 따르고, 식 (3.1)의 연결함수로 항등(identity) 연결함수를 사용하였다:

$$\mu_{ij}^z = \beta_0^z + \beta_1^z y_{ij}^O + u_i^z + v_j^z, \quad (3.6)$$

여기서  $\beta_0^z, \beta_1^z$ 은 고정 효과(fixed parameter)를 나타낸다. 평균 기온은 강수량의 유무에 크게 영향을 받음이 알려져 있으므로 (Furrer와 Katz, 2007), 평균 모형에 강수량 유무 변수인  $y_{ij}^O$ 도 설명변수로 사용하였다. 공간 상관성을 나타내는 변량효과  $u^z$ 와 시간 상관성을 나타내는  $v^z$ 에 대한 모형은 식 (3.2)와 (3.3)과 같은 공분산을 가지는 다변량 정규 분포를 따른다고 가정한다.

### 3.5. 제안된 일기산출 모형 알고리즘

미래 일기 시나리오 정보를 반영한 제안하는 다중 지점 일기산출의 과정은 다음과 같다.

단계 1: 산출하려는 연도(year)와 월(month)을 정한다. 해당 기간의 예측된 미래 일기 시나리오의  $K$ 개 관측 지점의 일평균 강수량 벡터를  $y^P = (y_1^P, \dots, y_K^P)^T$ 라고 하자. 1976년의 해당 월에 해당하는 기간 동안 관측된 강수량 자료의 일평균 강수량을 다음과 같이 구한다.

$$\bar{y}_{1976} = \left( \frac{1}{T} \sum_{j=1}^T y_{1j}, \dots, \frac{1}{T} \sum_{j=1}^T y_{Kj} \right)^T.$$

위와 같은 방법으로 1977년부터 2005년까지에 해당하는 관측된 강수량 자료의 일평균 강수량  $\bar{y}_{1977}, \dots, \bar{y}_{2005}$ 를 구한다.  $\bar{y}_k$ 와  $y^P$ 와 거리(Euclidean distance)  $r_k = \|\bar{y}_k - y^P\|_2$ 를 구한다.  $r_k$  값이 작은 순서대로  $k$ 값을  $k_1, k_2, k_3$ 라고 하였을 때, 이 세 가지 연도 가운데 랜덤하게 한 연도를 선택한다.

단계 2: 단계 1에서 구한 연도에 해당하는 관측된 강수량 자료를 가지고, 3.2절, 3.3절을 통해 필요한 모수들을 추정한다.

단계 3: 단계 1과 같은 방법으로 미래 일기시나리오에 해당하는 일평균 기온 벡터와 거리가 가장 가까운 값을 가지는 3개의 일평균 기온 자료로부터 랜덤하게 한 연도를 선택하고, 3.4절로부터 필요한 모수들을 추정한다.

단계 4: 단계 2에서 추정된 모수들을 이용하여 강수 발생 유무에 대한 가상 자료를 생성한다. 즉, 식 (3.2), (3.3)으로부터 추정된 공분산 행렬  $\hat{\Sigma}_u, \hat{\Sigma}_v$ 를 이용하여 만든 두 다변량 정규 분포  $N(0, \hat{\Sigma}_u)$ 와  $N(0, \hat{\Sigma}_v)$ 로부터 각각  $u^O$ 와  $v^O$ 를 랜덤하게 생성(generation) 한다. 생성된  $u^O$ 와  $v^O$  및 추정된  $\beta^O$ 를 가지고 식 (3.5)을 이용하여, 모든  $i = 1, \dots, K, j = 1, \dots, T$ 에 대하여, 조건부 확률  $\mu_{ij}^O$ 를 구한다. 조건부 확률  $\mu_{ij}^O$ 인 이항분포로부터 강수 발생 유무 변수  $y_{ij}^O$ 들을 랜덤하게 모두 산출한다.

단계 5: 단계 4와 유사하게, 추정된 모수 및 랜덤하게 생성한 변량효과를 가지고, 식 (3.6)으로부터 구한 조건부 평균  $\mu_{ij}^A$ 를 가지는 감마 분포에서 강수량  $y_{ij}^A$ 들을 랜덤하게 생성한다. 이때,  $y_{ij}^A$ 는 산출된 강수 발생 유무 변수  $y_{ij}^O = 1$ 인  $(i, j)$ 들에 대해서만 산출한다.

단계 6: 단계 3에서 추정된 모수 및 변량효과와 단계 4에서 생성된 강수 발생 변수  $y_{ij}^O$ 들을 가지고, 식 (3.6)을 통해 모든  $i = 1, \dots, K, j = 1, \dots, T$ 에 대하여, 조건부 평균  $\mu_{ij}^z$ 를 구한다. 구한  $\mu_{ij}^z$ 과 추정된  $\phi$ 를 분산으로 가지는 정규분포  $N(\mu_{ij}^z, \phi)$ 로부터 일평균 기온 변수  $z_{ij}$ 들을 랜덤하게 모두 산출한다.

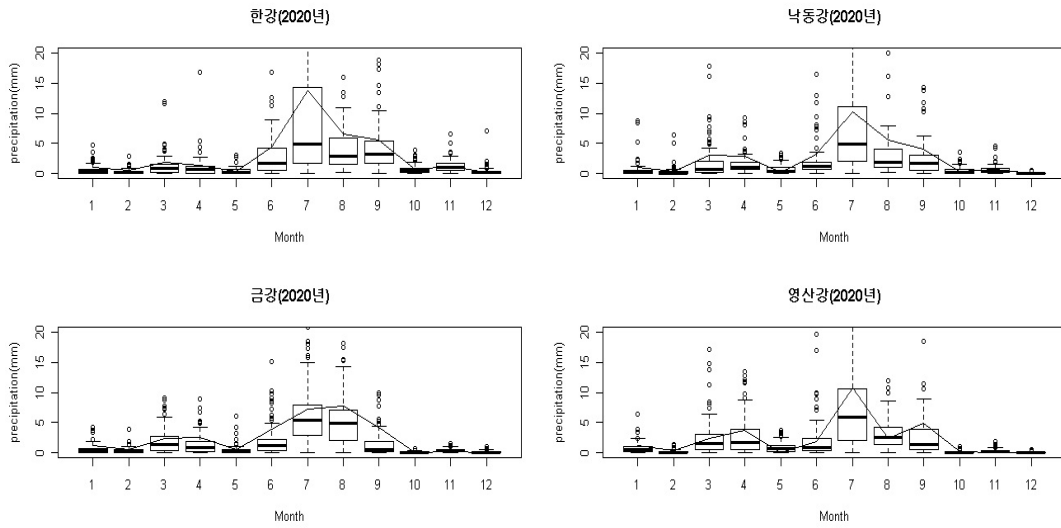


그림 4.1. 4대강 유역별 미래 일 강수 예측치 (2020년, 온실가스 배출 시나리오 B1 적용)

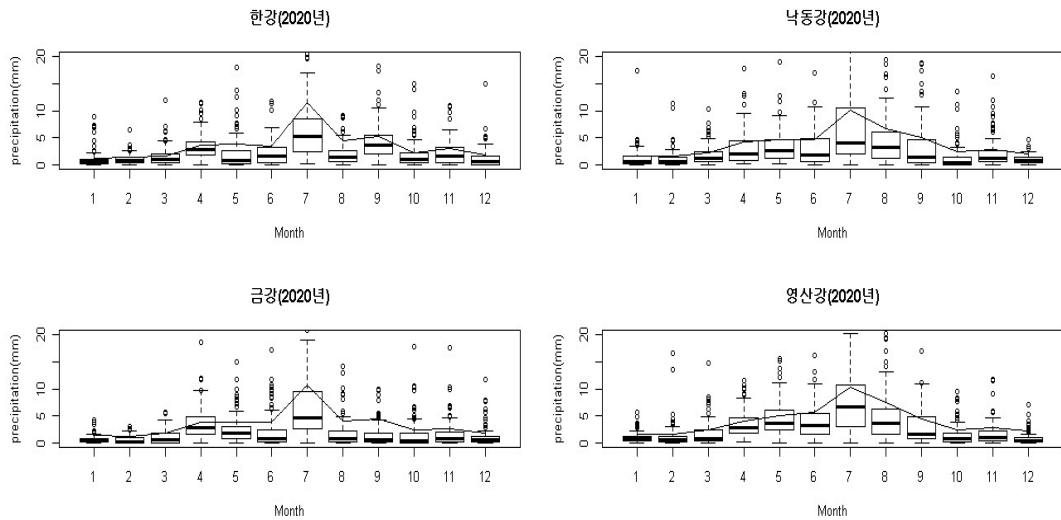


그림 4.2. 4대강 유역별 미래 일 강수 예측치 (2020년, 온실가스 배출 시나리오 A1B 적용)

#### 4. 실증 예제

3.5절에서 제안한 알고리즘 절차를 따라서, 우리나라 4대강 유역별로 2011년부터 2100년까지 90년간 (32,872일)의 일별 강수량 및 평균 기온 예측값을 산출하였다. 또한 3가지 온실가스 배출 시나리오에 따라 일기 변수 예측값들을 산출하였으며, 각각 100번씩 반복하여 산출하였다. 예시로 그림 4.1~4.3에 2020년 1년간에 해당하는 일평균 강수량 산출 예측값들의 상자그림과 함께 전지구 모델에서 제공하는 월별 일평균 강수 예측값(실선)을 나타내었다. 또한 그림 4.4~4.6에 2020년 1년간에 해당하는 일평균 기온 산출 예측값들의 상자그림과 전지구 모델에서 제공하는 월별 일평균 기온 예측값(실선)을 나타내

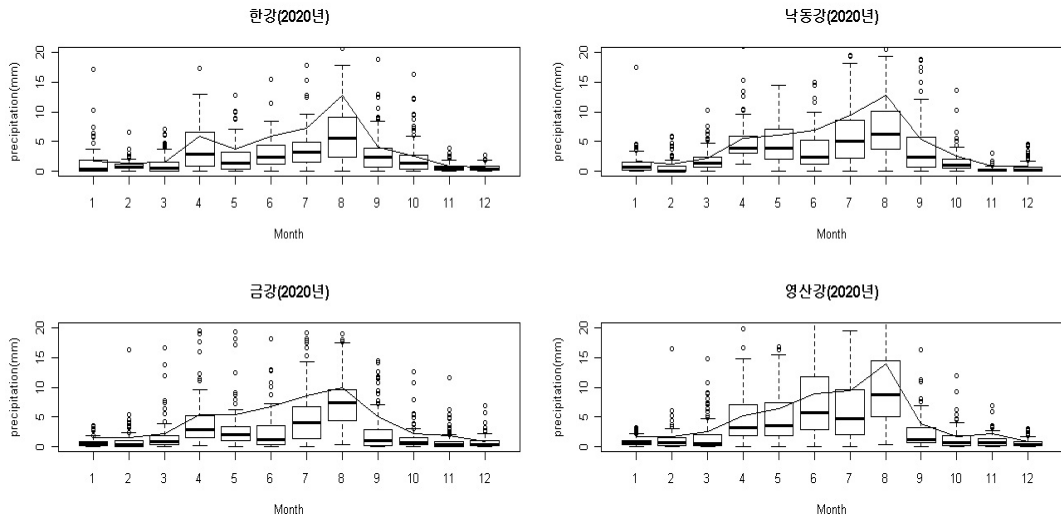


그림 4.3. 4대강 유역별 미래 일 강수 예측치 (2020년, 온실가스 배출 시나리오 A2 적용)

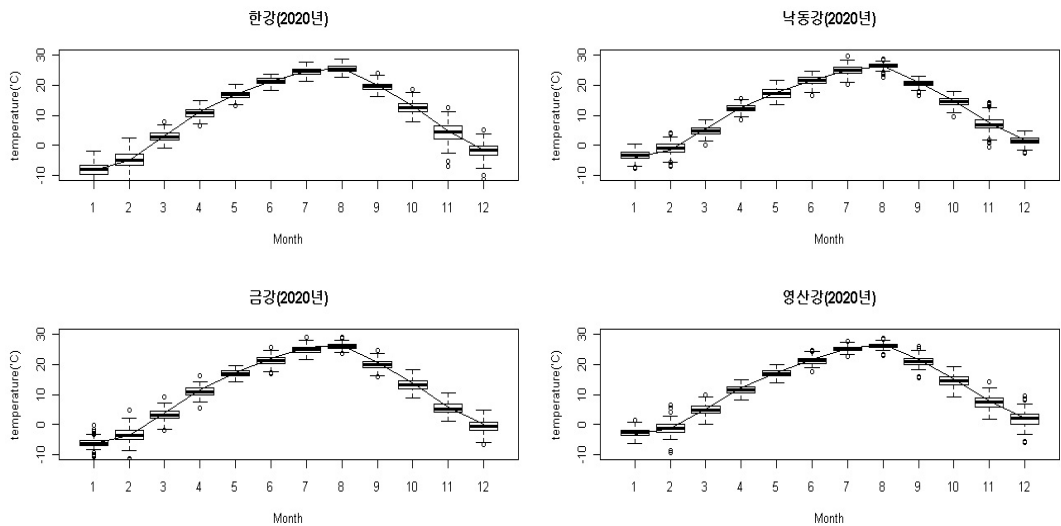


그림 4.4. 4대강 유역별 미래 일 기온 예측치 (2020년, 온실가스 배출 시나리오 B1 적용)

었다. 일평균 강수량의 경우 모든 시나리오에서 제안된 일기산출 모형으로부터 산출된 강수량 예측값들이 여름철 기간 동안에는 전지구 모델자료의 예측값에 비해 과소 산출하는 경향을 보이는 하지만, 대체로 전 지구 모델의 결과에 부합하는 것을 알 수 있다. 일평균 기온의 경우 모든 시나리오에서 제안된 일기산출 모형으로부터 산출된 기온 예측값들이 전 지구 모델의 결과와 매우 비슷한 양의 값들을 산출함을 알 수 있다. 이는 본 연구에서 제안된 일기산출 모형이 미래 온실가스 배출 시나리오의 정보를 잘 반영한다고 볼 수 있다.

그림 4.7은 온실가스 배출 시나리오별로 54개 지점에서 산출한 일기 산출값을 가지고 서울의 일기 산출값과 다른 53개 관측지점에서의 일기 산출값의 표본 상관계수(sample correlation)을 한반도 지도상에



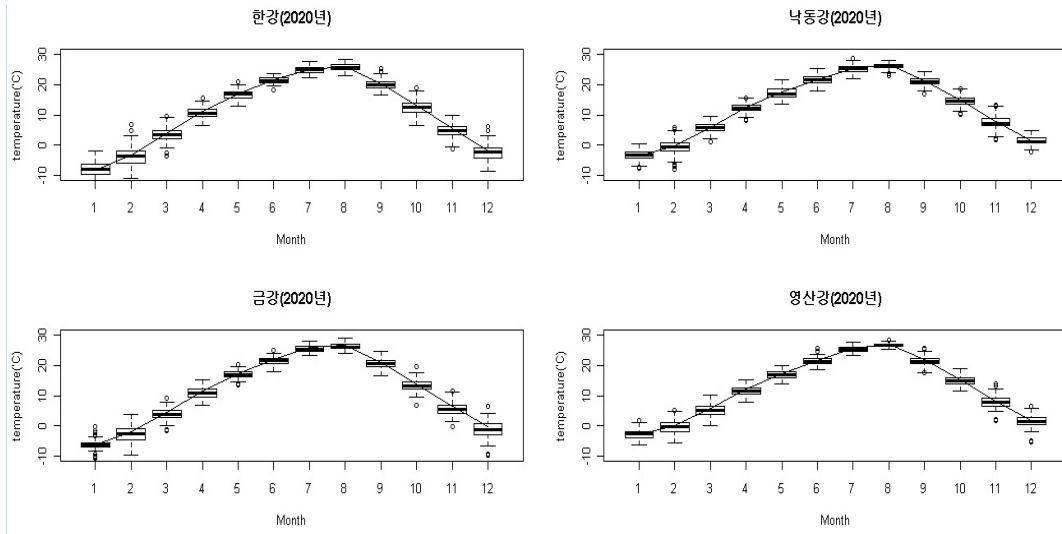


그림 4.5. 4대강 유역별 미래 일 기온 예측치 (2020년, 온실가스 배출 시나리오 A1B 적용)

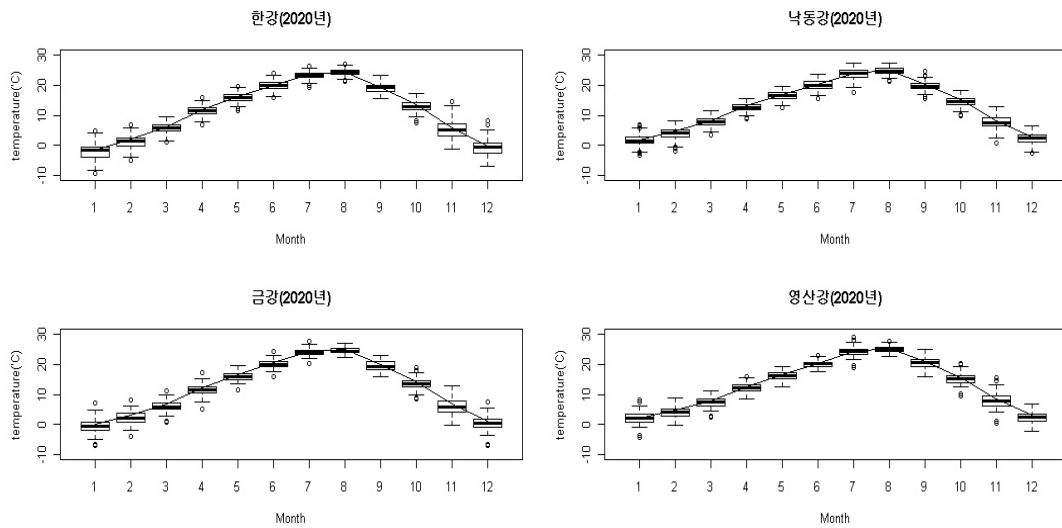


그림 4.6. 4대강 유역별 미래 일 기온 예측치 (2020년, 온실가스 배출 시나리오 A2 적용)

표현한 것이다. 세 가지 시나리오에 따른 결과는 큰 차이를 보이지 않았으며, 강수량의 경우 서울과 인접한 지역에서만 높은 상관성을 보이지만, 거리가 먼 지역과는 큰 상관성이 없는 것으로 확인된다. 그러나 기온의 경우 우리나라 전역에서 서울과 높은 양의 상관관계를 보임을 알 수 있다.

5. 결론

본 연구에서는 시공간 상관성이 있는 우리나라 과거 일기 자료들을 다단계 일반화 선형 모형을 이용하여 모형화 하였다. 또한 우리나라 실제 관측 일기 자료를 가지고 구축한 통계 모형과 미래 온실가스 배출

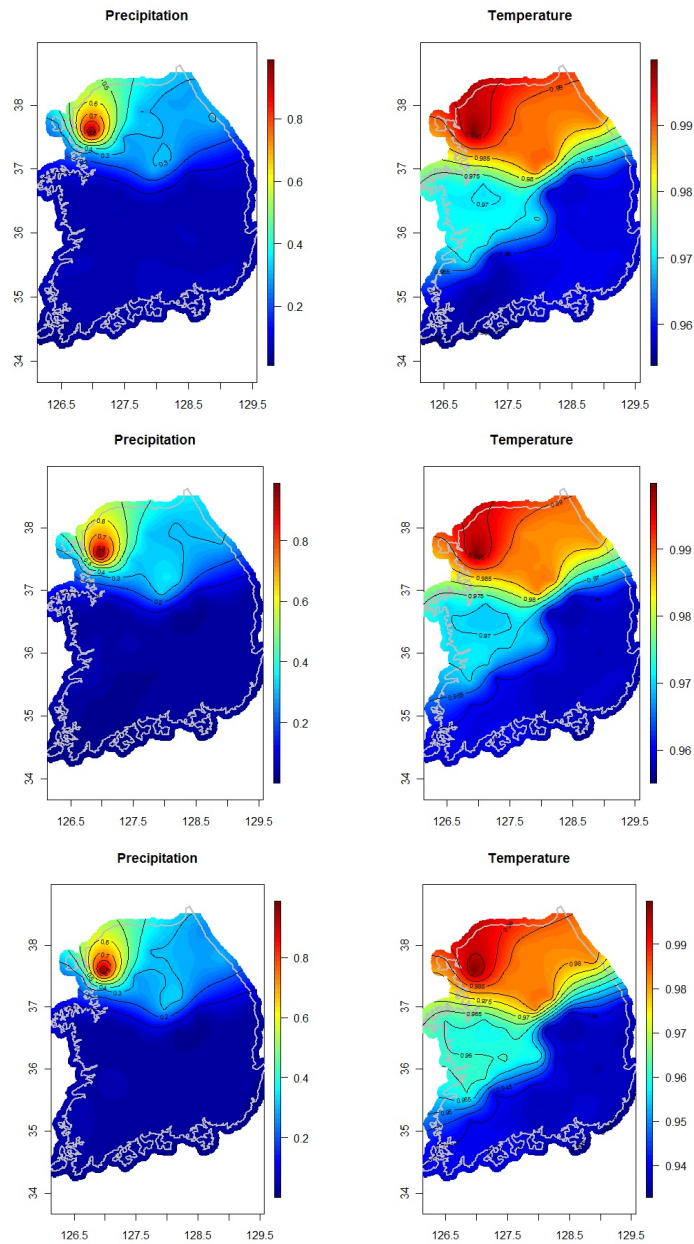


그림 4.7. 2020년의 서울과 다른 관측지점에서의 강수 예측값의 상관도(좌)와 기온 예측값의 상관도(우) (위에서부터 온실가스 배출 시나리오 B1, A1B, A2 적용)

시나리오 정보를 함께 이용하여 온실가스 배출 정도에 따른 우리나라의 향후 90년간의 미래 강수량, 기온 예측값들의 일별 자료 산출이 가능해졌다. 본 연구에서 제안된 일기산출 모형은 전 지구 모델로부터 구한 시나리오 정보는 월단위의 강수량, 기온 예측값만 사용 되었음에도, 과거 일별(daily) 자료들과 가

장 가까운 자료를 검색함으로써, 일 단위의 미래 평균 강수량과 기온 예측값을 제공해준다는 점에서 향후 한반도 기후모델 연구 또는 수자원 개발과 같은 다양한 연구 분야에서 모의실험의 입력자료로써 사용될 수 있다.

## 참고문헌

- 강문성, 박승우, 진영민 (1998). 기상자료 미계측 지역의 추계학적 기상발생모델, <한국농공학회지>, **40**, 57-67.
- Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998). Model-based geostatistics, *Applied Statistics*, **47**, 299-350.
- Furrer, E. M. and Katz, R. W. (2007). Generalized linear modeling approach to stochastic weather generators, *Progress in Physical Geography*, **23**, 329-357.
- Jang, M. J., Lee, Y., Lawson, A. B. and Browne, W. J. (2007). A comparison of the hierarchical likelihood and Bayesian approaches to spatial epidemiological modelling, *Environmetrics*, **18**, 809-821.
- Kim, T., Ahn, H., Chung, G. and Yoo, C. (2008). Stochastic multi-site generation of daily rainfall occurrence in south Florida, *Stochastic Environmental Research and Risk Assessment*, **22**, 705-717.
- Lee, D., An, H., Lee, Y., Lee, J., Lee, H-S. and Oh, H-S. (2010). Improved multisite stochastic weather generation with applications to historical data in South Korea, *Asia-Pacific Journal of Atmospheric Sciences*, **46**, 497-504.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion), *Journal of the Royal Statistical Society B*, **58**, 619-678.
- Lee, Y. and Nelder, J. A. (2001). Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions, *Biometrika*, **88**, 987-1006.
- Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society A*, **135**, 370-384.
- Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources Research*, **17**, 182-190.
- Richardson, C. W. and Wright, D. A. (1984). WGEN: A model for generating daily weather variables, *US Department of Agriculture*, (ARS-8).
- Wilks, D. S. (1998). Multisite generalizations of a daily stochastic precipitation generation model, *Journal of Hydrology*, **210**, 178-191.
- Wilks, D. S. and Wilby, R. L. (1999). The weather generation game: A review of stochastic weather models, *Progress in Physical Geography*, **23**, 329-357.

# Future Weather Generation with Spatio-Temporal Correlation for the Four Major River Basins in South Korea

Donghwan Lee<sup>1</sup> · Jaeyong Lee<sup>2</sup> · Hee-Seok Oh<sup>3</sup> · Youngjo Lee<sup>4</sup>

<sup>1</sup>Department of Statistics, Seoul National University

<sup>2</sup>Department of Statistics, Seoul National University

<sup>3</sup>Department of Statistics, Seoul National University

<sup>4</sup>Department of Statistics, Seoul National University

(Received December 12, 2011; Revised January 25, 2012; Accepted January 31, 2012)

---

## Abstract

Weather generators are statistical tools to produce synthetic sequences of daily weather variables. We propose the multisite weather generators with a spatio-temporal correlation based on hierarchical generalized linear models. We develop a computational algorithm to produce future weather variables that use three different types of green-house gases scenarios. We apply the proposed method to a daily time series of precipitation and average temperature for South Korea.

**Keywords:** Weather generator, Hierarchical generalized linear models, spatio-temporal correlation.

---

---

This work was funded by the Korea Meteorological Administration Research and Development Program under Grant RACS.2010-4010.

<sup>4</sup>Corresponding author: Professor, Department of Statistics, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea. E-mail: youngjo@snu.ac.kr