

## 은닉 마코프 모형을 이용한 한강유역 수문학적 가뭄의 확률론적 평가

### Probabilistic Assessment of Hydrological Drought Using Hidden Markov Model in Han River Basin

박 예 준\* / 유 지 영\*\* / 권 현 한\*\*\* / 김 태 웅\*\*\*\*

Park, Yei Jun / Yoo, Ji Young / Kwon, Hyun-Han / Kim, Tae-Woong

#### Abstract

Various drought indices developed from previous studies can not consider the inherent uncertainty of drought because they assess droughts using a pre-defined threshold. In this study, to consider inherent uncertainty embedded in monthly streamflow data, Hidden Markov Model (HMM) based drought index (HMDI) was proposed and then probabilistic assessment of hydrologic drought was performed using HMDI instead of using pre-defined threshold. Using monthly streamflow data (1966~2009) of Pyeongchang river and Upper Namhan river provided by Water Management Information System (WAMIS), applying the HMM after moving-averaging the data with 3, 6, 12 month windows, this study calculated the posterior probability of hidden state that becomes the HMDI. For verifying the method, this study compared the HMDI and Standardized Streamflow Index (SSI) which is one of drought indices using a pre-defined threshold. When using the SSI, only one value can be used as a criterion to determine the drought severity. However, the HMDI can classify the drought condition considering inherent uncertainty in observations and show the probability of each drought condition at a particular point in time. In addition, the comparison results based on actual drought events occurred near the basin indicated that the HMDI outperformed the SSI to represent the drought events.

**Keywords** : drought index, uncertainty, hidden markov model, posterior probability

#### 요 지

지금까지 많은 연구를 통하여 제안된 다양한 가뭄지수들은 사전에 정의된 등급을 통하여 가뭄을 평가하기 때문에 가뭄현상에 내재된 불확실성을 고려하지 못하고 있다. 본 연구에서는 월 유출량 자료에 내재되어 있는 불확실성을 고려하기 위해 은닉 마코프 모형(HMM) 기반의 가뭄지수(HMDI)를 제안하고, 이를 이용하여 수문학적 가뭄에 대한 확률론적 평가를 수행하였다. WAMIS에서 제공하는 한강유역의 평창강과 남한강상류의 월평균 유출량 자료(1966~2009)를 이용하여 3, 6, 12개월씩 누적시킨 후, HMM에 적용하여 은닉상태의 사후확률을 계산하였다. 연구방법의 검증을 위해 HMM을 이용하여 추정된 각 은닉상태 별 사후확률(HMDI)과 기준값에 의해 가뭄을 평가하는 방법 중 하나인 표준유출지수(SSDI)와 비교를 하였다. 분석결과, 기존 가뭄지수(SSDI)를 사용하였을 때는 하나의 지수로 특정 시점에서의 가뭄 상태를

\* 한양대학교 대학원 건설환경공학과 석사과정 (e-mail: swchz@hanyang.ac.kr)

Graduate Student, Dept. of Civil and Environmental Engineering, Hanyang Univ., Seoul 133-791, Korea

\*\* 전북대학교 공과대학 토목공학과 연구교수 (e-mail: 7924pooh@hanmail.net)

Research Professor, Dept. of Civil Engineering, Chonbuk National Univ., Jeonju 561-756, Korea

\*\*\* 전북대학교 공과대학 토목공학과 부교수 (e-mail: hkwon@jbnu.ac.kr)

Associate Professor, Dept. of Civil Engineering, Chonbuk National Univ., Jeonju 561-756, Korea

\*\*\*\* 교신저자, 한양대학교 공과대학 건설환경플랜트공학과 부교수 (e-mail: twkim72@hanyang.ac.kr Tel: 82-31-400-5184)

Corresponding Author, Associate Professor, Dept. of Civil and Environmental Engineering, Hanyang Univ., Ansan 426-791, Korea

판단하였지만, HMDI는 자료에 내재된 불확실성을 이용하여 가뭄의 상태를 분류하였고, 이는 특정 시점에서 가뭄 상태들이 나타날 확률로 표현되었다. 또한, 실제 가뭄사례와의 비교를 통해서 HMDI가 SSI에 비하여 가뭄에 대한 재현능력이 우수한 것으로 나타났다.

**핵심용어** : 가뭄지수, 불확실성, 은닉 마코프 모형, 사후확률

## 1. 서 론

가뭄은 정의하는 관점에 따라 크게 기상학적, 수문학적, 농업적, 사회경제적 가뭄과 같이 분류할 수 있으며 (American Meteorological Society, 2004), 보통 가뭄지수를 통하여 가뭄의 심도를 판단한다. 기상학적 가뭄은 일정 기간의 강수량이나 무강우 지속일수 등으로 정의되고, 수문학적 가뭄은 강수량의 부족으로 인하여 하천유출과 저수지 저수량의 부족현상이 주된 인자이며, 강수량에 따른 유출량 과부족이 주요 지표이다. 또한, 농업적 가뭄은 농작물의 성장과 직접 관련된 토양수분을 이용하여 주로 표시되며, 사회경제적 가뭄은 기상학적, 수문학적, 농업적 가뭄의 요소를 모두 고려하여 사회와 경제에 미치는 영향을 지표로 나타난 것이다.

현재 국내외에서 가뭄을 정량적으로 평가하는 방법은 주로 수문학적 또는 기상학적인 요인을 고려하여 산정된 가뭄지수를 이용하는 것이다. 가뭄지수 중 표준강수지수 (Standardized Precipitation Index, SPI; McKee et al., 1993), 파머가뭄지수 (Palmer Drought Severity Index, PDSI; Palmer, 1965), 지표수분공급지수 (Surface Water Supply Index, SWSI; Shafer and Dezman, 1982)가 널리 사용되고 있다. Modarres (2007)은 수자원의 계획과 관리를 위하여 SARIMA 모형을 유출량 자료에 적용하여 가뭄을 예측하였으며, SPI와 동일한 기준값을 이용하여 가뭄을 판단하였다. Shukla and Wood (2008)은 표준유출지수 (Standardized Runoff Index, SRI)를 이용하여 가뭄의 특성을 파악하였으며, SPI와의 비교를 통해 SRI가 계절적 특성과 수문학적 관점에서 가뭄을 잘 표현한다는 것을 보인 바 있다. 국내에서 Ahn et al. (2009)은 Surface Water Supply Index (SWSI)를 보완한 농업가뭄지수 (Agricultural SWSI, ASWSI)를 이용하여 농업용 저수지의 수문학적 가뭄을 평가하였고, Kwak et al. (2012)은 Copula 이론을 이용하여 수문학적 가뭄의 결합확률분포를 유도하였으며 재현기간을 산정하여 가뭄을 분석하였다. Park et al. (2011)은 기존의 가뭄지수들은 수리시설물들의 가용능력을 고려하지 못하는 한계가 있다는 점을 지적하고 수자원

가용능력을 고려한 수문학적 가뭄지수 (Water Availability Drought Index, WADI)를 개발하였다. 현재 우리나라의 기상청, 소방방재청, 한국농어촌공사 그리고 한국수자원공사에서 구축한 가뭄관련 정보시스템은 공통적으로 강수량의 부족에 기초한 표준강수지수 (SPI)와 파머가뭄지수 (PDSI)를 이용하고 있다.

SPI나 PDSI와 같은 가뭄지수들은 특정 기준값 (threshold)을 기준으로 가뭄의 심도를 표현하고 있다. 그러나 강수량의 변화가 작은 지역에서는 강수량의 작은 차이가 가뭄의 등급을 결정하는데 큰 영향을 미칠 수 있다 (Lloyd-Hughes and Saunders, 2002). 또한 미리 정해진 기준값을 이용하는 해석은 여러 지역의 가뭄 심도를 상대적으로 비교할 수 있는 장점이 있지만, 가뭄이 발생할 위험이 있는 지역을 구분하지는 못한다. 가뭄은 특성상 시작과 끝을 알기가 힘들뿐 아니라 피해지역을 명확하게 파악하기가 어려워 정량적인 평가에 많은 불확실성이 따른다 (Bae et al., 2012; Yoo et al., 2014). 또한 다양한 가뭄지수를 이용하여 가뭄을 판단할 때, 가뭄등급을 결정함에 있어 혼란이 발생할 수 있다 (Mishra and Singh, 2010). 특히, SPI는 강수량 부족기간을 고려하지 않고 가뭄을 판단하기 때문에 수개월에 걸친 가뭄을 효과적으로 표현하지 못할 수도 있다 (Byun, 2009).

은닉 마코프 모형 (Hidden Markov Model, HMM)은 1960년대 후반에 음성인식을 위해 개발되었다. Rabiner (1989)는 Baum-Welch와 Viterbi 알고리즘을 사용하여 성공적인 음성인식 시스템을 구현하였다. HMM을 정보추출에 이용하기 위해 Miller et al. (1999)은 검색하고자 하는 정보와 어느 정도 유사한지 판단할 수 있는 HMM 정보검색 시스템을 개발하였다. Seymore et al. (1999)은 HMM의 은닉상태 간 전이구조를 파악하는 연구를 수행한 바 있다. Deng et al. (2005)은 음성인식 분야에서 기존 방법의 불확실성을 감소시키기 위해 HMM을 이용하였고, 예측된 값과 실제 값과의 차이가 50% 이상 감소했음을 보인 바 있다. 현재는 음성인식 분야뿐만 아니라, 기상학과 생물학 등 여러 분야에 HMM이 응용되고 있으며, 수자원공학 분야에 응용된 사례는 다음과 같다. Thyer and Kuczera (2003)는

수문시계열의 장기 지속성을 모의하기 위해 HMM을 적용하였고, Robertson et al. (2004)은 HMM을 이용하여 강우와 기후인자와의 연관성을 분석하였다. Kwon et al. (2013)은 우리나라의 복잡한 강수패턴을 인지하고 강수량의 다양한 분포특성을 고려할 수 있는 혼합분포를 이용한 동질성 HMM을 제안하고 일 강수량 모의에 적용한 바 있다. 최근에는 가뭄지수의 불확실성을 고려하기 위하여 HMM을 이용한 가뭄해석이 수행된 바 있다(Mallya et al., 2013; Yoo et al., 2014).

본 연구는 강수량 자료를 활용한 Yoo et al. (2014)의 후속연구로, 가뭄지수를 이용하여 미리 정의된 기준값으로 인한 가뭄해석의 불확실성을 줄이고, 유출량 자료에 내재되어 있는 특성을 재현하기 위하여 HMM을 이용한 가뭄해석 방법을 제안하였다. 또한 실제 가뭄 사례와의 비교를 통해 적합성을 검증하였다. 대상 자료로 WAMIS에서 제공하는 월평균유출량 자료를 이용하였으며, HMM을 이용하여 자료에 내재된 특성을 활용하여 가뭄의 은닉상태를 분류하였다. 또한 분류된 가뭄의 은닉상태를 분석하여 상태별 확률밀도함수를 도출하고 가뭄지수와 비교분석을 수행하였다.

## 2. 연구방법 및 연구대상

### 2.1 은닉 마코프 모형(HMM)

HMM은 관찰되지 않는 은닉상태(hidden state)를 가지며, 각 상태들은 직접적인 관찰이 가능하지 않지만, 일정 확률을 가지기 때문에 관찰이 가능한 결과를 보여준다. 즉, 자료의 은닉상태  $X_t$ 는  $K$ 개의 무작위 변수( $X_t = \{x_1, x_2, \dots, x_k\}$ )로 구성되고, 어떤 시간  $t$ 에서 관측된 결과  $O_t$ 는  $t$ 에서의 은닉상태  $X_t$ 에 영향을 받는다. HMM은 Fig. 1과 같이 간략하게 나타낼 수 있으며, 본 연구에서는 다음과 같은 기본가정을 바탕으로 하였다. 가뭄의 상태  $X_t$ 와 관측값  $O_t$ 는 다른 시간의 상태와 관측결과와 조건부 독립이고, 가뭄의 상태  $X_t$ 는 바로 이전의 상태인  $X_{t-1}$ 의 영향을 받으며, 나머지 상태는 서로 독립인 1차 마코프 모형을 따

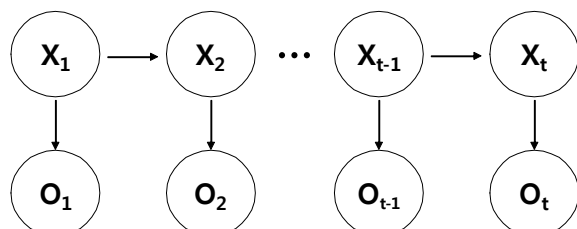


Fig. 1. Graphical Representation of HMM

른다. 이를 수식으로 표현하면 Eq. (1)과 같다. 방출확률 밀도함수(emission probability distribution)는 어떤 상태  $X_t$ 에서  $O_t$ 를 출력할 확률이며 Eq. (2a)와 같다. 이는 각 상태에서 특정한 값이 나올 확률을 모의하는 출력 확률로 구성되고, 이를 이용하여 각 은닉상태의 확률밀도함수를 알 수 있다. 또한 본 연구의 자료가 평균값을 기준으로 정규화된 자료이기 때문에 방출확률밀도함수는 정규분포를 따르는 형태로 가정하였으며, Eq. (2b)와 같다.

$$P(X_t|X_1, X_2, \dots, X_{t-1}) = P(X_t|X_{t-1}) \quad (1)$$

$$P(O_t|O_{1:t-1}, X_{1:t}) = P(O_t|X_t) \quad (2a)$$

$$P(O_t|X_t = k) = N(O_t|\mu_k, \sigma_k^2) \quad (2b)$$

HMM은  $\lambda = \{\Pi, \Gamma, Y\}$ 로 표현할 수 있으며, 여기서  $\Pi$ 는 초기확률(initial state probability),  $\Gamma$ 는 천이확률(transition probability) 그리고  $Y$ 는 관측확률(observation probability)을 나타내면, Eq. (3)과 같이 표현할 수 있다.

$$\Pi = \pi_i = P(x_i|i), \quad 1 \leq i \leq K \quad (3a)$$

$$\Gamma = \gamma_{ij} = P(x_{t+1} = j|x_t = i), \quad 1 \leq i, j \leq K \quad (3b)$$

$$Y = b_j(h) = P(o_t = v_h|x_t = j), \quad 1 \leq h \leq M, 1 \leq j \leq K \quad (3c)$$

여기서, 관측값  $O_t$ 는 관찰이 가능한 기호들로 표현할 수 있으며( $O_t = \{v_1, v_2, \dots, v_h\}$ ),  $M$ 은 관찰기호의 개수 그리고  $K$ 는 은닉상태의 수를 나타낸다.

$\Pi$ 는 처음에  $X_i$ 라는 은닉상태가 선택될 확률이며, Eq. (3a)와 같고,  $\Gamma$ 는 모형에서 은닉상태간의 천이확률을 나타낸 것으로, 이전의 은닉상태에서 현재의 은닉상태로 천이될 확률을 행렬로 나타낸 것이며, Eq. (3b)와 같다. 또한  $Y$ 는 은닉상태  $X_j$ 에서 관측값  $O_t$ 가 나올 확률이며, Eq. (3c)와 같다.

HMM을 실제 모형에 적용할 때 문제가 되는 것으로 알려진 모형의 변수 산정은 Baum-Welch 알고리즘으로 해결이 가능하다(Bilmes, 1997; Moon, 2002; Park, 2007). 따라서 본 연구에서는 Baum-Welch 알고리즘을 이용하여  $P(O_t|X_t)$ 의 우도함수를 최대화 시키는 변수를 모형의 최적변수를 가정하였다. Baum-Welch 알고리즘은 은닉 마코프 모형에 특화된 Expectation-Maximization (EM)으로 Eq. (4)와 같이 간략하게 나타낼 수 있다.

$$P(O_t|X_t' = \lambda') \geq P(O_t|X_t = \lambda) \quad (4)$$

Baum-Welch 알고리즘은 초기 모형을 구성하고, 초기 모형과 관찰된 자료 기반의 새로운 모형을 구성하는 방법이다. 알고리즘 계산을 위한 변수는 전향변수( $\alpha_t$ , Eq. (5a)), 후향변수( $\beta_t$ , Eq. (5b)), 시간  $t$ 에서 상태  $i$ 에 있고, 시간  $t+1$ 에서 상태  $j$ 에 있을 확률  $\xi_t$ (Eq. (5c))와 시간  $t$ 에서 주어진 관찰열을 가지고 상태  $i$ 에 있을 확률  $\gamma_t$ (Eq. (5d))와 같다. Expectation 단계(Eqs. (5c)~(5d))로  $\xi_t$ 와  $\gamma_t$ 를 추정하고, Maximization 단계에서  $\xi_t$ 와  $\gamma_t$ 로 새로운  $\lambda' = \{\Pi', \Gamma', Y'\}$ 를 산정하며 새로운  $\lambda'$ 에 의하여 Eq. (4)가 성립하고, 변화가 없이 수렴하는 부분에서 최적의 변수를 찾는다.

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, x_t = i | \lambda) \quad (5a)$$

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | x_t = i, \lambda) \quad (5b)$$

$$\xi(i, j) = \frac{\alpha_t(i) \gamma_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) \gamma_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \quad (5c)$$

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (5d)$$

## 2.2 연구 대상지역 및 자료

본 연구의 대상지역은 한강유역 내의 평창강, 주천강 그리고 오대천 유역으로 Fig. 2에 도시하였다. 이들 유역은 한강유역의 최상류에 해당하여 가뭄의 영향을 독립적으로 분석하기 용이한 지역이다(Kwak et al., 2012). 대상 자료는 WAMIS에서 제공하는 1966년부터 2009년까지 기

록된 월평균 유출량 자료(Fig. 3)이며, 월별 특성을 고려하기 위해 3, 6, 12개월 누적 월 유출량 시계열을 구축하였다. 또한 유출량의 과·부족량을 구분하기 위하여 자료를 각 월의 평균과의 차이로 시계열을 재구축하고, 표준유출지수(Standardized Streamflow Index, SSD)와 은닉 마코프 가뭄지수(Hidden Markov Drought Index, HMDI)를 산정하였다.

전체적인 연구절차는 다음과 같다. (1) 누가 유출량 시계열 자료 구축(3, 6, 12개월), (2) 과부족량 시계열 구축, (3) 두 가지 가뭄지수 즉 SSI와 HMDI 산정, (4) 두 가뭄지수의 결과 비교 및 실제 가뭄사례와 비교분석 수행.

## 2.3 가뭄지수 산정

본 연구에서는 한강유역의 수문학적 가뭄의 특성을 파악하기 위하여 유출량 자료를 활용하였으며, 결과의 비교를 위한 가뭄지수로 표준유출지수(SSI)를 선택하였다. SSI는 SPI의 개념과 유사하며, 유출량의 부족에 기초한 지수이다. SSI 산정을 위하여 누가된 유출량 시계열 자료를 Gamma 분포에 적합시킨 후, 누적분포함수를 산정하고, 이를 다시 정규분포의 확률변량으로 역변환하여 SSI를 산정하였다. SSI의 가뭄판단을 위한 기준값을 정하기 위해, 대표적인 표준지수(Standardize Index) 중 하나인 SPI의 가뭄판단 기준을 SSI의 가뭄판단 기준으로 이용하였으며, 이는 Table 1과 같다. HMM을 이용하여 가뭄의 시작과 끝을 구분하기 위해 각 월의 평균을 임계수준(threshold)으로 정하고, 구분된 과·부족량을 이용하여 임계수준 이하의 유량을 가뭄으로 정의하였다. 즉, 유출량이 임

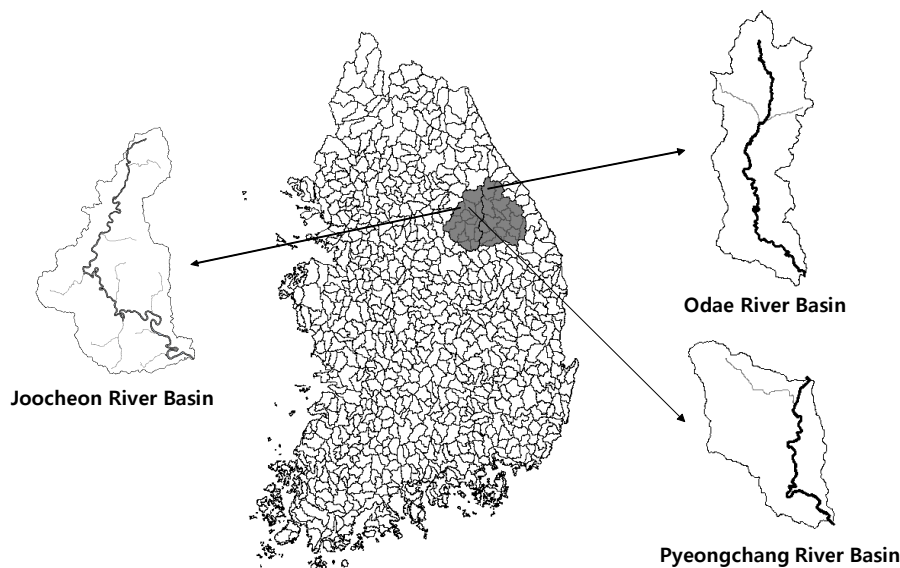


Fig. 2. Study Watersheds in this Study

계수준보다 낮을 때가 가뭄의 시작이고, 임계수준 이상일 때를 가뭄의 끝으로 정의하였다. 또한 HMM을 이용하여 산정된 은닉상태의 확률을 은닉 마코프 가뭄지수(HMDI)라고 명명하였고, 가뭄판단 기준은 SSI와 동일하게 Table 1을 적용하였다.

### 3. 연구결과 및 고찰

#### 3.1 은닉상태의 개수(K) 결정

HMM을 적용하기 위해서는 우선 은닉상태의 수를 결정해야 한다. 기존 연구에서는 은닉상태의 수를 결정하기 위해,

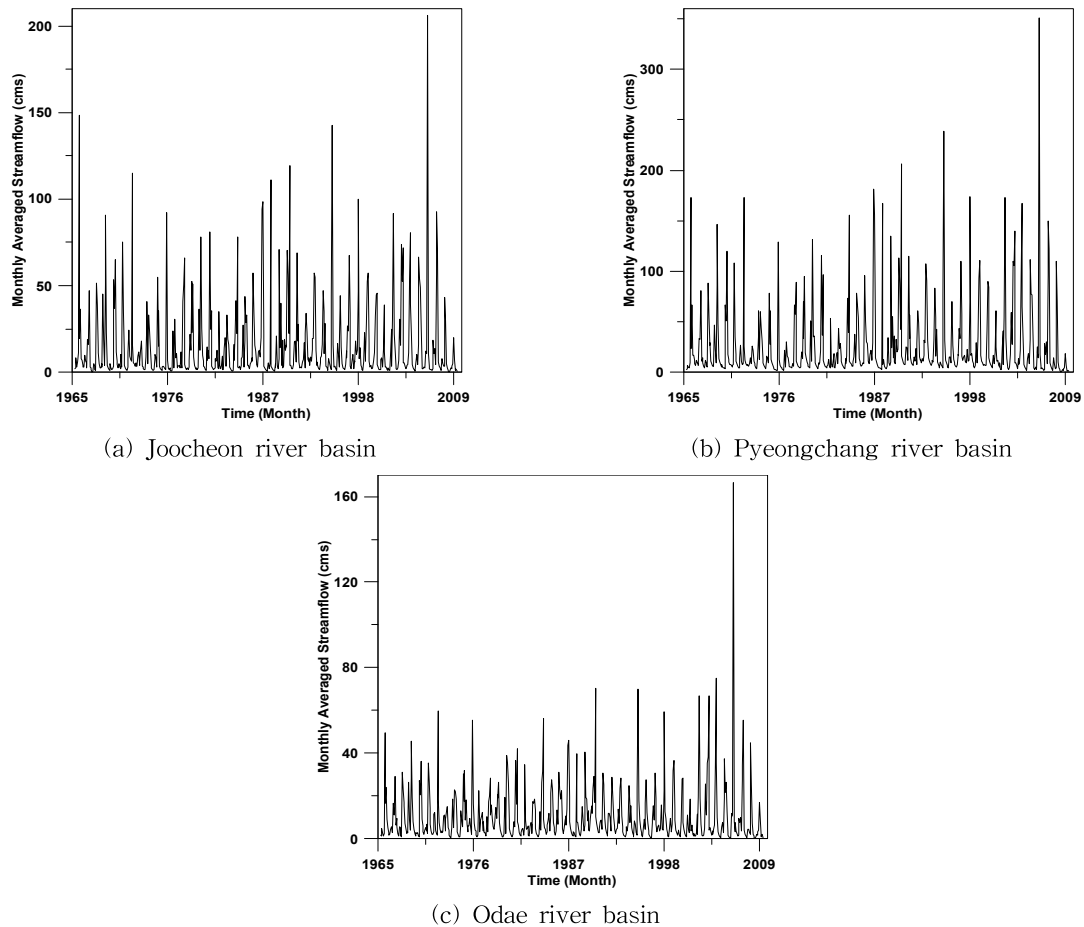


Fig. 3. Time Series of Monthly Averaged Streamflow

Table 1. Drought Classification Based on SPI-value (also SSI-value)

State	Range	Condition
D1	$-2.0 > Z$	Extreme dry
D2	$-2.0 \sim -1.5$	Severely dry
D3	$-1.5 \sim -1.0$	Moderately dry
D4	$-1.0 \sim 1.0$	Near normal
D5	$1.0 \sim 1.5$	Moderately wet
D6	$1.5 \sim 2.0$	Very wet
D7	$2.0 < Z$	Extremely wet

Akaike Information Criterion (AIC) (Letinic et al., 2010) 또는 Bayesian Information Criterion (BIC) (Robertson et al., 2004; Suk and Chung, 2008)를 적용하였다. Matsuyama (2011)은 은닉상태의 수가 증가할 경우 모형이 과적합되는 문제가 발생할 수 있어, 대수우도함수(log-likelihood)가 수렴하는 부분에서 은닉상태의 수를 결정하는 것이 효율적이라고 주장한 바 있다. Kwon et al. (2013)은 대수우도함수가 수렴하는 부분을 최적의 은닉 상태수로 결정하였다.

본 연구에서는 은닉상태의 수를 결정하기 위해 모형의 적합성 정도를 평가하는데 사용되는 대수우도함수와 AIC, BIC를 사용하여 여러 은닉상태의 수를 대상으로 평가하였다. 은닉상태의 수(K)를 2에서 12까지 늘려가며 우도함수를 산정하였으며, 대상 유역의 은닉상태의 수를 결정하기 위한 대수우도함수는 Fig. 4와 같고, AIC와 BIC 값은 Fig. 5와 같다. 모형이 과적합(overfitting)되지 않도록 하기 위해 우도함수가 수렴하는 부분에서 AIC와 BIC가 최

솟값을 갖는 상태의 수를 최적의 은닉상태의 수를 정하였다. 우도함수는 대체로 은닉상태가 7~11개일 때 증가폭이 적었고, AIC와 BIC 두 값이 가장 작은 값을 갖는 상태 수 K는 8개로 결정되었다.

### 3.2 은닉마코프 모형을 이용한 가뭄 분석

3.1절에서 결정된 은닉상태의 수를 이용하여 HMM에 적용한 결과, 대상유역의 은닉상태별 평균 유출량의 과부족은 Fig. 6과 같다. 각 상태에 따라 평년대비 유출량이 부족한 정도를 알 수 있었고, 유역에 평균유출량이 부족할 때를 가뭄으로 판단하였다. Fig. 6에서 첫 번째 상태는 유출량이 가장 부족한 상태, 즉 유출량을 기준으로 하였을 때 가뭄에 가장 큰 영향을 미치는 상태이며 상태의 번호가 증가할수록 낮은 상태에 비해 가뭄에 영향을 덜 미치는 상태라고 할 수 있다.

Fig. 7은 대상유역의 은닉상태 별 발생빈도를 나타낸

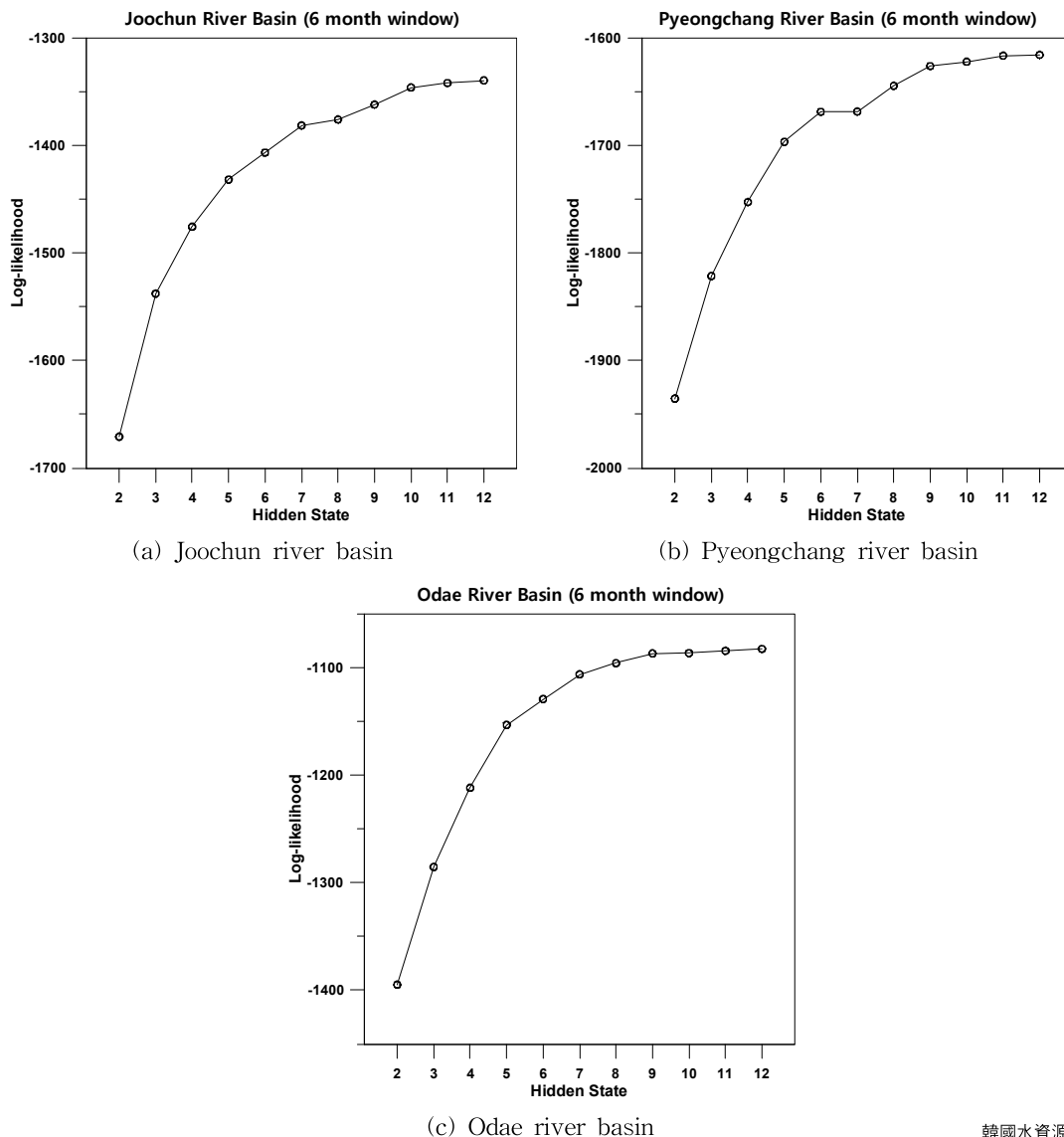


Fig. 4. Log-likelihood Functions for Choosing the Number of States

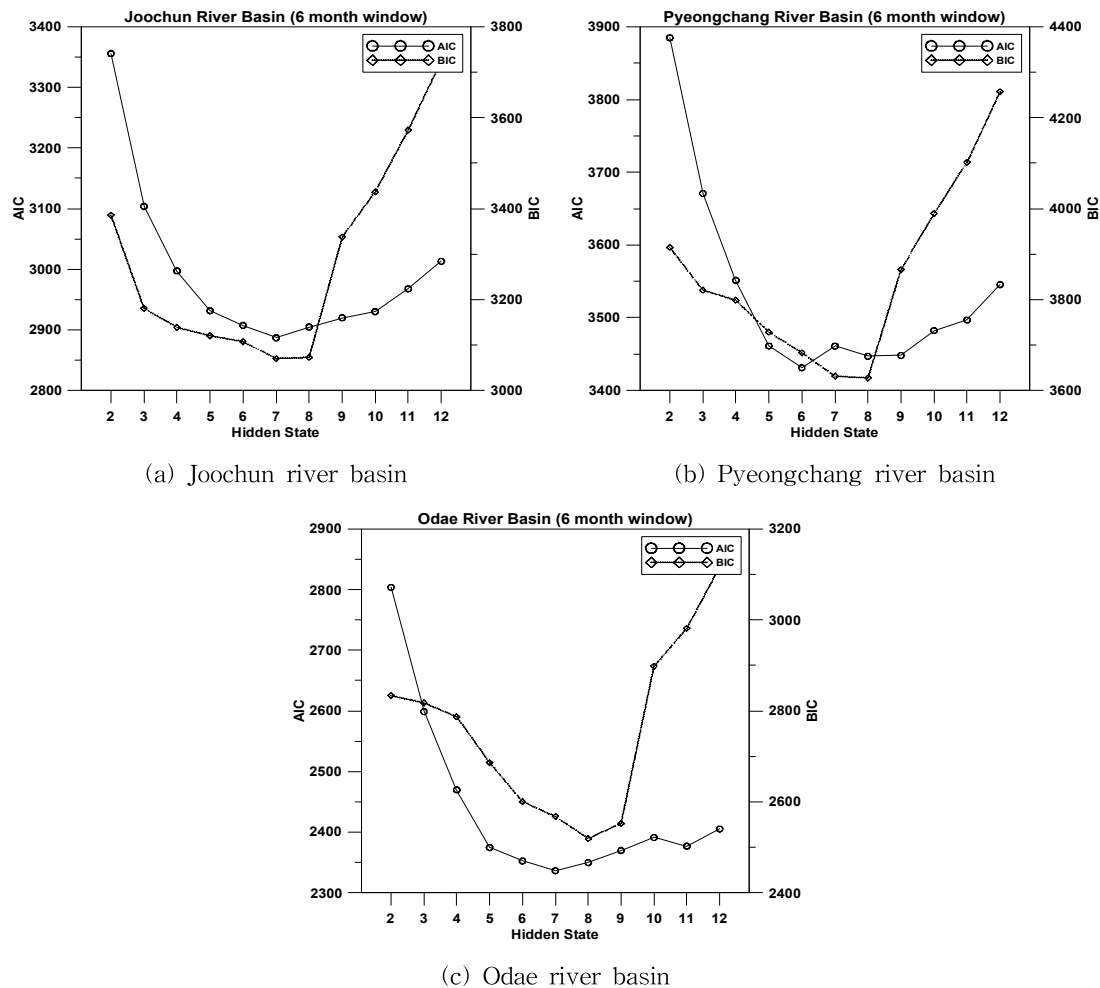


Fig. 5. The Values of AIC and BIC for Choosing the Number of States

것이다. Fig. 7에서 주천강의 경우 은닉상태 1~3은 상대적으로 유출량이 부족한 상태를 나타내며 약 48%의 비율로 추정되었고, 은닉상태 6~8은 상대적으로 유출량이 과다한 상태를 나타내며 약 37%의 비율을 가지는 것으로 추정되었다. 또한 평창강은 유출량이 부족한 상태가 약 33%를 나타내었고, 유출량이 과다한 상태가 약 35%의 비율을 갖는 것으로 추정되었다. 오대천은 유출량이 부족한 상태가 약 23%, 과다한 상태가 49%의 비율을 가지는 것으로 추정되었다.

HMM 이용하여 산정된 각 은닉상태의 평균과 표준편차를 통하여 방출확률밀도함수를 추정하였다. Fig. 8은 주천강, 평창강, 오대천 유역의 방출확률밀도함수이며, 왼쪽 y-축은 각 상태의 확률이며, 오른쪽 y-축은 유출량자료 전체의 확률이다. HMDL1은 은닉상태 중 유출량이 가장 부족한 첫 번째 상태를 나타내며, HMDL4는 평년대비 부족 유출량이 정상상태에 가장 가까운 네 번째 상태를

나타낸다. 방출확률밀도함수를 통하여 가뭄에 영향을 미치는 은닉상태의 발생확률을 알 수 있었으며, 모든 유역에서 정규분포에 가까운 특성을 보이고 있음을 확인할 수 있다.

### 3.3 HMDI와 SSI에 의한 결과 비교

특정 기준값을 이용하는 SSI와 달리, HMDI은 각 시점의 은닉상태별 사후확률을 추정할 수 있고, 이를 통하여 가뭄의 상태를 정의할 수 있다. Fig. 9~11은 주천강, 평창강, 오대천 유역의 은닉상태별 사후확률을 나타낸 것이고, 상단의 그래프는 각 시점에서 SSI에 의하여 추정된 가뭄의 상태이며, 그 아래는 HMM에 의하여 추정된 가뭄의 상태(HMDI)이다. 또한 왼쪽 y-축은 HMDI의 발생확률이며, 오른쪽 y-축은 유출량의 과부족량을 나타낸다. Figs. 9~11에 나타난 바와 같이, SSI로 추정된 결과는 각 시점에서 하나의 가뭄 상태만을 보여주고 있는 반면 HMDI로

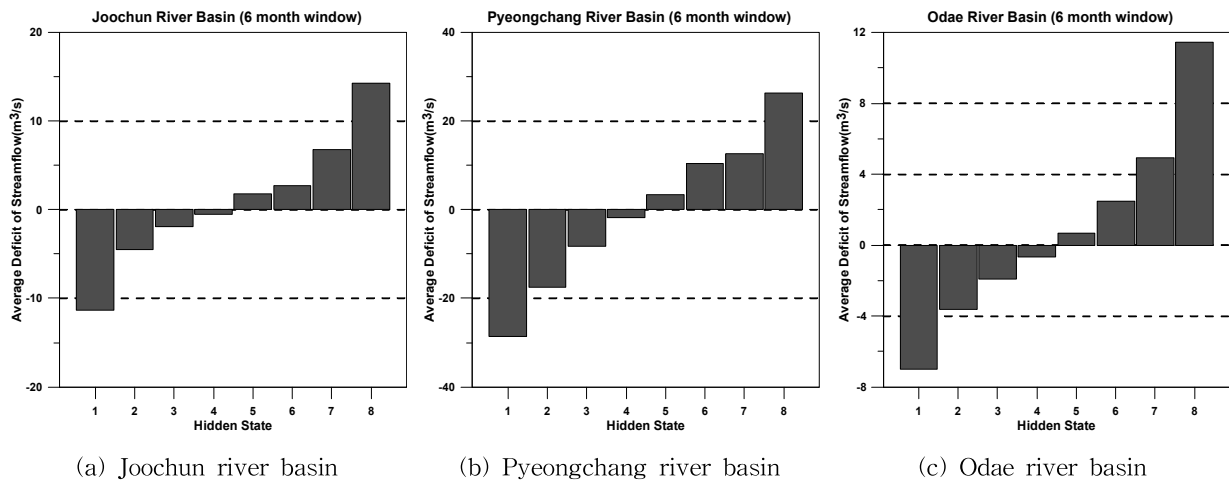


Fig. 6. Average Deficit of Streamflow Corresponding to Hidden State

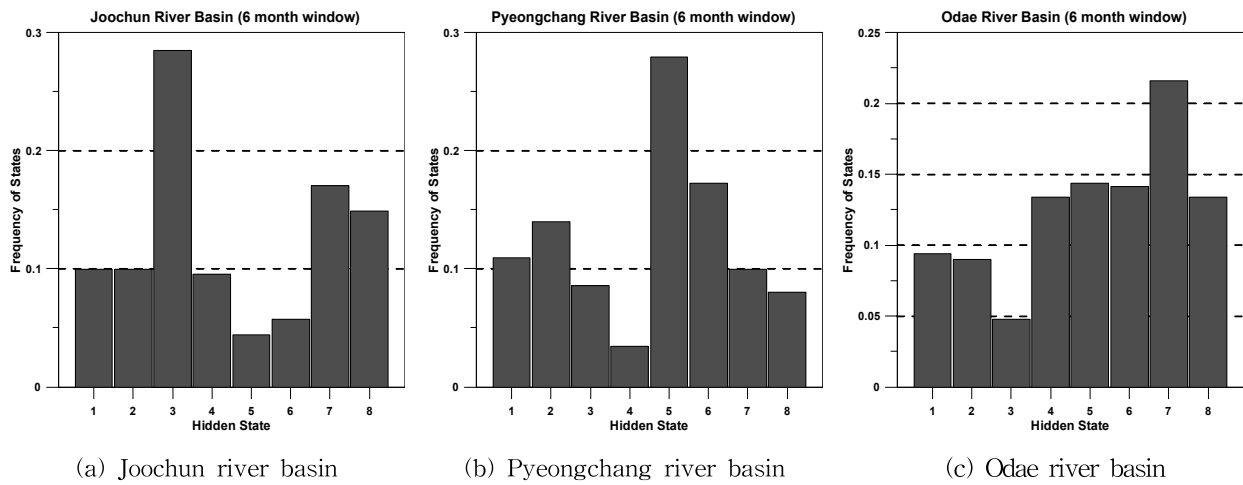


Fig. 7. Frequency of Hidden States

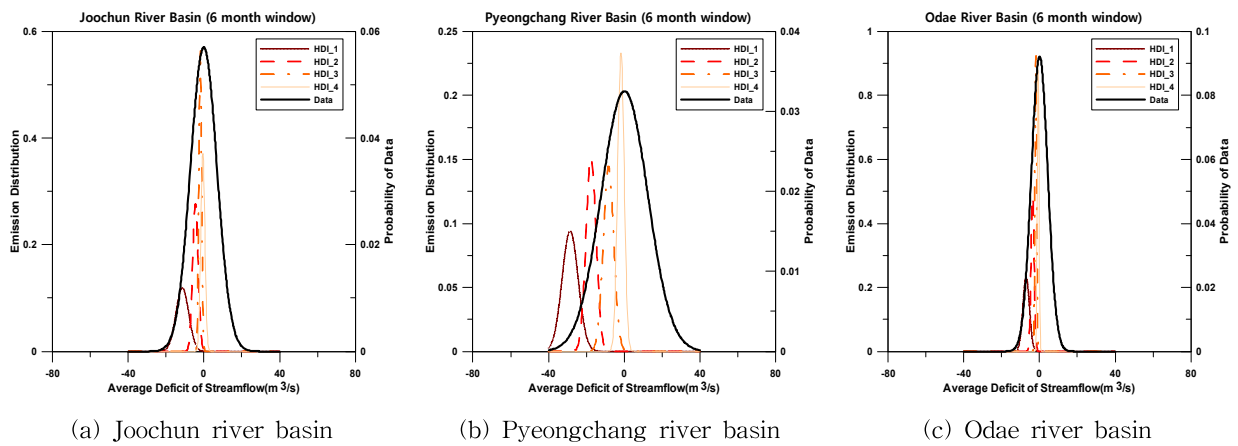


Fig. 8. Emission Probability Distribution for Hidden States



추정된 결과는 각 시점에서 여러 상태의 가뭄을 확률의 형태로 보여주고 있다.

본 연구방법의 적합성을 검토하기 위하여 대상유역에 실제로 큰 가뭄이 발생했던 1967~1968년, 1973년, 1976~

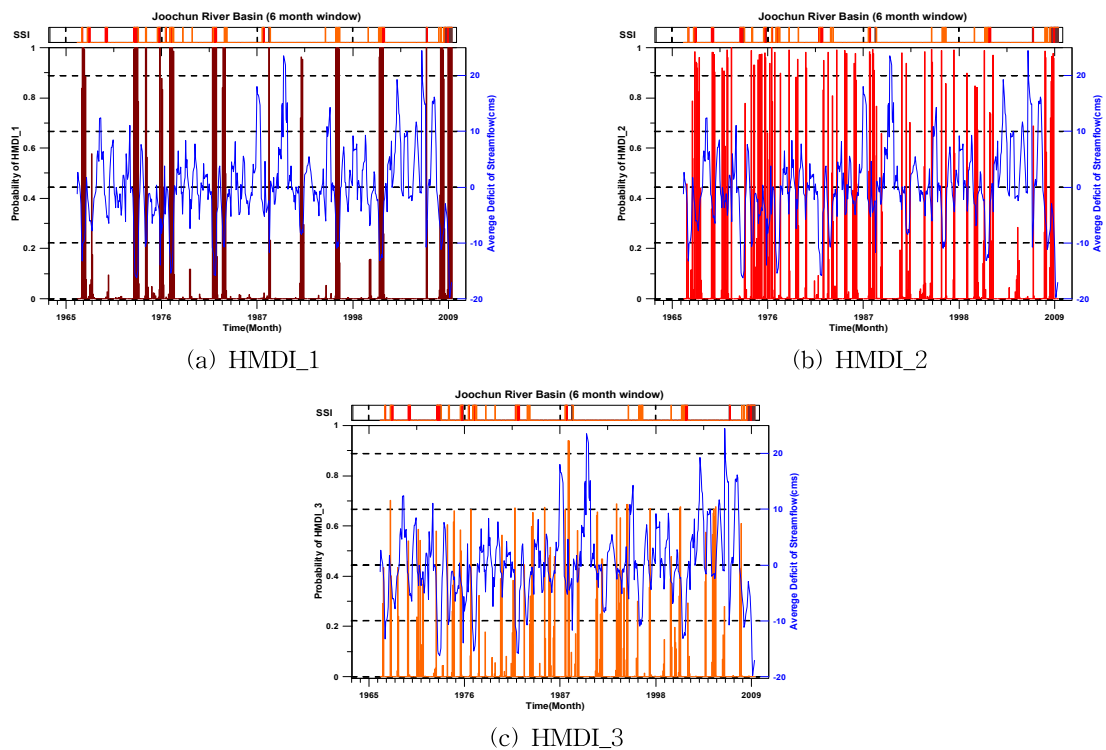


Fig. 9. Posterior Probability of HMDI Compared with SSI for Joochun River Basin

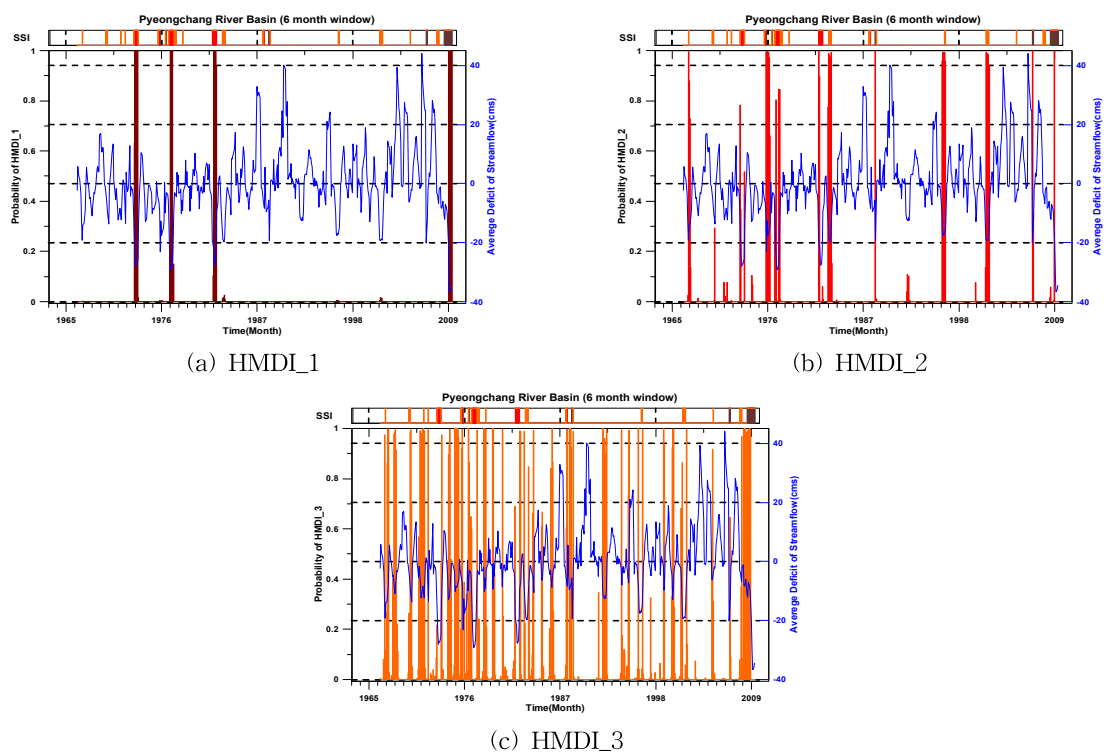


Fig. 10. Posterior Probability of HMDI Compared with SSI for Pyeongchang River Basin

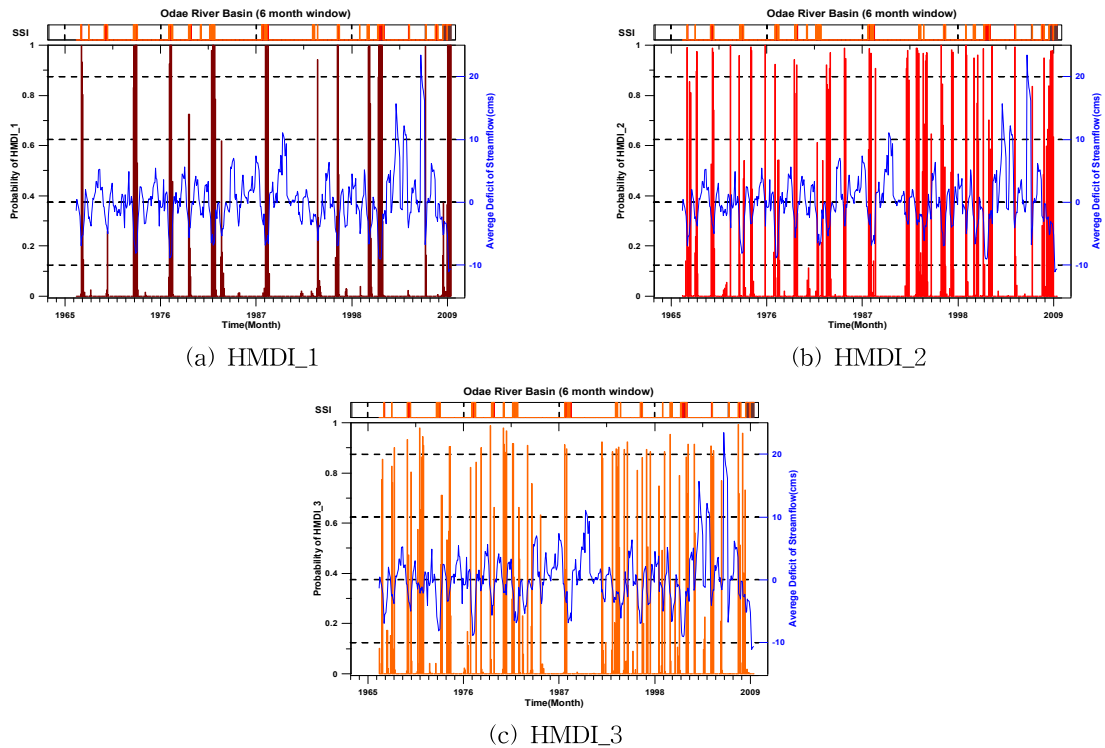


Fig. 11. Posterior Probability of HMDI Compared with SSI for Odae River Basin

1978년, 1981~1982년, 2001년, 2008~2009년을 중심으로 가뭄분석결과를 비교하였다. 실제 가뭄사례를 비교한 결과, 세 유역 모두에서 HMDI가 SSI에 비해 사후확률이 상대적으로 크게 나타나는 것으로 확인되었다. 1967~1968년의 경우, HMDI\_2는 약 23%, HMDI\_3은 약 31%의 확률을 나타낸 반면 SSI 지수는 약한 가뭄인 상태가 10%로 나타났다. HMM에 의하여 추정된 가뭄지수 HMDI는 실제 큰 가뭄이 있었던 시기에서 대부분 가장 심한 가뭄상태인 HMDI\_1부터 약한 가뭄상태인 HMDI\_3까지의 발생 확률을 모두 나타내었다. 하지만 HMDI와 달리 SSI의 경우에는 가장 극심한 가뭄상태인 SSI\_1부터 약한 가뭄상태인 SSI\_3까지 모두 나타나지 않고 하나의 상태 혹은 두 개의 상태를 나타내는 경우가 많았다. 2008~2009년에서 SSI에 의한 상태의 확률은 연속적이지 않은 것을 볼 수 있는데, 이는 연속적인 가뭄의 상황에서 가뭄의 상태가 극심한 상태와 약한 상태만 추정이 되었다는 것이다. 또한 1973년, 1981~1982년에서 가뭄의 상태는 연속적이지 않고, 하나의 가뭄상태(SSI\_2)만 보이는 결과를 나타내었다. 2001년, 2008~2009년은 비교적 큰 가뭄이 발생했을 때이며, SSI와 달리 HMDI의 경우, 가뭄의 상태는 HMDI\_1, HMDI\_2, HMDI\_3이 고르게 나와 있다. 이 결과는 연속적인 가뭄의 특성을 고려하였을 때, 심한 가뭄

이었던 시기에 HMDI가 SSI에 비해 더 가뭄을 잘 표현하였다고 판단된다. 즉, 과거 발생했던 실제 가뭄의 표현과 가뭄의 지속성을 고려하였을 때 SSI를 이용한 방법보다 HMDI를 이용한 방법이 실제 가뭄에 대한 재현능력이 우수하였다.

#### 4. 결 론

본 연구에서는 특정 기준값을 이용하여 가뭄을 해석하는 방법이 아닌, 자료에 내재되어 있는 특성을 활용할 수 있는 은닉 마코프 모형(HMM)을 이용하여 수문학적 가뭄의 상태를 확률론적으로 해석하는 방법을 제안하였다.

은닉상태의 개수를 결정할 때, 은닉상태의 개수와 정교한 정해진 값이 아닌 대수우도함수와 AIC 및 BIC를 통하여 가장 적합한 상태라고 판단되는 값으로 정하였다. HMM을 이용하여 추정된 은닉상태의 평균, 분산과 같은 특성을 알 수 있었으며, 이를 이용하여 가뭄의 상태를 분류하였다. 또한 사후확률을 통하여 모든 시점의 은닉상태 별 확률을 추정하여 불확실성을 고려한 해석을 수행하였다. 또한 특정 기준값에 근거한 가뭄해석 방법 중 하나이며 유출량 자료를 이용하는 표준유출지수(SSI)를 산정하여 실제 가뭄이 발생한 시기와의 비교를 통해 본 연구의 결과를 검토하였

다. 그 결과, SSI는 미리 정해진 기준값을 이용하여 가뭄의 판단 기준을 분류하고, 각 시점에서의 가뭄상태만 알 수 있는 반면 은닉 마코프 가뭄지수(HMDI)는 미리 정의된 기준값을 사용하지 않고, 유출량 자료에 내재된 특성을 이용하여 가뭄의 상태를 확률의 형태로 재현할 수 있었다. 실제 대상유역에 큰 가뭄이 있었다고 알려진 1967~1968년, 1973년, 1976~1978년, 1981~1982년, 2001년, 2008~2009년 등의 가뭄년에 대해서 두 방법의 직접적인 비교를 통하여 HMDI가 SSI에 비해 가뭄의 심도와 수문학적 지속성을 더 잘 표현하였다. 즉, 과거의 가뭄년을 대상으로 한 검증에서 HMDI는 기존의 가뭄지수에 의한 방법보다 가뭄을 더 효과적으로 재현하였고, 사후확률을 통한 불확실성을 고려한 가뭄해석이 가능하였다.

본 연구에서 제안한 HMM을 이용한 가뭄해석방법은 기존의 가뭄지수를 이용한 해석방법과 비교하여 은닉상태의 사후확률 추정 가능성이 높기 때문에, 여러 관점에서 가뭄을 분석할 수 있다. 또한 가뭄을 확률론적으로 해석함으로써 가뭄의 판단 기준을 제시하는데 도움이 될 것으로 판단된다. 본 연구에서 제안한 방법은 가뭄 예측 혹은 미래의 물 부족에 대한 확률을 예측하는데 적용할 수 있으며 효율적인 수자원 관리에 많은 도움이 될 것이다.

## 감사의 글

이 연구는 2013년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원(한국연구재단-2013R1A1A2013160)을 받아 수행되었으며, 이에 감사드립니다.

## References

- Ahn, S.R., Park, J.Y., Jung, I.K., Na, S.J., and Kim, S.J. (2009). "Hydrological drought assessment of agricultural reservoirs based on SWSI in Geum River Basin." *Journal of the Korean Society of Agricultural Engineers*, Vol. 51, No. 5, pp. 35-49 (in Korean).
- American Meteorological Society (2004). Annual Report 2004, American Meteorological Society, Boston.
- Bae, D.-H., Son, K.-H., Hong, J.-Y., Kim, G.-S., Chung, J.S., Jung, U.-S., and Kim, J.-J. (2012). "Development of real-time drought monitoring and prediction system on Korea & East Asia region." *Atmosphere. Korean Meteorological Society*, Vol. 22, No. 2, pp. 267-277 (in Korean).
- Bilmes, J. (1997). *A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*, Technical Report ICSI-TR-97-021, ICSI.
- Byun, H.-R. (2009). "Comparative analysis of the drought diagnosis and related systems." *Journal of KOSHAM*, Vol. 9, No. 2, pp. 7-18 (in Korean).
- Deng, L., Wu, J., Droppo, J., and Acero, A. (2005). "Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion." *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 3, pp. 412-421.
- Kwak, J.W., Kim, D.G., Lee, J.S., and Kim, H.S. (2012). "Hydrological drought analysis using copula theory." *Journal of Korean Society of Civil Engineer*, Vol. 32, No. 3B, pp. 161-168 (in Korean).
- 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 Engineers*, Vol. 33, No. 5, pp. 1861-1870 (in Korean).
- Letinic, K., Sebastian, R., Barthel, A., and Toomre, D. (2010). "Deciphering subcellular processes in live imaging datasets via dynamic probabilistic networks." *Bioinformatics*, Vol. 26, No. 16, pp. 2029-2036.
- Lloyd-Hughes, B., and Saunders, M.A. (2002). "A drought climatology for Europe." *Int. J. Climatol.*, Vol. 22, No. 13, pp. 1571-1592.
- Mallya, G., Tripathi, S., Kirshner, S., and Govindaraju, R. (2013). "Probabilistic assessment of drought characteristics using hidden markov model." *Journal of Hydrologic Engineering*, Vol. 18, No. 7, pp. 834-845.
- Matsuyama, Y. (2011). "Hidden markov model estimation 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.
- McKee, T.B., Doesken, N.J., and Kieist, J. (1993). "The relationship of drought frequency and duration to time scales." *8th Conference on Applied Climatology*, 17-22 January 1993, Anaheim, California, pp. 179-184.
- Miller, D.R.H., Leek, T., Schwartz, R.M. (1999). *A hidden*

- markov model information retrieval system*. Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 214-221.
- Mishra, A.K., and Singh, V.P. (2010). "A review of drought concepts." *J. Hydrol.*, Vol. 391, No. 1-2, pp. 202-216.
- Modarres, R. (2007). "Streamflow drought time series forecasting." *Stoch Environ Res Risk Assess*, 21, pp. 223-233.
- Moon, H.K. (2002). *Emotional Classification from Text Data by Hybrid Naive Bayes HMM*, Master's Thesis, Seoul National University (in Korean).
- Palmer, W.C. (1965). *Meteorological Drought*, Research Paper No. 45, U.S. Weather Bureau, Washington, D.C.
- Park, H.-J. (2007). *Korean Stock Forecasting Using Hidden Markov Model*, Master's Thesis, Sungkyunkwan University (in Korean).
- Park, M.J., Shin, H.J., Choi, Y.D., Park, J.Y., and Kim, S.J. (2011). "Development of a hydrological drought index considering water availability." *Journal of the Korean Society of Agricultural Engineers*, Vol. 53, No. 6, pp. 165-170 (in Korean).
- Rabiner, L.R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-286.
- Robertson, A.W., Kirshner, S., and Smyth, P. (2004). "Downscaling of daily rainfall occurrence over northeast Brazil using a hidden markov model." *J. Climate*, Vol. 17, No. 22, pp. 4407-4424.
- Seymore, K., McCallum, A., and Rosenfeld, R. (1999). "Learning hidden markov model structure for information extraction." *In AAAI 99 Workshop on Machine Learning for Information Extraction*, pp. 37-42.
- Shafer, B.A., and Dezman, L.E. (1982). "Development of a surface water supply index (SWSI) to assess the severity of drought conditions in snowpack runoff areas." *Proceeding of the Western Snow Conference*, pp. 164-175.
- Shukla, S., and Wood, A.W. (2008). "Use of a standardized runoff index for characterizing hydrologic drought." *Geophysical Research Letters*, Vol. 35, L02405, doi:10.1029/2007GL032487.
- Suk, S.-Y., and Chung H.-Y. (2008). "A speech and character combined recognition engine for mobile devices." *International Journal of Pervasive Computing and Communications*, Vol. 4, No. 2, pp. 232-249.
- Thyer, M., and Kuczera, G. (2003). "A hidden markov model for modelling long-term persistence in multi-site rainfall time series 1. Model calibration using a Bayesian approach." *Journal of Hydrology*, Vol. 275, No. 1-2, pp. 12-26.
- Yoo, J.-Y., Kwon, H.-H., Kim, T.-W., and Lee, S.-O. (2014). "Probabilistic assessment of drought characteristics based on homogeneous hidden markov model." *Journal of the Korean Society of Civil Engineers*, Vol. 34, No. 1, In print (in Korean).

논문번호: 14-015	접수: 2014.02.08
수정일자: 2014.04.08	심사완료: 2014.04.08