

불확실성, 식자우환?



김 상 옥
 서울대학교 건설환경공학부
 BK21 SIR 사업단 박사후연구원
 plethor1@snu.ac.kr



이 길 성
 서울대학교 건설환경공학부 교수
 kilselee@snu.ac.kr

1. 서론

인간 또는 자연의 거동을 예측 또는 추정하는 사회과학, 자연과학 및 각 종 분야의 공학의 예측결과에는 필연적으로 불확실성이 포함되는 것이 일반적인 상식이다. 예측된 결과의 불확실성에 따라 평가되는 정확성의 문제는 해당 분야의 적용 학문의 종류에 따라 다양하게 평가될 수 있다. 일반적으로 자연현상을 다루는 학문의 예측결과는 그렇지 않은 경우보다 불확실성의 범위가 넓기 때문에 관련 정책의 입안이나 해당 구조물의 설계 등에 있어서 불확실성을 거론하는 것이 오히려 혼란을 주는 경우가 있다. 즉 최종적으로 선정된 결과에 포함되어진 불확실성을 이해하고 그에 따른 대비책 또는 차선책 등을 마련하는 것이 마땅함에도 불구하고, 불확실성에 따른 다양한 대안은 식자우환을 염려하여 삭제되거나 축소되는 경우가 많은 것이 현실이다.

수자원공학 분야는 특히 불확실성과 밀접한 관계

를 가지고 있다. 수자원과 관련된 설계 및 시공은 공학적인 측면에서 일정 정도 이상의 정확도를 가지고 수행되지만 물이 가지고 있는 다양한 무작위성(randomness) 때문에 여타의 다른 공학 분야보다 불확실성이 많이 포함될 수밖에 없다. 또한 그러한 이유로 수자원공학자는 최종 단계에서 어떤 확정적인(deterministic) 값을 제시함에 있어서 신뢰도(reliability) 또는 신뢰구간(confidence interval)을 함께 제시하는 것이 바람직하다 할 수 있다.

위와 같은 불확실성 분석에 대한 필요성으로 인하여 최근 20년간 수자원공학 분야에서는 불확실성을 바탕으로 하는 많은 이론적인 연구가 진행되고 있으며, 특히 홍수와 관련된 홍수빈도분석 등을 중심으로 확률통계적인 방법들이 제시된 바 있다.

그러나 위와 같은 이론적 연구의 진행에도 불구하고 수자원공학 분야에서 발생될 수 있는 불확실성을 최종적 결정 단계에서 고려하기 위한 표준적인 절차는 마련되어 있지 않은 것으로 파악되며, 이에 따라 아직까지도 확정적인 최종 값만을 중심으로 대부분의 수자원 관련 계획이 마련되고 있는 상황이라 할 수 있다. 그러므로 본 고에서는 먼저 불확실성의 이해를 위하여 불확실성의 발생 원인별 분류 및 불확실성의 표현방법을 요약하고, Pappenberger and Beven (2006)이 제시한 불확실성 분석결과를 사용하지 못하게 되는 원인들을 살펴보고자 한다. 또한 최근 들어 사용되고 있는 불확실성을 감소시킬 수 있는 몇 가지 통계학적 방법들을 제시하고 이 방법들을 이용한 수자원공학 분야에서의 응용사례들을 제시하고자 한다.

2. 불확실성의 발생 원인별 분류 및 정량화 방법

불확실성은 여러 학자들에 의해 다양한 용어가 사용되어 분류되었다. Ang and Tang (1975)은 불확실성을 통계적인 방법을 통하여 추론될 수 있는 객관적 불확실성과 정량화 될 수 없는 주관적 불확실성으로 분류한 바 있으며, Yevjevich (1972)는 불확실성을 어떤 절차가 가지고 있는 고유한 무작위성에 근거한 불확실성과 그 외의 원인으로 인해 발생하는 불확실성으로 구분한 바 있다. 특히 수자원공학과 관련하여 발생할 수 있는 불확실성의 원인을 Yevjevich (1972)가 제시한 범주 내에서 간략히 요약하면 다음과 같은 몇 가지 원인에 의한 불확실성으로 분류할 수 있다.

- (1) 자연적 불확실성(natural uncertainty): 자연 현상이 가지고 있는 본연의 무작위성에 근거한 불확실성
- (2) 모형에 의한 불확실성(model uncertainty): 자연현상을 수학 또는 물리학적으로 표현하는데 있어서 실제의 자연현상과 모형과의 차이에 의한 불확실성
- (3) 매개변수에 의한 불확실성(parameter uncertainty): 모형이 가지고 있는 매개변수의 추정에 있어서 발생할 수 있는 오차에 의한 불확실성
- (4) 자료에 의한 불확실성(data uncertainty): 자료를 측정하고 관리하는 데 있어서 발생하는 불확실성으로 주로 측정오차, 비주기적 측정으로 인한 오차, 통계분석을 위한 표본 추출에서 발생하는 오차 등에 의한 불확실성
- (5) 관리적 측면에 의한 불확실성(operational uncertainty): 모형 또는 설계과정에는 포함되지 않은 유지 관리 측면에서 발생할 수 있는 불확실성

위에서 언급한 불확실성은 무작위적인 것에 근거하는 무작위오차(random errors)와 모형화 과정에

서 발생하는 계통오차(systematic errors)에 의해 발생된다. 이와 같은 오차에 의한 대부분의 불확실성은 각각 분리되어 나타나는 것이 아니라 그림 1과 같이 중복되어 나타나는 것이 일반적이다.

또한 분류된 불확실성은 정량적으로 표현되어야 현실적으로 이해되고 이용될 수 있는데, 불확실성의 정량적 표현을 위해서는 확률통계적인 방법이 주로 사용된다. 불확실성의 정량화를 위해서 가장 먼저 생각해 볼 수 있는 방법은 통계학에서 사용되는 모멘트들을 이용하는 것이다. 특히 2차 모멘트인 분산(variance)을 이용하여 간단하게 확률변수의 불확실성을 표현할 수 있으며, 간혹 변동계수(coefficient of variation)나 왜곡도(skewness)도 불확실성을 나타내기 위한 일종의 지표로 활용되기도 한다.

불확실성의 정량화를 위해서 사용되는 방법 중 가장 이상적인 방법은 알고자 하는 확률변수의 분포를 나타내는 확률밀도함수(probability density function, PDF)를 구하는 것이다. 그러나 구하고자 하는 확률변수의 확률밀도함수를 정확히 유도하는 것은 현실적으로 어려운 일이므로 흔히 사용되는 확률밀도함수들 중에서 가장 적합도가 높은 것을 채택하여 불확실성을 표현하게 된다.

위에서 제시한 불확실성의 정량적 표현 방법 이외

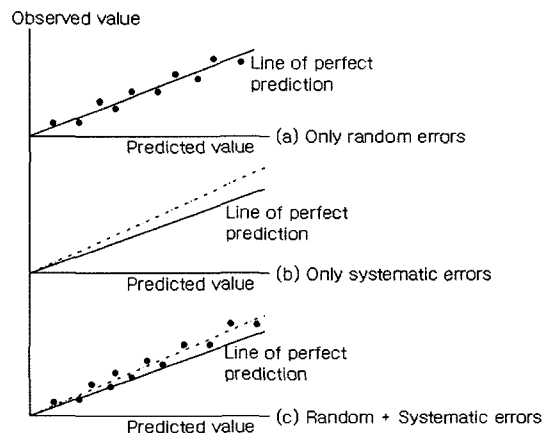


그림 1. 불확실성을 발생시키는 오차의 종류

에 가장 흔히 사용되는 불확실성의 표현방법은 통계적 지표 중 하나인 신뢰구간(confidence interval)을 사용하여 확률변수가 가지는 참값에 대한 상한 및 하한의 구간을 제시하는 것이다. 신뢰구간을 이용한 불확실성의 표현은 몇 가지의 통계적 지표(평균, 표준편차 등)들을 이용하여 확률변수가 가질 수 있는 구간을 표시할 수 있기 때문에 대부분의 통계분석에서 가장 흔하게 이용되고 있는 방법이다. 그러나 신뢰구간을 산정하기 위해서는 오차가 정규분포를 따른다는 정규성(normality) 가정이 필요한데, 이 가정은 특히 추출하고자 하는 자료의 개수가 적은 경우에는 최종적으로 산정된 신뢰구간의 상한 및 하한값이 과다추정되는 경우가 많아 현실적인 측면에서 이용하기 곤란한 경우가 발생된다 (Bickel and Doksum, 1977). 이러한 신뢰구간 산정에 있어서의 문제점은 수자원공학과 관련되어서는 두드러지게 나타나는데, 예를 들어 홍수빈도분석 등을 수행하는 경우 우리나라에서는 이용할 수 있는 자료의 수가 크게 제한적이므로 정규성 가정을 만족시키기 위해 Box-Cox변환과 같은 방법이 사용될 수 있다.

3. 불확실성 분석결과가 응용되기 어려운 이유

앞 절에서는 불확실성 분석의 필요성과 불확실성을 정량적으로 나타낼 수 있는 방법과 그 어려움에 대하여 간단히 언급하였다. Pappenberger and Beven (2006)은 지난 20년간 불확실성의 표현과 관련된 이론적 문제를 해결하고자 하는 다양한 연구에도 불구하고 수자원공학 분야의 응용 또는 실무적인 분야에서 연구결과의 반영 실례를 찾아보기 어려움을 지적하였으며, 이에 대한 이유를 다음과 같이 요약하여 제시한 바 있다.

- (1) 불확실성 분석 결과는 정책입안자들에게 이해되기 어렵다.
- (2) 최종 결정시에는 불확실성 분석 결과를 이용할

수 없다.

- (3) 불확실성을 나타내는 구간의 범위가 넓어 정책 결정시 이용이 어렵다.
- (4) 불확실성 분석 과정은 너무 주관적이다.
- (5) 불확실성 분석 자체를 수행하기 너무 어렵다.

(1)과 (2)의 이유는 유사한 맥락에서 파악할 수 있는 항목들이며, 정책의 입안 또는 결정과 관련된 문제라 할 수 있다. (1)과 (2)를 바꾸어 표현하면 정책의 입안 또는 결정시에는 불확실성에 따른 다양한 대안의 제시보다는 한 가지의 확정적인 대안이 최종 결정에 유효할 수 있다는 것이다. 예를 들어 필자가 내일 우산을 가지고 나가야 할지 말지를 결정해야 하는 문제에 있어서 기상예보는 이 결정에 지대한 영향을 미치는 인자일 것이다. 기상예보가 이진적인(binary) 정보인 '내일은 비가 온다 또는 내일은 비가 오지 않는다' 만을 제공한다면 우산을 가지고 가야할지 말아야 할지에 대한 결정은 아주 간단한 문제이다. 그러나 기상예보가 '내일 비가 올 확률은 50% 이다.' 라고 하는 경우에는 이 문제는 상당히 복잡한 문제가 될 수 있다. 먼저 필자는 주어진 확률값이 얼마나 되는지를 생각할 것이고, 내일 실외에서 보내는 시간이 얼마나 많을 지도 생각할 것이며, 최종 결정에 필자의 경험도 반영할 수 있을 것이다. 결정의 문제를 복잡하게 만든 확률적 기상예보는 필자에게 부적절한 것이었는지, 바람직한 것이었는지 생각해볼 필요가 있다. 기상예보는 자연현상을 다루는 분야 중의 하나이므로 그 과정에서 당연히 불확실성이 개입된다는 것을 인정한다면 기상예보를 수행하는 주체는 이를 포함하여 확률적 예보를 하고 최종적인 결정은 필자의 상황에 따라 필자가 결정하는 것이 바람직한 결정구조라 생각된다. 오히려 이진적인 정보만으로 우산 문제를 결정하였을 때 낭패를 보는 경우의 확률은 산술적으로 또는 심층적으로 증가될 수 있다.

물론 수자원공학과 관련된 많은 정책과 사업들이 국가적 대형 사업임을 감안할 때, 제시된 불확실성이 가지는 범위 내에서의 다양한 대안을 구성하고 이중

에서 하나를 결정하는 일은 확정적인 분석에 의한 결정보다 훨씬 어렵고 복잡한 문제일 것이다. 그러나 결정구조가 복잡해짐에 따라 결정이 유예된다든지, 지체된다든지 할 것을 염려하여 다양한 대안을 삭제 또는 축소하여 결정권자에게 제시하는 것은 공학 또는 과학자로서의 올바른 모습은 아닐 것이다. 오히려 불확실성에 따른 여러 가지 대안을 제시함과 동시에 각 대안의 명확한 장단점과 정책결정자의 당시의 상황도 학제적 연구를 통하여 함께 제시하는 것이 올바른 공학자의 모습이라고 생각된다. 즉, 불확실성을 제시하는 것이 정책결정자를 어렵게 만든다고 생각하기 보다는 정책결정자로 하여금 보다 유연한 대안을 선정할 수 있도록 하는 결정과정 중의 하나라고 생각하는 것이 올바른 것이다.

(3), (4), (5)의 문제는 과학자 또는 공학자와 더욱 밀접한 관계의 문제이다. 불확실성이란 앞서 제시한 여러 가지 종류의 오차들 간의 선형 또는 비선형 관계에 의해 만들어지는 총체라고 할 수 있다. 그러므로 자연적 현상에 의해 발생하는 본연적인 오차는 감소시킬 수 없다 하더라도 계통오차(systematic error)는 실무적 관점의 노력과 이론적 부분의 개발 여하에 따라 충분히 감소시킬 수 있는 여지가 있다.

특히 유량이나 유속 등의 측정 시 발생하는 오차, 자료 관리에서 발생하는 오차, 이상치 오차 등은 수자원 공학과 관련된 결과산정에서 오차를 가장 많이 감소시킬 수 있는 여지를 주는 항목들이다. 또한 불확실성을 표현하는 방법들 중에서 가장 많이 사용되는 방법인 신뢰구간을 표시함에 있어서도 불확실성을 감소시킬 수 있는 통계적 방법의 이론적 개발 또는 적용과 검증은 불확실성에 대한 응용성을 확보할 수 있는 항목이다. 이와 관련하여 Reis and Stedinger (2005), Seidou et al. (2006)은 홍수빈도분석에서 Lee and Kim (2008)은 저수빈도분석에서 베이지안 방법을 사용함으로써 유역에서 측정된 기존 수문자료를 사용한 객관적 사전분포를 이용하여 빈도분석에 있어서의 불확실성을 효과적으로 감소시킨 연구를 진행한 바 있다.

4. 베이지안 방법론에 의한 불확실성의 감소

위에서 언급한 불확실성의 감소 중 매개변수의 추정 시 발생하는 오차 및 신뢰구간의 산정에서 발생하는 오차는 베이지안 방법론을 사용하여 효과적으로 감소시킬 수 있다. 베이즈의 정리는 두 개의 사건 A, B가 서로 종속적일 경우 A의 사건에 의해 B 사건의 확률이 달라진다는 것이다. 베이즈의 정리를 수식으로 나타내면 식 (1)과 같고, 여기서 각각의 확률 사건을 연속 확률밀도함수로 나타내면 베이즈의 정리는 식 (2)와 같이 표현될 수 있다.

$$\Pr(B_j|A) = \frac{\Pr(B_j)\Pr(A|B_j)}{\sum_{j=1}^n \Pr(A|B_j)\Pr(B_j)} \quad (1)$$

$$\pi(\theta|x_1, x_2, \dots, x_n) = \frac{f(x_1|\theta) \cdots f(x_n|\theta)\pi(\theta)}{\int \theta f(x_1|\theta) \cdots f(x_n|\theta)\pi(\theta) d\theta} \quad (2)$$

식 (2)에서 좌변의 $\pi(\theta|x_1, x_2, \dots, x_n)$ 는 사후분포 (posterior distribution), 우변 분자의 $\pi(\theta)$ 는 사전분포(prior distribution)라 명명되며, 우변의 분모는 상수로서 주변분포(marginal distribution)이고, 우변의 분자의 $f(x_1|\theta) \cdots f(x_n|\theta)\pi(\theta)$ 는 발생할 수 있는 모든 가능성을 고려한 우도함수(likelihood function)이다. 그러므로 분석하고자 하는 자료를 나타낼 수 있는 확률밀도함수가 결정되면 이로부터 우도함수를 유도할 수 있고, 적절한 사전분포를 부여함으로써 사후분포로부터 확률밀도함수의 모수를 추출하고 모수의 불확실성을 탐색할 수 있으며 가정조건이 필요하지 않기 때문에 불확실성의 범위를 산정하는 데 있어서 기존의 방법보다 감소된 불확실성의 범위를 산정할 수 있는 것으로 알려져 있다. 그러나 Bayesian 방법을 이용하여 사후분포를 계산하는 것은 쉬운 작업이 아니다. Bayesian 방법의 초기 연구 단계에서는 위와 같은 Bayesian 방법의 적용에 필요한 계산을 위하여 수학적으로 유도가 가능한 몇 가지

의 공액사전분포(conjugated prior distribution)만을 이용하는 경우가 많았으며, 이는 베이저안 방법의 적용에 있어서 많은 논란을 불러일으킨 바 있다.

그러나 최근 계산능력의 하드웨어부분의 발전과 주변 확률분포함수의 적분이 필요없는 Metropolis-Hastings 알고리즘, Gibbs sampling 알고리즘, 주표집(importance sampling) 알고리즘과 같은 Bayesian MCMC 방법에 입각한 Bayesian 계산방법의 소프트웨어부분의 발전으로 인하여 공액사전분포를 사용하지 않아도 식 (2)의 계산이 가능하게 됨으로서 최근 들어 수자원공학 분야에서도 다시 활발히 적용되고 있는 실정이다.

5. 맺음말

수자원공학에서 필수적으로 다루어야 할 불확실성에 대한 정의와 분류, 정량화 방법, 불확실성의 실무 응용이 어려운 점을 제시하였으며, 이에 대한 해결점 중 매개변수 추정과 신뢰구간에서 발생할 수 있는 불확실성을 감소시키기 위하여 최근 많이 다루어지고 있는 베이저안 방법론을 간단히 소개하였다.

해외의 경우 불확실성 분석과 관련된 연구가 진행되고 있는 상황이며, 국내에서도 건교부 (2006)는 수자원장기종합계획을 보완하는 과정에서 용수 수요의 불확실성을 고려하여 저수요, 기준수요, 고수요로 구분하고 이에 따른 각각의 분석결과가 제시된 바 있으나 아직까지 국내외적으로 불확실성에 대한 분석 결과를 실무분야에서 활발히 이용하지는 못하고 있는 실정이다.

본 고에서는 위와 같은 불확실성에 대한 실무 응용이 어려운 점을 크게 정책입안자들에게 너무 복잡한 대안을 제시하는 것을 아닐까 하는 염려와 제시된 불확실성의 범위가 너무 큰 것은 아닐까 하는 걱정으로 정리하여 보았다.

미국 캘리포니아 수자원국에서 연구총괄 담당업무

를 맡고 계신 정일환(Francis I. Chung) 박사님은 최근 한국에서의 그의 강연에서 공학자의 본분은 본인이 알고 있는 수공학적 지식을 최대한 이용하여 정책 결정자가 상황에 따라 누를 수 있는 여러 가지의 버튼(물론 양심적 버튼이다)을 꾸준히 준비하는 것이라고 언급한 바 있다. 이는 사업시행과 관련되어 준비된 하나의 버튼보다는 다양한 대안을 제시할 수 있는 여러 개의 준비된 버튼이 있을 때, 실제로 사업화될 수 있는 여지가 더 많을 수 있다는 것을 간접적으로 표현한 것이고 첫 번째 걱정에 대한 좋은 해답이 될 수 있으리라 생각된다.

또한 불확실성의 발생 원인을 이해한다는 것은 최종 결과에 미치는 오차들을 총체적으로 파악할 수 있다는 의미이므로, 오차의 발생 원인별 관리를 수행할 수 있다는 의미이며 오차를 감소시킴으로써 궁극적으로 불확실성을 감소시킬 수 있다는 것을 뜻한다. 그러므로 두 번째 걱정은 지속적인 오차의 관리와 이론적 부분의 개발로써 해소될 수 있으리라 판단된다.

마지막으로 불확실성을 반영한 다양한 대안을 마련할 수 있는 능력을 배양하는 것과 불확실성을 감소시키기 위한 지속적인 오차 관리 및 연구개발은 그간 불확실성 분석에 대한 사고를 '識字憂患'에서 '識字憂喜'로 바꾸기 위해 꾸준히 노력되어야 할 점이라는 것을 분명히 하고 싶다.

참고문헌

1. 건설교통부 (2006). 수자원장기종합계획 보고서.
2. Ang, A.H.S., and Tang, W.H. (1975). Probability Concepts in Engineering Planning and Design Vol. I: Basic Principles, John Wiley and Sons, NY.
3. Bickel, P.J., and Doksum, K.A. (1977). Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day, Inc., San

- Francisco, CA.
4. Lee, K.S., and Kim, S.U. (2008). "Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method", *Hydrological Processes*, Vol. 22(12), pp. 1949–1964.
 5. Pappenberger, F. and Beven, K.J. (2006). "Ignorance is bliss: Or seven reasons not to use uncertainty analysis" *Water Resources Research*, Vol. 42, W05302.
 6. Reis, D.S., and Stedinger, J.R. (2005). "Bayesian MCMC flood frequency analysis with historical information." *Journal of Hydrology*, Vol. 313, pp. 97–116.
 7. Seidou, O., Ouarda, T.B.M.J., Barbet, M., Bruneau, P., and Bobee, B. (2006). "A parametric Bayesian combination of local and regional information in flood frequency analysis." *Water Resources Research*, Vol. 42, W11408.
 8. Yevjevich, V. (1972). *Probability and Statistics in Hydrology*. Water Resources Publications, Highlands Ranch, CO. 📖