

Chow-Liu Tree 모형과 동질성 Hidden Markov Model을 연계한 다지점 일강수량 모의기법 개발

Development of Multi-Site Daily Rainfall Simulation Based on Homogeneous Hidden Markov Chain Model Coupled with Chow-Liu Tree Structures

권 현 한* / 김 태 정** / 김 윤 기*** / 이 동 루****

Kwon, Hyun-Han / Kim, Tae Jeong / Kim, Oon Ki / Lee, Dong Ryul

Abstract

This study aims to develop a multivariate daily rainfall simulation model considering spatial coherence across watershed. The existing Hidden Markov Model (HMM) has been mainly applied to single site case so that the spatial coherences are not properly addressed. In this regard, HMM coupled with Chow-Liu Tree (CLT) that is designed to consider inter-dependences across rainfall networks was proposed. The proposed approach is applied to Han-River watershed where long-term and reliable hydrologic data is available, and a rigorous validation is finally conducted to verify the model's capability. It was found that the proposed model showed better performance in terms of reproducing daily rainfall statistics as well as seasonal rainfall statistics. Also, correlation matrix across stations for observation and simulation was compared and examined. It was confirmed that the spatial coherence was well reproduced via CLT-HMM model.

Keywords : daily rainfall, Hidden Markov Chain Model, Chow-Liu Tree, spatial correlation, correlation matrix

요 지

본 연구에서는 유역의 공간상관성을 고려한 다지점 일단위 강수량을 동시에 모의할 수 있는 일강수량 모의기법을 개발하였다. 기존 Hidden Markov Chain Model(HMM)은 단일지점 강수모의에 적용되어 왔으나 관측지점간의 유역상관성을 충분히 고려하지 못하는 문제점을 가지고 있다. 따라서 본 연구에서는 Chow-Liu Tree (CLT) 모형을 적용하여 다변량(multivariate) 형태로써 유역내에 위치한 강우관측소간의 상호종속성을 고려하기 위하여 기존의 동질성 HMM 강수모의기법과 CLT 알고리즘을 결합한 동질성 CLT-HMM 모형을 개발하였다. 본 연구에서 개발된 동질성 CLT-HMM 모형을 사용하여 장기간의 수문자료를 보유하고 있는 기상청 산하의 한강유역 강수네트워크에 대해서 적합성을 검토하였다. 동질성 CLT-HMM 모형을 적용하여 모의된 결과를 보면 일강수량의 계절적 특성뿐만 아니라 일강수량 모의 시 강수시계열의 통계적인 특성들까지 우수하게 모의하였다. 추가적으로 상관행렬(correlation matrix)을 이용하여 기상관측소간의 공간상

** 전북대학교 공과대학 토목공학과 부교수, 방재연구센터 (e-mail: hkwon@jbnu.ac.kr)
Associate Professor, Department of Civil Engineering, Chonbuk National University, Jeonju 561-756, Korea
** 전북대학교 공과대학 토목공학과 석사과정, 방재연구센터 (e-mail: kim.t.j@jbnu.ac.kr)
Master Course, Department of Civil Engineering, Chonbuk National University, Jeonju 561-756, Korea
*** 정읍시청 재난안전관리과 (e-mail: kog6858@korea.kr)
Disaster & Safety Administration, Jeongeup City Hall, Jeongeup
**** 교신저자, 한국건설기술연구원 수자원연구실 연구위원 (e-mail: dryi@kict.re.kr)
Corresponding Author, Korea Institute of Construction Technology Water Resources Research Division

관 재현성을 검토한 결과 관측지점들 사이의 공간상관성도 비교적 우수하게 재현하는 것을 확인할 수 있었다.

핵심용어 : 일강수량, Hidden Markov Chain 모형, Chow-Liu Tree, 유역상관성, 상관행렬

1. 서 론

최근 기후변동성 증가와 극치수문사상의 발생빈도가 증가하는 실정으로 인하여 신뢰성 있는 수자원계획 수립이 어려워 수자원관련 재난피해 또한 빈번하게 발생하고 있다. 이러한 극치강우사상과 연계하여 발생하는 집중호우는 일상생활이나 산업활동에 상당한 피해를 초래하고 있으며, 이와 반대로 봄에 발생하는 극한가뭄은 농작물 생산의 커다란 피해를 주고 있는 실정이다. 이러한 기상현상으로 인한 재해의 예방을 위해서 사전에 위험을 인지하고 그 규모를 예측할 수 있는 여러 수문해석기법들이 개발 및 적용되고 있다. 수공구조물의 설계 및 유역단위의 수자원계획을 수립하기 위한 가장 기본적인 자료 중 하나는 대상유역의 강수량자료이며, 이와 같은 배경으로 유역단위의 장기적인 강수시나리오를 도출하기 위한 다양한 강수모의기법 개발이 이루어져 왔다.

특히 수공학 분야에서 저수지 및 댐 설계와 같은 장기적인 수자원계획을 수립하기 위해서는 장기간의 유출량 자료가 필수적으로 요구된다. 그러나 현실적으로 장기간의 유출량 자료를 획득하기가 어려우며 이러한 점에서 강수모의기법을 이용하여 유출량 자료를 확충하는 간접적인 방법이 주로 이용되고 있다(Kim, 2013). 그러나 강수모의기법을 적용하는데 있어서 우리나라의 경우 강우의 대부분이 여름철인 6~9월에 집중되며, 특히 강수의 변동성이 매우 큰 특성으로 인하여 기존 일강수량 모의기법 적용시 관측 강수량의 특성을 재현하는데 어려움이 따른다.

유역단위에서 강우사상은 시·공간적 변동성이 매우 크며 특히 산악지형과 해안가 등이 많은 우리나라에서는 지형적인 영향을 더 많이 받는 것이 일반적이다. 이러한 시·공간적 특성을 고려하는 가장 근본적인 이유는 강수모의시 유역단위에서 극치강수량 재현과 더불어 현실성 있는 유출분석을 위함이다. 즉, 기존 강수모의기법의 경우 지점별로 무작위로 강수모의가 이루어지기 때문에 유역단위에서 면적강수량을 효과적으로 복원하기가 어려운 단점이 있다. 이러한 점에서 본 연구의 목적은 유역단위에서 나타나는 강수지점간의 상관특성을 고려한 다지점 강수모의기법을 개발하는 것이며 궁극적으로 장기유출시나리

오를 도출하기 위한 강수시나리오로 활용하기 위함이다.

일강수량을 모의하는 방법은 여러 가지 방법이 있지만, 국내외적으로 Markov Chain을 중심으로 많은 연구가 이루어져 왔다. Heo (1997)은 선형 또는 다중회귀식을 사용하여 관측소간의 공간적인 가중치를 주어 각 지점의 일강수량을 모의하였다. Kang and Seoh (2000)는 자료의 유역내 각 지점의 공간적 특성을 고려하여 강수량을 모의 발생하기도 하였다. 또한 Cha and Moon (2003)은 Multivariate 핵밀도함수를 사용한 일강수량 모의 발생기법을 개발하여 서울지역에 적용하였다. Yoo and Park (2006)은 Markov Chain에 근거하여 지점간의 공간상관을 적절히 고려할 수 있는 일강우의 다지점 모의 발생방법을 제안하였다. 이러한 연구동향에 기인하여 최근 유역내 다수의 강우지점(multisite)의 강수를 동시에 모의하는 시공간적 강수모의 기법에 대한 많은 연구가 제안되고 있다(Wilby, 1994; Bates et al., 1998; Wilks, 1999; Hughes et al., 1999; Bellone et al., 2000; Mehrotra et al., 2004; Mehrotra et al., 2006; Robertson et al., 2004). 국외적으로 활발한 다지점 강수모의기법 연구동향에 맞추어 최근에 국내에서는 Keem et al. (2008)은 다지점 일강수 발생모형을 개발하여 낙동강 유역 강수관측망에 적용하였고, Jeong (2009)은 다지점의 일단위 강수량을 동시에 모의할 수 있는 추계학적 강수모의모형을 제시하여 우리나라 중부지역에 위치한 17개 관측지점에 적용한바 있다. Kwon and So (2011)는 불연속 Kernel-Pareto 분포를 이용한 일강수량 모의기법을 개발하여 한강유역에 적용한바 있다. 또한 Kwon et al. (2013)은 불연속 Kernel-Pareto 분포를 이용한 다지점 강수모의기법과 Bayesian HEC-1 모형을 연계하여 대청댐 유역에 적용한 바 있다. 일강수를 모의하는데 있어서 최근의 주된 관심사는 강우 및 무강우일을 특정 일기상태(weather state)를 통하여 묘사하기 위해서 종관기후학적(synoptic climatological) 대기 패턴을 고려할 수 있는 모형을 구성하는 것이다. 이러한 점을 고려하여 특정 일기상태를 고려한 강수모의 기법의 개발이 이루어지고 있다(Hughes and Guttorp, 1994; Kwon et al., 2009a; Kwon et al., 2009b).

본 연구에서는 일기상태를 고려하고 각 일기상태별로 Mixture 확률분포를 적용할 수 있는 일강수량 모의기법

을 적용하였다. 본 연구에서 일기상태는 은닉상태(hidden state)의 형태를 가지는 것으로 고려하였으며 은닉상태 안에서의 천이확률(transition probability)은 시간에 따라 변화하지 않는 것으로 가정하였다. 즉, 시간에 따른 천이확률 및 Mixture 분포의 변동성은 고려하지 않는 정상성 가정을 기본으로 하는 동질성 Hidden Markov Chain (HMM) 모델을 적용하였다. HMM은 각각의 은닉상태마다 조건부 확률분포(conditional probability distribution)를 부여하고 이들 조건부 확률분포로부터 강우를 모의하는 방법이다.

본 연구에서는 대상유역에서 관측된 강수량 자료를 바탕으로 강수발생과 강수량에 대해 각 관측소지점의 강수특성과 다지점(multi-site) 강우관측소 사이의 공간적 상관특성을 보다 정량적으로 재현하기 위하여 조건부 확률분포에 근간을 둔 Chow-Liu Tree (CLT) 방법을 도입하여 유역내에 위치한 다수의 강우관측소의 공간상관성을 효과적으로 고려하는 모형으로 확장하고자 한다. 본 연구에서는 구축된 모형을 1973년부터 2012년까지 총 40년간의 기상청관할 한강유역 14개 관측소 일강수량 자료를 사용하여 대상유역 강수사상의 시·공간적 특성과 계절적 변동성을 평가하고, 추가적으로 대상유역의 강수발생 여부와 강수량의 공간상관성 재현기법인 상관행렬(correlation matrix)을 적용하여 본 연구에서 적용된 CLT-HMM 모형의 적합성 여부를 검증하였다. 검증에 필요한 모의 강수량은 총 100 번의 반복실행에 의한 앙상블 형태의 강수시계열로부터 추출하여 사용하였다.

본 논문의 구성은 1장에서는 논문의 배경 및 목적에 대해서 언급하였으며, 2장에서는 동질성 HMM모형과 CLT 모형에 대한 분석방법의 설명을 위주로 기술하였다. 3장에서는 대상유역 소개 및 관측자료와 대상유역의 강우관측소에 대해 CLT-HMM 모형의 적용결과를 정리하여 나타내었으며 마지막으로 결론 및 토의를 4장에 수록하였다.

2. 분석 방법

2.1 동질성 Hidden Markov Chain Model

기존 HMM 모형은 주로 음성인식분야에서 사용을 시작하였으며 점차적으로 확대되어 생물학과 경제학 등으로 적용분야가 확장되었다. 수공학 분야에서는 1990년대 HMM을 사용하여 미국 워싱턴 지역 강우지점의 강수발생을 모의하는데 사용하였으며, 또한 은닉지점의 비동질성 변화와 은닉지점과 입력변수에 따라 매개변수가 변동하는 동적 HMM이 적용된 사례가 있다.

R_t 는 시간 t 에서 차원 M 을 갖는 강수량 벡터라고 하자. $R_{1:T} = (R_1, \dots, R_T)$ 는 시간 T 의 연속 벡터계열 자료를 나타낸다. $S_{1:T} = (S_1, \dots, S_T)$ 를 시간 t 에 해당하는 은닉상태(hidden state)를 나타내며 은닉상태의 수 K 값 중 하나의 값을 갖는다고 가정한다. 예를 들어 은닉상태의 수 K 가 총 5개라 하면 1에서부터 5까지 은닉상태가 존재하며 시간 t 에 해당하는 날짜에 1에서 5까지 값 중 하나가 부여된다. HMM 모델은 두 가지 조건부 독립 가정(conditional independence assumptions)을 이용하여 $R_{1:T}$ 와 $S_{1:T}$ 의 결합 확률분포를 정의한다. 두 가지 조건부 독립 가정은 다음과 같다.

첫째, 개별적인 관측벡터 R_t 는 주어진 은닉상태 S_t 에서 모형의 모든 다른 변수들과 조건부적으로 독립이라고 가정한다.

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

둘째, 시계열의 바로 이전의 은닉상태에만 상관성이 있다고 보는 경우를 1차 HMM이라고 하며 다음과 같이 나타낼 수 있다.

$$P(s_t | s_{1:t-1}) = \begin{cases} P(s_t | s_{t-k:t-1}) & t > k, \\ P(s_t | s_{1:t-1}) & t \leq k. \end{cases} \quad (2)$$

동질성 HMM은 조건부 확률 $P(s_t | s_{t-1})$ 을 추정하는데 있어서 Fig. 1과 같은 정상성 가정을 기본으로 이루어진

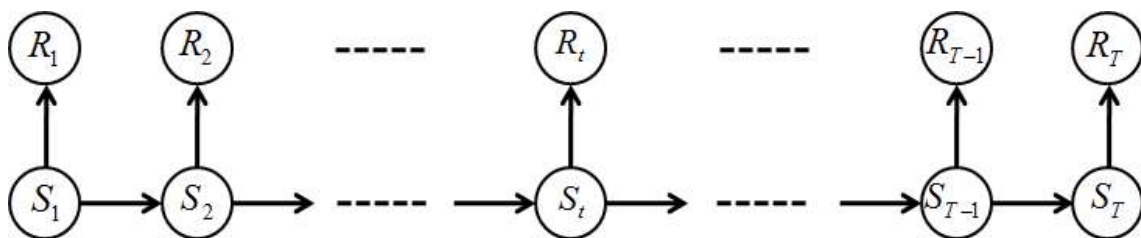


Fig. 1. Bayesian Network Representation of a Homogeneous HMM Satisfying Conditional Independence Assumptions

다. 즉, 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 정의한 후, 여기서 Υ 는 은닉상태 S_t 에 조건부로 결정되는 강수량 r_t 의 발생확률 $F_i(r) = P(r_t|S_t = i)$ 로 정의할 수 있으며 최종적으로 강우량자료, 은닉상태, 확률밀도함수 매개변수의 결합확률은 아래의 Eq. (3)과 같이 나타낼 수 있다.

$$P(r_{1:T}, S_{1:T} = s_{1:T} | \Pi, \Gamma, \Upsilon) = \left[\pi_{s_1} \prod_{t=2}^T \gamma_{s_{t-1}s_t} \right] \left[\prod_{t=1}^T F_{s_t}(r_t | r_{t-1}) \right] \quad (3)$$

2.2 Hidden Markov Model과 Chow-Liu Tree를 연계한 다지점 강수 모의

기존의 HMM은 단일지점 강수모의 기법에 일반적으로 적용되어 왔으나 관측지점간의 유역 상관성을 현실적으로 고려하지 못하는 문제점을 가지고 있었으며 이러한 문제점으로 인하여 실질적으로는 매우 제한적인 형태로 적용되었다.

Chow-Liu Tree (CLT)는 다변량 형태의 다지점의 상호 종속성을 고려하기 위해 개발되었으며, 앞서 언급한 목적에 맞추어 HMM과 CLT를 결합한 CLT-HMM 다지점 강수모의 모형을 적용하였다. CLT모형은 결합 확률밀도함수를 근사적으로 해석하는데 효율적인 방법으로서 Chow and Liu (1968)에 의해서 처음 제안되었다. 즉, CLT 방법은 결합 확률밀도함수를 추정하는데 있어서 2차(second order) 조건부확률분포와 주변확률밀도함수의 곱(product)을 통하여 이루어진다. 예를 들어, 강수 지점간의 상관성 구조가 Fig. 2와 같다면 4차원의 결합확률분포 $P(X_1, X_2, X_3, X_4, X_5)$ 는 $P(X_5 | X_4)P(X_4 | X_2)P(X_3 | X_2)P(X_2 | X_1)$ 로 추정이 가능하다. Fig. 2에서 보면 각 향간의 곱은 지점간의 종속성

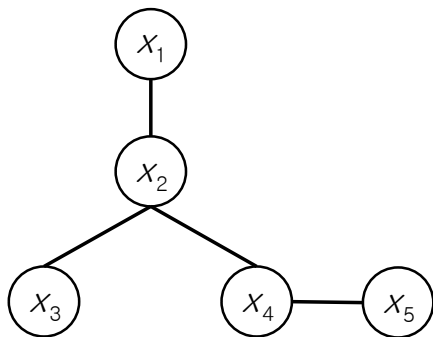


Fig. 2. A first-order Dependency of Chow-Liu Tree Representing Coherences between Weather Stations

으로 표현되며 1차(first order) 상관 Tree형태를 가지고 있다. 결국 CLT 알고리즘은 여러 차수의 종속성 중에서 실제로 상관성이 큰 종속 Tree만을 선별하고 조건부확률로서 결합확률을 추정하는 방안이라 할 수 있다. 따라서 3차 이상의 종속성을 고려하지 않기 때문에 완전한 형태의 결합 확률밀도함수를 추정하기 보다는 근사적인 해석방법으로 활용이 되고 있다. Bayesian Network 형태의 CLT 모형은 Pearl (1988)에 의해서 제안되었으며 Kirshner et al. (2004)는 CLT 모형을 강수지점간의 상관성을 복원하기 위한 목적으로 강우모의기법에 처음 적용되었다. 최근 기후 변화에 따른 영향을 검토하기 위해서 CLT를 이용한 강수 모의기법 연구 및 적용이 이루어지고 있다(Kirshner et al., 2004; Kwon et al., 2011). CLT 이론에 대해서 Kirshner et al. (2004)과 Kwon et al. (2011)에서 제시된 방법론을 중심으로 요약하여 정리하였다.

동질성 HMM모형은 시간적인 구조를 모의할 수 있는 반면에 CLT는 다변량 형태의 종속성을 모의할 수 있는 방법으로서 유역단위에 다지점 형태의 강우네트워크에 모형을 적용할 수 있는 장점을 가지고 있으며, CLT-HMM 모형에서 M개의 강수지점과 시간 T의 강우벡터가 K개의 은닉상태 수를 갖는다고 가정하면 매개변수 집합 θ 는 $K \times K$ 의 천이확률 행렬 Γ , $K \times 1$ 의 은닉상태의 초기확률 Π , 은닉상태 i 에 해당하는 매개변수 (ϕ_i, v_i) 를 가지는 Tree형태의 $T_i(r|r_{prev})$ 로 구성된다.

HMM과 연계한 CLT 모형의 매개변수 θ 에 대한 우도는 $P(r_{nt}|s_{nt}, r_{n,t-1}, \theta) = P(r_{nt}|s_{nt}, \theta)$ 와 Chow-Liu Tree $T_i(r_{nt}|r_{n,t-1}) = T_i(r_{nt})$ 를 이용하여 다음과 같이 계산될 수 있다. 본 연구에서는 이전 강우사상이 은닉상태를 통하여 모의되기 때문에 이전 강우사상 $r_{n,t-1}$ 을 누락이 가능하다.

i 의 값을 갖는 은닉상태 S_{t-1} 에 대해서 확률분포 $P(R_t|\theta)$ 는 천이확률 행렬 Γ 의 i 번째 행과 같은 값을 갖는 Mixture 계수 $(\gamma_{n1}, \dots, \gamma_{iK})$ 로 이루어진 CLT 혼합분포를 나타낸다. 관련된 매개변수들은 Expectation Maximization (EM) 기법(Dempster et al., 1977)의 하나인 Baum-Welch 알고리즘을(Baum et al., 1970)을 이용하여 우도를 최대화하는 매개변수 θ 를 추정하였다.

$$\begin{aligned} L(\theta) &= P(R|\theta) = \prod_{n=1}^N \sum_{s_{n:1}:n} P(s_{n:1}:T, r_{n:1}:T|\theta) \\ &= \prod_{n=1}^N \sum_{s_{n:1}:T} P(s_{n1}|\theta) \prod_{t=2}^T P(s_{nt}|s_{n,t-1}, \theta) \prod_{t=1}^T P(r_{nt}|s_{nt}, r_{n,t-1}, \theta) \quad (4) \\ &= \sum_{n=1}^N \sum_{i_{n1}=1}^K \pi_{i_{n1}} T_{i_{n1}}(r_{n1}) \sum_{t=2}^T \sum_{i_{nt}=1}^K \gamma_{i_{nt}-i_{nt-1}} T_{i_{nt}}(r_{nt}|r_{n,t-1}) \end{aligned}$$

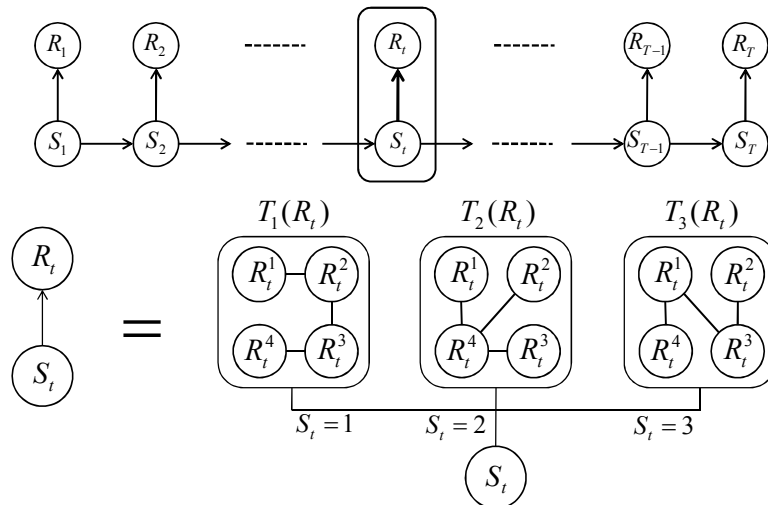


Fig. 3. Graphical Model Interpretation of a Proposed CLT-HMM (Kirshner and Robertson, 2004)

통계학에서 EM 알고리즘은 최우도(maximum likelihood) 또는 최대사후추정치(maximum a posteriori, MAP)를 기반으로 Hidden State와 같이 관측되지 않은 잠재변수(latent variable)와 연계된 매개변수를 추정하기 위한 목적으로 학습모형(learning model)에서 널리 이용되고 있다.

즉, 불완전한 입력 정보를 통하여 완전한 해를 구하는 목적으로 개발된 모형이 EM 기법으로서 상대적 간편성과 매개변수 추정이 상대적으로 효율적인 방법으로서 HMM이나 Bayesian Networks와 같이 확률적인 모델을 기반으로 음성인식분야, 패턴인식 분야에 활용되고 있다(Matsuyama, 2011). 다만 초기값에 대해서 민감도가 매우 크기 때문에 다수의 매개변수의 추정을 통하여 해의 수렴을 검토하는 과정이 반드시 필요하다.

일강수량 모의를 위해서 우리나라의 강수분포 분석에 적합한 확률분포로 일반적으로 사용되고 있는 Gamma 분포가 이용되며 무강수를 제외하고 강수에 대해서만 확률분포를 적용하는 Delta 함수가 도입된다(Khalil et al., 2010; Kwon et al., 2011; Kwon et al. (2013). 본 연구에서는 Delta-Gamma 분포를 활용하여 강수량을 모의하였다. 이렇게 혼합분포를 활용함으로써 강수량의 다양한 확률분포 특성을 고려할 수 있는 장점이 있다.

Delta-Gamma분포의 확률밀도함수와 매개변수를 추정하기 위한 우도함수는 다음 Eq. (5)와 같다. 여기서, 첨자 M, i, c 는 각각 관측소의 개수, 은닉상태 번호, 혼합분포의 개수를 나타내며 p 는 혼합분포의 가중치를 의미한다. 또한 α, β 는 Gamma 분포의 매개변수를 나타낸다.

$$P(r_t|S_t = i) = \prod_{m=1}^M P(r_t^m|S_t = i) = \prod_{m=1}^M a_{im} \quad (5)$$

여기서,

$$a_{im} = \begin{cases} p_{im0} & r_t^m = 0, \\ \sum_{c=1}^{num_2 - 1} \frac{p_{im1}}{p_{im1}} \frac{\beta_{imc}^{\alpha_{imc}} (r_t^m)^{\alpha_{imc} - 1} \exp(-\beta_{imc} r_t^m)}{\Gamma(\alpha_{imc})} & r_t^m > 0 \end{cases}$$

본 연구에서는 강수량을 모의하기 위해서 각 은닉상태별로 혼합 Gamma 분포의 매개변수, 은닉상태간의 천이확률 및 천이확률 모의를 위한 Logistic 모형의 매개변수 또한 앞서 언급한 EM 기법을 통하여 추정하였다.

3. 모형 적용

3.1 대상지역 및 자료

본 연구에서는 장기간의 자료를 보유하고 있는 한강유역 강수량에 대하여 CLT-HMM모형의 적합성을 검토하고자 한다. 한강유역은 북위 36°30'~38°55', 동경 126°24'~129°02'에 걸쳐 한반도의 중앙부에 위치한 강으로 강원도, 충청북도, 경기도, 서울특별시를 거쳐 서해로 유입되는 한반도 중부지역의 강으로 유량을 기준으로 할 경우 남한에서 가장 규모가 큰 강이다. 한강의 유역면적은 26,018 km²이며, 유로연장은 514.8 km이며 직할하천 15개, 하천연장 813.0 km, 지방하천 12개 하천연장 553.0 km, 지방 관리 하천 678개 하천연장 5,890 km로 총 하천 705개 하천 연장 7,256 km로 구성되어 있으며 한강의 본류인 경기도 양평군 양수리에서 남한강과 북한강이 만나서 이루어진다.

CLT-HMM 모형의 적용을 위하여 본 연구에서는 강우자료의 확보가 용이하고 과거로부터 관측 자료의 신뢰

성이 확보되어 있는 한강유역의 기상청관할 관측소의 관측자료를 모형의 분석 및 검증을 위한 입력 자료로 사용하였다. 강우분석을 위한 14개 관측소의 관측된 자료를 확인한 결과 이들 관측소간의 관측기간이 상이하고 관측자료 중에서 관측년수가 30년 이상인 10개의 지점(대관령, 제천, 충주, 원주, 양평, 이천, 인제, 춘천, 홍천, 서울지점)을 분석 대상지점으로 선정하고, 30년 미만의 관측자료가 존재하는 4개의 지점(영월, 철원, 동두천, 문산지점)을 분석대상에서 제외하였다.

분석 대상지점으로 선정한 총 10개의 관측소의 관측 자료를 동일한 관측년도가 모든 관측소에서 확보될 수 있도록 1973~2012년까지의 총 40년의 관측 강수량 자료를 활용하였다. Table 1은 한강유역에 존재하는 강우관측소를 나열한 것으로 강수분석에 사용된 총 14개의 관측지점에 대한 기상관측소의 정보를 나타낸다.

3.2 모형 적용 결과

한강유역의 일강수량을 모의하기에 앞서, 1973년부터 2012년까지 한강유역의 총 10개 기상관측소의 일강수량 자료를 대상으로 모형을 적용시켰으며, 우리나라의 연 강수량이 집중되는 6월부터 9월까지의 강수량을 모형에 적용하였다. 은닉상태 수 K 를 지점별로 변화하면서 우도를 산출하고, 그 과정에서 K 가 증가함에 따라 우도함수의 증가로 일정한 은닉상태에서는 수렴하는 것을 확인하였다. 하지만 은닉상태를 지나치게 증가시키는 경우 모형의 매개변수가 과도하게 산정(overfitting)되는 경우가 발생하

기 때문에 최적의 은닉상태 수 $K=5$ 는 Log-likelihood를 기준으로 결정하였다. CLT-HMM 모형에서는 기상상태를 은닉상태로 가정하고 최적화 과정을 통해 유역내의 강우네트워크에 상태를 결정하게 되며 강우의 특성을 효과적으로 분석할 수 있는 장점이 있다. 즉, 강우계열에 1번부터 5번까지 은닉상태를 부여할 수 있으며 이를 그림으로 나타내면 Fig. 4와 같다.

Fig. 5는 분석된 은닉상태를 토대로 은닉상태별 강우량 및 강우빈도를 분석한 결과로서 상태 1의 경우 강우의 빈도는 매우 낮지만 여름강수량 전체의 45%를 차지하는 특징을 가지고 있으며 집중호우의 특성을 갖는 것으로 유추할 수 있다. 상태 5의 경우 강우의 빈도는 50%를 차지하지만 강우량은 5% 미만으로 매우 작은 강우강도의 강수형태임을 판단할 수 있다.

일강수량 모의기법을 통하여 모의된 강우량이 홍수빈도해석과 같은 수문해석에 사용되기 위해서는 일강수량의 통계적 특성치가 실제 관측자료와 유사하게 모의되어야 하며, 이러한 과정을 통하여 수자원 설계 및 계획에 활용이 가능할 것으로 판단된다. 아래의 그림들은 한강유역의 강수량에 대한 1973년부터 2012년까지 6월부터 9월 동안의 일강수량 모의결과를 나타내며 각각의 관측지점별로 총 100번의 반복 모의를 실시하였으며 총 100개의 강우모의 앙상블에서 추정된 강우 통계치를 비교하여 나타내었다.

첫 번째로 검토해야 할 사항은 계절강수량의 복원능력이다. 즉, 목표로 하는 계절의 일강수량을 모의한 후 계절

Table 1. List of Meteorological Observation Stations Used in This Study

Station_ID	Station_No.	Name	Latitude	Longitude	Data Period
1	100	Daegwallyeong	37.7	128.8	1973~2012
2	101	Chuncheon	37.9	127.7	1973~2012
3	108	Seoul	37.6	127.0	1973~2012
4	114	Wonju	37.4	128.0	1973~2012
5	131	Cheongju	37.0	127.4	1973~2012
6	202	Yangpyeong	37.5	127.5	1973~2012
7	203	Icheon	37.3	127.5	1973~2012
8	211	Inje	38.1	128.2	1973~2012
9	212	Hongcheon	37.7	127.9	1973~2012
10	221	Jecheon	37.2	128.2	1973~2012
-	121	Yeongwol	37.1	128.5	1995~2012
-	95	Cheorwon	38.1	127.3	1988~2012
-	98	Dongducheon	37.9	127.1	1998~2012
-	99	Munsan	37.9	126.8	2002~2012

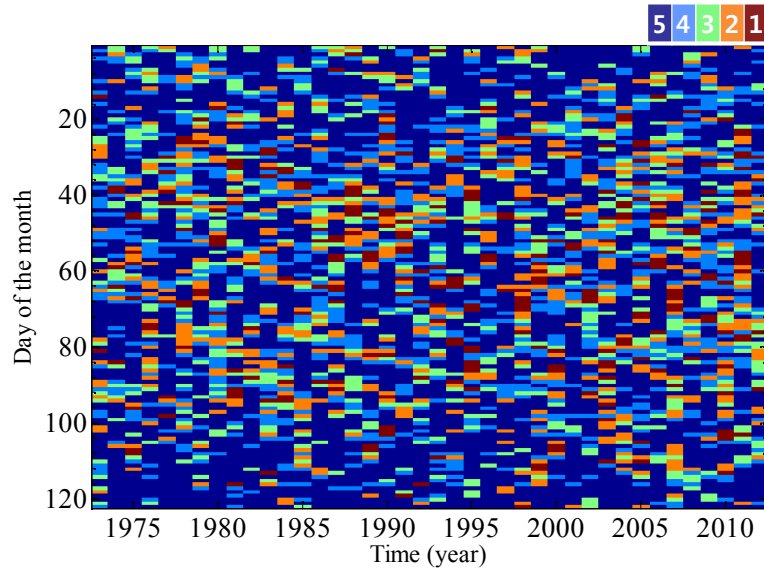


Fig. 4. The Estimated State Sequence Via Vieterbi Algorithm for JJAS (June–September) Season at Han River Watershed

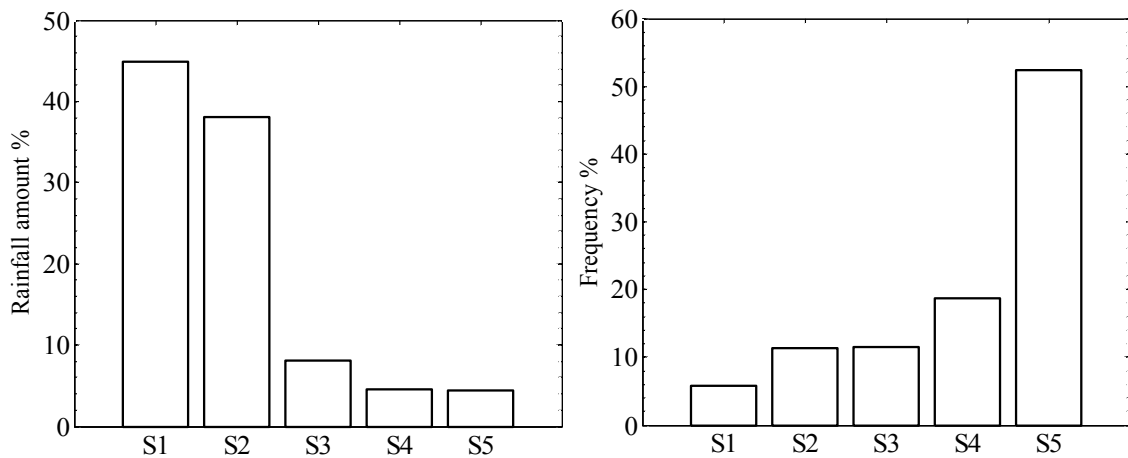


Fig. 5. The Estimated Percentage of Amount and Frequency of Seasonal Rainfall according to Each State at Han River Watershed

별로 합산했을 때 복원여부를 판단하는 것으로서 기상학적으로 경년변화의 모의능력과 연관성이 크며 Fig. 6에 CLT-HMM을 사용하여 모의된 강우량의 평균과 표준편차를 관측값과 비교하여 나타내었다. Fig. 6에서 나타나는 바와 같이 총 10개 지점의 계절모의강수량이 실측치와 매우 유사하게 모의됨을 확인할 수 있다. 모의결과는 유역내 10개 지점의 강수량을 동시에 모의한 후 비교한 결과로서 본 연구에서 제안하고 있는 CLT-HMM 다지점 모형이 유역내의 상관성을 고려하면서 효과적으로 강수량 모의가 가능한 점을 확인할 수 있었다.

특히, 일강수량에서 중요한 특성 중 하나는 모형을 통

해서 일강수계열의 확률적 특성이 얼마나 현실성 있게 재현될 수 있는가이다. 이러한 관점에서 CLT-HMM을 통하여 모의된 결과값을 boxplot를 사용하여 관측값과 비교하여 Fig. 7에 나타내었다. Fig. 7을 통하여 도출된 결론은 천이확률의 값이 부분적으로 과소 추정되고 있지만 전체적으로 신뢰구간에 위치하는 등 합리적인 모의결과를 나타내고 있다. 이러한 원인은 명확하지는 않으나 강수계열의 비정상성의 존재로 판단된다. 즉, 강수의 증가현상이 원인으로 고려될 수 있다. 그러나 이에 대한 분석은 본 연구의 범위를 벗어나는 연구로 추후연구로 진행이 필요할 것으로 판단된다.

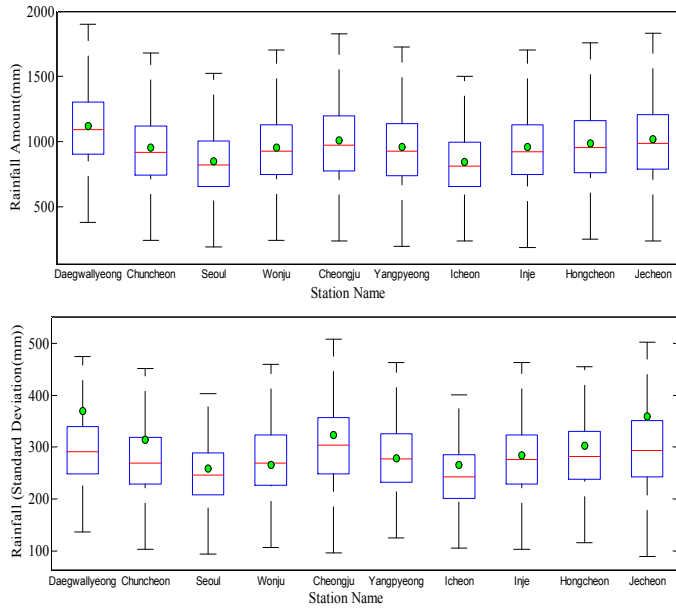


Fig. 6. A Comparison of Seasonally Rainfall Mean and Standard Deviation for Ten Weather Stations

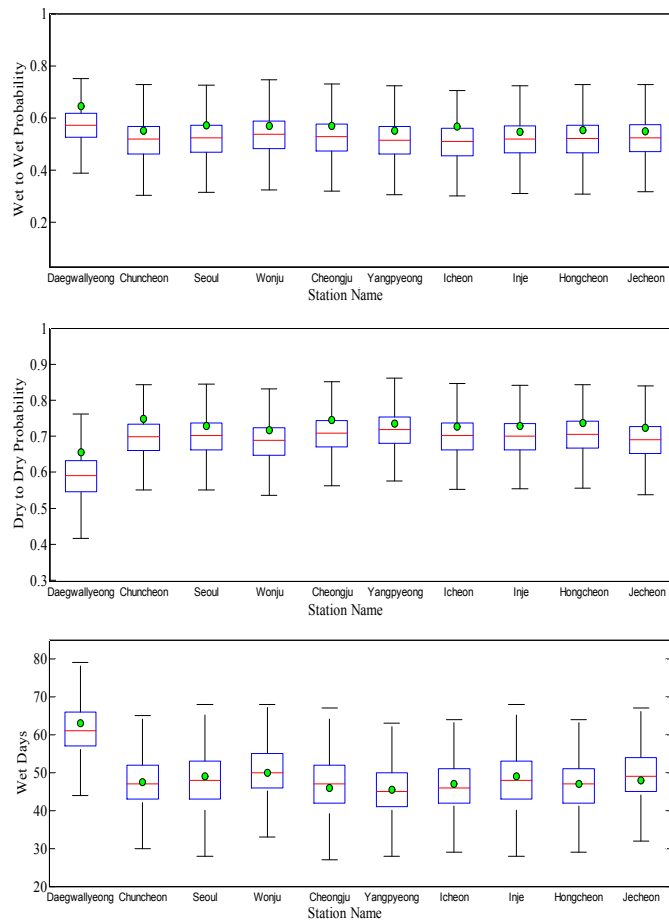


Fig. 7. A Comparison of Rainfall Wetdays and Rainfall Transition Probability between Observation and Simulation for Ten Weather Stations

Tables 2~5는 CLT-HMM을 통하여 모의된 일강수량 자료를 대상으로 평균 강우량과 천이확률 그리고 강우일수를 도출하여 관측강우량과 정량적으로 비교하였다. 모든 강수특성치에 대해서 관측값과 거의 유사한 통계적 특성을 가지고 재현되고 있음을 확인할 수 있다.

유역단위의 강우모의기법의 유역상관성 복원여부의 적합성을 평가하기 위해서 총 10개의 관측지점간의 상관행렬을 계산하여 Fig. 8에 나타내었다. Fig. 8의 상단과 하단은 총 10개의 관측지점에서 추정된 강수량에 대해서 각각 강우발생(occurrences)시계열과 강우량(amounts) 계열로 구분하여 지점간의 상관계수를 비교·분석한 결과를 나타내며 최대오차가 0.15정도로서 강우네트워크의 유역상관성 또한 효과적으로 복원하는 것을 확인할 수 있었다.

일강우량 모의결과값의 무강우 기준(threshold)을 0.1 mm로 설정하여 모든 지점에 대하여 관측강우량의 평균값과 표준편차를 비교하였다. Fig. 9에서 확인할 수 있듯이 CLT-HMM 강우모의기법을 통하여 얻은 결과값은 평균값의 경우 모든 지점에서 관측강우량의 평균값과 일치하는 것을 확인하였다. 표준편차의 경우는 1번 지점인 대관령 지점을 제외하고는 대체로 관측강우량의 표준편

차와 유사하게 모의되는 결과를 얻었다.

4. 결 론

우리나라는 연강수량의 편차가 매우 크게 나타나는 특징이 있어 연 강수량의 약 70% 이상이 6월부터 9월까지의 기간(약 122일) 동안 연간 강수량의 약 3분의 2가 집중되는 반면 10월부터 3월까지의 연간 강수량의 약 5분의 1 이하로 이와 같은 현상으로 인하여 홍수와 가뭄과 같은 기상재해가 빈번히 발생하고 있다. 본 연구에서는 동질성 Hidden Markov Chain 모형과 Chow-Liu Tree 모형을 연계한 추계학적 다지점 일강수량 모의기법을 구축하였으며, 구축된 CLT-HMM 모형을 6~9월까지 우리나라 우기의 일강수량을 대상으로 한강유역에 위치한 기상청관할 총 10개 관측소지점의 강우네트워크에 적용하여 모형의 적합성 및 적용성을 평가하였다. 본 연구를 통하여 얻은 결론은 다음과 같다.

첫째, 다지점 일강우량 모의기법인 CLT-HMM 모형의 적합성을 평가하기 위해서 한강유역의 총 10개의 강우관측소의 관측자료를 대상으로 연구를 진행하였다. 유역내 강

Table 2. A Comparison of Rainfall Amount (mm) Observation and CLT-HMM

Station	1	2	3	4	5	6	7	8	9	10	Mean
Observation	1031.4	936.2	813.2	949.5	962.8	916.2	838.5	923.0	915.6	954.1	924.1
CLT-HMM	1095.9	922.1	820.6	925.9	975.2	927.4	812.0	923.0	954.2	986.6	934.3

Table 3. A Comparison of Wet to Wet Probability Observation and CLT-HMM

Station	1	2	3	4	5	6	7	8	9	10	Mean
Observation	0.65	0.55	0.57	0.57	0.57	0.55	0.57	0.55	0.55	0.55	0.57
CLT-HMM	0.57	0.51	0.52	0.53	0.52	0.51	0.51	0.51	0.52	0.52	0.52

Table 4. A Comparison of Dry to Dry Probability Observation and CLT-HMM

Station	1	2	3	4	5	6	7	8	9	10	Mean
Observation	0.66	0.75	0.73	0.72	0.75	0.74	0.73	0.73	0.74	0.72	0.72
CLT-HMM	0.59	0.70	0.70	0.68	0.71	0.71	0.70	0.70	0.70	0.69	0.69

Table 5. A Comparison of Wet to Wet Days Observation and CLT-HMM

Station	1	2	3	4	5	6	7	8	9	10	Mean
Observation	63.0	47.5	49.0	50.0	46.0	45.5	47.0	49.0	47.0	48.0	49.2
CLT-HMM	61.3	47.4	47.8	50.2	47.2	45.5	46.7	48.0	46.8	49.3	49.0

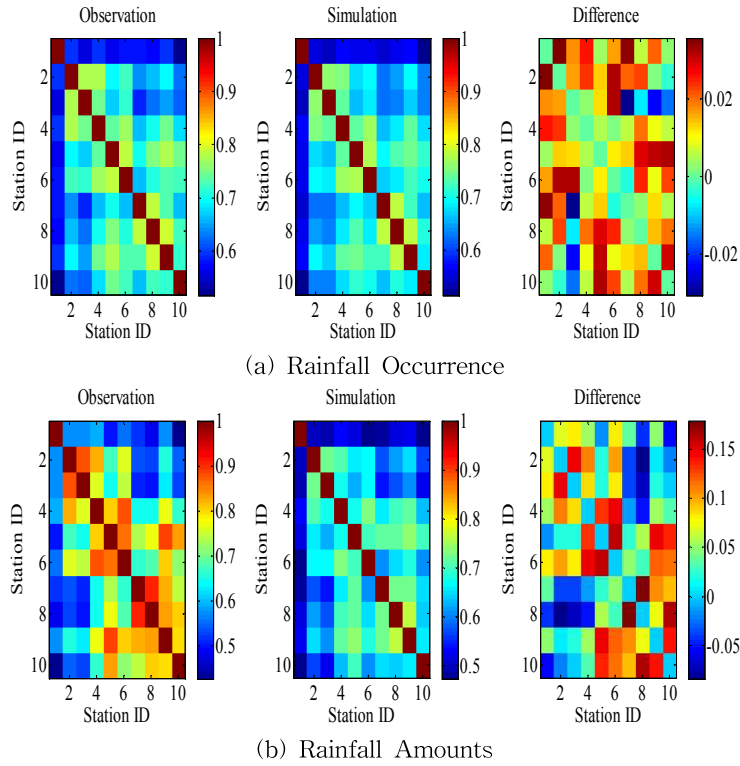


Fig. 8. A Comparison of Correlation Matrix Rainfall Occurrences (a) and Amounts (b) between Observation and CLT-HMM Simulation

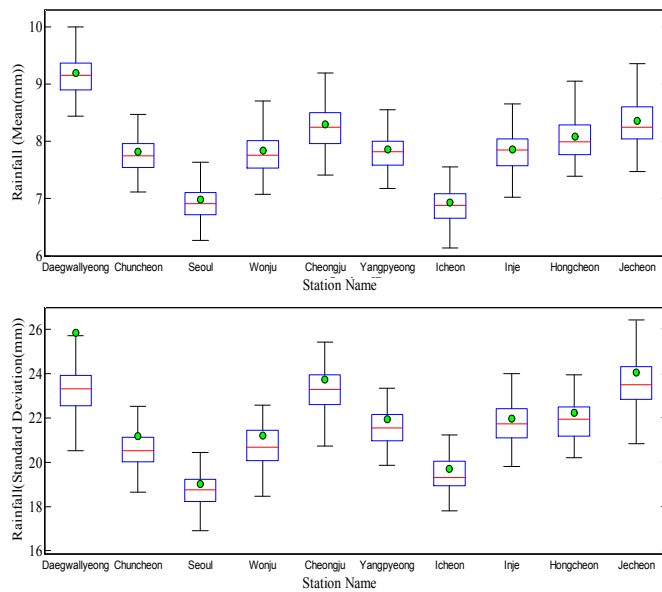


Fig. 9. A Comparison of Daily Rainfall Statistics (mean and standard deviation) between Observation and Simulation for Ten Weather Stations

우지점간의 공간상관성 재현여부는 강우발생(rainfall occurrence)과 강우량(rainfall amount)을 분리하여 평가를 하였다. 각각 CLT-HMM모형으로부터 얻은 모의값과

관측강우량을 사용하여 강우지점간의 상관행렬을 추정하여 비교한 결과 지점간의 상관행렬이 효과적으로 재현되는 것을 확인하였으며 다지점 강우모의 모형으로서 유역

단위 수문모형의 입력 자료로 활용이 가능할 것으로 판단된다.

둘째, 일강우의 통계적 특성을 효과적으로 재현하고 있는지 검토한 결과 유역에 존재하는 강우네트워크의 강우 발생패턴을 효과적으로 복원하면서 일강우량 통계 특성이 또한 효과적으로 복원할 수 있는 것을 확인하였다. 이러한 도출된 결과를 바탕으로 유역단위에서 면적강우량 추정시 효과적으로 재현이 가능할 것으로 기대되며, 이를 토대로 모의자료 기반의 홍수빈도분석 및 강우빈도해석에 입력 자료로 활용이 가능할 것으로 판단된다.

셋째, 본 연구에서 적용한 CLT-HMM 모형은 기상학적 상태를 고려할 수 있는 일기상태(weather state) 기반의 강우모의기법으로서 강우의 발생 및 강우의 양적 특성을 기상학적으로 분석할 수 있었다. 기상학적 진단결과 우리나라의 경우 집중호우와 관련된 일기상태와 무강우 또는 적은 강우와 연계된 일기상태를 효과적으로 분리할 수 있었으며 이러한 기상학적 연계성을 고려하여 강우예측 모형 개발 시 활용이 가능할 것으로 판단된다.

앞서 결과에서 언급했듯이 본 모형은 정상성 일강수량 모의기법으로서 강우계열의 비정상성을 고려하지 못하는 단점이 있다. 즉, 강우의 증가현상 및 주기 등이 존재하는 경우 동질성 CLT-HMM 모형을 적용하는데 한계가 있으며 이러한 점에서 추후 연구로서 비정상성 기반의 다지점 일강수량 모의기법의 연구가 필요할 것으로 판단된다.

감사의 글

본 연구는 한국건설기술연구원 주요사업(수문레이더 기반 홍수예경보 및 폭설 추정 플랫폼 개발)의 연구비 지원에 의해 수행되었습니다. 저자들은 전북대학교 방재연구센터에 소속되어 연구를 수행하였습니다.

References

Bates, B.C., Charles, S.P., and Hughes, J.P. (1998). "Stochastic downscaling of numerical climate model simulation." *Environmental Modeling & Software*, Vol. 13, No. 2-3, pp. 325-331.

Baum, L.E., Petrie, T., Soules, G., and Weiss, N. (1970). "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chain." *The Annals of Mathematical Statistics*, Vol. 41, No. 1, pp. 164-174.

Bellone, E., Hughes, J.P., and Guttorp, P. (2000). "A hidden Markov model for relating synoptic scale patterns to precipitation amounts." *Journal of Climate Research*, Vol. 15, No. 1, pp. 1-12.

Cha, Y.I., and Moon, Y.I. (2003). "Multivariate Kernel Density function and Simulation of Daily Precipitation." *Journal of University of Seoul Industrial Technical Laboratory*, pp. 119-118.

Chow, C.K., and Liu, C.N. (1968). "Approximating discrete probability distributions with dependence trees." *IEEE Transactions on Information Theory*, Vol. 14, No. 3, pp. 462-467.

Dempster, A., Laird, N., and Rubin, D. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society*, Vol. 39, pp. 1-38.

Heo, J.H. (1997). "Introduction to Statistical Hydrology (V)." *Journal of Korean Water Resources Association*, Vol. 30, No.1, pp. 88-96.

Hughes, J.P., and Guttorp, P. (1994). "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena." *Water Resources Research*, Vol. 30, No. 5, pp.1535-1546.

Hughes, J.P., Guttorp, P., and Charles, S.P. (1999). "A non-homogeneous hidden Markov Model for precipitation occurrence." *Journal of the Royal Statistical Society*, Vol. 48, No. 1, pp. 15-30.

Jeong, D.I. (2009). "Development of a Stochastic Precipitation Generation Model for Generating Multi-site daily Precipitation." *Journal of Korean Society of Civil Engineer*, Vol. 29, No. 5, pp. 397-408.

Kang, K.S., and Seoh, B.H. (2000). "Development and Application of Multi-site Daily Rainfall Generation Model." *Journal of Korean Society of Civil Engineer*, Vol. 2000, No. 3, pp. 19-22.

Keem, M.S., Ahn, J.H., Han, S.H., and Kim, S.D. (2008). "Multi-site daily Precipitation Generator : Application to Nakdong River Basin Precipitation Gage Network." *Journal of Korean Society of Water Quality*, Vol. 24, No. 6, pp. 725-740.

Khalil, A.F., Kwon, H.-H., Lall, U., and Kaheil, Y.H. (2010). "Predictive downscaling based on non-homogeneous hidden Markov models." *Hydrological Sciences*

- Journal*, Vol. 55, No. 3, pp. 333-350.
- Kim, O.G. (2013). *Development of Daily Rainfall Simulation Technique Based on Multisite Hidden Markov Chain Model*, Ph.D dissertation, University of Cbonbuk, Jeonju, Jeollabuk, South Korea.
- Kirshner, S., and Robertson, A.W. (2004). *Conditional Chow-Liu tree structures for modeling discrete-valued vector time series* Tech. Rep. ICS-TR 04-04, Informational and computer Science, University of California, Irvine, CA
- Kirshner, S., Smyth, P., and Robertson, A.W. (2004). "Conditional Chow-Liu Tree Structures for Modeling Discrete-Valued Vector Time Series." *Appears in Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence (UAI2004)*, pp. 317-324.
- Kwon, H.-H., and Kim, B.S. (2009a). "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., 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. 3, pp. 277-284.
- 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 Korea Water Resources Association*, Vol. 46, No. 3, pp. 59-71.
- Kwon, H.-H., Kim, T.J., Hwang, S.-H., and Kim, T.-W. (2013). "Development of Daily Rainfall Simulation Model Based on Homogeneous Hidden Markov Chain." *Journal of the Korean Society of Civil Engineer*, Vol. 33, No. 5. DOI: <http://dx.doi.org/10.12652/Ksce.2013.33.5.000>.
- Kwon, H.-H., Lall, U., and Obeysekera, J. (2009b). "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.
- Kwon, H.-H., Sivakumar, B., Moon, Y.I., and Kim, B.S. (2011). "Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model." *Stochastic Environmental Research and Risk Assessment*, Vol. 25, No. 4, pp. 567-581.
- Matsuyama, Y. (2011). "Hidden Markov model Based on alpha-EM Algorithm : Discrete and continuous alpha-HMMs." *International Joint Conference on Neural Network, San Jose*, Vol. 7, No. 5, pp. 808-816.
- Mehrotra, R., Sharma, A., and Cordery, I. (2004). "Comparison of two approaches for downscaling synoptic atmospheric patterns to multisite precipitation occurrence." *Journal of Geophysical Research*, Vol. 109, No. D14107, DOI: 10.1029/2004JD004823
- Mehrotra, R., Srikanthan, R., and Sharma, A. (2006). "A Comparison of three stochastic multi-site precipitation occurrence generators." *Journal of Hydrology*, Vol. 331, No. 1-2, pp. 280-292.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA, Morgan Kaufmann.
- Robertson, A.W., Kirshner, S., and Smyth, P. (2004). "Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model." *Journal of Climate*, Vol. 17, No. 22, pp. 4407-4424.
- Wilby, R.L. (1994). "Stochastic weather type simulation for regional climate change impact assessment." *Water Resources Research*, Vol. 30, No. 12, pp. 3395-3403.
- Wilks, D.S. (1999). "Multisite downscaling of daily precipitation with a stochastic weather generator." *Journal of Climate Research*, Vol. 11, pp. 125-136.
- Yoo, C.S., and Park, S.H. (2006). "Analysis of Spatial Distribution of the rainfall observatory in Korea." *Journal of Korean Society of Civil Engineer*, Vol. 2006, No. 10, pp. 1164-1167.

논문번호: 13-059	접수: 2013.07.15
수정일자: 2013.09.04/09.11	심사완료: 2013.09.11