

Multi-Site Stochastic Weather Generator for Daily Rainfall in Korea

Minjung Kwak^a · Yongku Kim^{b,1}

^aDepartment of Statistics, Yeungnam University

^bDepartment of Statistics, Kyungpook National University

(Received March 7, 2014; Revised April 28, 2014; Accepted May 8, 2014)

Abstract

A stochastic weather generator based on a generalized linear model (GLM) approach is a commonly used tools to simulate a time series of daily weather. In this paper, we propose a multi-site weather generator with applications to historical data in South Korea. The proposed method extends the approach of Kim *et al.* (2012) by considering spatial dependence in the model. To reduce this phenomenon, we also incorporate a time series of seasonal mean precipitations of South Korea in the GLM weather generator as a covariate. Spatial dependence was incorporated into the model through a latent Gaussian process. We apply the proposed model to precipitation data provided by 62 stations in Korea from 1973–2011.

Keywords: Daily precipitation, generalized linear model, multisite stochastic weather generator, spatial process, overdispersion.

1. 서론

Stochastic weather generator는 비교적 큰 규모의 기상이나 기후 정보를 이용하여 상대적으로 작은 규모의 일일 평균기온이나 일일 강우량을 생성해 내는 통계적 모형이다. 또한 이 모형은 기상 및 기후 정보를 날씨 정보로 downscaling하는 통계기법중 하나이기도 하다 (Wilks와 Wilby, 1999). 하지만 이 모형을 통해서 특정한 날의 날씨를 추정하는 것은 거의 불가능하므로 주로 생성된 날씨를 이용한 계절별 특성치 통계량이 실제 관측된 계절별 통계량과 일치하도록 모형을 구성한다 (Maraun 등, 2010; Wilks, 2010).

최근에는 이러한 Stochastic weather generator를 위해서 일반화선형모형(Generalized linear model; GLM)이 많이 고려되는데, 이 모형은 기후의 계절별 특성을 설명하고 일일 날씨에 영향을 미칠 수 있는 여러가지 현상 (예, 엘니뇨 현상, 지구온난화 현상 등)을 연관지어 설명하는데 매우 유용하다 (McCullagh와 Nelder, 1989; Furrer와 Katz, 2007). 하지만 이러한 GLM에 기초한 Stochastic weather generator는 모형을 통해 얻어진 계절별 특성값들의 분산이 실제 관측된 계절별 특성값들의 분산보다

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning(No. 2012-011113).

¹Corresponding author: Department of Statistics, Kyungpook National University, 80 Daehakro, Bukgu, Daegu 702-701, Korea. E-mail: kim.1252@knu.ac.kr

상대적으로 작게 추정되는 문제점을 가지고 있다 (Buishand, 1978; Katz와 Parlange, 1998). 이러한 현상을 overdispersion이라고 하는데 최근에는 GLM을 이용한 weather generator가 관측값 만큼의 충분히 큰 기후 변동을 만들어 내도록 하는 것이 상당히 중요한 이슈가 되어왔다 (Katz와 Zheng, 1999).

Kim 등 (2012)은 이러한 overdispersion 현상을 줄이기 위해 기존의 GLM weather generator에 계절별 총 강우량과 계절별 평균 강우량을 공변수로 사용하였다. 즉 계절별 특성치를 이용한 큰 규모의 경향성을 모형에 포함함으로써 추가적인 변동을 모형에 소개하였다. 단, 계절별 특성치 통계량을 그대로 모형에 포함할 경우 필요이상의 큰 기후 변동을 발생시키는 underdispersion 현상을 막기 위해서 계절별 통계량을 smoothing 처리하여 모형에 사용하였다. 본 논문에서는 LOESS(locally weighted scatterplot smoothing; Cleveland, 1979; Hastie와 Tibshirani, 1990) 기법을 이용하여 계절별 통계량을 smoothing 처리하였다. 참고로 다른 여러가지 smoothing 기법이 고려될 수도 있다. Wilks (1989)는 일일 강우량에 관한 모형에 월별 강우량을 이용하였고 Hansen와 Mavromatis (2001)는 overdispersion 현상을 줄이기 위해서 GLM weather generator의 계수를 임의로 조정하였다.

지점 정보에 기초한 Stochastic weather generator는 그 지점에 대한 날씨의 특징은 잘 표현하지만 여러 지역의 날씨를 동시에 생성하는데에는 한계가 있다. 특히 날씨의 공간적인 상관관계를 설명하기에 여전히 충분하지 않다. 이러한 한계를 극복하기 위해서 본 논문에서는 공간구조를 가지는 Stochastic weather generator를 소개하였다. 여기에서는 공간구조를 기존 모형에 적용하는 여러가지 방법 중 다층구조를 통해 Stochastic weather generator의 모형계수에 공간구조를 소개하는 방법을 제안하였다. 그리고 제안된 모형을 우리나라의 62개 지역의 일일 강우량에 적용하였다. 모형 적합을 위해서 1973년부터 2011년까지 39년 동안의 우리나라 각 지역의 일일 강우량 자료가 사용되었고, overdispersion 현상을 줄이기 위해 앞서 소개한 방법과 유사한 기법이 사용되었다. 본 논문은 다음과 같이 구성된다.

2장에서는 일일 강우량에 관한 GLM weather generator를 간단히 소개하였다. 여기에서 overdispersion 현상을 보완하기 위한 계절별 평균 강우량을 공변수로 사용되었다. 그리고 3장에서는 다층구조를 통한 공간구조를 가지는 GLM weather generator를 보여주고, 4장에서는 우리나라 62개 지점에서 관측된 일일 강우자료에 제안된 모형의 적용하여 모형의 적합성을 검증하였고, 마지막으로 5장에서 이 논문에서 제안한 방법의 결론을 제시하며 좀 더 보완해야 할 문제에 관해 논의하며 마무리 하겠다.

2. 일일 강우량에 대한 GLM weather generator

Furrer와 Katz (2007)에 의해서 소개된 일반화 선형모형에 기초한 stochastic weather generator는 Richardson (1981)이 처음으로 제안한 모형 중 가장 간단한 형태이다. 더 자세한 내용은 Furrer와 Katz (2007)나 웹사이트 <http://www.image.ucar.edu/~eva/GLMwgen/>를 참조하면 된다. Furrer와 Katz (2007)에 의해 소개된 강우 발생과 강우 강도에 관한 GLM weather generator는 궁극적으로는 Stern과 Coe (1984)의 모형과 같은 것인데 그들은 모형의 모수가 일년 주기를 가지는 chain-dependent process를 이용하여 일일 강우량에 관한 GLM 모형을 만들었다. 앞에서 잠깐 언급한 바와 같이 downscaling 기법은 영향평가에 사용되는 보다 현실적인 날씨를 생성하기 위한 효과적인 방법 중 하나이다. 규모가 큰 기상현상을 어떻게 규모가 작은 날씨와 연관시키는지에 관한 의문이 서로 다른 두 현상간의 관계를 이해하는 기회를 제공한다. 특히 통계적 downscaling은 큰 규모의 기상현상과 작은 규모의 날씨 사이에 밀접한 관계가 있다는 가정에서 시작한다. Kim 등 (2012)은 LOESS smoothing된 계절별 기상 통계치를 기존의 GLM weather generator에 추가변수로 고려함으로써 일일 날씨 통계량에 충분한 불확실성을 넣어 “overdispersion” 현상을 제거하였다. 본 논문에서는 제안된 모형을 통해 downscaling 기법을 GLM weather generator에 자연스럽게 도입하고자 한다.

이 접근법은 큰 규모의 설명변수를 작은 규모의 종속변수에 연결하는 것에 중점을 두고 있는데 예를 들어, 큰 규모의 기후나 대류현상 등을 나타내는 여러가지 지수가 일일 강우모형에서 설명변수로 사용된다. 본 논문에서 제안된 GLM weather generator는 계절별 총 강우량을 이용하여 작은 규모의 일일 강우량으로 분해하는 역할과 동시에 overdispersion을 해결하였다. 다음은 일일 강우발생과 강우량 생성 모형이다.

$J_t(s)$ 를 s 지점에서 t 시점의 강우발생 상태를 나타낸다고 하자 (즉, 비가 오면 $J_t(s) = 1$, 그렇지 않으면 $J_t(s) = 0$ 이 된다). $p_t(s)$ 는 어떤 특정 s 지점에 비가 올 확률 ($p_t(s) = P\{J_t(s) = 1\}$)이고, $1 - p_t(s) = P\{J_t(s) = 0\}$ 이 된다)을 정의한다. 그리고 어떤 상태가 바로 이전의 상태에 영향을 받는 1차 마코프 체인(Markov chain)과 같이, 어느 특정일에 비가 올 확률은 그 전날의 강우발생 유무 $J_{t-1}(s)$ 에 영향을 받는다. 따라서 s 지점의 t 시점에서의 강우발생에 대한 모형은 다음과 같이 표현할 수 있다.

$$\ln\left(\frac{p_t(s)}{1 - p_t(s)}\right) = \alpha(s) + \beta_1(s)J_{t-1} + \beta_2(s)C_t + \beta_3(s)S_t + \beta_4(s)C_t J_{t-1}(s) + \beta_5(s)S_t J_{t-1}(s) + \beta_6(s)I_t P_t^S(s) + \beta_7(s)(1 - I_t)P_t^W(s). \quad (2.1)$$

강우강도에 관한 모형은 강우가 발생했다는 가정(즉, $J_t(s) = 1$)하에서 평균을 $\mu_t(s)$ 로 가지는 감마분포를 가정하는데, 이 평균 강우강도에 대한 모형은 다음과 같다.

$$\ln(\mu_t(s)) = \alpha(s) + \gamma_1(s)C_t + \gamma_2(s)S_t + \gamma_3(s)I_t P_t^S(s) + \gamma_4(s)(1 - I_t)P_t^W(s), \quad (2.2)$$

여기에서 $C_t = \cos(2\pi(t - 181)/365)$ 이고 $S_t = \sin(2\pi(t - 181)/365)$ 이다. $\beta_1(s)$ 은 그 전날의 강우발생 유무에 따라 그 날의 강우발생 확률이 어떻게 달라지는지를 나타내고, $\beta_2(s)$ 와 $\beta_3(s)$ 는 계절에 따른 강우발생 차이를 고려하여 발생할 확률에 대한 주기를 나타내는 SINE 곡선의 위상과 진폭을 결정한다. 그리고 $\beta_4(s)$ 와 $\beta_5(s)$ 는 이러한 주기 효과가 그 전날의 강우발생 유무에 영향을 받도록 허용해 준다. $P_t^S(s)$ 와 $P_t^W(s)$ 는 s 지점의 연평균 강우량을 LOESS기법을 적용한 계절별 특성치이다. 앞에서 언급했듯이 좀더 현실적인 일일강우량을 생성하기 위해 추가한 설명변수로서 downscaling역할과 동시에 overdispersion 문제를 해결하는 변수이다. I_t 는 계절별 지시변수이다. 즉, I_t 가 1이면 우기(4월-9월)의 특성치 $P_t^S(s)$ 가 모형에 영향을 미치고, I_t 가 0이면 건기(10월-3월)의 $P_t^W(s)$ 가 적용된다. 그리고 계수 $\gamma_1(s)$ 와 $\gamma_2(s)$ 는 평균 강우량에 대한 주기를 나타내는 SINE 곡선의 위상과 진폭을 결정한다. 참고로, 분석의 간편성을 위해서 모든 지역의 강우모형에 smoothing 모수로서 0.3을 사용하였다. 또한 추가 분석에서 smoothing 모수의 변화에 따라 overdispersion 결과가 robust함을 확인하였다.

Figure 2.1은 1961년부터 2011년까지 51년 동안의 서울지역의 계절별 총강우량은 우기와 건기에 따라 각각 LOESS smoothing한 결과를 보여준 것이다. 참고로 LOESS 기법은 선형회귀모형 추정의 간편함과 비선형모형 추정의 유연함을 동시에 가지는 smoothing 기법 중 하나인데 특히 이상치에 대한 robust한 강점을 가진 것으로 알려져있고 계산적인 측면에서도 강점을 가지고 있다.

3. 공간구조를 가지는 GLM weather generator

2장에서 소개된 기존의 GLM weather generator에 공간구조를 적용하기 위해서 베이지안 모형에서 많이 사용되는 다층구조를 이용하였다. 이를 위해 Stochastic weather generator의 모형계수에 공간구조를 가지는 다변량 정규분포를 가정하였다. 공간모형을 고려할 때는 관측자료가 없는 지점에 대한 예측이 필요하므로 식 (2.1)-(2.2)에서 해당 지역의 강우량을 LOESS기법을 적용한 계절별 특성치로 사용하

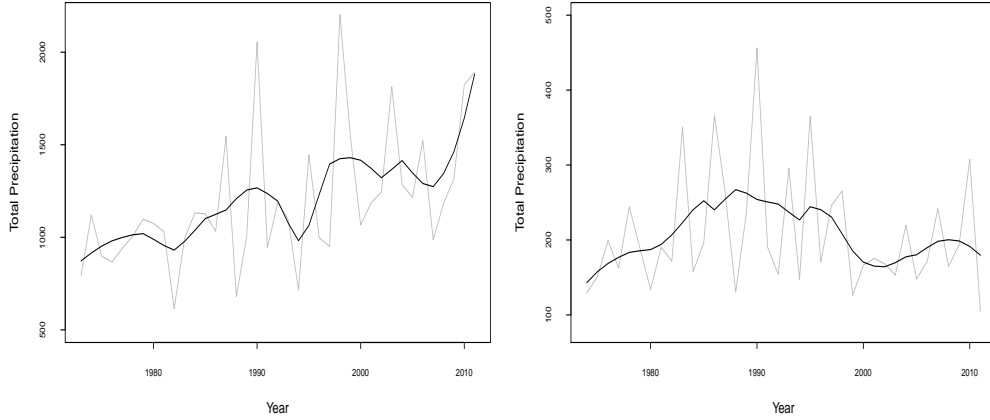


Figure 2.1. Optimal smoothed aggregated covariates of seasonal total precipitation during summer (left) and winter (right) seasons at Seoul. Dashed lines: corresponding observed values of the data series

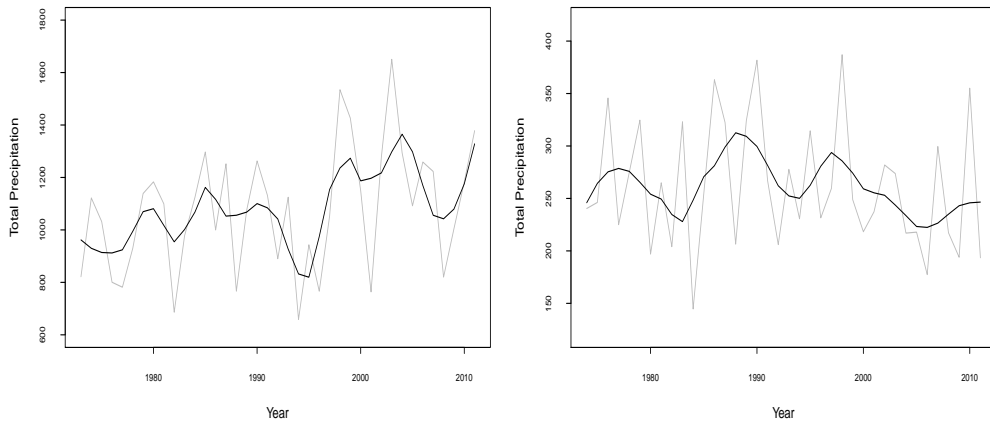


Figure 3.1. Optimal smoothed aggregated covariates of global seasonal total precipitation during summer (left) and winter (right) seasons. Dashed lines: corresponding observed values of the data series

는 대신 전국평균 강우량을 LOESS처리한 계절별 특성치 P_t^S 와 P_t^W 를 모형에 적용하였다.

$$\ln\left(\frac{p_t(s)}{1-p_t(s)}\right) = \alpha(s) + \beta_1(s)J_{t-1}(s) + \beta_2(s)C_t + \beta_3(s)S_t + \beta_4(s)C_tJ_{t-1}(s) + \beta_5(s)S_tJ_{t-1}(s) + \beta_6(s)I_tP_t^S + \beta_7(s)(1-I_t)P_t^W, \quad (3.1)$$

$$\ln(\mu_t(s)) = \delta(s) + \gamma_1(s)C_t + \gamma_2(s)S_t + \gamma_3(s)I_tP_t^S + \gamma_4(s)(1-I_t)P_t^W. \quad (3.2)$$

Figure 3.1는 1961년부터 2011년까지 51년 동안의 전국평균 계절별 총강우량을 우기와 건기에 따라 각각 LOESS smoothing한 결과를 보여준 것이다.

이제 GLM weather generator의 모형계수에 공간구조를 가지는 사전분포를 가정하고자 한다. 즉, 랜덤 효과에 해당하는 $\alpha(s)$ 와 $\delta(s)$ 에 대해서 평균이 μ_α 와 μ_δ 이고 공분산 함수로 $C_\alpha(\cdot, \cdot)$ 와 $C_\delta(\cdot, \cdot)$ 를 각각

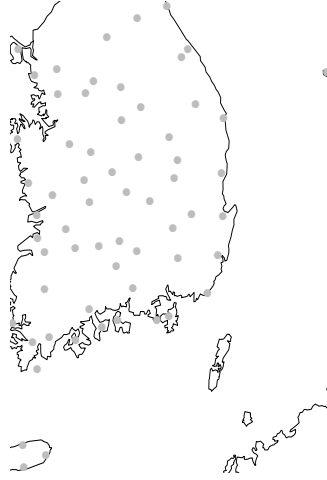


Figure 4.1. Locations of 62 observing stations

가치는 정규확률과정을 가정한다. 그리고 이러한 구조를 다른 모형계수인 β 와 γ 로 확장할 수 있다. 즉,

$$\begin{aligned}\beta_i(\cdot) &\sim MVN(\boldsymbol{\mu}_{\beta_i}, \boldsymbol{\Sigma}_i), \quad \text{for } i = 1, 2, \dots, 7, \\ \gamma_i(\cdot) &\sim MVN(\boldsymbol{\mu}_{\gamma_i}, \boldsymbol{\Psi}_i), \quad \text{for } i = 1, 2, 3, 4.\end{aligned}$$

여기에서 $\boldsymbol{\Sigma}_i$ 와 $\boldsymbol{\Psi}_i$ 는 공간구조를 가지는 공분산 행렬이고 분산값과 상관행렬의 곱의 형태로 나타난다. 즉, $\boldsymbol{\Sigma}_i = \sigma^2 R_{\rho, \nu}(s, s')$.

가장 일반적으로 사용되는 공분산함수는 Matérn covariance function인데 다음의 형태를 가진다.

$$R_{\rho, \nu}(s, s') = \frac{(d(s, s')/\rho)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu \left(\frac{d(s, s')}{\rho} \right),$$

여기에서 ρ 는 척도모수이고 ν 은 smoothness의 정도를 나타내는 모수이다. 이 모수를 적절히 조절하면 Matérn의 특수한 형태의 공분산 구조를 얻을 수 있다. 특히 $\nu = 0.5$ 일 경우에는 일반적으로 잘 알려진 지수모형을 가진다. 더 일반적인 시공간모형을 고려하기 위해서는 다음과 같은 공분산 함수를 고려할 수도 있다.

$$C_\alpha^t(s_1, s_2) \propto \exp \left(\frac{-\|s_1 - s_2\|}{A(t)} \right)$$

$\|s_1 - s_2\|$ 는 s_1 과 s_2 지점 사이의 거리이다. $A(t) = \exp(\alpha_0 + \alpha_1 C_t + \alpha_2 S_t)$ 이다. 일반적으로 공간적인 랜덤효과를 보기 위해서 $\alpha(\cdot)$ 와 $\delta(\cdot)$ 에 관한 추론에 초점을 맞추어진다. 본 논문에서는 Matérn 공분산 함수를 고려하였다.

4. 전국 76개 지역의 강우량 자료에 모형 적합

이 논문에서는 Korea Meteorological Administration(KMA)로부터 얻은 1973년 1월 1일부터 2011년 12월 31일까지로 39년동안 우리나라 76개 지역에서 측정된 일별 강우량 관측자료를 사용하였다 (Figure 4.1 참조). 그 중에서 100개 이상의 결측치가 포함된 지역 14군데를 제외한 나머지 62개 지역의 자

Table 4.1. Estimated coefficients (Coef.) and standard error (SE) values for all components of the stochastic weather generator with aggregated climate statistics as covariates at Seoul. Note: in precipitation models for convenience to make the results easy to present in a compact format, daily mean rate is used as a covariate instead of precipitation total at Seoul.

Covariate category	Precipitation Occurrence			Precipitation intensity (mm)		
	Term	Coef.	SE	Term	Coef.	SE
Constant	μ	-2.68	0.079	μ	1.74	0.118
Summer	$I_t P_t^S$	0.18	0.014	$I_t P_t^S$	0.09	0.019
Winter	$(1 - I_t) P_t^W$	0.32	0.083	$(1 - I_t) P_t^W$	0.28	0.115
Autocorrelation	J_{t-1}	1.06	0.041	-	-	-
Seasonality	C_t	0.65	0.059	C_t	0.74	0.073
	S_t	0.21	0.031	S_t	0.31	0.034
Interaction	$C_t J_{t-1}$	0.18	0.057	-	-	-
	$S_t J_{t-1}$	0.11	0.054	-	-	-

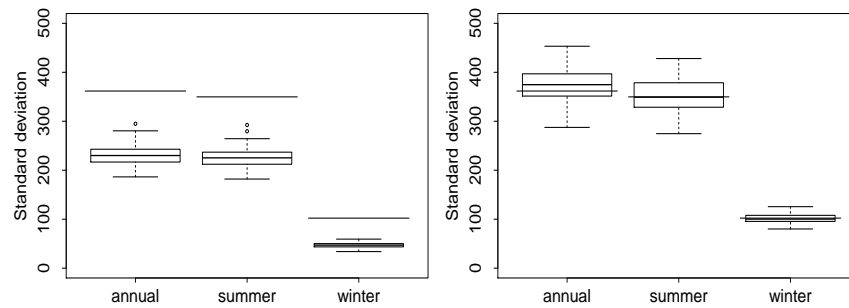


Figure 4.2. Boxplots of simulated Sd of annual, summer and winter total precipitation(mm) based on the original model (left) and with smoothed seasonal covariates (right). Horizontal solid lines: corresponding observed values of the data series.

료를 사용하였다. 여기서는 복잡한 데이터구조와 계산을 고려하여 1년을 365일로 적용하였고 윤달의 경우 2월 29일을 제외하였다. 계절특성치를 구하기 위해 우기와 건기로 나누었는데 우기는 4-9월, 건기는 1-3월과 10-12월이다.

우선 지역별 모형 적합도를 확인하기 위해서 서울지역 39년동안의 강우량 자료를 사용하였다. AIC와 BIC 관점에서 기존의 모형에 비해서 Kim 등 (2012)이 제안한 모형이 더 잘 적합됨을 확인할 수 있었다. Table 4.1에서는 기상청에서 제공한 서울지역 자료를 이용한 강우발생 유무와 강우량에 대한 GLM weather generator에 사용된 모수의 추정량과 추정오차를 각각 보여준다. 참고로 분석에는 공개소스 통계 프로그램인 R이 사용되었다. 분석표에서 나타난 바와 같이 사용된 모든 변수들이 통계적으로 유의함을 알 수 있다.

제안된 모형의 유효성을 평가하기 위해 추정된 모형에서 생성된 값들을 이용한 여러 가지 계절별 통계 특성치와 실제 관측된 자료들의 특성값들을 비교하였다. 이를 위해서 실제 관측기간과 동일한 39년치의 일일 기온을 모형으로부터 생성하는데, 이러한 작업을 500회 반복하였다. 참고로 일일 기온 생성에 필요한 일일 강우발생 유무는 GLM weather generator에서 동시에 생성하여 일일 기온모형에 변수로 입력된다. Figure 4.2는 일일 기온 모형에 LOESS smoothing된 계절별 평균 기온을 사용한 모형이 그렇지 않은 모형보다 얼마나 overdispersion 면에서 향상되었는지를 계절별 총강우량의 분산을 통해서 비교하였다. 기존의 모형에 비교해서 overdispersion 현상이 개선 되었음을 확인할 수 있었다. 앞에서 잠깐

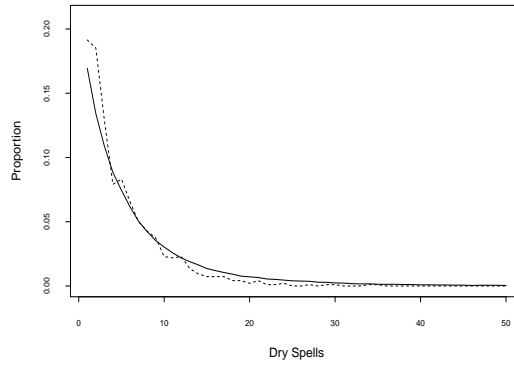


Figure 4.3. Distributions of observed dry spells (dashed line) and simulated dry spells (solid line) during summer season over the 39 years.

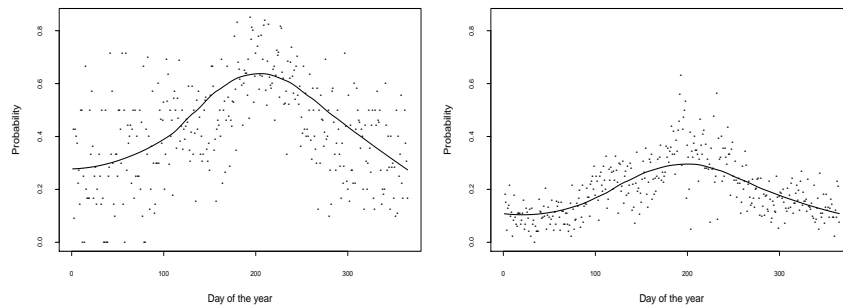


Figure 4.4. Modeled transition probabilities $p_{11}(t)$ (left) and $p_{01}(t)$ (right) with smoothed aggregated covariates. Dots: empirical transition probabilities, *i.e.* frequencies of observed transitions calculated separately on each day of the year.

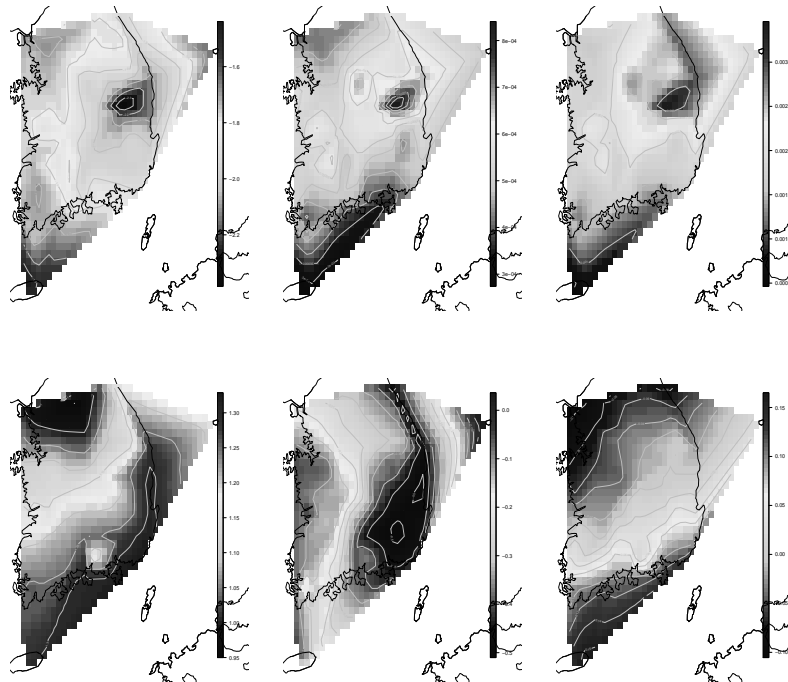
언급한 바와 같이 여기에서 사용된 smoothing 모수값은 결과에 아주 민감하지는 않으므로 적당한 수준에서 결정하여 사용하면 된다. 그리고 smoothing을 적게 하면 할수록 너무 많은 변동이 모형에 전해져서 오히려 underdispersion 현상이 발생할 수 있으므로 적당한 smoothing 모수값을 정하는 것은 반드시 필요하다.

추정된 모형을 평가하기 위해서 서울지역의 건조기간(dry spell)에 대한 상대도수를 비교하였다. 건조기간을 정의하는 기준은 명확하게 정해져 있지 않지만 기상학에서는 어느 지점에 강우가 거의 없는 기간으로서 일반적으로 2주 내지 1개월 정도의 기간이라고 정의하였다. 참고로 이 논문에서는 강우가 연속적으로 발생하지 않은 기간($J_t(s) = 0$)을 건조기간으로 두었다. 다시 말해 강우가 발생하지 않은 날부터 그 다음 강우량이 발생한 전날까지의 기간이다. Figure 4.3에서는 39년간 우기에 해당하는 일일강우량 자료를 이용하여 건조기간 상대도수를 점선으로 나타내었고, 실선은 500번의 반복실험과정을 거쳐 얻은 데이터로 건조기간을 나타낸 것이다. 그래프를 살펴보면 건조기간이 길수록 그 빈도는 급격히 떨어지고 20일 이상 지속적으로 비가오지 않은 빈도는 극히 드물다. 강우모형에서 얻은 데이터와 실제로 서울지역에서 관측된 자료의 건조기간 비율이 거의 일치하는 것으로 보아 모형의 적합성을 한번 더 확인할 수 있다.

추가적으로 일일 강우발생에 대한 전이확률을 확인하였다. 전 날 비가 왔을 때 다음 날 비가 올 조건부 확률은 $p_{11}(t) = P\{J_t = 1 | J_{t-1} = 1\}$ 이고, 전 날 비가 안왔을 때 다음 날 비가 올 조건부 확률은

Table 4.2. Estimated parameters of spatial covariance functions for all coefficients in the precipitation occurrence and intensity models

Precipitation Occurrence				Precipitation intensity			
Coefficient	Spatial Parameters			Coefficient	Spatial Parameter		
	scale	range	smoothness		scale	range	smoothness
α	1.02	0.63	0.31	α	0.99	0.44	0.29
β_1	1.00	0.45	0.33	γ_1	1.02	0.34	0.24
β_2	1.05	0.42	0.32	γ_2	1.01	0.21	0.31
β_3	0.98	0.54	0.45	γ_3	1.04	0.46	0.59
β_6	1.01	0.49	0.46	γ_4	1.08	0.55	0.35
β_7	1.00	0.61	0.55				

**Figure 4.5.** Posterior means of coefficients ($\alpha, \beta_1, \beta_2, \beta_3, \beta_6,$ and β_7) in the precipitation occurrence model.

$p_{01}(t) = P\{J_t = 1 | J_{t-1} = 0\}$ 으로 표현할 수 있다. Figure 4.4에서 점으로 표현한 것은 실제 39년치 강우량의 전이확률을 나타낸 그래프이고 실선은 모형에서 생성된 값의 전이확률의 평균을 나타낸 곡선이다. 그림에서 알 수 있듯이 모형에서 생성된 값과 실제값의 패턴이 비슷하다.

3장에서 소개한 공간구조를 가지는 GLM weather generator를 우리나라 76개 지역에서 39년간 측정된 일별 강우량 관측자료에 적용하였다. 참고로 Table 4.2에서는 각 모형의 계수에 대한 공분산 함수의 모수에 대한 추정값들이 주어졌다. Figure 4.5에서는 추정된 강우발생모형의 주요계수 $\alpha, \beta_1, \beta_2, \beta_3, \beta_6$ 그리고 β_7 에 대한 사후평균 값들이다. 사후평균과 함께 추정오차도 함께 구해질 수 있는데, 추정오차는 모든 계수가 비슷한 공간적 특성을 보임에 반해서 추정 평균값들은 각각 다른 공간적 특성을 보임을 알 수 있다. Figure 4.6에서는 강우량에 대한 모형의 결과를 보여준다. 여기에서도 강우발생모형과 비슷한 경향을 보여 준다. 이렇게 얻어진 공간정보를 이용하여 관측자료가 없는 지역에 대한 강우모형을 설계

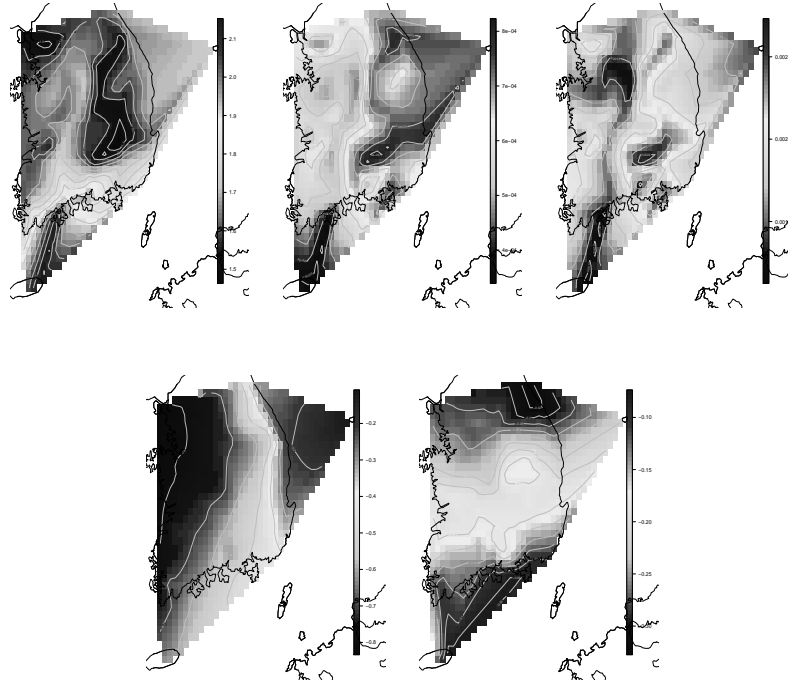


Figure 4.6. Posterior means of coefficients (δ , γ_1 , γ_2 , γ_3 , and γ_4) in the precipitation intensity model.

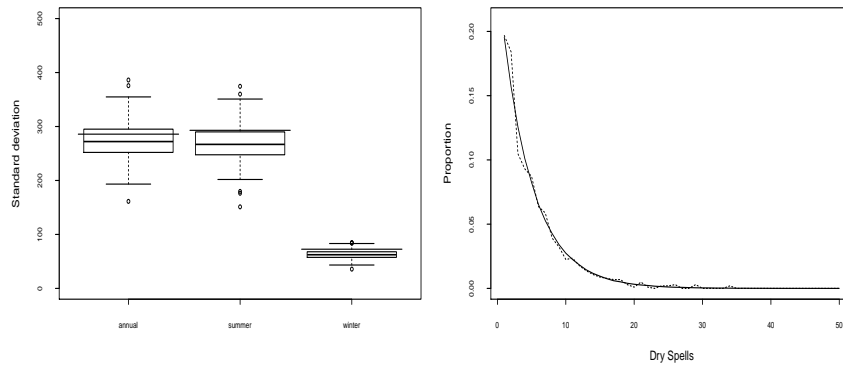


Figure 4.7. Model checking in the views of overdispersion and dryspell.

할 수 있다.

마지막으로 제안된 공간구조를 가지는 GLM weather generator가 실제 지역적 특성을 잘 설명하는지를 확인하기 위해서는 실제로 강우량을 관측하지 않은 지역에 대해 강우발생 모형과 강우량 모형이 적절하게 예측할 수 있을 것인가를 점검해 볼 필요가 있다. 본 논문에서는 수도권 한 지역의 강우 자료를 제거한 후, 주변 지역의 정보를 이용해 추정된 그 지역의 계절별 특성치 통계량이 실제 관측된 계절별 통계량과 얼마만큼 일치하는지 확인하였다. 여러가지 계절별 특성치 통계량 중 overdispersion 현상을 확인할 수 있는 계절별 총강우량의 분산과 건조기간(dry spell)에 대한 상대도수를 비교하였다 (Figure 4.7).

5. 결론

본 논문에서 우리는 공간구조를 가지는 GLM weather generator를 소개하였고 제안된 모형을 우리나라 76개 지역에서 39년간 측정된 일별 강수량 관측자료에 적용하였다 기존의 모형을 통해 smoothing된 계절별 총강수량이라는 장기 현상이 일일 강우발생이라는 단기 현상에 어떻게 영향을 미치는지를 알아보았다. 이를 통해 일반적으로 GLM weather generator가 가지는 계절별 총강수량의 overdispersion 현상을 효과적으로 제거할 수 있었다. 그리고 다층구조를 이용하여 기존의 GLM weather generator에 공간구조를 소개하였다. 이를 위해 Stochastic weather generator의 모형계수에 공간구조를 가지는 다변량 정규분포를 가정하였다. 이를 통해서 모형계수의 공간적 특성을 확인할 수 있었다. LOESS에서 적절한 smoothing 정도를 결정하는 과정이나 계절별 총강수량이 모형에 재사용이라는 관점에서 여전히 문제점을 가지고 있으므로 overdispersion 현상을 제거할 수 있는 다른 접근법에 관한 추가적인 연구가 필요하고 또한 공간모형을 기존의 GLM weather generator에 적용하는 방법에 대한 연구도 계속되어야 하겠다.

References

- Buishand, T. A. (1978). Some remarks on the use of daily rainfall models, *Journal of Hydrology*, **47**, 235–249.
- Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, **74**, 829–836.
- Furrer, E. M. and Katz, R. W. (2007). Generalized linear modeling approach to stochastic weather generators, *Climate Research*, **34**, 129–144.
- Hansen, J. W. and Mavromatis, T. (2001). Correcting low-frequency variability bias in stochastic weather generators, *Agricultural and Forest Meteorology*, **109**, 297–310.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman and Hall, New York.
- Katz, R. W. and Parlange, M. B. (1998). Overdispersion phenomenon in stochastic modeling of precipitation, *Journal of Climate*, **11**, 591–601.
- Katz, R. W. and Zheng, X. (1999). Mixture model for overdispersion of precipitation, *Journal of Climate*, **12**, 2528–2537.
- Kim, Y., Katz, R. W., Rajagopalan, B., Furrer, E. M. and Podestá, G. (2012). Reducing overdispersion in stochastic weather generators using a generalized linear modeling approach, *Climate Research*, **53**, 13–24.
- Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, N., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M. and Thiele-Eich, I. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, **48**, doi:10.1029/2009RG000314.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. 2nd ed., Chapman and Hall.
- Richardson, C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation, *Water Resources Research*, **17**, 182–190.
- Stern, R. D. and Coe, R. (1984). A model fitting analysis of daily rainfall data, *Journal of the Royal Statistical Society A*, **147**, 1–34.
- Wilks, D. S. (1989). Conditioning stochastic daily precipitation models on total monthly precipitation, *Water Resources Research*, **25**, 1429–1439.
- Wilks, D. S. (2010). Use of stochastic weather generators for precipitation downscaling, *Wiley Interdisciplinary Reviews: Climate Change*, **1**, doi:10.1002/wcc.85
- Wilks, D. S. and Wilby, R. L. (1999). The weather generator game: A review of stochastic weather models, *Progress in Physical Geography*, **23**, 329–357.

시공간구조를 가지는 확률적 강우 모형

곽민정^a · 김용구^{b,1}

^a영남대학교 통계학과, ^b경북대학교 통계학과

(2014년 3월 7일 접수, 2014년 4월 28일 수정, 2014년 5월 8일 채택)

요약

일반화 선형모형(GLM)에 기초한 확률적 날씨 발생기(Stochastic weather generator)는 일일 날씨를 생성하는데 가장 일반적으로 사용되는 방법이다. 본 논문에서는 다층구조를 이용하여 기존의 GLM weather generator에 공간구조를 소개하였다. 계절별 총강우량의 overdispersion 현상을 효과적으로 제거하기 위해서 smoothing된 계절별 총강우량을 모형에 포함하였고 공간구조를 소개하기 위해서 Stochastic weather generator의 모형계수에 공간구조를 가지는 다변량 정규분포를 가정하였다. 그리고 제안된 공간구조를 가지는 GLM weather generator 모형을 우리나라 76개 지역에서 39년간 측정된 일별 강우량 관측자료에 적용하였다.

주요용어: 강우모형, 일반화선형모형, 다중지점 날씨 발생기.

이 논문은 2012년도 정부(미래창조과학부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행된 연구임 (No. 2012-011113).

¹교신저자: (702-701) 대구광역시 북구 대학로 80, 경북대학교 통계학과. E-mail: kim.1252@knu.ac.kr