

동질성 Hidden Markov Chain 모형을 이용한 일강수량 모의기법 개발

권현한* · 김태정** · 황석환*** · 김태웅****

Kwon, Hyun-Han*, Kim, Tae Jeong**, Hwang, Seok-Hwan***, Kim, Tae-Woong****

Development of Daily Rainfall Simulation Model Based on Homogeneous Hidden Markov Chain

ABSTRACT

A climate change-driven increased hydrological variability has been widely acknowledged over the past decades. In this regards, rainfall simulation techniques are being applied in many countries to consider the increased variability. This study proposed a Homogeneous Hidden Markov Chain(HMM) designed to recognize rather complex patterns of rainfall with discrete hidden states and underlying distribution characteristics via mixture probability density function. The proposed approach was applied to Seoul and Jeonju station to verify model's performance. Statistical moments(e.g. mean, variance, skewness and kurtosis) derived by daily and seasonal rainfall were compared with observation. It was found that the proposed HMM showed better performance in terms of reproducing underlying distribution characteristics. Especially, the HMM was much better than the existing Markov Chain model in reproducing extremes. In this regard, the proposed HMM could be used to evaluate a long-term runoff and design flood as inputs.

Key words : Markov chain model, Homogeneous, Hidden markov chain model, Precipitation

초록

최근 기후변화 영향으로 인해 수문변동성이 크게 증가되고 있으며 이러한 변동성을 고려하기 위한 방안으로서 강수량 모의발생 기법에 대한 중요성이 대두되고 있다. 본 연구에서는 복잡한 강수발생 패턴을 인지하고 강수량의 다양한 분포특성을 고려할 수 있는 혼합분포를 이용한 동질성 Hidden Markov Chain(HMM) 모형을 제안하였다. HMM 모형의 개선효과를 검증하기 위해서 기존 Markov Chain 모형과 비교 하였으며 서울관측소 및 전주관측소를 대상으로 연구를 진행하였다. 계절강수량 및 일강수량 등 다양한 시간규모에서 모형의 적합성을 평가하기 위해서 천이확률, 평균, 분산, 왜곡도 및 첨예도 등을 비교하였으며 HMM 모형이 기존 Markov Chain 모형에 비해서 개선된 모의능력을 확인할 수 있었다. 특히, HMM 모형은 극치강수량을 재현하는데 있어서 기존 Markov Chain 모형에 비해서 월등한 모의능력을 보여주었다. 이러한 점에서 장기유출량 및 확률홍수량 등을 산정하기 위한 입력자료로 활용이 충분히 가능할 것으로 판단된다.

검색어 : Markov chain 모형, 동질성, Hidden markov chain 모형, 강수

1. 서론

최근 기상변동성 증가 및 기후변화 영향으로 인해 수문변동성이 크게 증가되고 있으며 이러한 변동성을 고려하기 위한 방안으로서 강수량 모의발생 기법에 대한 중요성이 대두되고 있다. 수문변동성으로 인해 발생하는 불확실성에 효과적으로 대응하고 신뢰성 있는

* 정회원 · 전북대학교 토목공학과 조교수, 방재연구센터 (hkwon@jbnu.ac.kr)

** 정회원 · 전북대학교 토목공학과 석사과정 (kim.t.j.@jbnu.ac.kr)

*** 정회원 · 한국건설기술연구원 수자원환경본부 수석연구원 (sukany@kict.re.kr)

**** 정회원 · 교신저자 · 한양대학교 건설환경플랜트공학과 부교수 (Corresponding Author · Hanyang University · twkim72@hanyang.ac.kr)

Received April 16, 2013/ revised June 11, 2013/ accepted July 17, 2013

수공구조물 설계 및 수자원계획을 수립하는데 있어서 기본적으로 요구되는 사항은 장기간의 수문자료 확보라 할 수 있다. 최근 수자원 분야에서 강수모의 기법은 다양한 목적으로 활용되고 있으며 일반적으로 일강수량을 모의하는 이유는 상대적으로 짧은 자료연한을 갖는 수문자료계열의 통계적인 특성을 고려하여 장기간 모의함으로써, 장기적인 목적을 가지고 수행되는 수자원계획 수립 및 불확실성을 반영한 평가를 위함이다. 그러나 우리나라를 포함한 여러 나라에서는 충분한 수문자료 확보가 어렵기 때문에 통계적인 방법을 도입하여 자료를 확충하고 이를 통하여 수문변동성을 평가할 수 있는 방안이 일반적으로 채택되고 있다. 특히, 최근 도시화와 산업화로 인한 화석연료의 사용이 급증하면서 지구온난화로 기인하는 기후변화가 나타나고 있으며 이와 더불어 엘니뇨와 라니냐 등의 대규모 기상변동성이 나타나고 있다. 이러한 점을 감안하여 수자원계획 수립 시 발전된 해석기법을 통하여 수문 자료를 확충하고 이를 이용하여 수문기상학적 변동성을 평가하고자 하는 연구가 활발하게 진행되고 있다.

강수모의기법은 매개변수의 시간적 변동성 및 매개변수의 추정 방법에 따라 구분할 수 있다. 첫째, 매개변수에 시간적 변동성의 고려 여부에 따라 동질성 모형과 비동질성 모형으로 구분된다. 모형별로 정리하면 동질성 모형은 과거의 통계적 특성이 미래에도 유지된다는 가정을 중심으로 강수모의를 위한 매개변수들이 고정되어 사용된다. 반면 비동질성 모형은 동역학적(dynamic) 형태를 가지며 시간에 따라 강수모의를 위한 매개변수들이 변화하게 되며 기후변화 연구를 위해서 사용되는 Downscaling 기법이 대표적인 비동질성 강수모의기법이라 할 수 있다. 최근 기후변화에 따른 수자원의 영향 평가를 위한 Downscaling 기법의 발전 및 적용이 국내외적으로 매우 활발하게 이루어지고 있다(권현한 등, 2009). 둘째, 매개변수의 추정방법에 따라 매개변수적 방법과 비매개변수적 방법으로 나눌 수 있다. 일반적인 Markov Chain 모형은 천이확률 행렬을 통하여 강수와 무강수 등의 강수발생(rainfall occurrence) 패턴을 모의한 후 강수량(rainfall amount)을 특정 확률분포로 가정하여 강수량을 모의 발생시키는 방법을 사용하였다. 반면 비매개변수적 방법은 천이확률 뿐만 아니라 강수량을 모의하는데 있어서 특정 확률분포형의 가정 없이 자료 자체의 분포특성을 핵밀도함수(kernel density function)를 통해 도출하는 방안으로서 강수의 통계적 특성을 보다 현실적으로 모의할 수 있는 장점이 있다.

일강수량을 모의하는데 있어서 강수계열의 단기간 기억(memory)을 활용한 2-State 1차 Markov Chain 모형이 가장 일반적으로 사용된다. 여기서, 천이확률은 강수의 발생을 묘사하기 위해서 이용되며 강수량은 주로 감마분포(gamma distribution), 지수분포(exponential distribution) 등에 의해서 모의된다. Katz(1996)는 기존 Richardson 모형의 통계적 특성을 간략화하고 매개변수의

변동성을 고려하여 강수량을 모의하는 모형을 제안하였다. Wilks (1992)는 동일한 개념으로서 강수량을 모의하기 위한 분포형으로 지수분포 대신에 감마분포를 활용하여 기후변화 영향에 따른 강수량의 변동성을 평가하였다. 여기서, 기후변화에 따른 매개변수의 변동성을 고려하기 위해서 참조기간(baseline climate)을 중심으로 추정된 비율을 월 단위(monthly)로 매개변수에 반영함으로써 기후변화 영향을 고려하였다. 즉, 매월 현재 기후조건을 기준으로 미래 발생할 기후조건을 평균과 분산을 조정하여 미래의 일강수량을 추정하였다.

이와 같은 일강수량 모의결과를 통하여 물수지분석 등 장기간의 수문학적 변동성을 평가하기 위해서는 일반적으로 이상불형태의 일강수량자료가 입력 자료로 활용되게 되며, 2-State 1차 Markov Chain 모형이 국내외에서 주로 이용되고 있다. 기존 수문자료 확충방법으로서 일 강수량 및 수문자료의 모의발생 또는 보완을 위해 선형(linear) 또는 다중회귀방정식(regression equation)이 이용되었다(허준행, 1997). Markov Chain 이론에 의한 강수량모의발생을 이용하여 강수량을 확충 후 이를 강우-유출모형에 이용하기도 하였다. 이창훈 등 (1995)은 회귀분석과 주성분 분석을 이용하여 결측된 월강수량 자료를 보완하는데 Markov Chain 모형을 이용하였다. 김병식 등(1999)은 일 강수량 모의 발생을 통한 갈수유량 계열의 산정 및 빈도분석 연구를 수행하였으며, 정영훈 등(2005)은 기존 2-State 1차 Markov Chain 모형을 이용하여 추정된 일강수량 자료를 NWC-PC 모형의 입력자료로 활용하여 갑천 유역에 갈수유량에 따른 필요유량을 산정하였다. 또한 권현한 등(2011)은 불연속 Kernel-Pareto 분포를 이용한 일강수량 모의 기법을 개발하여 한강유역에 적용한바 있다. 또한 권현한 등(2013)은 불연속 Kernel-Pareto 분포를 이용한 다지점 강수모의기법과 Bayesian HEC-1 모형을 연계하여 대청댐 유역에 적용한 바 있다.

기존 Markov Chain을 사용한 일강수 모의기법은 관측된 강수시계열의 패턴을 모형화 하는데 있어서 실제 강수특성을 왜곡하는 현상이 있다. 일강수를 모의하는데 있어서 최근의 주된 관심사는 강수 및 무강수일을 특정 일기상태(weather state)를 통해 묘사하기 위해서 중관기후학적 대기 패턴을 고려할 수 있는 모형을 구성하는 것이다. 이러한 점을 고려하여 일기상태를 고려한 강수모의 기법의 개발이 이루어지고 있다(Hughes and Guttorp, 1994; Bellone et al., 2000; Kwon et al., 2009). 본 연구에서는 일기상태를 고려하고 각 일기상태별로 혼합분포(mixture distribution) 형태의 확률분포를 적용할 수 있는 일강수 모의기법을 개발하였다.

본 연구에서 일기상태는 은닉상태(hidden state)의 형태를 가지는 것으로 고려하였으며 은닉상태 안에서 천이확률은 시간에 따라 변화하지 않는 것으로 가정하였다. 즉, 시간에 따른 천이확률 및 Mixture분포의 변동성은 고려하지 않는 동질성 Hidden Markov

Chain(HMM) 모형을 개발하였다. HMM 모형은 은닉상태마다 조건부 확률분포를 부여하고 이들 조건부 확률분포로부터 일강수를 모의하는 방법으로 우리나라와 같이 계절에 따라 강수발생 특성이 복잡한 경우에 적합한 방법으로 활용이 가능하다. HMM 모형은 1960년대 음성처리를 위해 이용되기 시작했다. 초기의 HMM은 대부분 음성분야에 국한되어 사용되었으나 수문학분야에서는 HMM 모형은 미국 워싱턴 지역의 강수발생을 모델링하기 위해서 처음 이용되었다(Hughes and Guttorp, 1994). Bengioet al.(1995)는 은닉상태 천이확률의 동적 변화뿐만 아니라 출력자료가 은닉상태 및 입력 자료에 조건부로 변화하는 것을 허용하는 입출력 HMM 모형을 제안하였다. Bellone et al.(2000)은 강수량을 모델링하기 위해서 HMM모형을 이용하였으며 강수량을 모의하기 위한 확률분포형으로서 Gamma 분포를 사용하였다. 본 연구에서는 앞서 언급하였듯이 Gamma 분포, Exponential 분포를 동시에 고려할 수 있도록 혼합분포를 활용하여 모형을 구성하였다.

결론적으로 일반적인 Markov Chain 모형에 비해서 관측지점의 통계적 특성을 효과적으로 복원하며 종관기상학적 패턴을 파악하는데 유리한 강수모의기법을 개발하는 것이 본 연구의 목적이다. 즉, 기존 Markov Chain 모형 기반의 강수모의기법을 대체할 수 있는 단지점 HMM 일강수량 모의기법을 개발하고자 하며 기상청 산하의 서울관측소와 전주관측소의 강수량을 대상으로 모형의 적합성을 평가하였다. 본 논문의 구성은 다음과 같다. 1장에서는 논문의 배경 및 목적에 대해서 언급하였으며, 2장에서는 기존의 정상성 2-State Markov Chain 모형과 HMM 모형에 대하여 기술하였다. 3장에서는 HMM 모형 적용 결과를 요약하여 나타내었으며 마지막으로 결론 및 토의를 4장에 수록하였다.

2. Markov Chain 기반 일강수량 모의기법

2.1 정상성 Markov Chain Model

일반적으로 무작위 시계열 자료간의 계열 상관성과 미래의 상태를 현재와 과거의 상태들과의 상관관계 분석만으로 추계학적으로 추출하는 과정을 Markov 과정이라고 하고 한정적인 상태공간에서의 Markov 과정을 Markov Chain라 한다.

국내의 경우 일 강수량자료는 많게는 100여년 정도에서부터 적게는 10여년 정도까지 다양한 길이로 존재하고 있다. 그러나 현재 우리나라의 대부분의 유역에서 관측된 수문자료는 매우 제한적이고 결측자료가 존재하여 수문현상을 분석하는데 어려움이 있는 실정이다. 본 논문에서 취급할 일강수량 자료계열은 유출량 자료계열에 비하여 변동성이 크고 수문학적인 지속성 즉 자기상관성 (autocorrelation)이 크지 않은 시계열 자료로서 이와 같은 경우 Markov Chain을 적용하면 그 특성을 잘 묘사하는 것으로 알려져

있다(Nord, 1975). Markov Chain의 기본 개념은 짧은 과거의 기억을 통해 미래 상태를 예측하는 것(Hann 등, 1976)으로서 다음과 같은 조건부 확률로 나타낼 수 있다.

$$p(R_{n+1} | R_n, R_{n-1}, \dots, R_1) \tag{1}$$

여기서 상태 R_n 는 현재의 강수 상태, R_{n-1}, R_n, R_1 은 과거의 강수 상태를 나타내며 미래의 강수 상태는 과거의 상태와는 독립 (independence)이며 단지 가장 근접한 현재의 강수 상태에 의해서만 영향을 받는다면 이때 Markov Chain의 성질은 다음과 같은 조건부 확률로 나타낼 수 있다. 이러한 과정을 1차 Markov Chain이라 한다.

$$p(X_{n+1} | X_n) \tag{2}$$

여름과 같이 우기기간 내의 강수량 간에는 작은 계열 상관(serial correlation)이 존재할 수 있으나 종속성이 있다고 보기에는 무리가 많다. 따라서 강수량을 모의하기 위하여 연속하는 사상 간의 독립성을 가정하고 이론적인 확률분포를 적합 시키는 방안들이 주로 활용되고 있다. 2-State 의 1차 Markov Chain모형에서 각 상태공간에 존재하는 강수사상의 계열인 벡터(S)와 사상의 발생(강수일)과 강수 발생(무강수)의 확률을 나타내는 천이확률(P_{tr})은 다음과 같은 식으로 표현 될 수 있다.

$$P_{tr,i,j} = \frac{S_{i,j}}{\sum_{j=1}^2 S_{i,j}} \tag{3}$$

여기서, 행렬의 원소 $P_{tr,i,j}$ 는 상태 i 에서 상태 j 로 천이되는 확률을 나타내며 상태 S_i 에서는 반드시 $S_{1,1}, S_{1,2}, S_{2,1}, S_{2,2}$ 중 한 곳으로 천이 되므로 천이확률의 각 행의 합은 반드시 1이 된다. 식 (3)으로 표현되는 천이확률(transition probability)의 각 원소(element)는 모두 0보다 크거나 같으며, 각 행의 원소의 합이 1일 때의 행렬을 확률행렬(probability matrix)이라 정의 한다. 각각의 확률행렬은 하나의 Markov 과정을 정의하며, 강수 발생여부를 판단하게 된다. 따라서 천이확률 P_{tr} 이 결정되면 Markov Chain의 초기 상태만을 통해 강수발생을 모의할 수 있다. 2-State 1차 Markov Chain 모형에 의한 강수발생 과정을 Table 1과 같은 천이확률 행렬(transition probability matrix)로 나타낼 수 있다.

Table 1에서 $0 \leq a \leq 1, 0 \leq b \leq 1$ 로서, a 는 무강수일 다음에 강수일이 발생할 확률이고 b 는 강수일 다음에 무강수일이 발생할 확률이며 이를 천이확률 식으로 나타내면 아래 식과 같다.

Table 1. Transition Probability Matrix for State 2-1 Markov Chain

t-1 (day) \ t(day)	Dry	Wet
Dry	1-a	a
Wet	b	1-b

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} \quad (4)$$

Markov Chain 모형에 의한 강수발생과정에는 초기 확률과 천이확률 행렬이 필요하게 된다. Markov Chain 모형에서 초기 확률은 무작위난수로 부터 초기 확률을 추출하여 강수발생 여부를 결정한다. 마지막으로 과거 강수량자료로부터 추정된 확률밀도함수의 매개변수를 이용하여 Monte Carlo 모의를 실시하여 강수일에 해당하는 강수량을 결정한다.

2.2 Hidden Markov Chain Model

2.2.1 기본 모형 구성

HMM은 조건부 독립가정을 기본으로 관측 자료로부터 은닉상태의 확률론적 과정(stochastic process)을 관측이 가능하도록 발생시켜 다른 확률론적 과정을 통하여 모델링하는 이중의 확률론적 과정이다. 우선 시간의 차원을 갖는 강수량자료를 시간 T의 연속 벡터로 나타내면 다음과 같다.

$$R_{1:T} = (R_1, \dots, R_T) \quad (5)$$

또한 시간 t에 해당하는 은닉상태(hidden state)를 나타내는 연속 계열순서(sequence)는 다음과 같다.

$$S_{1:T} = (S_1, \dots, S_T) \quad (6)$$

HMM 모형은 두 가지 조건부 독립 가정들을 이용하여 $R_{1:T}$ 와 $S_{1:T}$ 의 결합확률(joint probability) 분포로 정의하고 두 가지 조건부 독립 가정은 다음과 같다.

첫째, 주어진 은닉상태와 k일 이전의 강수량으로 이루어진 모형에서 관측강수 벡터는 다른 모든 변수와 조건부적으로 독립이라고 가정한다. 즉, 다음과 같이 나타낼 수 있다.

$$P(r_t | r_{1:t-1}, s_{1:t}) = P(r_t | s_t). \quad (7)$$

둘째, 시계열의 바로 이전의 은닉상태에만 상관성이 있다고 보는 경우를 1차 HMM이라고 하며 일반적으로 Markov Chain이

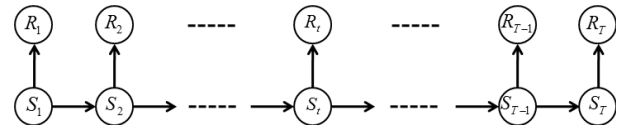


Fig. 1. Bayesian Network Representation of a Homogeneous HMM Satisfying Conditional Independence Assumptions

라고 하면 1차 Markov Chain을 말하는 것으로 즉 은닉상태의 확률분포는 시간적으로 오직 이전 은닉상태에 의존한다.

$$P(s_t | s_{1:t-1}) = \begin{cases} P(s_t | s_{t-k:t-1}) & t > k, \\ P(s_t | s_{1:t-1}) & t \leq k. \end{cases} \quad (8)$$

위의 Eq. (8)에서 추가적으로 명시된 언급이 없으면 일반적으로 k=1의 은닉상태 확률분포는 다음과 같다.

$$P(s_t | s_{1:t-1}) = \begin{cases} P(s_t | s_{t-1}) & t \geq 2, \\ p(s_1) & t = 1. \end{cases} \quad (9)$$

동질성 HMM은 조건부 확률 $P(s_t | s_{t-1})$ 을 추정하는데 있어서 Fig. 1과 같은 정상성 가정을 기본으로 이루어진다. 앞서 언급된 은닉상태에 조건부확률 $P(s_t | s_{t-1})$ 가 시간에 따라서 일정하다는 정상성 가정에 기본을 두고 있다. 동질성 HMM에 대해서 $\Pi = (\pi_1, \dots, \pi_K)$ 는 초기 확률벡터라 하고 $\Gamma = (\gamma_{11}, \dots, \gamma_{KK})$ 를 천이확률을 갖는 행렬(transition probability matrix)로 정의하자.

여기서 Υ 를 은닉상태 S_t 에 조건부로 주어지는 강수량 r_t 의 발생확률 $F_t(r) = P(r_t | S_t = i)$ 은 각각의 은닉상태별로 결정되는 출력(emission) 확률밀도함수의 매개변수들로 구성된다.

$$P(r_{1:T}, S_{1:T} = s_{1:T} | \Pi, \Gamma, \Upsilon) = \left[\pi_{s_1} \prod_{t=2}^T \gamma_{s_{t-1}s_t} \right] \left[\prod_{t=1}^T F_{s_t}(r_t | r_{t-1}) \right] \quad (10)$$

2.2.2 은닉 상태 분포(hidden state distribution)

일반적으로 관측된 강수량 자료를 이용하여 은닉상태의 확률밀도 함수를 추정하는 것이 보편적인 방법이다. 다시 말하면 관측 강수량을 이용하여 은닉상태의 확률분포를 구하고자 한다.

$$P(s_{1:T} | r_{1:T}, x_{1:T}) = \frac{P(s_{1:T}, r_{1:T} | x_{1:T})}{P(r_{1:T} | x_{1:T})} = \frac{P(s_{1:T}, r_{1:T} | x_{1:T})}{\sum_{s_{1:T}} P(s_{1:T}, r_{1:T} | x_{1:T})} \quad (11)$$

자료의 우도 $P(r_{1:T}|x_{1:T})$ 를 계산하기 위해서 Forward-backward 과정을 통한 반복법을 이용하였다. 각 은닉상태 S_t 에 대해서, 반복적으로 선행 은닉상태(α_t)와 이틀 상태를 뒤따르는 은닉상태 (β_t)를 계산하는 과정을 수행하였으며 다음과 같이 수식으로 나타 낼 수 있다.

$$\alpha_t(i) = P(S_t = i, r_{1:t}|x_{1:t}) \quad (12a)$$

$$\beta_t(i) = p(r_{t+1:T}|S_t = i, r_t, x_{t+1:T}) \quad (12b)$$

$$\alpha_1(i) = P(S_1 = i|x_1)P(r_1|S_1 = i)$$

$$\alpha_{t+1}(i) =$$

$$P(r_{t+1}|S_{t+1} = j, r_t) \sum_{i=1}^K P(S_{t+1} = j|S_t = i, x_{t+1}) \alpha_t(i)$$

$$\beta_T(i) = 1$$

$$\beta_t(i) =$$

$$\sum_{j=1}^K P(S_{t+1} = j|S_t = i, x_{t+1}) P(r_{t+1}|S_{t+1} = j, r_t) \beta_{t+1}(j)$$

강수량 자료에 대한 우도는 다음과 같이 계산된다.

$$P(r_{1:T}|x_{1:T}) = \sum_{i=1}^K P(S_T = i, r_{1:T}|x_{1:T}) = \sum_{i=1}^K \alpha_T(i) \quad (12c)$$

일강수량 모의를 위해서 우리나라의 강수분포 분석에 적합한 확률분포로 일반적으로 사용되고 있는 Gamma 또는 Exponential 분포가 이용되며 무강수를 제외하고 강수에 대해서만 확률분포를 적용하는 Delta 함수가 도입된다. 본 연구에서는 Delta-exponential 분포, Delta-gamma 분포와 두 가지를 혼합하여 Delta-exponential-gamma 등을 활용하여 강수량을 모의하였다. 이렇게 혼합분포를 활용함으로써 강수량의 다양한 확률분포 특성을 고려할 수 있는 장점이 있다. Delta-exponential 분포와 Delta-gamma분포의 확률 밀도함수를 추정하기 위한 우도함수는 다음과 같다. 여기서, M , i , c 는 각각 관측소의 개수, 은닉상태 번호, 혼합분포의 개수를 나타내며 p 는 혼합분포의 가중치를 의미한다. 또한 λ , α , β 는 Exponential 분포와 Gamma 분포의 매개변수를 나타낸다.

$$P(r_t|S_t = i) = \prod_{m=1}^M P(r_t^m|S_t = i) = \prod_{m=1}^M a_{im} \quad (13)$$

$$\text{여기서 } a_{im} = \begin{cases} p_{im0} & r_t^m = 0, \\ \sum_{c=1}^C p_{imc} \lambda_{imc} \exp(-\lambda_{imc} r_t^m) & r_t^m > 0 \end{cases}$$

$$P(r_t|S_t = i) = \prod_{m=1}^M P(r_t^m|S_t = i) = \prod_{m=1}^M a_{im} \quad (14)$$

여기서

$$a_{im} = \begin{cases} p_{im0} & r_t^m = 0, \\ \sum_{c=1}^C p_{imc} \frac{\beta_{imc}^{\alpha_{imc}} (r_t^m)^{\alpha_{imc}-1} \exp(-\beta_{imc} r_t^m)}{\Gamma(\alpha_{imc})} & r_t^m > 0 \end{cases}$$

3. 적용

1974년부터 2012년 까지 서울관측소와 전주관측소 등 2개의 관측소의 일강수량 자료를 대상으로 모형을 적용시켰다. 모형 적용 과정은 주로 서울관측소를 대상으로 설명하고 적합성을 평가하는 과정은 전주관측소를 포함하여 제시하였다. HMM 모형의 적합성을 평가하기 위해서 Gamma 확률밀도함수 기반의 2-State 1차 Markov Chain 모형과 비교하여 분석하였다.

모형의 적용에 앞서 은닉상태 수를 결정하기 위하여 HMM 모형의 적합성 정도를 평가하는데 사용되는 대수우도함수(log-likelihood function)를 사용하여 서로 다른 은닉상태 수 K 를 가지고 평가하였다. 은닉상태 수 K 를 분기별로 2~7까지 변화하면서 우도를 산출하였고 그 과정에서 K 가 증가함에 따라 우도함수 또한 증가하고 일정한 은닉상태 수 이상에서는 수렴하는 결과를 확인할 수 있었다. 하지만 은닉상태 수를 무한히 증가할 경우 모형이 과적합(overfitting)되는 문제점이 있으므로 우도가 수렴하는 부분에서 최적의 은닉상태 수를 결정하였다. 본 연구에서는 여름 계절에 대해서 HMM 모형을 적용하였다. 즉, 서울 및 전주관측소 일강수량에 대해서 6월부터 9월(JJAS, June-July-August-September)의 여름 강수량을 대상으로 연구를 진행하였다. 논문의 지면관계상 전주지점의 경우 서울지점과 유사한 결과를 보여주는 그림 및 표는 생략하고 설명으로 대체하였다. 강수 계열을 계절로 구분한 것은 강수사상의 계절적 특성을 보다 효과적으로 반영한 모형을 구축하기 위함이다.

최적 은닉상태 수 결정은 Fig. 2와 같이 대수우도함수를 이용하여 이루어진다. Fig. 2는 각 관측소의 JJAS 계절에 해당하는 은닉상태 수에 따른 대수우도함수를 나타냈으며 최적우도를 가지는 은닉상태 수를 선택하여 강수모의 분석을 실시하였다.

Fig. 3과 Fig. 4는 우리나라의 우기기간에 해당하는 6월에서부터 9월까지의 서울관측소에서 관측된 강수량의 강수특성을 분석한 결과이며 5개의 은닉상태 수를 사용하여 각각의 은닉상태의 강수발생 빈도와 강수량의 특징을 파악하기 위한 그림이다. Fig. 3은 매년 6월부터 9월까지 총 122일에 대해서 은닉상태를 분류한 그림으로서 다섯 번째 은닉상태(S5)가 가장 높은 빈도를 차지하는

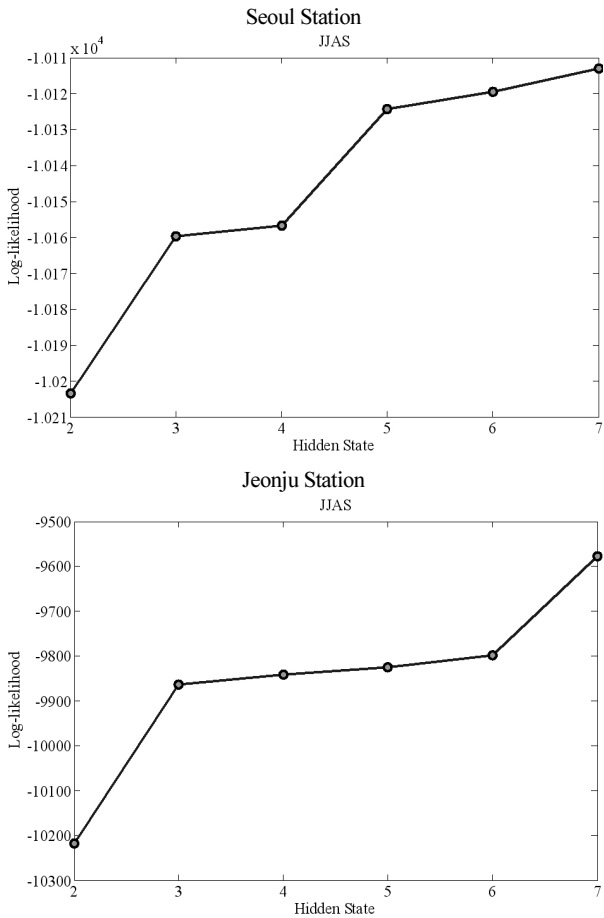


Fig. 2. Log-likelihood Estimations for Amounts Models Using Mixture of two Exponential Distributions for Each Station Rainfall on Wet Days

기상상태로 판단되며 은닉상태 S1, S3 및 S4는 상대적으로 낮은 빈도를 나타내고 있다. Fig. 4의 왼쪽 그림은 모의기간에 대하여 발생된 강수의 총량을 나타내며 오른쪽에 위치한 그림은 각각의 은닉상태에 따른 강수 발생빈도를 나타내고 있다.

첫 번째 은닉상태(S1)의 경우에는 전체 강수량의 78%정도를 차지하고 있는데 비하여 강수사상의 발생빈도는 10%정도를 나타내고 있는 것으로 보아서 집중호우의 형태를 가지는 기상상태로 판단할 수 있다. 네 번째 은닉상태(S4)와 같은 경우는 강수량이 거의 0%에 가깝지만 강수사상의 발생빈도는 58%를 나타내고 있음을 확인할 수 있다. 이는 거의 강수가 발생하지 않거나 무강수에 가까운 기상상태로 판단해 볼 수 있다. 즉, 강수량은 대부분 은닉상태 S1, S2, S3이 차지하고 있으며, 그중 S1은 낮은 빈도로 발생하지만 극치 강수량을 발생시키는 강수사상의 특징을 보여주고 있다.

본 연구의 가장 중요한 목적이 일강수량 자료계열의 특성치를 최대한 관측치와 유사하게 모의하는데 있으므로, 본 연구를 통해

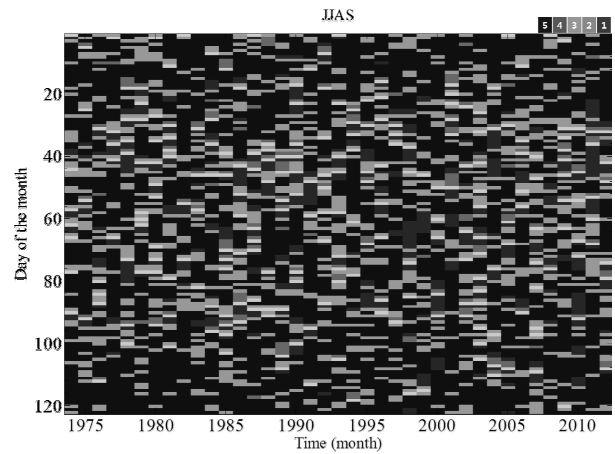


Fig. 3. The Estimated State Sequence Via Vieterbi Algorithm for JJAS Season at Seoul Station

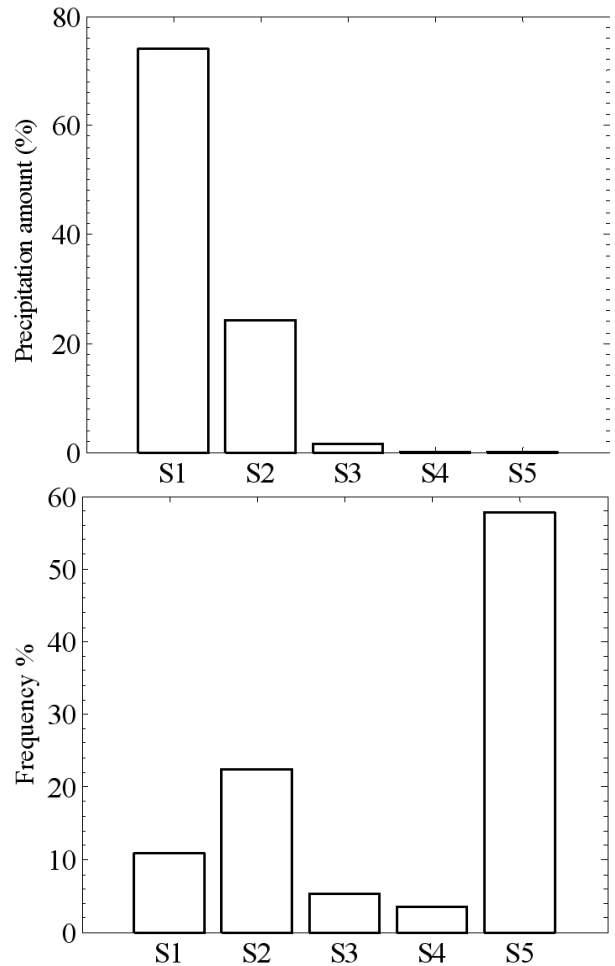


Fig. 4. The Estimated Percentage of Amount and Frequency of Seasonal Rainfall According to Each State at Seoul Station

추정된 연구 결과 또는 이러한 특성을 평가할 수 있는 지표들을 대상으로 관측치와 비교하였다. 즉, 계절 강수량의 총량을 얼마나 유사하게 재현하는지를 평가하기 위해서 계절 총 강수량을 관측치와 함께 나타내었다. 또한 일강수량의 특성치인, 천이확률 및 습윤일을 비교분석을 실시하였으며, Boxplot은 총 200개의 Ensemble로 추정되었으며 관측치(“▽”)와 중첩하여 나타내었다.

HMM 모형을 통해 모의된 강수량의 적합성을 평가하기 위해서 Fig. 5에서는 2-State 1차 Markov Chain 모형 결과와 계절적 강수 특성을 비교 검토하였다. 각 분기별로 실제 관측된 강수특성과 비교하여 계절 강수량, 강수에서 강수로 변화하는 천이확률, 무강수에서 무강수로 변화하는 천이확률, 강수일수에 대하여 분석한 결과를 나타내는 그림으로 HMM 강수모의 모형을 통하여 얻은 모의 강수값은 Boxplot을 사용하여 나타내었으며 관측값은 ‘▽’으로 표시하고 Gamma 모형분포 기반의 2-State 1차 Markov Chain 모형 통한 결과값은 ‘△’으로 표시하였다.

Fig. 5에서 볼 수 있듯이 HMM모형과 기존 Markov Chain 모형 모두 계절 단위에서 강수의 통계 특성을 효과적으로 모의하는 것을 확인할 수 있다. 전주지점의 경우에도 동일한 결과를 확인할 수 있었다. 기존 Markov Chain 모형의 경우 관측된 천이확률을 그대로 재현하여 강수를 모의하기 때문에 천이확률과 강수일수가 관측치와 유사하게 모의되는 것은 당연하다 하겠다. 추가적으로

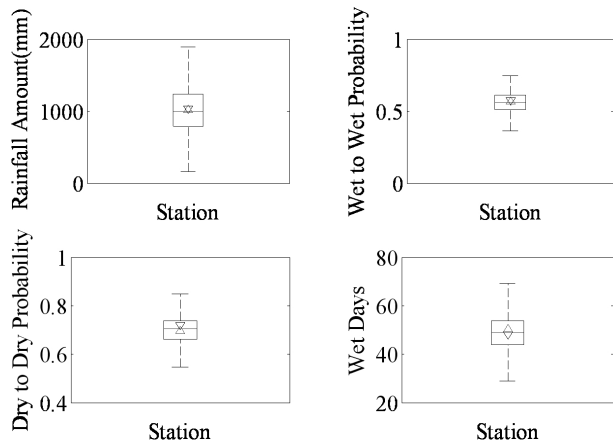


Fig. 5. A Comparison of Seasonally Averaged Rainfall Statistics for Seoul Weather Stations for JJAS Season

Table 2. A Comparison of Seasonally Averaged Rainfall Statistics for Seoul Weather Station for JJAS Season

Statistics	Observation	HMM	Gamma
Rainfall Amount(mm)	1031.58	1014.92	979.87
Wet to Wet	0.57	0.57	0.54
Dry to Dry	0.72	0.70	0.69
Wet Days	48	49.55	49.06

서울관측소의 강수총량과 천이확률 및 강수 발생일을 Table 2에 정리하였다.

Fig. 6에서는 서울관측소의 JJAS 계절의 강수량에 대한 1974년부터 2012년까지 모의 결과를 평균과 표준편차를 대상으로 비교 평가하였다. 그림에서 확인할 수 있듯이 HMM모형에 의하여 모의된 계절강수량의 평균값과 표준편차 모두 관측치(“▽”)와 유사한 결과를 나타내는 반면 기존 Markov Chain 모형을 통한 모의강수값(“△”)은 평균값은 유사하게 재현하지만 표준편차는 모의값과 관측값이 불일치하는 것을 확인할 수 있다. 전주지점에 대해서 동일하게 HMM 모형이 기존 Markov Chain 모형에 비해 개선된 모의결과를 나타내었다. 이는 HMM 모형이 강수 모의시 극치 강수량의 재현성이 기존 Markov Chain 모형에 비하여 우수하다는 점을 의미한다.

이러한 점을 좀 더 자세히 평가하기 위해서 일강수량계열을 대상으로 평균, 표준편차, 왜곡도, 첨예도 등 4차 통계모멘트까지

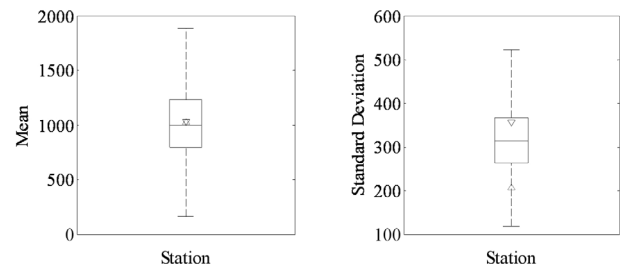


Fig. 6. A Comparison of Mean and Standard Deviation of Seasonal Rainfall(JJAS) Between Observation and Simulation. Each Panel Shows Mean, Standard Deviation and Skewness. The Box Plot is Derived by 200 Ensemble Series and the Symbol ▽ Indicates Observation While △ Represents Results Derived From the Existing Markov Chain Model

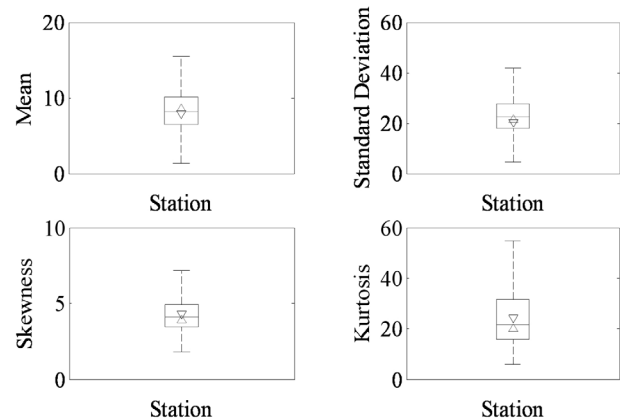


Fig. 7. Basic Statistics of Daily Rainfall Between Observation and Simulation. Each Panel Shows Mean, Standard Deviation, Skewness and Kurtosis for Seoul Station. The Box Plot is Derived by 200 Ensemble Series and the Symbol ▽ Indicates Observation While △ Represents Results Derived From the Existing Markov Chain Model

Table 3. Seasonally-averaged Statistical Moments of Daily Rainfall Between Observation and Simulation

Statistics	Observation	HMM	Gamma
Mean	7.84	8.32	7.00
Standard Deviation	20.22	22.45	18.56
Skewness	4.28	4.17	3.53
Kurtosis	24.04	24.16	16.57

Table 4. Statistical Moments of Daily Rainfall Between Observation and Simulation for Each Month

Statistics	June			July			August			September		
	Obs.	HMM	Gam.	Obs.	HMM	Gam.	Obs.	HMM	Gam.	Obs.	HMM	Gam.
Mean	3.9	4.76	5.2	11.35	12.51	7.99	9.15	11.19	8.42	3.31	5.14	5.68
Standard Deviation	11.05	12.41	12.77	23.28	26.26	15.17	19.91	24.33	20.47	10.69	15.08	15.45
Skewness	3.27	3.29	2.78	2.73	2.77	2.17	2.55	2.79	2.72	3.39	3.55	2.94
Kurtosis	12.98	14.02	9.11	10.3	10.92	7.93	8.57	10.94	10.14	14.34	15.63	10.58

Table 5. Daily Rainfall Statistics of Seoul Station for Each Month

Statistics	June			July			August			September		
	Obs.	HMM	Gam.	Obs.	HMM	Gam.	Obs.	HMM	Gam.	Obs.	HMM	Gam.
Rainfall Amount(mm)	142.3	142.7	159.9	393.6	387.6	237.8	342.9	346.8	273.5	152.6	154.1	177.6
Wet to Wet	0.45	0.44	0.36	0.64	0.66	0.5	0.63	0.6	0.61	0.42	0.41	0.47
Dry to Dry	0.72	0.73	0.65	0.63	0.62	0.59	0.67	0.65	0.63	0.76	0.77	0.79
Wet Days	10	10.08	11.04	16	16.16	14.12	14	14.47	15.1	9.0	8.72	8.76

재현 특성을 비교 검토하여 Fig. 7에 나타내었다. Boxplot은 총 200개의 Ensemble로부터 추정되었으며 관측치(“▽”)와 중첩하여 나타내었다. 그림에서 보면 서울관측소에서 HMM 모형과 기존 Markov Chain 모형은 일강수량의 통계특성을 효과적으로 재현하고 있음을 확인할 수 있다. 그러나 서울지점의 경우 기존 Markov Chain 모형은 3차모멘트 이상에서 작은 차이지만 관측값에 비해 과소 추정되는 것을 확인할 수 있다. 반면 HMM 모형은 관측치와 거의 일치하는 결과를 나타내고 있는 등 전체적으로 기존 Markov Chain 모형에 비해서 개선된 모의 결과를 나타내었다. 3차 모멘트 이상에서 작은 차이지만 극치강수량을 재현하는데 있어서 큰 차이를 나타낼 수 있다는 점에서 모의된 강수량을 대상으로 연최대치 자료만을 추출하여 빈도해석을 수행하여 관측치와 비교하였다. Table 3 and 4는 일강수량의 대한 평균, 표준편차, 왜곡도, 첨예도를 계절기준과 월 기준으로 평균하여 비교하였다. 본 연구에서 제시하고 있는 HMM 모형이 기존 Markov Chain 모형에 비해 우수한 모의능력을 나타내었다. Table 5는 월별 일강수량 총량, 월단위로 평균된 천이확률 및 강우일수를 관측값과 비교하여 나타내었으며 HMM 모형이 기존 모형에 비해 다양한 측면에서 개선된 효과를 확인할 수 있었다. 특히, 7~8월의 장마철 강수계열의 경우 기존

모형의 경우 강우량이 크게 과소 추정되는 문제점이 발생하고 있으나 HMM 모형은 강수량 및 강우일수에 보다 현실성 있는 모의가 가능하였다.

Fig. 8과 9는 각각 서울지점과 전주지점의 관측값 및 모의값으로부터 추출된 극치자료계열에 대해서 핵밀도함수와 Gumbel 확률밀도함수를 이용한 100년 빈도 강수량을 비교한 그림이다. 그림에서 보면 기존 Markov Chain의 경우 관측값의 확률밀도함수와 매우 다른 형태를 나타내고 있으며 이로 인해 추정된 100년 빈도 확률강수량도 관측값에 비해 크게 과소 추정되는 결과 보이고 있다. 반면 본 연구에서 제시하고 있는 HMM 모형은 확률밀도함수가 관측값과 매우 유사한 특성을 나타내고 있으며 Gumbel분포를 통해서 추정된 100년 빈도 확률강수량의 경우도 관측값과 거의 일치하는 결과를 얻을 수 있었다. 이러한 결과는 HMM 모형구조가 가지는 장점에 기인하는 결과로서 판단된다. 즉, 강수계열에 내재되어 있는 복잡한 여름강수량 패턴을 인지함과 동시에 다양한 분포형태를 고려할 수 있는 혼합분포를 강수모의에 도입함으로써 일강수량 통계특성, 계절강수량 통계특성, 극치강수량 통계특성 등 다양한 시간규모에서 기존 모형에 비해서 개선된 모의능력을 나타낸다고 할 수 있다.

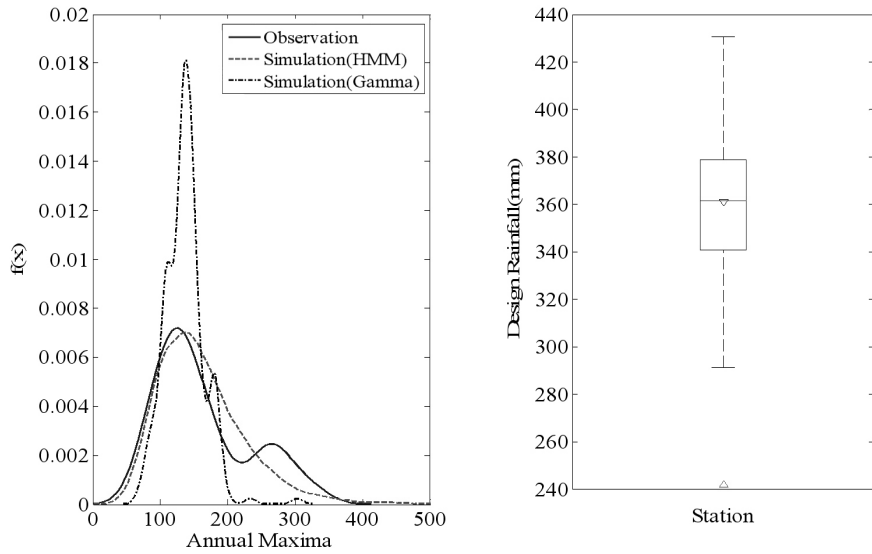


Fig. 8. A Comparison of Kernel Density Function of Annual Extreme Rainfall Between Observed and Simulated Rainfall for Seoul Station

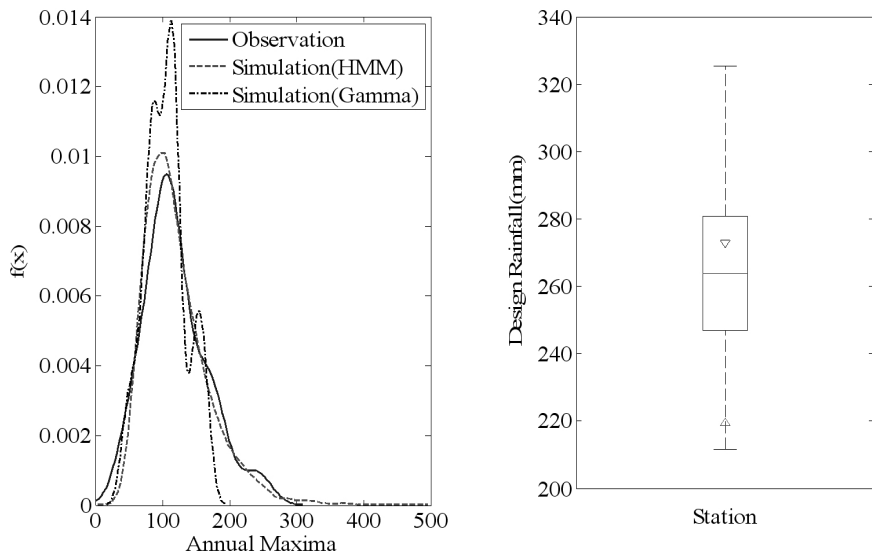


Fig. 9. A Comparison of Kernel Density Function of Annual Extreme Rainfall Between Observed and Simulated Rainfall for Jeonju Station

4. 결론

수자원 계획, 설계, 운영 및 관리를 위해서는 최소 30년 이상의 수문자료가 필수적으로 요구된다. 특히 대규모 수자원 계획에서는 50년 이상의 수문자료를 활용하여 장기적인 관점에서 계획을 수립하는 것이 일반적이며 신뢰성 있는 분석을 위해서는 최소한 계획기간을 상회하는 자료연한이 필요하다. 그러나 우리나라를 포함하여 많은 나라에서 결측치가 없는 30년 이상의 장기관측 자료는 찾아보기 어려운 실정이다. 이러한 문제점에 기인하여 단기간의 수문자료를 확충하여 수자원계획에 이용하고자 하는 일강수량 모의기법에

관한 연구들이 다수 진행되어 왔다. 그러나 기존 연구들에서는 일강수량의 평균 및 천이확률 등과 같이 일반적인 통계치를 재현하는 부분에 초점을 맞추어져 연구가 진행되었으나 분산, 왜곡도 등 극치강수량 재현을 위한 연구는 상대적으로 미진한 실정이다. 이러한 점에서 본 연구에서는 강수의 발생 패턴을 자동으로 인지하고 강수 모의에 활용하며 강수량의 분포특성을 전역적으로 고려할 수 있는 혼합분포를 도입한 HMM 모형을 개발하였다.

본 연구에서는 기존 정상성 Markov Chain 모형을 개선한 동질성 HMM 모형을 개발하여 서울관측소와 전주관측소의 강수자료를 사용하여 모형의 적합성 및 적용성을 평가하였다. 본 연구를 통해

얻은 결론은 다음과 같다.

첫째, 모의된 일강수량을 계절강수량으로 변환하여 관측값과 모의값을 비교한 결과 관측치의 통계적 특성을 유지하는 것을 확인할 수 있었으며 특히 계절강수량의 분산을 재현하는데 있어서 본 연구에서 제안한 HMM 모형이 기존 Markov Chain 모형에 비해 우수한 재현 특성을 나타내었다.

둘째, 일강수계열에 대해서 4차모멘트까지 통계적 특성을 검토한 결과 관측값과 거의 유사하게 모의가 되는 것을 확인할 수 있었으며 계절강수량과 동일하게 3차모멘트이상에서 기존 Markov Chain 모형에 비해서 개선된 모의 능력을 확인할 수 있었다.

셋째, 마지막으로 극치강수량 재현효과를 검토하기 위해서 모의 자료와 관측값에 대해서 빈도해석을 수행하여 비교 검토하였다. 검토결과 본 연구에서 제안한 HMM 모형은 기존 관측값으로 추정된 확률밀도함수와 매우 유사한 분포특성을 나타내고 있는 반면 기존 Markov Chain 모형은 매우 왜곡된 분포 특성을 보여주었다. 이러한 점으로 인해 추정된 100년 빈도 강수량의 경우에도 HMM 모형은 관측값을 기준으로 추정된 확률강수량과 거의 동일한 값을 갖는 반면에 기존 Markov Chain 모형은 과소 추정되는 것을 확인할 수 있었다.

위에서 언급된 내용을 중심으로 판단해보면 본 연구에서 제시한 HMM 모형이 기존 Markov Chain 모형에 비해서 극치강수량 재현부분에 특히 적용성이 우수한 것으로 판단되며 장기유출량분석 및 대규모유역의 확률홍수량을 추정하는데 있어서 입력자료로 활용이 가능할 것으로 판단된다. 추후 연구로서 유역단위에 면적강수량 산정을 위한 다지점(multi-site) 강수모의 기법으로 모형을 확장하는 것이 필요할 것으로 판단된다. 이와 더불어 기후변화 연구를 위해서 기후변화 시나리오를 입력 자료로 활용할 수 있는 비정상성 모형으로 확장이 필요할 것으로 판단된다.

감사의 글

본 연구는 한국건설기술연구원 주요(시드)사업(지역특성을 반영한 상세 격자강수량 생산기술 개발)의 연구비 지원에 의해 수행되었습니다. 저자들은 전북대학교 방재연구센터에 소속되어 연구를 수행하였습니다.

References

Kim, B. S., Kang, K. S. and Seoh, B. H. (1999). "Low flow frequency analysis of streamflows from the stochastically generated daily rainfall series." *Journal of Korean Water Resources Association*,

Vol. 32, No. 3, pp. 265-279.
 Bellone, E., Hughes, J. P. and Guttorp, P. (2000). "A hidden markov model for downscaling synoptic atmospheric patterns to precipitation amounts." *Climate Research*, Vol. 15, pp. 1-12.
 Haan, C. T., Allen, D. M. and Street, J. O. (1976). "A markov chain model of daily rainfall." *Water Resour. Res.*, Vol. 12, No. 3, pp. 443-449.
 Hughes, J. P. and Guttorp, P. (1994). "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena." *Water Resour. Res.*, Vol. 30, No. 5, pp. 1535-1546.
 Kwon, H. H. and Kim, B. S. (2009). "Development of statistical downscaling model using nonstationary markov chain." *Journal of Korean Water Resources Association*, Vol. 42, No. 3, pp. 213-225.
 Kwon, H.-H., Kim, J.-G. and Park, S.-H. (2013) "Derivation of flood frequency curve with uncertainty of rainfall and rainfall-runoff model." *Journal of Korean Water Resources Association*, Vol. 46 No. 1, pp. 59-71.
 Jung Y. H., Yi, C. S., Kim, H. S. and Shim, M. P. (2005). "Estimation of needed discharge considering frequency based low flow in gabcheon basin." *Journal of Korean Society of Civil Engineer*, Vol. 25, No. 2, pp. 97-105.
 Heo, J.-H. (1997). "Introduction to statistical hydrology(v)." *Journal of Korean Water Resources Association*, Vol. 30 No. 1, pp. 88-96.
 Katz, R. W. (1996). "Use of conditional stochastic models to generate climate change scenarios." *Climate. Change*, Vol. 32, pp. 237-255.
 Kwon, H.-H. and So, B.-J. (2011). "Development of daily rainfall simulation model using piecewise kernel-pareto continuous distribution." *Journal of Korean Society of Civil Engineer*, Vol. 31, No. 3B, pp. 277-284.
 Kwon, H.-H., Lall, U. and Obeysekera, J. (2009). "Simulation of daily rainfall scenarios with interannual and multidecadal climate cycles for south florida." *Stochastic Environmental Research and Risk Assessment*, Vol. 23, No. 7, pp. 879-896.
 Lee, C. H. and Kim, S. (1995). "Estimation of mean annual and monthly precipitations in south korea by the regression analysis." *Journal of Korean Society of Civil Engineer*, Vol. 15, No. 5, pp. 1255-1266.
 Nord, J. (1975). "Some applications of markov chains, proceedings fourth conference on probability and statistics in atmospheric science." *Tallahas*, pp.125-130.
 Wilks, D. S. (1992). "Adapting stochastic weather generation algorithms for climate change studies." *Climate Change*, Vol. 22, pp. 67-84.
 Bengio, Y. and Frasconi, P. (1995). "Diffusion of context and credit information in markov chain models." *Journal of Artificial Intelligence Research*, Vol. 6, No. 95, pp. 249-270.