



Development of bias correction scheme for high resolution precipitation forecast

Uranchimeg, Sumiya^a · Kim, Ji-Sung^b · Kim, Kyu-Ho^b · Kwon, Hyun-Han^{a*}

^aDepartment of Civil Engineering, Chonbuk National University

^bHydro Science and Engineering Research Institute, Korean Institute of Science and Technology

Paper number: 18-017

Received: 14 February 2018; Revised: 30 March 2018; Accepted: 30 March 2018

Abstract

An increase in heavy rainfall and floods have been observed over South Korea due to recent abnormal weather. In this perspective, the high-resolution weather forecasts have been widely used to facilitate flood management. However, these models are known to be biased due to initial conditions and topographical conditions in the process of model building. Therefore, a bias correction scheme is largely applied for the practical use of the prediction to flood management. This study introduces a new mean field bias correction (MFBC) approach for the high-resolution numerical rainfall products, which is based on a Bayesian Kriging model to combine an interpolation technique and MFBC approach for spatial representation of the error. The results showed that the proposed method can reliably estimate the bias correction factor over ungauged area with an improvement in the reduction of errors. Moreover, it can be seen that the bias corrected rainfall forecasts could be used up to 72 hours ahead with a relatively high accuracy.

Keywords: Bias correction, Numerical forecast model, UM3.0

고해상도 강수량 수치예보에 대한 편의 보정 기법 개발

오랑치맥 솜야^a · 김지성^b · 김규호^b · 권현한^{a*}

^a전북대학교 공과대학 토목공학과, ^b한국건설기술연구원 수자원·하천연구소

요 지

최근 이상기후로 인한 집중호우 발생빈도와 이로 인한 국지적인 홍수 피해가 증가하고 있다. 이러한 점에서 홍수피해 예방측면에서 수치예보 정보 활용이 요구되고 있다. 그러나 수치예보모델은 초기 조건 및 지형적 요인으로 인해 시공간적 편의가 존재하며 실시간 예측정보로 활용하기 전에 모형결과에 대한 편의보정이 요구된다. 본 연구에서는 관측지점 기준으로 편의 보정계수를 산정하는 과정에서 모든 관측소간의 상관성을 거리의 함수로 고려하여 미계측지점의 편의 보정계수를 공간적으로 확장할 수 있는 Bayesian Kriging 기반 MFBC 기법을 개발하였다. 본 연구에서 개발한 방법은 미계측 유역에 대해서도 보정계수를 효과적으로 추정하는 것이 확인되었으며, 비교적 고해상도로 72시간(3일) 정도까지 예측강우 정보를 활용하는 것이 가능할 것으로 판단된다.

핵심용어: 편의 보정, 수치예보 모델, UM3.0

1. 서 론

최근 이상기후로 인한 집중호우 발생빈도가 증가하고 있으며 대우역 보다는 도심지역 및 국지적인 지역의 홍수로 인

한 피해가 증가하고 있다. 이러한 점에서 자동기상관측장비 (Automatic Weather System, AWS)와 종관기상관측장비 (Automated Synoptic Observing System, ASOS) 등 기상관측 기상정보, 레이더강수정보, 수치기상예측정보 등 다양한 다중정보를 활용한 효율적인 방재관리가 요구되고 있다. 특히, 선행시간 확보관점에서 수치기상예측정보의 활용성을 재고하고자 하는 연구들이 다수 진행되고 있으며, 특히 수치

*Corresponding Author. Tel: +82-63-270-2426
E-mail: hkwon@jbnu.ac.kr (H.-H. Kwon)

모델의 연산능력 및 물리적인 고려사항 등에 대한 개선으로 인해 예측정보로서 활용성이 증가하고 있다. 그러나 수치예보 모형의 기본가정 및 지형 등으로 인한 강제효과 등으로 인한 시공간적 편이가 여전히 주요 문제로 인식되고 있으며 이러한 편이를 보정하는 연구들도 진행되고 있다(Berg *et al.*, 2012; Bordoy and Burlando, 2013; Hey *et al.*, 2000; Landerink *et al.*, 2007; Themeßl *et al.*, 2011).

수치 예보 모델의 해상도는 컴퓨터 계산 능력에 따라 점차 증가하고 있다. 고해상도 모델은 저해상도의 모델보다 좀 더 상세하게 강수 현상과 연관된 대기현상을 재현할 수 있다(Jee and Kim, 2017). 국내외적으로 단기 예보 모델로서 슈퍼컴퓨터 기반 고해상도 예보 모델을 많이 활용하고 있다. 수치기상 예보모델은 고해상도로 갈수록 신뢰성 있는 장기 예보를 제공하는데 어려움이 있다(Buizza, 2010). 이는 미세한 오차가 발생 하더라도, 실제 기상학적 관점에서 시공간적으로 변동성이 크게 발생할 개연성이 크며, 이로 인해 모델에서 발생하는 불확실성은 더욱 커질 수 있다. 또한 모든 예보 모델들은 예보 시간이 길어질수록 오차가 커지는 것이 일반적이다(Buizza, 2010; Jee and Kim, 2017).

국내에서 수치기상정보를 수자원 분야에 활용하기 위하여 많은 선행연구들이 진행되었다. Yoon *et al.* (2017)은 우리나라 기상청 국지예보모델(Local Data Assimilation and Prediction System, LDAPS)과 일본 기상청의 중규모모델(Meso-Scale Model, MSM)에서 생산된 강우예측정보의 활용성을 평가하였으며 두 모델의 강우예측은 국지적인 강우사상에 비해 광역적 강우사상 예측에 있어 적합성을 갖는 것으로 평가하였다. Son *et al.* (2017)의 연구에서는 기상청 범주형 확률장기예보를 다양한 평가지표를 활용하여 예측력을 검증하였고, 추가적으로 범주형 정보를 활용하여 예측 강수량을 생산할 수 있는 방안을 제시하였다. Lee *et al.* (2015)의 연구에서는 WRF 모형을 이용하여 과거 호우사상에 대한 재현 능력을 평가하였다. 분석결과 강우강의 분포는 효과적으로 재현하였지만, 강수량의 과대추정과 시간분포의 오차가 큰 것으로 분석되었다. 국외에서는 수치기상예측정보를 홍수예측에 활용하는 연구들이 진행되고 있다. Yucel *et al.* (2015)의 연구에서는 WRF-Hydro 모형을 이용하여 강우예측과 홍수예측을 동시에 수행할 수 있는 분석시스템을 구축하였으며 위성강수의 활용성도 평가하였다. WRF 모델과 자료동화기법을 연계하는 경우 예측오차의 36.9%를 개선할 수 있는 것으로 분석되었으며, 미세측유역으로 모형의 확장성을 제시하였다. Siddique *et al.* (2015)의 연구에서는 미국의 대표적인 수치기상예보시스템인 GEFS과 SREF로부터 생산된 예측정보의 정확성을

평가하였으며, 기상학적 관점에서 6일 선행시간까지는 예측 정확성이 있는 것으로 제시하였다. 두 예측시스템 모두 단기 예측정보 차원에서 유사한 예측성을 갖는 것으로 분석되었다. Wu *et al.* (2017)의 연구에서는 대만 산악지역의 홍수예측을 위하여 1-6시간의 선행시간에 대하여 앙상블 단기예측정보를 생산하였으며, 물리기반의 앙상블 홍수예측 전략에 대해서도 언급하였다.

한국 기상청(Korean Meteorological Administration, KMA)에서는 영국 기상청(United Kingdom Meteorological Office, UKMO)으로부터 도입한 통합모델(Unified Model, UM)을 현업 화하여 운영하고 있다. 기상청에서는 다양한 시공간적 규모의 예측 모델을 운영 중이며 이 중에서 3 km의 공간해상도를 가지며 1시간단위로 167시간까지 예측 정보를 제공할 수 있는 UM3.0 모델이 대표 현업모델로서 다양한 목적으로 활용되고 있다. 본 연구에서는 UM3.0의 예측 정보의 활용성을 개선하고자 편의 보정 기법 개발 및 불확실성 추정 방안을 수립하고자 하며, 이를 통해 UM 3 km 예보모델에 대한 수문 기상학적 활용성도 개선하고자 한다. 분석에 활용한 자료 기간은 2014년 1월 1일부터 2015년 12월 31일까지 2년간의 예보 모델 강수량과 기상청 관할의 관측소 강수량 자료를 활용하였다.

본 연구의 목적은 UM3.0 모델 기반 강수량의 편의 보정 및 불확실성 추정 방안을 제시하고자 한다. 수문기상학적 자료들에 대한 편의 보정 방법으로 일반적인 선행회귀분석, 다항식회귀분석, Quantile Mapping 등의 방법이 있으며 본 연구에서 보정하고자 하는 UM 3 km 예보모델의 예보 기간이 2년, 공간해상도가 3 km라는 특성과 편의 보정 방법의 적용성 및 계산속도 등을 고려하여 Mean Field Bias Correction (MFBC) 방법을 기본적으로 선택하였다. 격자자료에 대한 편의보정 시 가장 큰 어려운 점은 보정계수 산정이 관측지점단위에 대해서만 이루어진다는 점이며, 이를 전 격자자료에 적용하기 위한 방안 연구는 매우 미진한 실정이다. 이러한 점에서, 본 연구에서는 공간적 편의 보정 및 불확실성 평가를 위해서 앞서 언급한 편의 보정 방법인 MFBC 방법과 매개변수 추정단계에서부터 공간적 내삽 기법인 Kriging 기법을 결합할 수 있는 방법론을 도입하였으며, Bayesian 이론을 바탕으로 추정된 보정계수의 불확실성 또한 추정이 가능한 공간적 편의 보정 기법을 개발하였다. MFBC 방법은 주로 레이더 강수량의 보정방법으로 주로 활용되고 있으며(Borga, 2002; Lee *et al.*, 2015; Seo *et al.*, 1999), 격자형태의 원격자료를 지점기준의 기상자료와 비교를 통해서 산정되는 평균편의보정 방법이라 할 수 있으며, 수치모델결과에 대한 편의보정을 위해 사용된

사례는 많지 않다(Durai and Bhradwaj, 2014; Woodcock and Engel, 2004).

본 장에서는 연구배경 및 선행연구 등에 대해서 살펴보았다. 2장에서는 분석에 사용된 자료들에 대해서 정리하였으며, 3장에서는 본 연구에서 제안된 방법론의 이론적인 배경을 제시하였다. 4장에서 2년간의 수치예보자료를 대상으로 방법론의 적합성을 평가하였으며, 최종적으로 결론을 제시하였다.

2. 대상자료

2.1 UM3.0 수치예측 모델 강수량

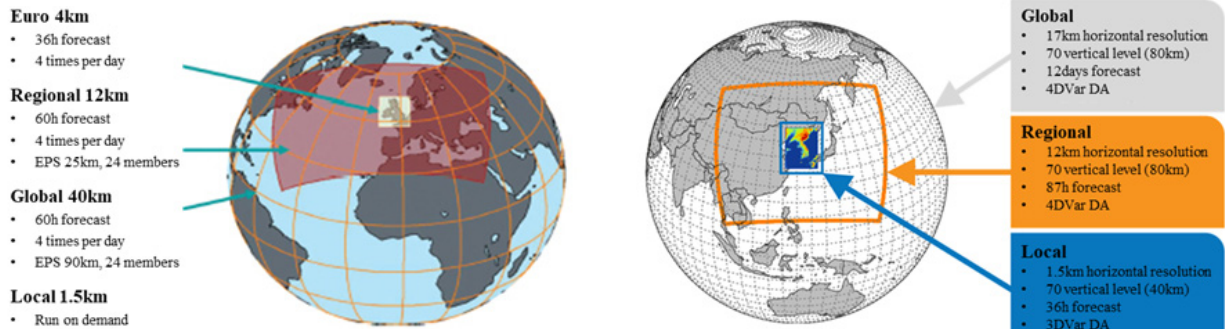
통합모델(UM)은 영국에서 개발된 모델로 현업용 NWP (Numerical Weather Prediction), 계절예보 및 기후모델링 등과 같은 여러 모델들의 기능을 하나의 구조 안에서 조합한 기상수치모델이다. 현재 기상청에서는 영국으로부터 도입한 UM 모델 운영을 통하여 전주기 예보모델(Global Data Assimilation and Prediction System, GDAPS)부터 지역 예보모델(Regional Data Assimilation and Prediction System), 국지 예보모델(LDAPS)까지 다양한 기상예측정보를 제공하

고 있다. Fig. 1은 4가지 기상예측 모델에 대한 영역 범위 및 예측시간에 대한 정보이다.

본 연구에서는 위와 같이 다양한 공간적 해상도를 갖는 모델 중에서 3 km의 공간해상도를 갖는 UM 3 km (UM3.0) 예보 모델의 예측 강수량을 활용하였다. UM3.0은 17 km 해상도의 전지구 UM, 12 km 해상도의 지역 UM의 순차적 2단계 공간상세화를 거쳐서 3 km 공간 해상도를 갖는 모델로 구축되었으며, 1시간의 시간해상도를 가지고 1일 2회(00UTC, 12UTC) 모의를 수행한다. 또한 30~37시간 정도의 비교적 짧은 예측 시간을 갖는 다른 모델들과 다르게 예보시작시점기준 7일이라는 긴 예측시간을 갖는다. UM3.0 모델의 공간적인 제공 범위는 위도 32.99~39.77°N, 경도 124.96~131.75°E로 Fig. 1(c)에서와 같이 한반도 전체를 포함한다. UM3.0 수치예보 자료에 대한 평가는 및 보정은 2014~2015년까지 2년간의 단일면 강수량 자료를 이용하였다.

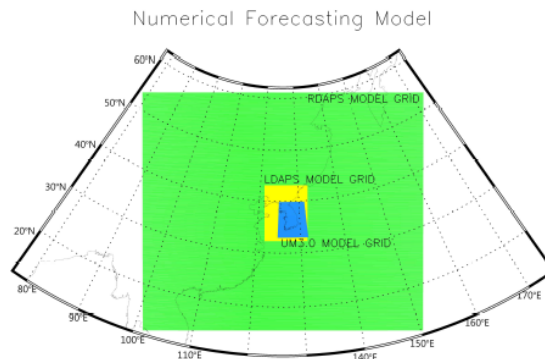
2.2 기상관측소 자료(AWS 및 ASOS)

예보 강수량을 검증하기 위해 기상청에서 설치, 운영 중인 자동기상관측장비(AWS)와 종관기상관측장비(ASOS) 관측소 자료를 이용하였다. 대상 연구 기간인 2014년 1월 1일부



(a) UM model setup in UK

(b) UM model setup in Korea



(c) Spatial domain of UM3.0 in Korea

Fig. 1. Forecast information of UM models

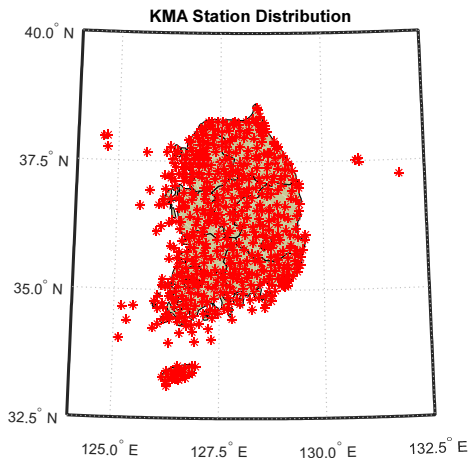


Fig. 2. Spatial distribution of AWS and ASOS

터 2015년 12월 31일까지 2년간 결측 값이 5% 이하인 442개 지점을 선택하였다. 선택된 442개 지점의 위치는 Fig. 2와 같으며, 각 관측소별 검증을 위해서는 각 관측소 위치 주변 9개 격자의 강우량을 평균한 값을 이용하였다.

3. 연구방법

본 연구에서는 고해상도의 수치예보 강수량 자료의 편의 보정 기법을 개발하였다. 시공간적 고해상도 자료의 편의 보정은 시계열 자료의 편의 보정 기법과 달리 상당히 큰 용량의 자료를 다루기 때문에 본 논문에서는 편의 보정 기법으로 MFBC

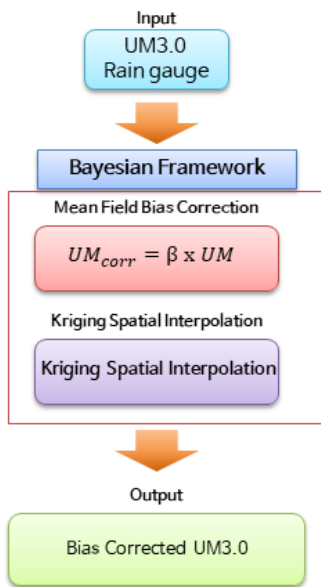


Fig. 3. Flowchart of Bayesian Kriging based MFBC method

기법을 채택하였다. 편의 보정 기법의 공간적 확장을 위하여 Kriging 내삽 기법을 MFBC 기법과 혼합하였으며 모형의 매개변수를 보다 효과적으로 산정하기 위해서 Bayesian 이론을 적용하여 모든 격자의 보정계수에 대한 사후분포를 산정하였다. 본 연구의 흐름도는 Fig. 3과 같으며 각 연구방법에 대한 이론적 배경은 다음 절에서 설명하였다.

3.1 Mean Field Bias Correction (MFBC)

예보모델 UM3.0의 강수량의 Mean Field Bias (B_s)는 관측 강수량에 기록되는 단위시간 누적 예보강수량 합에 대한 단위 누적 관측 강수량의 합의 비로써 산정되며 계산식은 다음 Eq. (1)과 같다(Anagnostou *et al.*, 1998).

$$B_s = \frac{\sum_{i=1}^n G_s(u_i)}{\sum_{i=1}^n R_s(u_i)} \quad s = 1, 2, \dots, n \quad (1)$$

여기서, s 는 예보강수의 기준지속시간, n 은 기준시간에 따른 관측 강수량과 예보모델 쌍의 개수, $R_s(u_i)$ 와 $G_s(u_i)$ 는 특정 지점 u_i 에 대한 기준시간 예보모델과 관측 누적강수량이다. 강수량은 이산 변량이며 이러한 변량에 대한 편의보정과정은 연속 변량에 대한 편의보정보다 비교적 복잡하다. 수치예보 자료의 경우, 과소평가(0,1)와 과대평가(1,0)가 문제이다. 여기서 (0,1)의 의미는 강수가 관측 자료에 기록이 되지 않았지만 수치예보 자료에 강수가 발생했을 경우를 의미하며, 반대로 (1,0)은 관측 자료에 강수가 기록이 되었지만 수치예보 자료에는 없는 경우를 의미한다. 일반적으로 (0,0), (1,0), (0,1), (1,1)의 4가지 경우가 있다. 본 연구에서는 관측 자료 및 수치예보 자료에 동시에 강수가 발생한 (1,1) 경우만 고려하였다. 또한, 본 연구에서 활용하는 예보모델과 관측 자료가 1시간 자료이기 때문에 기준시간은 1시간 단위로 하여 MFBC 기반의 보정계수를 추정하였다.

3.2 Kriging interpolation 기법

공간자료는 크게 점 자료(point data)와 면 자료(areal data)로 나눌 수 있으며, 점(point)으로 표현되는 공간자료를 모델링하는 일반적인 형태는 식과 같다.

$$y(s) = \mu(s) + \omega(s) + \epsilon(s) \quad (2)$$

여기서, $y(s)$ 는 지리적 위치 s 에서의 종속변수를 의미하며, $\mu(s)$ 는 종속변수의 평균 구조(mean structure) 또는 전역적

경향(global trend)을 의미한다. 평균 구조는 공변량과 모수와 의 선형관계, 즉 $\mu(s) = x^T(s)\beta$ 로 표현하는 것이 일반적이며 (Banerjee *et al.*, 2014), 공분산은 매개변수 행렬로 구성된다. 종속변수에 영향을 미치는 체계적 요인들(systematic components)을 공분산을 통하여 설정하였다면, 변동성은 오차항으로 반영되며, 비공간자료의 경우 독립적 오차항(백색잡음, white noise) $\epsilon(s)$ 로 표현한다.

공간자료의 특징은 공간적 자기상관성으로, 가까운 곳에 위치한 관측 값들은 서로 유사한 값을 갖기 마련이다(Tobler, 1970). 따라서 Eq. (2)에서 오차항은 두 부분으로 구분할 수 있는데, $\omega(s)$ 는 관측 값들 간의 공간적 자기상관성을 나타내는 임의 효과항(spatial random effect term)이며, $\epsilon(s)$ 는 비공간적인(non-spatial) 순수 오차항(pure error term)을 나타낸다. 점(point) 자료를 다루는 공간 모델링 분야에서 자료의 공간적 상관성은 크리깅(kriging) 기법을 이용하여 모형에 반영할 수 있다. 크리깅은 공간 보간을 위한 대표적인 기법으로, 관찰되지 않은 지점의 예측값을 주변관찰지점 값의 가중선형 조합으로 산출하는 방법이다(Isaaks and Srivastava, 1989).

3.3 Bayesian 모형

Bayesian 추론(Bayesian inference)은 확률을 발생 빈도나 어떤 시스템의 물리적 속성이라고 여기는 것과는 다른 해석적 접근방법이다. Bayesian 확률은 사전확률(prior probability)을 기준으로 자료의 증거에 의해서 정보가 갱신되며 최종적으로 조건부로 바뀌는 사후확률(posterior probability)을 추정하는 과정이다(Gelman *et al.*, 2004).

매개변수 벡터 θ 와 주어진 확률변수 y 가 있을 때, 두 확률변수들의 결합확률분포(joint probability distribution)를 생성하는 모형에 대해서, 결합확률밀도함수(joint probability density function)는 매개변수의 사전분포(prior distribution) $p(\theta)$ 와 관측값과 매개변수의 함수인 우도함수(likelihood function) $p(y|\theta)$ 의 곱으로 표현할 수 있다.

$$p(\theta | y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y | \theta)}{p(y)} \propto p(y|\theta)p(\theta) \quad (3)$$

Bayes 정리에서 각각의 항의 의미를 다시 정리해 보면, 다음과 같다. $p(\theta)$ 는 θ 의 사전 확률 또는 경계 확률이다. '사전'이라는 것은 아직 사건 y 에 관한 어떠한 정보도 고려하지 않음을 의미한다. $p(\theta | y)$ 는 y 가 주어졌을 때 θ 의 조건부 확률이다. 이것은 y 의 특정 값에 의해 결정되기 때문에 사후 확률이라고 불린다. $p(y|\theta)$ 는 θ 가 주어졌을 때 y 의 조건부 확률이며

우도(likelihood)를 나타낸다. $p(y)$ 는 y 의 사전 확률 또는 경계 확률이며, 정규화 상수의 역할을 한다. Eq. (3)에서 $p(y)$, 즉 정규화 상수역할을 하는 주변확률밀도함수를 추정하는 것이 필요한데, 변수가 늘어나는 경우 이론적으로 유도하는 과정은 복잡하여 수치적인 접근이 이루어진다.

본 연구에서는 Markov Chain Monte Carlo (MCMC) 기법을 이용하여 매개변수를 추정하였다. Bayesian MCMC 기법은 주어진 다변량 확률분포가 복잡하여 이를 따르는 iid 난수를 얻을 수 없는 경우에 사용가능한 기법으로서 iid 난수 대신 Markov Chain 난수를 추출하여 사용한다. Markov Chain을 통해 난수를 발생시킨다고 해서 정확하게 관심이 되는 확률분포를 따르지 않지만 이를 일정 시간동안 반복 후에 얻어지는 난수들은 추출을 원하는 분포에 수렴하게 된다. 따라서 Bayesian MCMC 기법은 복잡한 다변량 확률분포 및 매개변수의 추정을 요하는 문제에서 주로 사용되며 또한 Bayesian 통계 기법에서 사후분포의 추론의 이용될 수 있다(Gelman *et al.*, 2004; Kwon and Moon, 2007; Kwon *et al.*, 2008).

3.4 Bayesian Kriging 기법을 이용한 MFBC 방법

Kriging은 일반화최소제곱법(Generalized Least Squares, GLS), 최우추정법(MLE)과 같은 고전적 접근법으로도 추정이 가능하지만, Bayesian 접근법으로도 추정이 가능하다. Bayesian Kriging은 비교적 새로운 접근법으로, 역학(epidemiology), 보건지리(medical geography)분야에서 질병의 발생 패턴을 분석(Lai *et al.*, 2013; Slater and Michael, 2013; Scholte *et al.*, 2014) 하는 등 자연·환경 분야에서 최근 들어 활발하게 적용되고 있다. Eq. (2)에서 공간효과를 나타내는 $\omega(s)$ 는 공분산함수(stationary covariance function)를 따르는 임의의 항이고, $\epsilon(s)$ 는 정규분포 $N(0, \tau^2)$ 을 따르며, $y(s)$ 의 주변 공분산 행렬(marginal covariance matrix)은 Eq. (4)와 같이 표현할 수 있다(Banerjee *et al.*, 2014).

$$\sum \sigma^2 R(\phi) + \tau^2 I \quad (4)$$

여기서, σ^2 은 공간효과와 관련된 분산(spatial variance 또는 partial sill), R 은 공간적 상관성을 표현하는 상관행렬(correlation matrix), τ^2 은 $\epsilon(s)$ 의 분산(non-spatial variance 또는 nugget), I 는 항등 함수(s_1 이 s_2 와 일치했을 때 1이며 일치하지 않을 때 0)를 의미한다. 상관행렬 R 은 Eq. (5)와 같이 표현할 수 있다.

$$R_{ij} = \rho(s_i - s_j; \phi) \quad (5)$$

여기서, ρ 는 상관함수, $(s_i - s_j)$ 는 관측치 간의 거리, ϕ 는 감쇠 계수(decay parameter)를 나타낸다. 한편, 인근에 위치한 관측 값들은 서로 유사한 값을 갖기 마련이며, 이러한 양의 상관성(positive correlation)은 거리가 멀어짐에 따라 약화된다. 공간자료의 이러한 특징은 공분산함수를 이용하여 모형에 반영할 수 있다. 본 연구에서는 자료의 공간적 상관성이 자료들 간의 거리에 의해서만 결정된다는 가정(등방성, isotropy) 하에 지수 공분산함수(exponential covariance function)를 활용하여 $\omega(s)$ 를 설정하였다. 지수 공분산함수는 Eq. (6)으로 표현되며, 크리깅 기법에서는 Eq. (7)과 같이 베리오그램(variogram) 형태로 변형하여 사용하는 것이 일반적이다.

$$C(t) = \sigma^2 \exp(-\phi t), t > 0 \quad (6)$$

$$\gamma(t) = \tau^2 + \sigma^2(1 - \exp(-\phi t)), t > 0 \quad (7)$$

본 연구에서는 ϕ 에 사전분포로 균일분포(uniform distribution)를 따르게 하여 거리에 대한 영향을 상대적으로 가중할 수 있도록 하였으며, Eq. (8)과 같다. κ 가 1일 경우, Eq. (8)은 Eq. (6)과 같다.

$$f(t; \phi, \kappa) = \exp[-(\phi t)^\kappa], t > 0, \kappa \in [0, 2] \quad (8)$$

한편 추정해야 할 매개변수들을 θ 라 하면, $\theta = (\beta, \sigma^2, \tau^2, \phi)^T$ 로 나타낼 수 있고, 사전 확률분포 $p(\theta)$ 를 가정하면, Bayes 법칙에 의하여 우도함수는 Eq. (9)와 같다. 따라서 사후 확률분포를 $p(\beta|y)$ 라 하면, $p(\beta|y)$ 는 Eq. (6)과 같다. Eq. (10)은 컴퓨터를 이용한 적분계산이 필요하며, 앞서 언급한 MCMC 기법을 이용하여 추정할 수 있다.

$$p(\theta) = p(\beta)p(\sigma^2)p(\tau^2)p(\phi) \quad (9)$$

$$p(\theta|y) = \iiint p(\beta, \sigma^2, \tau^2, \phi|y) d\sigma^2 d\tau^2 d\phi \\ \propto p(\theta) \iiint f(y|\theta) p(\sigma^2) p(\tau^2) p(\phi) d\sigma^2 d\tau^2 d\phi \quad (10)$$

본 연구에서는 다음과 같은 두 가지 사항을 위해서 442개 강우관측소 중 100개 강우관측소를 매개변수 및 편의보정계수 산정에 이용하였다. 첫째, 교차검증(cross-validation)을 위한 목적으로 100개 자료만을 이용하여 편의보정계수를 추정함과 동시에 전체 격자에 대해서도 편의보정계수를 산정하였으며, 이를 이용하여 342개 지점에 대해서 교차검증을 실시

하였다. 둘째, Bayesian Kriging 기법 적용 시 대상 모든 지점의 상관성을 고려하기 위한 상관행렬이 계산되어야 하며 이 과정에서 계산상의 어려움이 생긴다. 이러한 점에서 공간적으로 균일하게 분포된 일부자료만을 활용하여 매개변수 추정에 활용하였다.

본 연구에서 수치예보 모델 UM3.0 강우예측 성능에 대한 평가를 RMSE 및 편의 등의 통계치를 이용하였다. 각 통계치 계산식을 다음과 같으며, 여기서 x 는 관측 값을 나타내며, y 는 UM3.0 수치예보 자료의 강우를 의미한다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (11)$$

$$Bias = \frac{\sum_{i=1}^n (x_i - y_i)}{n} \quad (12)$$

4. 연구 결과

본 연구에서 기상청 관할의 442개 강우관측소 바탕으로 편의 보정을 실시하였다. 여기서, UM3.0 수치예보 정보는 강우 관측소 위치에 제일 가까운 격자를 선택하여 강우 예보 정보로서 활용하였다. 본 연구에서 채택한 Bayesian Kriging 기반 MFBC 기법은 관측지점 기준으로 편의 보정계수를 산정하는 과정에서 모든 관측소간의 상관성을 거리의 함수로 고려하여 미계측지점의 편의 보정계수를 공간적으로 내삽시키는 과정을 포함한다. 기존 방법론에서는 지점별로 편의보정계수를 추정 후 2차적으로 공간분포시키게 되며 이러한 경우 미계측구역의 매개변수 추정 시 과대 또는 과소 추정되는 문제점이 발생한다. Bayesian Kriging 기법은 분석 대상 모든 지점의 서로간의 상관성을 고려하여 분석을 한다. 이 때 다수의 지점을 모두 활용하는 것은 계산상의 어려움을 가중시킨다. 즉, 자료 증가에 따라 분석시간이 배수로 늘어나기 때문에 바람직하지 않다. 이러한 점에서 본 연구에서는 기상청 관할의 100개 강우 관측소를 활용하였으며 Fig. 4에 관측소 위치를 나타냈다.

일반적으로 MFBC 기법의 편의 보정계수(β)를 관측 강수량과 예보 강수량 비율로 산정되며 Fig. 5에서 본 연구에 활용하고 있는 기상청 442개 강우관측소의 보정계수를 산정한 결과를 나타내었다. Fig. 5에서 나타난 결과는 기존 MFBC 방법을 통해 산정된 결과로서 기지의 관측강수량과 수치예보자료 기반의 결과로서 본 연구에서는 참값으로 가정하였으며,

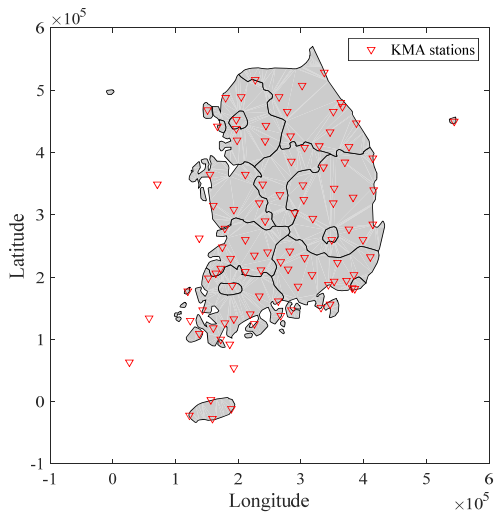


Fig. 4. Spatial distribution of the stations that used in estimate of β

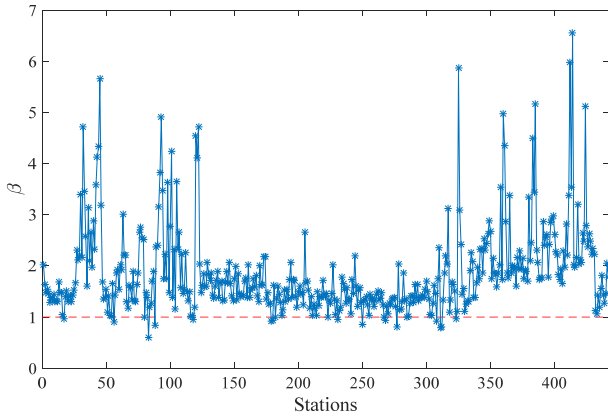


Fig. 5. β values estimates at the stations

Bayesian Kriging 기반 편의보정 계수와 비교 하였다. 산정된 β 값의 대부분이 1보다 큰 값을 가지고 있음을 확인 할 수 있었다. 즉, 이는 UM3.0 강수량이 전반적으로 적은 값을 가지고 있다는 것을 의미한다. Fig. 5를 살펴보면 3 이상의 β 값을 가진 관측소들이 있는 것을 확인할 수 있으며 이러한 큰 값을 가진 지역은 대부분 남부해안 지역과 제주도로 분류되었다.

Bayesian Kriging 기반의 MFBC으로 산정된 100개 기지의 지점과 나머지 342 지점에 대한 공간 보간된 편의 보정계수 β 는 Fig. 6과 같다. 즉, 기지의 지점 100개의 보정계수 값을 추정함과 동시에 공분산을 고려한 Bayesian Kriging 기법을 통해 미계측 지점의 β 값 또한 추정할 수 있다. 그림에서 보는 바와 같이 기지점의 경우 거의 동일하게 추정되고 있음을 확인할 수 있다. 이와 함께 95% 불확실성 구간 안에 참값으로 가정된 MFBC 값이 효과적으로 위치하는 것을 확인할 수 있다. 미지의 값으로 가정된 342개 지점의 경우에도 본 연구에서

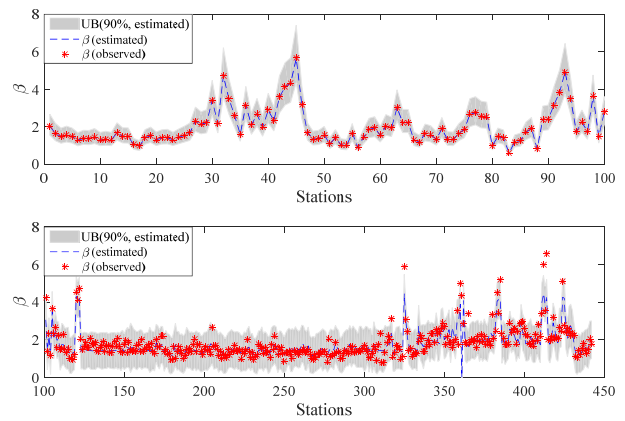


Fig. 6. β values estimates by Bayesian approach, β values estimated by Bayesian based MFBC method (top), β values estimated by Bayesian Kriging method (bottom)

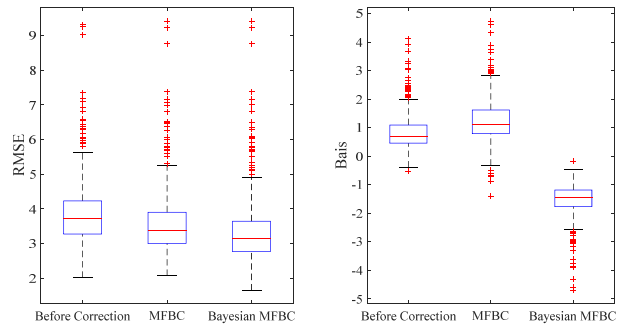


Fig. 7. Comparison of statistics before and after bias correction, during 2014~2015

제안한 Bayesian Kriging 방법의 결과는 참값과 거의 유사한 값을 가지는 것을 확인 할 수 있다. 다만 4를 넘어가는 보정계수에서 다소 과소 추정되고 있으며, 이는 지점간의 공분산을 고려하는 과정에서 주변 관측소와의 정보를 공유하면서 발생하는 사항으로 판단된다.

앞서 언급한 2014~2015년간의 강우시계열에 대해서 산정된 β 값을 이용하여 442개 관측소에서의 UM3.0 강우 자료를 편의 보정하였으며 보정 전·후 통계치를 Fig. 7에 나타냈다. 여기서 추정된 결과는 지점강수량과 수치예보자료 모두 강수가 발생한 경우를 기준으로 추정된 결과를 나타낸다. Bayesian MFBC 기법을 기존 MFBC 추정방법과 비교하였으며 Bayesian Kriging 기반의 MFBC 기법이 RMSE 측면에서 개선된 결과를 보여주고 있다. 반면 Bias의 경우에는 음의 값으로 보정 전보다 증가한 것으로 보이는 것은 앞서 Fig. 6에서 언급된 것처럼 4 이상의 큰 값에서 발생하는 과소추정으로 기인한 것으로 판단되나, RMSE를 기준으로 판단해보면 편의 보정 측면에서는 기존 방법론에 비해 개선된 결과라 할 수 있다.

Fig. 8에 나타난 결과는 2014~2015년의 연속적으로 발생한 시단위 전체 강수량 시계열에 대한 분석결과이며 Fig. 6에 산정된 β 값을 이용하여 기상청 관할의 442개 관측소에 대하여 예보시간(lead time)에 따른 편의 보정결과를 나타낸다. 여기서, 예보시간에 따라 강수량을 누적하지 않았으며 각 시점에 대한 편의 보정결과에 대한 통계치를 나타낸다. 선행예보 시간에 따른 오차의 증가가 미소하게 나타나고 있으나, 통계적으로 유의한 증가는 없는 것으로 평가되었다.

Fig. 9에서도 추정된 β 값을 이용하여 기상청 관할의 442개 관측소에 대하여 선행시간에 따른 편의 보정결과를 나타내고 있으나, Fig. 8과 다른 점은 보정된 강수량을 각각의 선행시간 까지 누적한 후 계산된 RMSE 통계치를 나타낸다. 즉, 선행예보 시간에 해당하는 강수량의 총량을 비교하여 통계치를 산정한 결과이며 보정 전 보다 개선된 결과를 확인할 수 있으며, 전반적으로 선행시간 증가에 따른 오차의 증가현상을 확인할

수 있다. Fig. 10은 Bayesian Kriging 기반의 MFBC 방법 β 값을 공간적으로 분포 시킨 결과를 나타내며, 각각 β 값의 5%, 50%, 95%의 값을 사후분포로부터 추출하여 도시하였다.

5. 결론

본 연구에서는 UM3.0 수치예보 자료에 대한 공간적 편의 보정 분석을 위해서 Bayesian Kriging 기반의 MFBC 기법을 도입하였다. UM3.0 등의 수치예보 자료는 일반적으로 시공간적으로 편의 보정을 거쳐 활용한다. 특히, 온도 같은 시공간적으로 연속적인 특성을 갖는 자료들은 시공간적 내삽을 통해 비교 자료를 확보할 수 있다. 하지만 강수량 같은 시공간적으로 불연속 특성을 가진 변량의 경우 시공간적 불연속 특성 때문에 단순히 내삽하기가 어렵다. 본 연구에서 이러한 문제를

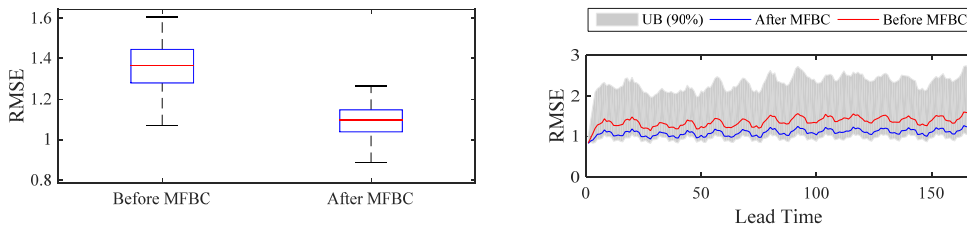


Fig. 8. Comparison of statistics of UM3.0 model precipitation forecast by lead time

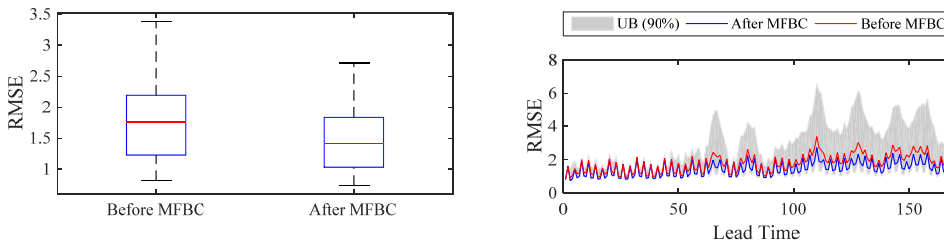


Fig. 9. Comparison of statistics of UM3.0 model cumulative precipitation forecast by lead time

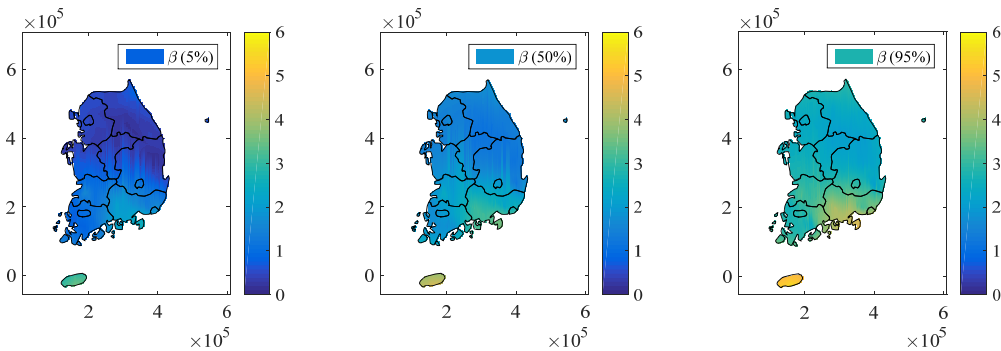


Fig. 10. Spatial distribution of β value estimates by Bayesian Kriging based MFBC method (left to right 5%, 50%, 95%)

개선하기 위해서 기상청 관할의 관측소 강수량과 관측소 위치에 근접한 격자의 강수량 자료를 이용하여 편의 보정계수를 추정함과 동시에 이를 Bayesian Kriging 기법을 적용하여 공간적으로 확장하였다. 기상청 관할의 442개 관측소에서 100개의 관측소 정보를 이용하여 모형을 구축하였으며 교차검정 관점에서 입력자료로 사용한 100개 관측소를 제외한 342개 관측소에 대해서 보정계수를 추정된 결과 미계측 지점에 대해서도 보정계수를 효과적으로 보간해주는 것을 확인하였다.

본 연구에서 산정된 편의 보정계수를 이용하여 UM3.0 강수 예보자료를 편의 보정하였다. 예보시간에 따른 편의 보정 기법 적용결과 시간에 따라 강수량을 누적하지 않을 경우 선행 시간에 따른 변화가 크지 않았다. 즉, 편의보정 후 예보시간에 따라 강수량을 누적하지 않고 각 시점에서 추정된 오차에 대한 통계치를 분석한 결과 선행예보시간에 따른 오차의 증가가 나타나고 있으나, 유의한 증가는 없는 것으로 평가되었다. 반면, 선행시간까지 강수량을 누적하여 오차를 분석할 결과 선행시간 60시간이 넘어서면서 bias가 급격히 증가하는 것을 확인하였다. 또한, 선행 시간에 따른 상관계수도 선행시간 60시간 이후에서는 0.6 이하로 떨어지는 특성을 확인할 수 있다.

본 연구 결과를 바탕으로 UM3.0 예보 모델 결과를 편의보정 하여 활용한다면 시공간적으로 1시간과 3 km이라는 비교적 고해상도로 7일까지 예보가 수행될 수 있을 것으로 판단되며, 선행 예보시간 관점에서 60시간까지는 오차의 증가가 크지 않은 점을 고려할 때 UM3.0 예보 모델의 강수량은 72시간(3일) 정도까지 예측강우 정보를 활용하는 것이 신뢰성 측면에서 유리할 것으로 판단된다.

향후 연구로서 강수량에 따른 편의보정계수에 차등을 주는 Mixture 분포 형태의 방법론 개발이 필요할 것으로 판단되며, 이러한 과정을 통해서 극치강수량 편의보정에 유리할 수 있을 것으로 기대된다.

감사의 글

본 연구는 국토교통부 물관리연구사업의 연구비지원(12 기술혁신C02)에 의해 수행되었습니다.

References

- Anagnostou, E. N., Krajewski, W. F., Seo, D. J., and Johnson, E. R. (1998). "Mean-field rainfall bias studies for WSR-88D." *Journal of Hydrologic Engineering*, Vol. 3, No. 3, pp. 149-159.
- Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC Press.
- Berg, P., Feldmann, H., and Panitz, H. J. (2012). "Bias correction of high resolution regional climate model data." *Journal of Hydrology*, Vol. 448, pp. 80-92.
- Bordoy, R., and Burlando, P. (2013). "Bias correction of regional climate model simulations in a region of complex orography." *Journal of Applied Meteorology and Climatology*, Vol. 52, No. 1, pp. 82-101.
- Borga, M. (2002). "Accuracy of radar rainfall estimates for streamflow simulation." *Journal of Hydrology*, Vol. 267, No. 1-2, pp. 26-39.
- Buizza, R. (2010). "Horizontal resolution impact on short-and long-range forecast error." *Quarterly Journal of the Royal Meteorological Society*, Vol. 136, No. 649, pp. 1020-1035.
- Durai, V. R., and Bhradwaj, R. (2014). "Evaluation of statistical bias correction methods for numerical weather prediction model forecasts of maximum and minimum temperatures." *Natural Hazards*, Vol. 73, No. 3, pp. 1229-1254.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*. Texts in Statistical Science Series.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. Vol. 2. Boca Raton, FL: CRC press.
- Hay, L. E., Wilby, R. L., and Leavesley, G. H. (2000). "A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States." *Journal of the American Water Resources Association*, Vol. 36, No. 2, pp. 387-397.
- Isaaks, E. H., and Srivastava, R. M. (2001). "An introduction to applied geostatistics. 1989. New York, USA: Oxford University Press. Jones DR, A taxonomy of global optimization methods based on response surfaces." *Journal of Global Optimization*, Vol. 23, pp. 345-383.
- Jakob Themeßl, M., Gobiet, A., and Leuprecht, A. (2011). "Empirical-statistical downscaling and error correction of daily precipitation from regional climate models." *International Journal of Climatology*, Vol. 31, No. 10, pp. 1530-1544.
- Jee, J. B., and Kim, S. (2017). "Sensitivity study on high-resolution WRF precipitation forecast for a heavy rainfall event." *Atmosphere*, Vol. 8, No. 6, p. 96.
- Kim, S. H., Kim, H. M., Kay, J. K., and Lee S. W. (2015). "Development and evaluation of the high resolution limited area ensemble prediction system in the Korea meteorological administration." *Atmosphere*, Vol. 25, No. 1, pp. 67-83.
- Kwon, H. H., and Moon, Y. I. (2005). "A study of long-memory analysis of rainfall time series." *Journal of The Korean Society of Civil Engineers*, Vol. 25, No. 4B, pp. 237-246.
- Kwon, H. H., and Moon, Y. I. (2007). "Development of statistical seasonal rainfall model considering climate information and typhoon characteristics." *Journal of The Korean Society of Civil Engineers*, Vol. 27, No. 1B, pp. 45-52.
- Kwon, H. H., Brown, C., and Lall, U. (2008). "Climate informed flood frequency analysis and prediction in Montana using

- hierarchical Bayesian modeling.” *Geophysical Research Letters*, Vol. 35, No. 5, pp. L05404.
- Lai, Y. S., Zhou, X. N., Utzinger, J., and Vounatsou, P. (2013). “Bayesian geostatistical modelling of soil-transmitted helminth survey data.” *Parasites & Vectors*, Vol. 6, No. 1, p. 359.
- Lee, D., Lee, J., and Kim, D. (2014). “Applicability of a space-time rainfall downscaling algorithm based on multifractal framework in modeling heavy rainfall events.” *Journal of Korea Water Resources Association*, Vol. 47, No. 9, pp. 839-852.
- Lee, J., Jang, S., and Kim, S. (2015). “Evaluation of the WRF model in reproducing the spatio-temporal pattern of a heavy rainfall event.” *Journal of Korean Society of Hazard Mitigation*, Vol. 15, No. 5, pp. 255-265.
- Lenderink, G., Buishand, A., and Deursen, W. V. (2007). “Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach.” *Hydrology and Earth System Sciences*, Vol. 11, No. 3, pp. 1145-1159.
- Mittermaier, M., and Roberts, N. (2010). “Intercomparison of spatial forecast verification methods: identifying skillful spatial scales using the fractions skill score.” *Weather and Forecasting*, Vol. 25, No. 1, pp. 343-354.
- Sandwell, D. T. (1987). “Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data.” *Geophysical Research Letters*, Vol. 14, No. 2, pp. 139-142.
- Scholte, R. G., Gosoniu, L., Malone, J. B., Chammartin, F., Utzinger, J., and Vounatsou, P. (2014). “Predictive risk mapping of schistosomiasis in Brazil using Bayesian geostatistical models.” *Acta Tropica*, Vol. 132, pp. 57-63.
- Seo, D. J., Breidenbach, J. P., and Johnson, E. R. (1999). “Real-time estimation of mean field bias in radar rainfall data.” *Journal of Hydrology*, Vol. 223, No. 3-4, pp. 131-147.
- Siddique, R., Mejia, A., Brown, J., Reed, S., and Ahnert, P. (2015). “Verification of precipitation forecasts from two numerical weather prediction models in the Middle Atlantic Region of the USA: A precursory analysis to hydrologic forecasting.” *Journal of Hydrology*, Vol. 529, pp. 1390-1406.
- Slater, H., and Michael, E. (2013). “Mapping, Bayesian geostatistical analysis and spatial prediction of lymphatic filariasis prevalence in Africa.” *PloS one*, Vol. 8, No. 8, pp. e71574.
- Son, C. Y., Jeong, Y. R., Han, S. H., and Cho, Y. H. (2017). “Assessment of predictability of categorical probabilistic long-term forecasts and its quantification for efficient water resources management.” *Journal of Korea Water Resources Association*, Vol. 50, No. 8, pp. 563-577.
- Tobler, W. R. (1970). “A computer movie simulating urban growth in the Detroit region.” *Economic Geography*, Vol. 46(sup1), pp. 234-240.
- Woodcock, F., and Engel, C. (2005). “Operational consensus forecasts.” *Weather and Forecasting*, Vol. 20, No. 1, pp. 101-111.
- Wu, M. C., and Lin, G. F. (2017). “The very short-term rainfall forecasting for a mountainous watershed by means of an ensemble numerical weather prediction system in Taiwan.” *Journal of Hydrology*, Vol. 546, pp. 60-70.
- Yu, W. S., Yoon, S. S., Choi, M. K., and Jung, K. S. (2017). “Performance comparison of rainfall and flood forecasts using short-term numerical weather prediction data from Korea and Japan.” *Journal of Korea Water Resource Association*, Vol. 50, No. 8, pp. 537-549.
- Yucel, I., Onen, A., Yilmaz, K. K., and Gochis, D. J. (2015). “Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall.” *Journal of Hydrology*, Vol. 523, pp. 49-66.