

Dempster-Shafer 이론을 이용한 강우빈도분석 및 불확실성의 정량화

Rainfall Frequency Analysis and Uncertainty Quantification

Using Dempster-Shafer Theory

서영민* · 지홍기** · 이순탁***

Seo, Young Min · Jee, Hong Kee · Soontak Lee

.....

요 지

Dempster-Shafer 이론은 미지의 매개변수 추정시 베이저안 기법의 제약을 완화시키기 위한 베이저안 접근법의 일반화로 해석될 수 있으며, 상호배타적인 싱글톤에만 확률이 할당되는 것이 아니라 가능한 결과의 부분집합들이 기본확률할당을 위한 대상으로 고려된다. 베이저안 접근은 우연적 불확실성 및 지식의 불확실성을 효율적으로 구분할 수 없으며, 특정도가 낮고 애매한 증거들을 다룰 수 없는 반면, Dempster-Shafer 증거추론은 이러한 문제들을 효율적으로 평가할 수 있다.

따라서 본 논문에서는 홍수위험평가 및 수자원 계획 수립시 가장 기본이 되는 강우빈도해석에서 확률분포의 매개변수에 대한 불확실성 고려한 확률강우량의 산정 및 불확실성의 영향을 평가하기 위하여 Dempster-Shafer 이론을 이용하여 불확실성을 고려한 강우빈도해석모델 구축 및 적용을 통해 홍수위험평가 및 수자원 계획 등에 있어서 불확실성 표현 및 처리기법을 제시하였다.

핵심용어 : Dempster-Shafer Theory, Evidence, Uncertainty, Likelihood, Inference

1. 서 론

홍수위험관리를 위한 홍수위험평가 및 의사결정과정에서 위험도와 불확실성의 규명에 대한 중요성이 증대되고 있다. 불확실성은 크게 우연적 불확실성(aleatory uncertainty)과 지식의 불확실성(epistemic uncertainty)으로 구분할 수 있다. 우연적 불확실성을 처리하기 위한 전통적 접근법으로는 과거자료에 근거한 확률론적 분석, 즉 빈도주의적 접근법(frequentist approach)이 주로 사용되고 있으며, 지식의 불확실성을 처리하기 위한 기존의 접근법으로는 베이저안 접근법(Bayesian approach)이 널리 사용되고 있다. 베이저안 접근법은 가능한 결과를 서로소인 증거체(body of evidence)로서 처리하여 싱글톤(singleton)에만 확률이 할당될 수 있으며, 추론을 위해 불충분이유원칙(principle of insufficient reason)이라는 다소 강한 가정을 필요로 하고 이에 따라 무지(ignorance)가 싱글톤들에 균등하게 분포되어야만 한다(Sadiq, Kleiner, and Rajani, 2006).

Dempster-Shafer 이론(Dempster-Shafer Theory, DST)은 미지의 매개변수 추정시 베이저안 기법의 제약을 완화시키기 위한 목적으로 Dempster(1966)에 의해 개발되었으며, 이후 Shafer(1976)에 의해 확장되어 현재의 Dempster-Shafer 이론에 이르고 있다. 베이저안 접근은 우연적 불확실성 및

* 정회원 · 영남대학교 대학원 · 박사수료 · E-mail : elofy@nate.com
** 정회원 · 영남대학교 건설시스템공학부 · 교수 · E-mail : hkjee@yu.ac.kr
*** 참여회원 · 영남대학교 석좌교수 · 공학 · 이학박사 · E-mail : leest@yu.ac.kr

지식의 불확실성을 효율적으로 구분할 수 없으며, 특정도가 낮고 애매한 증거들을 다룰 수 없는 반면, Dempster-Shafer 증거추론(evidence inference)은 베이저안 기법의 일반화로서 이러한 문제들을 효율적으로 평가할 수 있다(Sadiq, Kleiner, and Rajani, 2006).

따라서 본 논문에서는 Dempster-Shafer 이론을 이용하여 불확실성을 고려한 경우빈도해석모델 구축 및 적용을 통해 홍수위험평가 및 수자원 계획 등에 있어서 불확실성 표현 및 처리기법을 제시하였다.

2. Dempster-Shafer 이론

Θ 를 참값이 알려지지 않은 확률변수라고 하고 $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 를 Θ 에 대한 가능한 상호 배타적(mutually exclusive) 결과값을 나타낸다고 하자. 싱글톤을 포함한 Θ 의 모든 부분집합의 집합은 Θ 에 대한 D-S 프레임(D-S frame)을 정의하며, Dempster-Shafer 이론에서 불확실성의 완전한 표현은 D-S 프레임에 확률값을 할당함으로써 이루어질 수 있다. D-S 프레임에 대한 완전한 확률할당은 기본확률할당(basic probability assignment, BPA)(Shafer, 1976)으로서 정의되며, 이산 BPA 함수(discrete BPA function) $m(A)$ 는 식 (2.1)을 만족시켜야 한다.

$$\begin{aligned} 0 \leq m(A) \leq 1 \quad \text{for all } A \subset \Theta \\ \sum_{A \subset \Theta} m(A) = 1.0, \quad m(\emptyset) = 0 \end{aligned} \quad (2.1)$$

어떤 부분집합 $A \subset \Theta$ 에 대하여 믿음함수(belief function) $Bel(A)$ 및 가능성 함수(plausibility function) $Pl(A)$ 가 Θ 에 대한 BPA로부터 식 (2.2)와 같이 정의될 수 있으며(Shafer, 1976), $Bel(A)$ 와 $Pl(A)$ 는 식 (2.3)과 같이 각각 A 에 대한 확률의 하한 및 상한으로서 고려될 수 있다(Dempster, 1967; Wasserman, 1990a, 1990b).

$$Bel(A) = \sum_{B \subset A} m(B) \quad Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (2.2)$$

$$Bel(A) \leq P(A) \leq Pl(A) \quad A \subset \Theta \quad (2.3)$$

만약 Θ 에 대한 두 개의 BPA, $m_1(A)$ 와 $m_2(B)$ 가 두 개의 독립된 정보의 결과로부터 얻어진다면 그 BPA들은 Dempster 결합규칙(Dempster, 1967)에 의해 결합되어 식 (2.4)와 같이 새로운 BPA를 생성할 수 있다.

$$\begin{aligned} m(C) &= m_1(A) \oplus m_2(B) \\ &= (1-k)^{-1} \sum_{A \cap B = C} m_1(A)m_2(B) \quad \text{for } C \subset \Theta \\ &= \frac{\sum_{A \cap B = C} m_1(A)m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A)m_2(B)} \end{aligned} \quad (2.4)$$

여기서, $k = \sum_{A \cap B = \emptyset} m_1(A)m_2(B)$ 이고 $m(\emptyset) = 0$ 이다.

3. 표본우도기반 BPA 및 사전분포완화

X 가 어떤 확률변수라고 할 때 X 의 결과 x 가 어떤 확률분포 $f(x|\theta)$ ($\theta \in \Theta$ 는 매개변수)에 의해 가장 잘 설명되고 $\mathbf{x} = (x_1, x_2, \dots, x_m)$ 를 확률변수 X 에 대한 m 개의 독립이고 동일하게 분포된 관측치라고 하면 우도의 최소상계값(supremum likelihood value)에 대한 표본우도함수 $L(\theta)$ 의 비율로 정의되는 상대우도함수(relative likelihood function) $R(\theta)$ 는 식 (3.1)과 같다.

$$R(\theta) = \frac{L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} \quad L(\theta) = \prod_{j=1}^m f(x_j|\theta) \quad (3.1)$$

참값이 되는 각 매개변수 θ 의 가능성(plausibility) $Pl(\theta)$ 은 그것의 상대우도 $R(\theta)$ 와 같고 표본우도함수에 기반한 BPA 추정엔 consonant 특성을 가져야 한다고 가정할 경우 샘플링 정보로부터의 BPA 함수 $m_x(\cdot)$ 는 우도함수로부터 식 (3.2)와 같이 결정될 수 있다.

$$\begin{aligned} m_x(A_i) &= R(\theta'_i) - R(\theta_{i+1}') & \text{for } i = 1, \dots, n-1 \\ m_x(\Theta) &= R(\theta_n') \\ m_x(A) &= 0 & \text{otherwise} \end{aligned} \quad (3.2)$$

한편, 확률변수 Θ 에 대한 사전확률분포가 주어질 경우 사전확률분포에 대하여 분석자가 가지는 확신을 반영하기 위하여 0에서 1 사이의 어떤 인자 α 를 도입할 수 있다. 따라서 α 는 사전분포가 적절할 확률로 볼 수 있고 반면에 $1-\alpha$ 는 사전분포가 의미 없을 확률로 볼 수 있으며, 이 경우에는 미지의 매개변수 θ 에 대한 무지만이 존재한다. 이와 같은 방법으로 사전정보의 명확도를 완화시키는 방법을 사전분포완화(contamination of prior)라고 하며, 인자 α 에 의한 할인을 반영한 사전 BPA(prior BPA)는 식 (3.3)과 같이 나타낼 수 있다(Wuben Luo, 1993).

$$\begin{aligned} m_0(\theta'_i) &= \alpha P(\theta'_i) & \text{for } i = 1, 2, \dots, 10 \\ m_0(\Theta) &= 1 - \alpha \end{aligned} \quad (3.3)$$

4. Dempster-Shafer 이론의 적용

4.1 사전분포의 선정

Dempster-Shafer 이론을 이용하여 강우빈도분석에서 확률분포의 매개변수에 대한 불확실성을 평가하기 위해서는 먼저 확률사상에 적합한 확률분포형을 선정해야 한다. 본 연구에서는 위천 유역내 의성 관측소(기상청)을 대상으로 하였으며, 의성 관측소의 1973~2008년 동안의 시우량 자료를 수집하여 이로부터 적정 확률분포형으로 General Logistic(GLO) 분포를 채택하였다.

다음으로 선정된 GLO 분포의 매개변수들에 대한 사전분포를 구축하기 위하여 정보적 사전분포(informative prior distribution)를 적용하였다. GLO 분포의 각 매개변수들에 대한 사전분포를 구축하기 위해 의성 관측소(기상청) 인근의 18개 관측소를 대상으로 시우량을 수집한 후 지속시간별 최대강우량을 산정하고 GLO 분포에 적합시켰으며, 각 매개변수별로 각 관측소에 대한 매개변수값을 다시 확률분포에 적합시켜 각 매개변수, 즉 위치, 규모 및 형상매개변수에 대한 확률분포

를 각각 정규분포, Logistic 분포 및 삼각형 분포로 설정하였다.

4.2 D-S 통계학적 추론의 적용: 매개변수의 불확실성 추정

미지의 매개변수 추정에 대한 D-S 기법은 샘플링 정보 및 독립된 주관적 사전지식을 BPA로 나타내는 것과 Dempster 결합규칙을 이용한 BPA의 결합을 통해 결과 BPA를 생성하는 것으로 구성된다. 본 연구에서는 먼저 GLO 분포의 각 매개변수값에 대한 통계치, 즉 최소값, 최대값 및 모드를 이용하여 각 매개변수에 대한 믿음함수 및 가능성 함수를 작성하였으며, 다음으로 GLO 분포의 각 매개변수들에 대한 사전분포들에 있어서 각 위치매개변수를 앞에서 구한 믿음함수 및 가능성 함수로 이루어진 구간으로 설정하였다. Fig. 4.1은 앞에서 구한 사전분포에 인자 α 를 적용한 사전분포를 나타낸 것으로서 α 값에 따른 사전분포의 변화를 나타내고 있다.

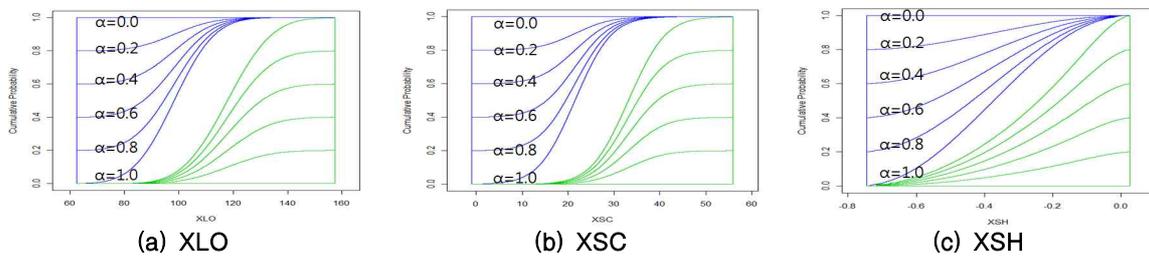


Fig. 4.1 Prior Contamination for Parameters of GLO Distribution(duration=24hr)

샘플링 데이터가 얻어지고 어떤 통계학적 모델이 가정되면 표본정보를 요약하는 표본우도함수를 구할 수 있다. 본 연구에서는 대상 관측소의 강우량 자료로부터 확률분포로서 GLO 분포를 채택하였으며, GLO 분포에 대한 우도함수를 작성하였다. 그리고 미지의 매개변수를 추정함에 있어서 표본증거가 consonant라는 가정을 충족시키기 위해 표본우도함수에 기반한 BPA가 consonant 특성을 가지도록 BPA값이 0이 아닌 부분집합들을 중첩되도록 구성하였다.

그리고 앞서 각 매개변수에 대한 사전분포와 샘플링 정보로부터 구축한 우도함수로부터 산정된 BPA들을 Dempster 결합규칙을 이용하여 결합함으로써 결과 BPA, 즉 사후 BPA(posterior BPA)를 산정하였다. Fig. 4.2는 Dempster 결합규칙을 이용하여 산정된 각 매개변수들에 대한 사후 BPA로서 α 값에 따른 사후 BPA의 영향을 나타내고 있다. 여기서, $\alpha=0$ 일 경우 사후 BPA는 표본우도함수만을 이용하여 얻어진 BPA와 동일하며, α 값이 1에 가까워질수록 사전분포의 영향이 증가하여 믿음함수와 가능성 함수로 이루어진 경계구간이 점차 감소함을 보여주고 있다.

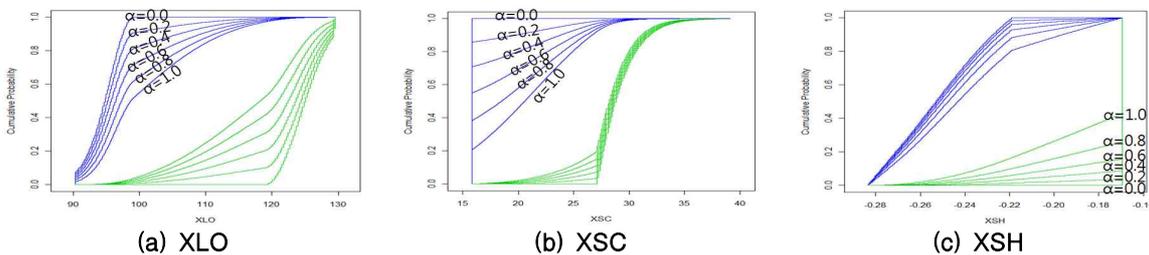


Fig. 4.2 Posterior BPA for Parameters of GLO Distribution(duration=24hr)

Fig. 4.3은 α 에 따른 각 매개변수의 기대치에 대한 구간의 변화를 나타낸 것으로서 α 값의 감

소는 사전정보의 불충분성(weakness)이 증가함을 나타내며, α 값이 증가함에 따라 각 매개변수의 기대치에 대한 상한과 하한의 차이가 감소함을 보여주고 있다.

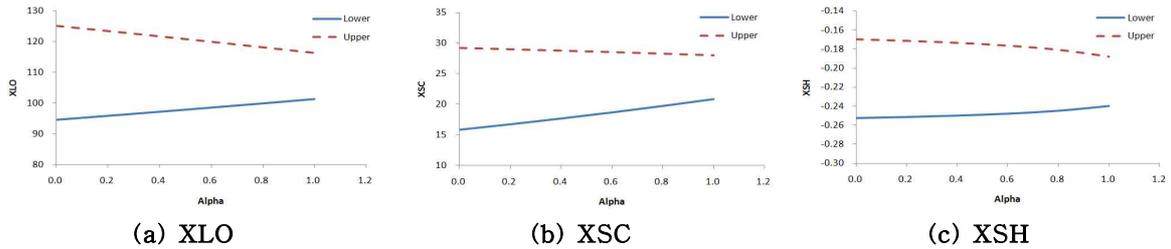


Fig. 4.3 Upper and Lower Expected Value of GLO Parameters As Functions of Alpha(duration=24hr)

5. 결 과

본 논문에서는 홍수위험평가 및 수자원 계획 수립시 가장 기본이 되는 강우빈도해석에서 확률분포의 매개변수에 대한 불확실성 고려한 확률강우량의 산정 및 불확실성의 영향을 평가하기 위하여 Dempster-Shafer 이론을 이용하여 불확실성을 고려한 강우빈도해석모델 구축 및 적용을 실시하였으며, 다음과 같은 결론을 얻었다.

- 1) 강우빈도해석에서 확률분포의 매개변수 추정에 대한 불확실성을 분석하기 위해 Dempster-Shafer 이론에 기반한 불확실성 평가모델을 구축하였으며, 이를 통해 확률강우량 산정시 확률분포의 매개변수에 대한 불확실성을 정량화할 수 있는 기법을 제시하였다.
- 2) Dempster-Shafer 이론에 기반한 강우빈도해석에서의 불확실성 평가시 사전확률분포에 대하여 분석자가 가지는 확신의 정도를 평가하기 위하여 사전분포완화기법을 적용하였으며, 이를 통해 사전확률분포에 대한 확신의 정도에 따른 매개변수의 민감도를 평가하였다.

참 고 문 헌

1. Dempster, A. P.(1966), New Methods for Reasoning Towards Posterior Distributions Based on Sample Data, Annals of Mathematical Statistics, 37, pp. 355-374.
2. Dempster, A. P.(1967), Upper and Lower Probabilities Induced by a Multivalued Mapping, Annals of Mathematical Statistics, 38, pp. 325-339.
3. Wasserman, L. A.(1990a), Belief Function and Statistical Inference, Canadian J. Statist., 18(3), pp. 183-196.
4. Wasserman, L. A.(1990b), Prior Envelopes Based on Belief Functions, Ann. Statist., 18, pp. 454-464.
5. Wuben Luo(1993), Imprecise Probability and Decision in Civil Engineering - Dempster-Shafer Theory and Application, The University and British Columbia.
6. Sadiq, Kleiner, and Rajani(2006), Estimating Risk of Contaminant Intrusion in Water Distribution Networks Using Dempster-Shafer Theory of Evidence, Civil Engineering and Environmental Systems, v.23, no.3, pp. 129-141.