신경망 기법과 GCM자료를 이용한 비정상성 확률강수량 산정

Nonstationary Probability Rainfall Estimation at Seoul Using Neural Networks and GCM Data

조현곤* · 김광섭**

Cho, Hyungon*, and Kim, Gwangseob**

Abstract

Estimation of probability precipitation for the design of hydraulics structure is an important factor in determining the strength and scale. Generally, we use probability precipitation through the assumption of stationarity. Regional impact of climate change causes the necessity of the nonstationary probability rainfall estimation. Several nonlinear regression methods were applied to deal with the nonstationary behavior of extreme rainfall effectively by modeling the variability of the location and scale parameters of future extreme probability distribution. These regression models have drawback of forcing future change shape according to the regression types. In this study, the variability of the probable maximum rainfall for different durations in target years at Seoul was analyzed with different approaches such as logistic regression, power regression, and neural networks using GCM data associated with different green house gas emission scenarios. The increasing ratio of the future hourly probability rainfall using NN is larger than that using Logistic and Power models, and the increasing ratio of the future 24hr probability rainfall using NN is smaller than that using Logistic and Power models.

Key words : Neural Network, Nonstationary, Nonlinear Model, Probable Rainfall

요 지

이수와 치수 목적의 수공구조물 설계를 위한 확률강수량 산정은 구조물의 강도와 규모를 결정하는 중요한 요소이다. 기후변화의 지역적 영향으로 인해 비정상성 확률강수량 산정의 필요가 대두되고 있다. 비정상성을 가지는 극치강수량 변동 특성을 효과적으로 대응하기 위하여 미래 극치확률분포형의 위치 및 형상계수의 변동특성에 대하여 각종 비선형 회귀모형을 적용하여 모형화하는 기법들이 제시되었다. 이러한 비선형회귀모형들의 적용은 극치 수렴특성이 적용 기법에 따라 사전에 정해지는 단점이 있다. 본 연구에서는 Logistic 회귀모형, Power 회귀모형을 이용하여 산정한 서울지점 목표연도별, 지속시간별 확률강수량 결과와 인공신경망 기법과 온실가스 배출시나리오별 GCM 자료를 이용한 목표연도별, 지속시간별 확률강수량을 비교 분석하였다. 본 연구에 사용된 인공신경망 기법에 따른 미래 확률강수량은 Logistic, Power 모형 적용에 따른 미래 확률강수량에 비하여 1시간 확률강수량은 증가율이 크게 나타났으며, 24시간 확률강수량에 대해서는 상대적으로 증가율이 작게 나타났다.

핵심용어 : 신경망 모형, 비정상성, 비선형 모형, 확률강수량

1. 서 론

세계 각 국에서는 기후변화에 의한 환경변화에 대응하고 자 기후변화로 인한 지역적 영향에 대한 평가, 예측, 정책수립 등을 위해 여러 분야에서 연구가 이루어지고 있으며 기후변

화 대응 시스템 구축은 현재 전세계적으로 직면하고 있는 중요한 과제가 되었다. 기후변화에 관련한 수문현상의 변화 특성 중에서 호우특성의 변화로 인한 수재해 특성의 변화가 현실화되고 있는 상황에서 미래의 극치강수량 특성변화를 분석하고 예측하기 위한 여러 연구가 수행되고 있는 상황이

*경북대학교 건축·토목공학부 박사과정(E-mail: Gonny@knu.ac.kr)

Ph.D. Candidate, Dept. of Arch. and Civil Eng., Kyungpook National Univ.

**교신저자, 정회원, 경북대학교 건축·토목공학부 교수(Tel: +82-53-939-6859, Fax: +82-53-950-6564, E-mail: kimgss@knu.ac.kr)

Corresponding Author, Member, Professor, Dept. of Arch. and Civil Eng., Kyungpook National Univ.

다. IPCC(2014) 5차보고서에 따르면 지구평균 기온이 112년(1901-2012) 동안 약 0.89°C 상승하였다. 이는 4차보고서 100년(906-2005) 동안 0.74°C 상승한 것과 비교해 볼 때 지구의 평균기온 상승이 가속화되고 있다는 것이다. 지구의 평균 기온뿐만 아니라 해수면 상승 또한 가속화되고 있으며 2100년에는 기온이 약 3.7°C 증가하며 해수면은 약 63 cm 증가할 것이라고 전망하였다. 또한 집중호우, 돌방홍수, 가뭄 등 이상기후 현상에 의한 인명 및 경제적 피해는 증가하고 있으며 자연재해 발생과 규모 또한 증가하고 있는 추세를 보이고 있다(KMA, 2015). 이러한 기후변화 추세를 고려하였을 때 수공구조물과 배수시설 등 수리·수문 시설 설계에 일반적인 방법으로 산정된 설계 강수량의 적용은 한계를 가진다. 최근에는 강수의 정상성을 가정한 수자원관리 시스템 운영에 있어 근래의 극치 강수량의 변화 특성을 비추어 볼 때 강수 정상성에 기초한 전통적인 빈도해석을 통해 산정된 설계 강수량의 적용에 대하여 타당성 재검토가 요구되고 있다(Kwon and Kim, 2009; Jang et al., 2011). 이러한 연구결과들로 인하여 극치강수량 표본자료의 평균, 분산 등 통계특성이 시간이 지남에 따라 변화하는 비정상성을 고려한 빈도해석의 필요성이 주장되고 있다. 비정상성 빈도해석은 매개변수나 분포가 변화하기 때문에 기존의 정상성을 가정한 빈도해석과 다른 접근법이 필요하다. 또한 특정 재현기간의 극치사상을 추정하기 위해 수문학적 시계열 자료를 모형화하는 통계학적 방법은 자료의 비정상성을 고려할 수 없다. 과거부터 현재까지 많은 연구자들에 의해 강수자료의 비정상성을 고려한 강수빈도해석을 위하여 극치분포의 매개변수 추정기법, 비정상성 빈도해석 기법 등 다양한 연구가 수행되었다(Leclerc and Ouarda, 2007; Hanel and Buishand, 2010; Villarini et al., 2010; Shin et al., 2008; Zhang et al., 2014).

국내외에서 강수 자료의 비정상성 빈도해석을 통한 확률강수량 산정과 관련한 연구가 활발히 이루어지고 있다. 국외의 연구동향은 He et al.(2006)은 Gumbel 분포 및 Log Pearson Type-III 분포의 매개변수 변화에 대하여 회귀분석을 통해 홍수빈도 해석을 하였으며 Renard et al.(2006)은 Bayesian 기법을 여러 가지의 극치분포형(exponential, Generalized Pareto(GP), Gumbel, Generalized Extreme Value)에 적용하였다. Sugahara et al.(2009)은 Peaks-Over-Threshold(POT)와 GP 분포를 이용하여 Brazil Sao Paulo 지역의 1933년-2005년까지 일강수량 극치 빈도해석을 하였으며 비모수적 추세검정을 위하여 Mann-Kendall 검정을 수행하였다. 국내의 연구동향은 Kwon et al.(2009)은 매년 강우자료가 추가되기 때문에 Gumbel 분포의 매개변수가 변화하고 결과적으로 확률강수량이 변화하는 점을 감안하여 초기 20년을 기준으로 1년씩 누적하면서 누적 평균과 매개변수의 상관관계를 이용하여 확률강수량을 산정하였다. Kim et al.(2012)이 초기 20년의 기준으로 1년씩 누적한 누적 평균 연 최대 강수량의 따라

임계값의 변화를 고려하여 추출된 자료의 재현기간별 확률강수량의 비교분석을 통해 임계값의 변화가 연구결과에 미치는 영향력을 제시하고 분포의 형태에 따른 임계값의 민감도 차이를 비교 분석하여 임계값의 변화에 영향을 적게 받는 분포를 제안하였다. Park et al.(2014)은 비연초과치 강수 자료를 활용한 빈도해석 방법을 제시하였으며, Shin et al.(2013)은 RCP 8.5 시나리오 자료를 정상성 빈도해석 결과와 GEV 분포를 이용한 비선형회귀를 통한 매개변수 추정과 산정된 빈도해석 결과를 비교하여 가까운 미래의 확률강수량의 경우 비정상성 빈도해석 결과가 적합하다고 하였다.

본 연구에서는 비정상성을 고려한 재현기간에 따른 확률강수량을 산정하기 위하여 Gumbel 분포를 이용하였으며 목표 연도에 해당하는 Gumbel 분포형의 매개변수를 추정하기 위하여 누적 평균 연 최대 강수량 변화에 대한 매개변수 변화 상관성을 연계하였다. 이를 위하여 선행연구들에서 적용된 각종 회귀모형은 사용된 기법에 따라 극치 수렴특성이 정해지며 목표연도 확률강수량이 과거 자료에 의해 결정되며 사용하는 회귀모형에 따라 주관적인 확률강수량을 산정하는 단점을 보완하고 AR5 GCM를 사용하여 미래 확률강수량 산정하여 기존의 회귀모형 적용에 따른 미래변화를 강제하는 단점을 개선하고 외삽하는 접근방법으로는 예측할 수 없는 온실가스 배출시나리오별로 확률강수량의 변화를 산정하고 분석하였다. 본 연구에서는 기존의 정상성을 가정한 확률강수량과 Logistic 회귀모형, Power 회귀모형을 적용하여 산정한 서울 지점 목표연도별, 지속시간별 확률강수량 결과와 인공신경망 기법과 온실가스 배출시나리오별 GCM 자료를 이용한 목표 연도별, 지속시간별 확률강수량을 비교 분석하였다.

2. 대상 지점 및 연구 방법

2.1 대상 지점 선정 및 자료 수집

빈도해석을 통한 확률강수량 산정은 강수분포의 극치를 다루고 있으며 극치분포의 테일(tail)을 사용함으로 자료의 보유기간이 확보되지 않으면 분포형 선정이나 매개변수 추정방법에 따라 확률강수량의 변동이 크게 나타나는 근본적인 한계를 가진다. 따라서 본 연구에서는 상대적으로 장기간의 가용 자료기간을 가진 서울 지점을 대상으로 1961년-2016년 기간 동안의 시간 단위의 강수자료를 이용하였다. 본 연구에 사용된 GCM 자료는 기상청에서 제공하는 HadGEM2-A0(134km)의 역학적 상세화를 통해 생성된 HadGEM3-RA(12.5km) 강수자료의 편의보정된 서울지점의 미래 강수량 자료로 IPCC RCP 2.6, 4.5, 6.0, 8.5 시나리오 별로 수집하였다. 수집된 GCM 자료 기간은 1956년~2099년(144년)이며 일 단위 자료로 인공신경망 구성의 입력자료를 구축을 위하여 강수 시자료 생산에 CM(Contraction Mapping)기법을 적용하였다. CM기법은 강수자료와 같은 자기유사 가정을 적용할 수 있는 자료에 대하여 자료 사이의

값을 일정 비율 이하로 축소시키는 함수이다. 함수 $f: D \rightarrow D$ 에 대하여, 임의의 $x, y \in D$ 에 대하여 $d(f(x), f(y)) \leq r \cdot d(x, y)$ 인 $0 \leq r < 1$ 가 존재하는 경우 f 를 (D, d) 위의 축약사상이라고 한다(Banach, 1922). Fig. 1은 도식화된 CM 기법의 downscaling 과정을 나타낸다.

Fig. 1. Schematic Diagram of the Downscaling Algorithm Using Contraction Mapping

2.2 확률분포함수 및 매개변수 추정

본 연구에서 1961년-2016년 기간 동안의 연 최대 강수량 분포의 적합분포로 일반화된 극치분포(Generalized Extreme Value Distribution, GEV)의 변형 중의 하나인 Gumbel 분포를 선정하였다. GEV 분포는 최대치 혹은 최소치 등 극치자료의 발생빈도 분석에 자주 이용되며 확률밀도함수는 다음과 같다(Eq. (1)).

$$f(x) = \frac{1}{\alpha} \left[1 - \beta \left(\frac{x - x_0}{\alpha} \right) \right]^{\frac{1}{\beta} - 1} \exp \left(- \left[1 - \beta \left(\frac{x - x_0}{\alpha} \right) \right]^{1/\beta} \right) \quad (1)$$

여기서 α , β , x_0 는 축척매개변수(scale parameter), 형상매개변수(shape parameter), 위치매개변수(location parameter)라 부르며 형상매개변수 $\beta=0$ 일 때를 Type- I (EV1)분포, Gumbel 분포라 하고 연 최대 홍수량 및 강수량 빈도해석에 일반적으로 쓰이고 있다. 국내에서는 확률강수량 산정 시 Gumbel 분포를 사용하고 있으며 설계홍수량 산정시에도 Gumbel 분포 사용을 기준으로 하고 있다. 따라서 본 연구에서는 연 최대 강수량의 적합분포를 Gumbel 분포로 선정하였으며 분포의 매개변수 추정은 표본의 크기가 작아도 상대적으로 영향을 적게 받으며 자료에 이상치가 있는 경우에도 왜곡특성이 크게 나타나지 않는 장점이 있는 확률가중모멘트법을 적용하였다. 다음은 Gumbel 분포의 확률밀도함수로 가용자료에 대한 축척매개변수와 위치매개변수를 산정하여 미래의 연 최대치 강수량의 변화에 따른 매개변수 상관성을 분석하여 확률밀도함수 변화를 제시하고자 한다(Eq. (2)).

$$f(x) = \frac{1}{\alpha} \exp \left[- \frac{1}{\alpha} (x - x_0) - \exp \left(- \frac{1}{\alpha} (x - x_0) \right) \right], \quad -\infty < x < \infty \quad (2)$$

2.3 비선형 회귀 모형

연 최대 강수량을 선형회귀 모형으로 추정할 경우 외삽기간 길어짐에 따라 물리적으로 타당하지 않은 매우 큰 값을 예측하는 단점을 가지므로 선형회귀모형의 적용에는 한계를 가진다. 본 연구에서 물리적으로 타당하지 않은 값의 외삽 한계를 극복하기 위하여 변수의 수렴특성을 외적으로 제한할 수 있는 비선형 회귀 모형인 Logistic 회귀모형과 Power 회귀모형을 이용하여 연 최대 강수량을 산정하였다.

2.3.1 Logistic모형

Logistic회귀모형은 독립변수의 선형결합 확률을 이용한 선형모형의 특수한 경우로서 Cox가 1958년 처음 제안하였다. 일반적인 회귀모형에서 독립변수가 $-\infty$ 에서 $+\infty$ 의 범위를 가지면 종속변수 또한 $-\infty$ 에서 $+\infty$ 의 범위를 가지는 것에 반해 독립변수와 종속변수를 0-1사이에 확률로 관계를 나타낸다. 또한 독립변수가 종속변수에 미치는 영향의 정도를 보다 정량적으로 표현할 수 있는 장점을 가지고 있다. Logistic 함수는 다음과 같다(Eq. (3)).

$$p(x) = \frac{e^{\delta_0 + \delta_1 x_1 + \dots + \delta_i x_i}}{1 + e^{\delta_0 + \delta_1 x_1 + \dots + \delta_i x_i}}, \quad i = 1, 2, \dots, N \quad (3)$$

여기서 δ 는 회귀계수, x 는 독립변수이다. 미래변화에 대한 Logistic 회귀모형 적용에 있어 주의해야할 점은 특정값으로의 수렴성이 높은 함수특성을 가지므로 대상값의 물리적 거동이 이와 상반될 때 적용에 주의를 요하는 한계를 가진다.

2.3.2 Power모형

Power 회귀모형 또는 log-log 회귀모형이라고 부르며 선형회귀와 비슷하나 종속변수와 독립변수에 로그 변환을 취한 후 선형회귀분석을 실시하여 회귀계수를 구한다는 점에서 다르다(Good and Hardin, 2012). 일반적으로 Power 회귀모형 분석은 선형회귀분석에 기초하여 독립변수와 종속변수의 관계가 선형이지 않거나, 한 변수 내 수치적 차이가 커서 실제의 선형관계보다 더 완전한 선형관계를 보이는 경우에 사용한다. 다음은 Power 회귀모형의 일반식이다(Eq. (4)).

$$y = \alpha x^\beta \quad (4)$$

여기서 회귀계수 α , β 는 변수에 로그를 취한 후 선형회귀를 실시한 $\ln y = \beta \ln x + \delta$ 로 산정한다. 본 연구에서 종속변수는 1961년-2016년 기간 동안의 지속시간별 연 최대강수량이며 독립변수는 각 연도를 취하였다. Power함수의 특성은 미래 거동에 대하여 점진적으로 증가하는 거동을 잘 구현할 수 있으나 미래 거동의 외삽형태를 과거 자료에 따라 정해진다. 전술한 Logistic 회귀모형뿐만 아니라 전체적인 회귀모형

의 한계인 회귀모형 선택에 따라 미래 거동형태를 주관적으로 결정하게되는 한계를 가진다.

2.4 인공신경망 모형(Artificial Neural Network, ANN)

인공신경망 모형은 두뇌가 의사결정을 하는 과정을 모방하여 개발된 통계학적 모형으로 1943년도에 McCulloch and Pitts에 의해 처음 제안되었으며 현재에는 의료, 신호처리 등 다양한 분야에서 활용되고 있다. 본 연구에서 시나리오에 따른 기후모형의 결과값을 이용하여 미래 극치강수 특성 변화에 사용되는 목표연도 평균 연 최대 강수량의 변화를 예측하기 위하여 인공신경망 기법을 적용하였다. 인공신경망은 다차원 자료의 상호관계를 표현한 자료처리 기법의 하나로 다층인공신경망(multi-layer neural network)의 경우 기본적으로 입력층(input layer), 은닉층(hidden layer), 출력층(output layer)으로 구성되며 각 층은 여러 개의 노드로 구성되고 각 층간의 노드는 전이함수와 가중치로 연결되어 있다. 학습과정을 통하여 전이함수의 최적 매개변수와 가중치를 결정한다. 전이함수는 일반적으로 Step 함수, Sigmoid 함수, Relu 함수 등이 있다. 학습은 선형회귀처럼 최소제곱법 등을 사용하여 각 노드마다 가중치와 오차를 산정한다. 본 연구에서 입력층 입력자료는 연강수량, 지속시간별 연 최대 강수량 등으로 구성하였으며 각 입력노드 값의 가중합이 은닉층의 노드값을 형성한다(Eq. (5)).

$$h_j = \sum_{i=1}^m w_{ji} x_i \quad (j=1, \dots, n) \quad (5)$$

여기서, x 는 입력노드값을 나타내며 w_{ji} 는 신경망 가중치이며 m 은 입력노드의 총 개수를 나타내며 n 은 은닉층의 총 노드수를 나타낸다. 은닉층 노드값을 전이함수의 종속변수로 하여 산정된 함수값에 가중치를 적용한 값들의 합을 출력값으로 하여 학습기간에 대한 연 최대치 강수량과의 적합성을 개선하는 학습과정을 수행하였다(Eq. (6)).

$$o_k = \sum_{j=1}^n w_{kj} f(h_j) = \sum_{j=1}^n w_{kj} f\left(\sum_{i=1}^m w_{ji} x_i\right) \quad (k=1, \dots, l) \quad (6)$$

3. 적용 결과 및 고찰

본 연구에서의 비정상성을 고려한 빈도해석의 절차는 서울 지점을 대상으로 강수관측자료를 이용하여 처음 1961년-1980년 동안의 각 연도별 최대강수량의 평균을 산정하고 이후 1년씩 누적하면서 연도별 최대강수량의 평균을 산정한 누적 평균 연 최대 강수량과 동일 기간 강수자료의 Gumbel 분포의 매개변수 계산하고 선형회귀 모형을 구축하였다. 관측 자료를 바탕으로 계산된 누적 평균 연 최대 강수량

자료를 이용하여 Logistic모형과 Power모형을 통하여 목표연도의 지속시간별 누적 평균 연 최대 강수량을 비선형 회귀 방법으로 산정하였다. 또한 GCM 자료를 CM기법으로 다운스케일한 후 예측된 강수의 지속시간별 연 최대 강수량을 계산하고 이를 인공신경망의 입력자료로 목표연도의 연 최대 강수량을 산정하였다. 산정된 누적 평균 연 최대 강수량을 이용하여 목표연도의 Gumbel 분포의 매개변수를 추정하고 확률강수량을 산정하였다. 사용된 지속시간은 1, 12, 24시간이며 목표연도는 2050, 2080, 2100년으로 선정하였다. 누적 평균 연 최대 강수량은 CMAMP(Cumulative Mean Annual Maximum Precipitation)라하고 이후 CMAMP로 표현한다.

3.1 관측자료의 누적 평균 연 최대 강수량 및 Gumbel 분포의 매개변수

지속시간별 연 최대치 자료에 대한 Gumbel 분포의 매개변수 추정을 위하여 관측기간 1961년-2016년에서 처음은 1961년-1980년을 동안의 매개변수를 추정하고 1년씩 누적하면서 위치매개변수와 축척매개변수를 추정하여 매개변수 변동과 누적 평균 연 최대 강수량의 상관관계를 분석하였다. Fig. 2는 지속시간 24시간에 대한 누적 평균 연 최대치 변화에 따른 위치매개변수와 축척매개변수의 상관관계를 나타낸다. 본 연구에서는 관측자료에 기반하여 극치분포형의 매개변수 변화가 지속시간별 누적 평균 연 최대치 자료의 변화에 따른 상관성에 기반하여 지속시간별 미래 누적 평균 연 최대치의 변화의 예측에 따라 해당 자료의 확률분포형의 변화를 예측하고자 하였으므로 전술한 바와 같이 목표연도의 누적 평균 연 최대치 비선형 거동을 Logistic 모형, Power 모형 및 GCM 자료와 인공신경망을 연계한 예측을 통하여 비교분석하였다.

3.2 Logistic 회귀모형을 이용한 목표연도 CMAMP 산정

관측으로부터 산정된 CMAMP를 이용하여 Logistic 회귀모형을 이용하여 목표연도별 CMAMP를 산정하였다. Fig. 3은 Logistic 회귀모형을 이용한 결과이며 대부분의 지속시간에 대하여 CMAMP 거동이 Logistic 회귀곡선에 잘 따르는 것으로 나타났다. 적합도 평가를 위하여 1980년-2016년 기간 동안 Logistic 회귀모형의 지속시간별 통계검정 결과 결정계수는 약 0.78-0.85로 나타났으며 목표연도별 CMAMP는 다음과 같다(Table 1).

Logistic 기법은 독립변수와 종속변수를 0-1사이에 확률로 관계를 나타내기 때문에 추정기간이 길어질수록 특정값으로 수렴한다는 단점이 있다. 이러한 방법론적 한계는 Table 1 결과에서 확인되듯이 지속시간 2, 6, 18, 24시간 등에서 목표연도 2050년과 2080년에는 CMAMP의가 증가하나 2080년과 2100년에는 차이가 없으며 다른 지속시간에서도 CMAMP의 차이는 미미하다.

Fig. 2. Sample Relationship between Annual Maximum (24hr precipitation) and Location Parameter, Relationship between Annual Maximum (24hr precipitation) and Scale Parameter

Table 1. Estimation of the Target Year CMAMP and the Coefficient of Determination using the Logistic Model

	Target year			R^2		Target year			R^2
	2050	2080	2100			2050	2080	2100	
1h	50.73	50.73	50.73	0.8593	12h	164.49	164.49	164.49	0.8006
2h	76.76	76.76	76.76	0.796	15h	176.85	176.85	176.85	0.8112
3h	97.70	97.70	97.70	0.8408	18h	189.97	189.97	189.97	0.8185
6h	130.92	130.92	130.92	0.7722	21h	198.96	198.96	198.96	0.8222
9h	152.48	152.48	152.48	0.8066	24h	206.53	206.53	206.53	0.8286

Fig. 3. Logistic Regression Curves of Average Annual Maximum Precipitation for Different Durations

3.3 Power 회귀모형을 이용한 목표연도 CMAMP 산정

전술한 Logistic 회귀모형과 같이 관측으로부터 산정된 CMAMP를 이용하여 Power 회귀모형을 이용하여 목표연도별 CMAMP를 산정하였다. Fig. 4는 Power 회귀모형을 이용한 결과이며 적합도 평가를 위한 Logistic 회귀모형과 비교결과 지속시간에 대하여 CMAMP 거동이 Logistic 회귀곡선에 비해 적합도가 낮게 평가되었다. 1980년-2016년 기간동안 Power 회귀모형의 지속시간별 통계검정 결과 결정계수는 약 0.55-0.7로 나타났으며 목표연도별 CMAMP는 다음과 같다(Table 2). Power 기법은 추정기간이 길어질수록 수렴하

는 Logistic 기법의 단점을 보완할 수 있다. 그러나 Table 2 결과에서 확인되듯이 지속시간 전체에 대하여 적합도가 Logistic 모형보다 낮으며 목표연도가 길어질수록 회귀값 또한 증가하는 거동을 강제한다는 단점을 가진다.

3.4 인공신경망 모형을 이용한 목표연도 CMAMP 산정

인공신경망 모형을 이용하여 목표연도의 CMAMP를 산정하기 위하여 다음과 같이 인공신경망 모형을 구축하였다. 은닉층의 노드는 RCP 시나리오별, 지속시간별로 10개에서 30개 사이로 구성하였으며 학습과정은 Levenberg-Marquardt

Fig. 4. Power Regression Curve of Average Annual Maximum Precipitation for Different Durations

Table 2. Estimation of the Target Year CMAMP and the Coefficient of Determination using the Power Model

	Target year			R^2		Target year			R^2
	2050	2080	2100			2050	2080	2100	
1h	53.18	55.31	56.43	0.7272	12h	165.92	170.04	172.19	0.5762
2h	78.51	80.78	81.96	0.5548	15h	174.66	178.74	180.87	0.5099
3h	99.76	102.71	104.26	0.5031	18h	187.88	192.83	195.42	0.4990
6h	130.93	134.18	135.87	0.5438	21h	196.55	201.71	204.40	0.4660
9h	151.39	154.96	156.82	0.5210	24h	204.85	210.37	213.26	0.4791

backpropagation algorithm을 이용하였다. 입력변수는 연강수량, 이동평균으로 구한 지속시간별 연 최대 강수량 9개, Block maxima 방법으로 구한 지속시간별 연 최대 강수량 9개로 총 19개로 구성하였으며 출력변수는 CMAMP 1개이다. 인공신경망 훈련을 위하여 자료의 70%를 이용하였으며 검정 과정에는 30%의 자료를 사용하였다. 각 지속시간에 대한 훈련을 하였으며 RCP 2.6, 4.5, 6.0, 8.5에 따라서 각각 학습과정을 수행하였다. 검정통계량은 MSE(Mean Square Error), R^2 두 가지로 평가하였으며 분석결과 RCP 2.6의 지속시간 1시간 CMAMP와 RCP 8.5의 지속시간 1시간 및 24시간을 제외한 결정계수가 평균 0.7-0.85의 값을 보여 인공신경망으로 산정된 예측값이 전반적으로 적합하다고 판단되었다(Table 3).

목표연도별로 추정된 CMAMP 값은 Table 4에서 제시하였다. 누적 평균 연 최대 강수량 회귀 모형을 구축하고 적합도 분석 결과 Logistic 회귀모형의 적합도가 Power 회귀모형보다 MSE, R^2 에서 더 우수한 검정량을 보였음에도 불구하고 미래의 극치거동이 특정값에 신속하게 수렴함으로 미래변화를 결정한다는 한계를 가진다. GCM 자료와 연계하여 목표연도 CMAMP 예측시 RCP 8.5 제외한 시나리오에서 상대적으로 낮은 값을 예측하였다.

3.5 Gumbel 매개변수 회귀 산정과 비정상성 확률강수량 산정

전술한 Logistic모형과 Power모형, 인공신경망 모형으로 생산된 목표연도별 CMAMP를 이용하여 CMAMP와 Gumbel 매개변수의 선형회귀를 통하여 목표연도별 Gumbel의 위치, 척도 매개변수를 추정하였다(Table 5). 정상성을 가정한 목표연도별 강수 지속시간에 따른 확률강수량을 재현기간 100년

까지 산정하고 비선형 회귀모형을 통해 추정된 Gumbel 분포의 매개변수를 이용하여 목표연도별 강수지속시간에 따른 확률강수량을 산정하였다(Fig. 5). Fig. 5에서 S는 1961년-2016년 기간 동안의 자료를 이용한 정상성을 가정한 확률강수량이며 L은 Logistic 회귀분석을 통한 확률강수량, P는 Power 모형을 이용한 확률강수량, C2, C4, C6, C8을 CM기법으로 다운스케일한 RCP 시나리오 2.6, 4.5, 6.0, 8.5 자료를 인공신경망을 통해 생산된 확률강수량이다. Logistic모형과 Power모형을 사용하여 산정된 강수량은 정상성을 가정한 확률강수량보다 대부분의 목표연도, 지속시간별 확률강수량에서 높게 산정되었다. GCM 자료를 사용하여 인공신경망 모형을 통해 산정된 확률강수량에서는 모의 결과에 의존하므로 RCP 2.6, RCP 4.5 시나리오 강수에서 정상성을 가정한 확률강수량보다 낮게 예측하는 구간도 발생하였다.

목표연도 2050년도 확률강수량 분석에서 RCP 8.5를 이용한 모형으로 산정된 확률강수량이 가장 높게 나왔으며 Power모형, Logistic모형을 이용한 순서로 확률강수량이 높았다. 지속시간 1시간을 제외한 모든 지속시간에서 정상성을 가정한 확률강수량 보다 높게 나오나 RCP 2.6과 RCP 4.5에서는 지속시간 12시간, 지속시간 24시간 강수에서도 정상성 보다 낮게 나왔다. Logistic 모형의 경우 지속시간 1시간에서는 정상성 대비 약 1.2% 감소하였으나 지속시간 12시간, 24시간 강수에서는 약 14-20% 증가하였으며 Power 모형 또한 Logistic 모형과 유사한 결과를 보였으나 목표연도가 늘어남에 따라 Logistic모형보다 더 큰 증가폭을 보였다. 인공신경망을 이용한 확률강수량 산정에서는 RCP 2.4, RCP 4.5에서는 약 3% 감소, RCP 6.0, RCP 8.5 시나리오에서 약 정상성 대비 5-15% 증가하였다. 목표연도 2080년도의 확률강수량 분석에서는 지속시간 1시간

Table 3. Test Statistics for ANN Results

RCP	D(h)	MSE	R^2 (train)	R^2 (test)	RCP	D(h)	MSE	R^2 (train)	R^2 (test)
2.6	1h	228.75	0.13	0.24	6.0	1h	133.86	0.81	0.57
	12h	977.29	0.73	0.70		12h	1581.54	0.85	0.75
	24h	2279.63	0.86	0.79		24h	3309.23	0.84	0.84
4.5	1h	84.88	0.94	0.72	8.5	1h	273.71	0.29	0.28
	12h	2244.23	0.82	0.76		12h	2327.04	0.58	0.59
	24h	1454.98	0.90	0.85		24h	8253.61	0.55	0.32

Table 4. Estimation of the Target Year CMAMP using the ANN Model

RCP	D(h)	2050	2080	2100	RCP	D(h)	2050	2080	2100
2.6	1	43.58	47.58	42.04	6.0	1	45.16	47.99	47.98
	12	117.76	157.70	117.58		12	145.77	151.82	163.72
	24	159.18	170.47	162.30		24	177.10	192.36	197.41
4.5	1	42.39	41.97	46.27	8.5	1	55.27	59.57	62.98
	12	141.92	145.18	154.34		12	170.61	162.81	169.90
	24	182.23	168.08	182.03		24	233.36	222.64	245.29

Table 5. Gumbel Parameters of Different Approaches at Target Years

Location parameter

D(h)	Logistic			Power			ANN(RCP 2.6)		
	2050	2080	2100	2050	2080	2100	2050	2080	2100
1	44.10	44.10	44.10	47.04	49.58	50.92	35.54	40.33	33.69
12	133.62	133.62	133.62	134.64	137.56	139.08	100.46	128.80	100.34
24	158.95	158.95	158.95	157.85	161.49	163.40	127.68	135.14	129.74
D(h)	ANN(RCP 4.5)			ANN(RCP 6.0)			ANN(RCP 8.5)		
	2050	2080	2100	2050	2080	2100	2050	2080	2100
1	34.11	33.61	38.76	37.43	40.82	40.81	49.54	54.69	58.77
12	117.61	119.92	126.42	120.34	124.63	133.08	137.97	132.43	137.46
24	142.90	133.56	142.77	139.52	149.60	152.93	176.67	169.59	184.55

Scale Parameter

D(h)	Logistic			Power			ANN(RCP 2.6)		
	2050	2080	2100	2050	2080	2100	2050	2080	2100
1	11.49	11.48	11.48	11.64	10.92	10.53	13.35	13.35	13.35
12	53.47	53.47	53.47	55.19	57.26	58.34	46.31	46.31	46.31
24	82.41	82.41	82.41	82.43	85.67	87.37	65.88	65.88	65.88
D(h)	ANN(RCP 4.5)			ANN(RCP 6.0)			ANN(RCP 8.5)		
	2050	2080	2100	2050	2080	2100	2050	2080	2100
1	13.35	13.35	13.35	13.35	13.35	13.35	13.35	13.35	13.35
12	46.31	46.31	46.31	46.31	46.31	46.31	46.31	46.31	46.31
24	65.88	65.88	65.88	65.88	65.88	65.88	65.88	65.88	65.88

Fig. 5. Sample IDF Curves for Different Target Years

의 증감률이 다른 지속시간별 확률강수량 보다 낮은 증가율 또는 감소를 보였다. Logistic 모형에서는 약 1% 감소하는 것으로 나타났으며 Power모형에서 약 정상성 대비 2% 증가율을 보였다. 인공신경망 모형에서는 RCP 8.5를 사용한 확률강수량이 약 5-18% 증가를 보여 가장 높은 증가율을 보였으며 RCP 6.5에서도 약 5%의 증가율을 보였으며 RCP 2.6, RCP 4.5에서는 평년과 비슷한 수준을 보였다. 2100년도에서 산정된 확률강수량 분석에서는 RCP 8.5에서 산정된 확률강수량이 가장 높게 증가율을 보이며 2080년도와 유사한 거동을 보였다.

목표연도별 지속시간에 따른 IDF곡선을 분석한 결과 Logistic, Power 모형은 목표연도 2050년의 지속시간 24시간, 목표연도 2100년의 지속시간 24시간 등 확률강수량 산정에서 95% 신뢰구간을 벗어나는 결과를 보였으며 RCP 8.5 시나리오를 사용한 인공신경망 모형은 지속시간 1시간 확률강수량 산정에서 신뢰구간을 벗어나는 결과를 보였다. 모든 목표연도의 지속시간 1시간 확률강수량은 Logistic모형에서는 감소를 Power모형에서는 정상성 대비 약 1-1.5% 증가하는 것이 확인 되었으나 인공신경망 모형을 통해 산정된 확률강수량은 RCP 2.6, 4.5 시나리오에서 지속시간과 상관없이 감소를 보였으며 RCP 6.0, 8.5에서는 약 5-20%의 강수량이 증가하는 것으로 나타났다. Logistic 모형과 Power 모형, 인공신경망 모형 모두 고려하는 강수 지속시간이 길어짐에 따라 정상성 대비 확률강수량의 증가율이 증가하는 것으로 나타났다. 그러나 Logistic 모형은 목표연도가 늘어남에 따라 정상성 대비 격차가 좁아지는 모습을 보였으며 Power모형은 증가폭이 더 커지는 모습을 보였다 (Table 6).

4. 결 론

본 연구에서는 1961년-2016년 기간 동안의 지속시간별 누적 평균 연 최대 강수량을 구하고 목표연도의 누적 평균 연 최대 강수량을 산정하기 위해 Logistic 회귀모형, Power 회귀모형을 구축하였다. 또한 GCM 자료와 연계한 인공신경

망 모형을 이용하여 1961년-2016년 동안의 누적 평균 연 최대 강수량을 훈련하여 목표연도의 누적 평균 연 강수량을 추정하였다. 관측자료로부터 누적 평균 연 최대 강수량과 Gumbel 분포의 위치, 축적매개변수의 선형관계를 이용하여 목표연도의 Gumbel 분포의 매개변수를 추정하고 이를 이용하여 지속시간별, 시나리오별 확률강수량을 산정하였다. 모형으로부터 산정된 확률강수량 분석을 위하여 정상성을 가정한 확률강수량과 비교 분석하였다. 본 연구의 주요 연구 결과는 다음과 같다.

누적 평균 연 최대 강수량 회귀 모형구축하고 적합도 분석 결과 Logistic 회귀모형의 적합도가 Power 회귀모형 보다 MSE, R^2 에서 더 우수한 검정량을 보였으나 기법의 특성상 미래 극치 거동을 신속하게 특정값에 수렴하게 하는 단점을 가진다. GCM 자료와 연계한 인공신경망 모형은 목표연도 누적 평균 연 최대 강수량을 RCP 8.5 시나리오를 제외하고는 상대적으로 낮게 예측하였다. 목표연도별 분석에서는 2050년 Logistic 모형의 경우 정상성 대비 약 15-25% 증가하였으며 Power 모형 또한 비슷한 결과를 보였다. 인공신경망 모형은 모든 RCP 2.6, 4.5 시나리오에서는 정상성 대비 감소를 RCP 6.0, 8.5에서는 증가하였다. 목표연도 2080년도의 확률강수량 분석에서는 Logistic 모형과 Power모형에서 정상성 대비 약 15-25%의 증가율을 보였으며 Power모형을 이용한 확률강수량이 Logistic 회귀모형보다 높은 증가율을 보였다. 인공신경망 모형에서는 RCP 8.5를 사용한 확률강수량이 7% 증가를 보였으며 시나리오별 증가율에서 가장 높은 증가율을 나타냈다. 2100년도에서 산정된 확률강수량 분석에서는 RCP 8.5, Power, Logistic 모형순으로 산정된 확률강수량의 증가율이 높게 나타났다. 지속시간에 따른 정상성 대비 확률강수량의 증감을 분석에서는 모든 목표연도에서 지속시간 24시간 확률강수량이 가장 큰 증가폭을 보였으나 RCP 2.6, 4.5에서는 감소하는 모습을 보였다. Logistic모형은 지속시간 1시간 확률강수량이 감소하는 모습을 보였으나 지속시간이 늘어남에 따라 확률강수량이 정상성 대비 증가하는 것으로 나타났으며 Power모형에서는 정상성 대비 지속시간 1시간 확률강수

Table 6. Percentage of Change Compared to the Stationary Probable Precipitation

Target Year	D(h)	Logistic	Power	ANN2.6	ANN4.5	ANN6.0	ANN8.5
2050	1	-1.27	2.57	-1.97	-3.49	0.05	12.94
	12	14.55	17.18	-5.65	-0.19	0.67	6.28
	24	21.51	21.26	-2.72	0.92	0.11	9.01
2080	1	-1.28	1.99	3.13	-4.02	3.66	18.42
	12	14.55	20.90	3.37	0.54	2.04	4.52
	24	21.51	25.42	-0.94	-1.31	2.52	7.31
2100	1	-1.28	1.69	-3.93	1.46	3.64	22.77
	12	14.55	22.84	-5.69	2.61	4.73	6.12
	24	21.51	27.60	-2.23	0.89	3.32	10.89

량은 약 2-35% 증가, 지속시간 12시간 이상부터는 약 25% 증가하는 것으로 나타났다.

적용한 기법에 따라 상이한 결과를 보였으나 근본적으로 비선형회귀모형들을 적용은 극치 수렴특성이 적용 기법에 따라 사전에 정해지는 단점이 있다. 본 연구에서 적용한 인공지능망 기법과 GCM 자료를 이용한 목표연도별, 지속시간별 확률강수량은 온실가스 배출 시나리오에 따른 GCM 결과에 의존하는 한계를 가지나 물리모형의 미래 예측치를 연계한 결과로 전술한 회귀모형 적용의 단점을 보완한 보다 물리적으로 타당한 확률강수량 예측이 가능하다고 판단되며 본 연구 접근법 또한 미래 수자원 관리를 위한 확률강수량 산정 확립에 활용할 수 있을 것으로 사료된다.

감사의 글

본 연구는 국토교통부 물관리연구사업의 연구비지원 (17AWMP-B079625-04)에 의해 수행되었습니다.

References

- KMA (2015) *Abnormal Climate Report*.
- Banach, S. (1922) Sur Les Opérations Dans Les Ensembles Abstraits Et Leur Application Aux Équations Intégrales. *Fundamenta Mathematicae*, Vol. 3, pp. 133-181.
- Cox, D.R. (1958) The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society, Series B*, Vol. 20, pp. 215-242.
- Good, P.I., and Hardin, J.W. (2012) *Common Errors in Statistics*. John Wiley & Sons.
- Hanel, M., and Buishand, T.A. (2010) On the Value of Hourly Precipitation Extremes in Regional Climate Model Simulations. *Journal of Hydrology*, Vol. 393, pp. 265-273.
- He, Y., Brdossy, A., and Brommundt, J. (2006) Non-stationary Flood Frequency Analysis in Southern Germany. *Proceedings of The 7th International Conference on Hydro Science and Engineering*, Philadelphia, USA.
- IPCC (2014) *Climate Change 2014: Synthesis Report*.
- Jang, S.W., Seo, L., Kim, T.W., and Ahn, J.H. (2011) Non-stationary Rainfall Frequency Analysis based on Residual Analysis. *Journal of the Korean Society of Civil Engineers*, Vol. 31, No. 5B, pp. 449-457 (in Korean).
- Kim, G.S., Jo, A.J., Jeon, H.C., and Kim, B.K. (2012) Sensitivity Analysis of Frequency Analysis using Annual Rainfall Exceedance Series. *Proceedings of 2012 Annual Conference*, Korean Society of Civil Engineers, pp. 805-808 (in Korean).
- Kwon, Y.M., and Kim, T.W. (2009) Derived I-D-F Curve in Seoul Using Bivariate Precipitation Frequency Analysis. *Journal of the Korean Society of Civil Engineers*, Vol. 29, No. 2B, pp. 155-162 (in Korean).
- Kwon, Y.M., Park, J.W., and Kim, T.W. (2009) Estimation of Design Rainfalls Considering an Increasing Trend in Rainfall Data. *Journal of the Korean Society of Civil Engineers*, Vol. 29, No. 2B, pp. 131-139 (in Korean).
- Leclerc, M., and Ouarda, T.B.M.J. (2007) Non-stationary Regional Flood Frequency Analysis at Ungauged Sites. *Journal of Hydrology*, Vol. 343, pp. 254-265.
- McCulloch, W.S., and Pitts, W. (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, Vol. 5, No. 4, pp. 115-133.
- Park, Y.J., Kwon, H.-H., Chung, E.S., and Kim, T.-W. (2014) Comparison of Design Rainfalls from the Annual Maximum and the Non-annual Exceedance Series. *Journal of the Korean Society of Civil Engineers*, Vol. 34, No. 2, pp. 469-478 (in Korean).
- Renard, B., Lang, M., and Bois, P. (2006) Statistical Analysis of Extreme Events in a Non-stationary Context Via a Bayesian Framework: Case Study with Peak-over-threshold Data. *Stochastic Environmental Research and Risk Assessment*, Vol. 21, pp. 97-112.
- Shin, H., Nam, W., Jung, Y., and Heo, J. (2008) Uncertainty Assessment of Regional Frequency Analysis for Generalized Logistic Distribution. *Journal of the Korean Society of Civil Engineers*, Vol. 28, No. 6B, pp. 723-729 (in Korean).
- Shin, J.Y., Park, Y.J., and Kim, T.W. (2013) Estimation of Future Design Rainfalls in Administrative Districts Using Nonstationary GEV Model. *J. Korean Soc. Hazard Mitig.*, Vol. 13, No. 3, pp. 147-156 (in Korean).
- Sugahara, S., da Rocha, R.P., and Silverira, R. (2009) Non-stationary Frequency Analysis of Extreme Daily Rainfall in Sao Paulo, Brazil. *International Journal of Climatology*,

Vol. 29, pp. 1339-1349.

Villarini, G., Smith, J.A., and Napolitano, F. (2010) Nonstationary Modeling of a Long Record of Rainfall and Temperature over Rome. *Advances in Water Resources*, Vol. 33, pp. 1256-1267.

Zhang, X., Xu, Y.P., and Fu, G. (2014) Uncertainties in

SWAT Extreme Flow Simulation under Climate Change. *Journal of Hydrology*, Vol. 515, pp. 205-222.

Received	■	November 1, 2017
-----------------	---	------------------

Revised	■	November 2, 2017
----------------	---	------------------

Accepted	■	November 20, 2017
-----------------	---	-------------------