

유역수문예측을 위한 통계적 기법의 활용



강 부 식 ▶▶

단국대학교 토목환경공학과 조교수
bskang@dankook.ac.kr



문 태 완 ▶▶

한국수자원공사 물관리센터 실장
twmoon@kwater.or.kr



황 만 하 ▶▶

한국수자원공사 수자원연구원 수자원환경연구소
수석연구원
hwangmh@kwater.or.kr

1. 서 론

강우, 기온, 증발, 유출등 수문변수의 예측을 위한 기법은 크게 통계적 모형과 역학적 모형이 있다. 일반적으로 통계적 기법은 마코프 연쇄모형과 같은 장기예측이나 수자원 계획을 위한 빈도분석에 주로 활용되어져 왔으나 최근에는 역학적 모형의 국부오차보정이나 확률예보를 이용한 앙상블 예측에도 적용되고 있다. 역학적 모형의 국부오차보정은 기상예측모형의 후처리과정이라고도 할 수 있는 데, 관측이 존재하는 과거기간에 대하여 기상예측모형의 계통오차특성을 찾아내어 보정을 가해주는 hindcast기법을 일컫는

다. 가장 단순하게는 Perfect-Prog방법을 이용하여 보정하는 방법이 있다. 이 방법은 과거기간의 예측인자에 대한 모의가 완벽했다는 것을 가정으로 하는 방법이지만, 실제로 예측인자에 대한 모의는 완벽할 수 없기 때문에 이를 보완하기 위하여 최근에는 출력통계법(MOS; Model Output Statistics)을 사용하고 있다. 이 방법은 과거기간 중 예측인자모의에 있어서의 계통적 오차를 고려하여 오차보정을 할 수 있다는 장점이 있다. 분위사상법이나 비선형추정기법인 인공신경망기법이 출력통계법의 범주에 속한다.

확률(Probabilistic) 예보 또는 범주(Categorical) 예보는 정성적 예보로서 이를 정량적으로 활용하는 기법이 앙상블 예측이다. 미국과 같은 선진국은 비교적 완성된 형태의 확률예보가 제공되고 있지만 국내는 아직 초보적 형태의 범주형 예보가 전국단위로 제공되고 있다. 그러나 앙상블 예측시스템의 구축은 기상예보의 주관적 해석에 기초한 활용방식에서 한걸음 더 나아가 객관적이고 정량적인 활용을 가능케 한다는 점에서 의의를 찾을 수 있다. 이 글에서는 통계적 기법을 이용한 유역강수예측모형의 후처리방법과 앙상블 유량예측기법의 국내외 연구동향 및 적용사례를 소개한다.

2. 강수예측모형의 후처리(Post-processing)

2.1 분위사상법(Quantile Mapping)

기상변수의 관측값과 모의값간에 계통적 오차(systematic bias)가 존재할 경우 두 자료군의 확률분포는 기본 형태가 같더라도 서로 다른 모수를 갖게

된다. 이상적인 모형이 구성되었다면 모형으로부터 모의된 자료는 관측값과 완전한 상관성을 갖거나 유사한 확률분포를 가져야 하나 실제로는 모형의 단순화 과정과 입력정보에 내재된 편이(bias)들로 인하여, 모의된 자료는 일정한 계통적 오차를 갖게 되며 이러한 계통적 오차는 모형내부에서 해결하기 용이하지 않은 경우가 많다. 따라서 오차보정을 모형의 외부에서 수행하는 경우가 있는데 분위사상법(Qunatile mapping)은 이러한 방법론의 일종으로서 관측값과 모의값이 동시에 존재하는 일정한 과거 기간을 선택하여, 관측값과 모의값의 누적확률분포를 이용하여 모의값의 확률분포를 관측값의 확률분포에 사상(mapping)시키는 방법이다. Y_j 를 j 번째 유역의 기상변수값이라하면 오차보정된 값 Z_j 는 다음과 같다.

$$Z_i = F_{oi}^{-1}(F_{si}(Y_i)) \quad (1)$$

여기서 $F_{si}(\cdot)$ 와 $F_{oi}(\cdot)$ 는 각각 해당변수의 모의값

과 관측값에 대한 경험적 확률분포이다.

그림 1은 분위사상법을 이용하여 모의값을 변환하는 과정을 설명하고 있다. 분위사상의 과정을 통하여 변환된 모의값은 결과적으로 관측값과 동일한 확률분포를 가지게 된다. 분위사상법을 이용하여 기상변수를 보정한 예는 Leung et al. (1999)나 Wood et al. (2002)를 들 수 있으며 전자는 기후변화모의에 후자는 장기수문예측에 각각 적용하였다.

그림 2는 2006년 1월 1일부터 5월 31일까지 금강 유역 상류부에 위치한 용담댐의 3개 소유역중 1번 소유역에서의 관측강우와 RQPM모의자료간의 사상(mapping)전·후의 분위도를 도시한 그림이다. 특히 강우예측모형이 예보선행시간에 따라 예측능력이 저하되는 특성을 고려하기 위하여 예보선행시간에 따라 사상을 시도할 수도 있다. 적용의 유연성을 위하여 경험적 확률분포 대신 Weibull확률분포를 가정한 확률분포식에 적합시켜 적용할 수 있다. 21C 프론티어 2단계 연구과업 중 수자원연구원이 주관한 「유역

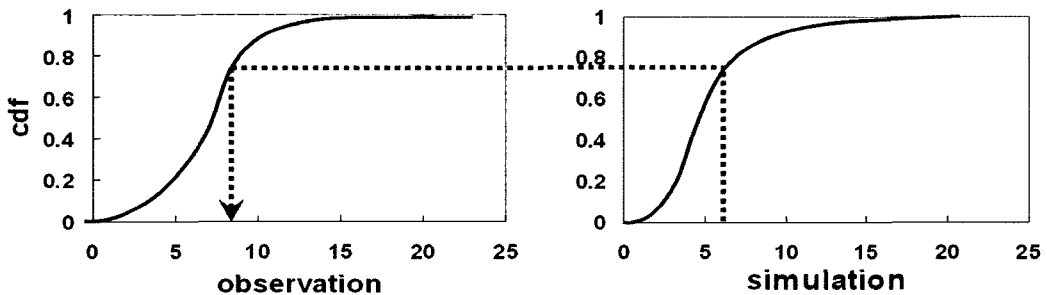
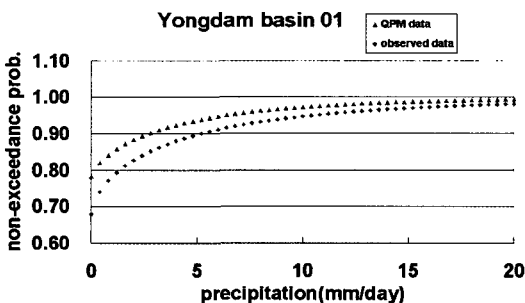
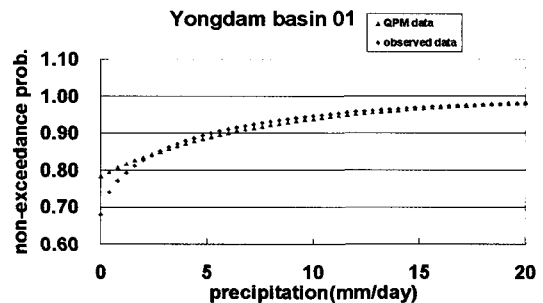


그림 1. 분위사상법을 이용한 모의값 변환



(a) 용담 1번 소유역의 mapping전 분위도



(b) 용담 1번 소유역의 mapping후 분위도

그림 2. 용담댐 1번 소유역에 대한 사상전후 분위도

물관리 운영기술개발」과제에서는 금강유역의 14개소 유역에 대하여 예측시간(1일~10일), 관측값, 계절(이수기와 홍수기)별로 확률분포식을 추정하여 사상(mapping)을 수행한바 있다. (과학기술부, 한국수자원공사, 2007).

2.2 인공신경망을 이용한 수치예보 후처리과정

대규모 대기모형에서의 예측은 모형의 초기조건의 불확실성, 물리적과정의 모수화에 있어서의 불완전한 지식, 지배방정식의 수치해석상의 부정확성(예를 들어 스펙트럴 모형과 격자모형), 이산화된 계산스케일 등으로 인하여 계산오차(bias)를 갖게 된다. 이러한 오차가 부정확한 예측을 야기하게 된다. 따라서 강우 수치예보모형에 내재되어 있는 오차를 규명하고, 정량화하고, 보정하는 것은 강우-유출모형에서의 수치예보활용에 있어서 유출예측의 신뢰도를 확보하기 위한 대단히 중요한 과정이다. 일반적으로 강우수치모형의 해상도는 계산장비의 제약을 받는다. 특정위치에서의 관측값과 모의값과의 선형 혹은 비선형 관계를 규정함으로써 강우수치모형 결과를 원하는 스케일로 상세화시킬 수 있다.

2.2.1 인공신경망 구성

수치예보보정을 위한 신경망모형은 다중 퍼셉트론의 feedforward 구조(그림 3)를 가지고 있다. 신호는 뉴런입력으로부터 은닉층을 통하여 전방으로 흐르게 되고 출력뉴런에 이르게 되는 구조를 가지고 있다. 뉴런은 층으로 배열되어 있으며, 입력층은 단순히 입력변수의 값을 넣게 된다. 은닉층과 출력층의 뉴런은 선형층의 모든 요소들과 연결되어 있다. 기본방정식은 다음과 같다.

$$z_k(t) = g\left(\sum_{h=0}^{n_h} w_{kh}^o f\left(\sum_{i=0}^{n_i} w_{hi}^h x_i(t) - \theta_i\right) - \theta_h\right) \quad (2)$$

여기서 x_i 는 입력층 요소 i 로의 입력, z_k 는 출력층

요소 k 로의 출력

$i = 1, 2, \dots, n_i$ 여기서 n_i 는 입력요소갯수

$h = 1, 2, \dots, n_h$ 여기서 n_h 는 은닉요소갯수

$k = 1, 2, \dots, n_k$ 여기서 n_k 는 출력요소갯수

w^{hi} 는 입력요소 i 와 은닉요소 h 사이의 연결강도를 조절하는 매개변수 또는 가중치

q_i and q_h 는 기준값(threshold)

w^{ok} 는 은닉요소 h 와 출력요소 k 사이의 연결강도를 조절하는 매개변수 또는 가중치

f 는 Logistic 함수형태의 활성화함수, g 는 $g(x) = x$ 형태의 항등 함수(Identity function)이다.

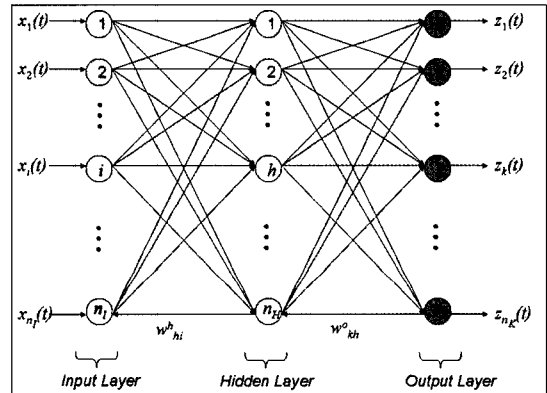


그림 3. 다중퍼셉트론 신경망 구조

각 수치예보 격자에 대하여 강우의 계절적 특성을 고려하는 방법으로, 강수기와 건조기에 대한 2종류의 신경망모형을 구축할 수 있다. 첫번째 모형은 강수확률(PoP; Probability of Precipitation)을 예측하는 것이다. 이러한 PoP 신경망은 RDAPS 예측모의에 적용하여 3시간 동안의 강수유무를 예측하는데 활용된다. 두번째는 정량적 강수예측(QPF) 신경망으로서 특정 기준값을 초과하는 PoP값에 해당하는 강수기에서 수치예보 예측모의에 적용된다. PoP(QPF) 신경망은 수치예보 예측모의와 수치예보 수행 6시간 혹은 12시간 전에 발생한 관측강수로서 예측인자(predictor)를 구성하며, 예측변수(predictand)는 수치예보의 격자에 해당하는 강우의 발생확률(또는 보정된 강우값)이다.

2.2.2 자료

표 1의 자료는 2001년 7월부터 2002년 6월까지의 각종 기상변수에 대한 수치예보 모의자료 및 관측자료이다. 예측모의는 기상청의 현업 중규모 강우수치예보모형인 RDAPS모형으로부터 30 x 30km 격자에서 수행되었다. RDAPS는 00UTC와 12UTC에서 하루에 2차례 48시간 예측을 수행한다(그림 4). 따라서 각 시간대는 예측수행시간에 따라 8차례 예측값이 갱신되게 된다. 이러한 예측결과는 기상청에 의하여

자연재해관련 기관, 수자원관리기관, 그리고 대학등에 실시간으로 공식 제공되고 있다. 지표에서의 기상 관측은 전국의 452개 무인 자동기상시스템(AWS)에서 매시간 기록된다. 그림 5는 RDAPS격자와 함께 남한내 AWS의 분포를 나타내고 있다.

2.2.3 강수기 QPF 신경망

다층퍼셉트론 신경망은 강수기 강수예측을 위해서도 다층퍼셉트론 신경망을 구성할 수 있다. 본 모형

표 1. RDAPS 및 관측자료 목록

| RDAPS variables | |
|---------------------|--|
| 1. | Geopotential height 500 hPa [m] |
| 2. | Geopotential height 750 hPa [m] |
| 3. | Geopotential height 1000 hPa [m] |
| 4. | Atmospheric thickness 500 hPa - 1000 hPa [m] |
| 5. | Wind direction at 500 hPa |
| 6. | Wind speed at 500 hPa [m/s] |
| 7. | Wind direction at 750 hPa |
| 8. | Wind speed at 750 hPa [m/s] |
| 9. | Wind speed at surface [m/s] |
| 10. | Temperature at 500 hPa [°K] |
| Observed variables: | |
| 1. | Hourly precipitation [mm/h] |
| 2. | Temperature [°C] |
| 3. | Wind direction [degree] |
| 4. | Wind speed [m/s] |
| Geographic data: | |
| 1. | Latitude, longitude, and elevation corresponding to the center of each RDAPS' grid box |
| 2. | Latitude, longitude, and elevation of each weather station |

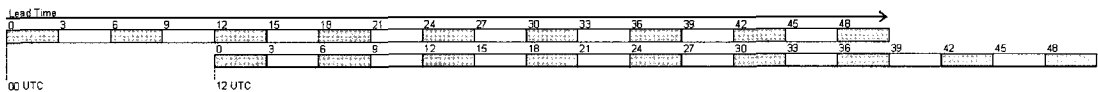


그림 4. RDAPS NWP 모형의 출력구조(3시간 누적강수를 48 예보시간까지 계산하며, 매 12시간마다 갱신된다.)

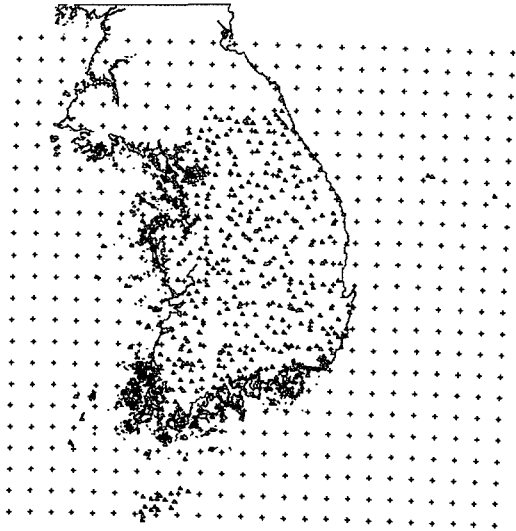


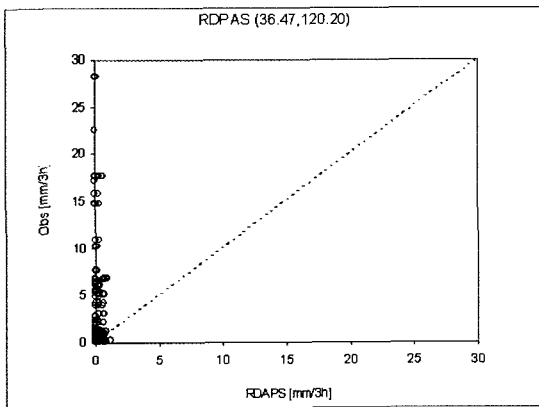
그림 5. 한반도를 포괄하는 RDAPS 격자 (+)와 지상관측소(Δ).

은 3층으로 구성되어 1층에는 13개의 유닛이, 2층에도 13개, 마지막층에는 1개의 유닛이 배열되어 있다. QPF신경망을 통하여 강우수치예보모의의 시스템오차가 감소하게 되는 데, 이러한 경향은 관측치와 RDAPS 모의, 그리고 QPF 신경망 보정값과의 비교를 통해 확인할 수 있다 (그림 6). 최소자승 선형회귀 결과 강우량 14mm/3시간 이하 구간에서는 과다 예측을 14mm/3시간 이상의 구간에서는 과소예측하는

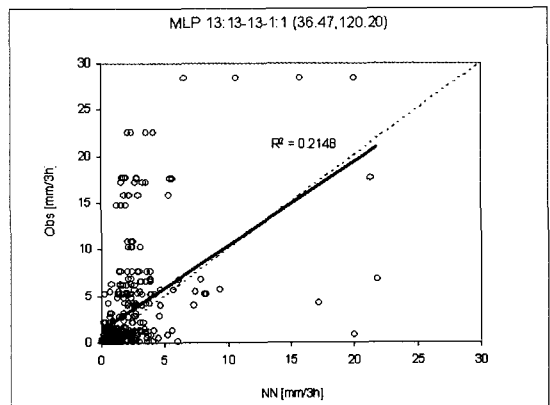
경향을 보이며 결정계수는 0.215이다. 그러나 이러한 낮은 결정계수에도 불구하고 그림 7은 QPF신경망모형이 RDAPS의 정확도를 개선시키고 있음은 뚜렷이 확인할 수 있다.

3. 앙상블(Ensemble) 유량예측기법

앙상블 유량예측기법(ESP; Ensemble Stream-flow Prediction)이란 현재의 유역상태를 초기조건으로 사용하고 과거의 온도나 강수 등의 시계열 앙상블을 강우-유출모형에 입력하여 유출량을 앙상블로 예측하는 기법이다. ESP는 결국 현재의 유역상태와 유역에서의 과거 강우 관측기록, 미래 강우예측에 대한 정보를 조합하여 그에 따른 유출 앙상블을 생산해 내게 된다. 유출 앙상블은 각 앙상블 트레이스(trace)가 갖는 가중치에 따라 예측확률분포를 달리 갖게 되고 경우에 따라서는 유량으로부터 2차적으로 유도되는 변수들의 확률분포로 전이되기도 한다. 미국 기상청(National Weather Service)에서는 현업에서의 장기 하천 유량예보를 위해 NWS 하천예보시스템(NWSRFS; NWS River Forecast System) 내에 ESP를 탑재시켜 계절 유량예측을 주기적으로 제공한다(NWS, 2003). 따라서 미국의 경우 ESP에 대한

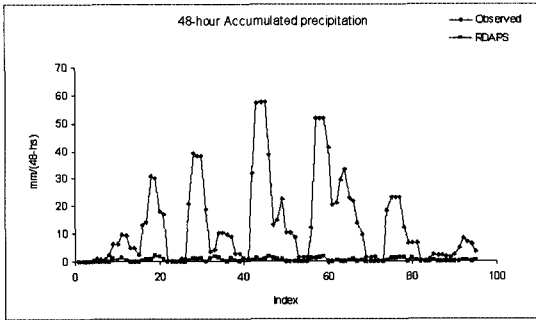


(a) RDAPS 예측

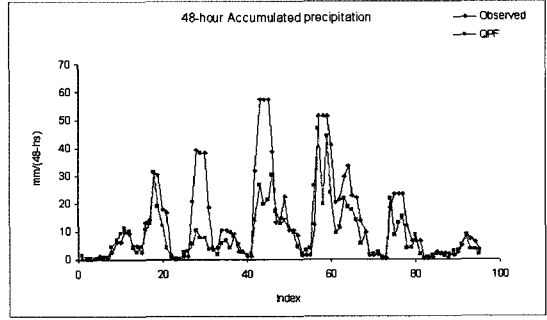


(b) 신경망 예측

그림 6. 3시간 누적강수에 대한 예측과 관측간의 산포도(홍수기)



(a) RDAPS 예측



(b) QPF 신경망 예측

그림 7. 48시간 누적강수에 대한 시계열도(홍수기)

연구는 비교적 활발히 수행되어 왔다(Brandon, 2005; Werner et al., 2004; Croley II, 2000; Day, 1985). 그동안 NWS의 현업 수치확률예보를 활용하여 ESP를 구성하는 연구가 주로 이루어졌으나, 최근에는 기후변동에 따른 장기 기상변화를 예측하기 위하여 ENSO나 PDO와 같은 기후인자를 사용하여 계절단위의 장기 유량예측을 시도하는 연구(Hamlet과 Lettenmeier, 1999; Werner 등, 2004)도 활발히 진행되고 있다.

3.1 시스템 구성

ESP는 과거 기상자료가 미래에서 발생가능한 사상을 대표한다는 가정을 전제로 유역의 현재상태를 초기조건으로 하고 과거강수를 입력자료로 사용하여 유출을 예측하는 시스템이다. 과거 N년의 강우시계열 중 특정년도의 강우시계열에 해당하는 유출시계열을 각각 강우 또는 유출트레이스라고 하고 N개의 트레이스를 총칭하여 앙상블이라 일컫는다. 각각의 트레이스는 가중치를 가지고 있는데, 가중치 설정 문제는 ESP에서 가장 중요하게 다루어지는 부분이다.

ESP 시스템에서 기상예측을 반영하기 위해서는 크게 2가지의 방법이 있다. 하나는 시스템의 입력이 되는 강우나 온도에 대한 사전조정이고, 다른 하나는 유출트레이스의 가중치를 보정하는 사후조정이다. 지금까지 국내 ESP에 대한 강우예측의 활용에 있어 이

론은 소개된 바 있으나, 현업화는 본격적으로 시도되지 못하였다. 선진국에서도 강우예측정보의 활용은 비교적 최근에 와서 시도되기 시작하였는데, APRFC(Alaska Pacific River Forecast Center)의 Alaska 기법이 대표적인 방법이다(Werner, 2004). 또한 강우앙상블의 모든 트레이스를 사용하지 않고 예보나 전망에 따라 일부 년도로 트레이스를 제한하는 방법도 사용되고 있는데(Hamlet and Lettenmaier, 1999), 이러한 노력들로 인하여 장기 유량의 예측능력이 개선되고 있다.

3.2 강우앙상블의 트레이스 가중치 설정

ESP의 강우앙상블을 구성하는 N개의 트레이스들은 앞에서 유도된 확률밀도함수로부터 발생확률을 추정할 수 있다. 이제 이렇게 하여 추정된 발생확률로부터 트레이스 가중값을 계산하여야 한다. 각 트레이스에 가중치, w_i 를 사용했다고 가정하면 누적확률분포의 추정치는 다음과 같이 표현이 된다.

$$\hat{P}[X \leq x] = \frac{1}{n} \sum_{i | x_i^* \leq x} w_i \quad (3)$$

여기서 $\hat{P}[X \leq x]$ 이다.

이후 단계에서 가중치(w_i) 설정에 대한 다양한 기법들이 이미 소개되어 있는데, 이중 가장 보편적인 방법은 Croley방법(Croley II, 2000)과 PDF-Ratio

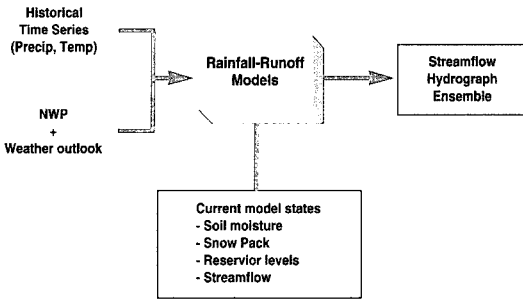


그림 8. 앙상블 유량예측의 전체적 구성

방법(Stedinger and Kim, 2002)을 들 수가 있다. Croley방법은 범주형 확률예보 (Categorical Probabilistic Forecast) 형태의 확률예보를 이용하여 표본재구성 (Sample restructuring)이나 표본가중치설정 (Sample weighting)으로 표본의 각 트레이스에 각기 다른 가중치를 부여하는 방법이며, PDF-Ratio방법은 과거 기상변수의 확률밀도함수와 현재 유역의 상태, 지구기상시스템 등과 같은 모든 수문학적 정보를 고려한 조건부 확률밀도함수를 이용하여 가중치를 구하는 방법이다.

Croley방법의 단점은 트레이스 가중값을 구하기 위하여 다수의 기상확률예보가 존재해야 한다는 점이고, 기상확률예보가 다양하지 않은 경우 각 트레이스에 동일한 가중치를 부여할 수밖에 없다는 점을 들 수 있다. PDF-Ratio방법은 Croley방법에 비하여 하나의 기상확률예보만으로도 가중값을 구할 수 있다는 장점이 있으나 유역의 상태변수와 기상변수간의 조건부 확률값을 찾아내기가 쉽지 않은 단점이 있다 (정대일과 김영오, 2002).

21C 프론티어 2단계 연구과업 중 「유역물관리 운영기술개발」과제에서는 국내의 기상청 강수전망을 이용하여 앙상블 멤버가중값을 적용한 바 있다. 가중값 산정 절차는 그림 9에 제시된 바와 같다.

4. 맺음말

수문예측을 위한 통계적 기법중 인공지능망을 이

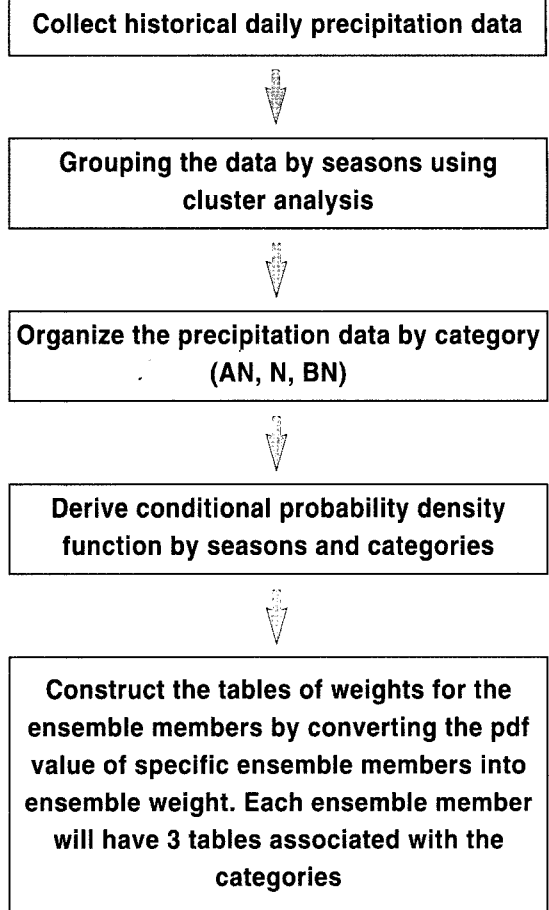


그림 9. 순(10-day)강수전망을 이용한 앙상블가중값산정 절차

용한 강우수치예보의 국부오차보정 그리고 범주형 확률예보의 정량적 활용을 위한 앙상블 유량예측기법에 대하여 기본적 이론과 적용사례를 살펴보았다. 강우수치예보는 동시적자계에서의 중규모기상모형을 이용한 역학적 상세화기법의 적용이 보편적 접근방식이지만 이경우에서도 중규모기상모형의 경계조건이 되는 대규모 강제력(large scale forcing)에 대한 계통오차의 보정은 필요한 과정으로 판단되며 본 연구에서 제시된 인공지능망등의 비선형 예측이론이 효과적일 것으로 판단된다.

앙상블유량예측의 경우 기존의 균등가중값을 적용할 경우 홍수기 유량예측의 신뢰도가 떨어지는 것이

문제로 지적되어 왔으나 강수전망을 활용할 경우 홍수기 유량예측을 개선할 수 있다. 향후 강수전망의 자료가 더욱 축적되면 보다 의미있는 앙상블 유량예측검증이 가능할 것이다. 또한 현재 기상청에서는 전국 및 주요도시별 강수전망이 제공되고 있지만, 주요 수계별 강수전망이 제공된다면 보다 신뢰성있는 앙상블 유량예측에 의한 과학적인 유역수자원관리기술확보가 가능해 질 것이다.

감사의 글

본 연구는 21세기 프론티어 연구개발사업중 수자원의 지속적 확보기술개발사업단의 세부과제 “유역물관리 운영기술개발(과제번호 1-6-2)”로 수행되었습니다.

참고문헌

과학기술부, 한국수자원공사 (2007). 실시간 물관리 운영 시스템 적용, 2단계 연구과제 최종보고서, 21세기 프론티어 연구개발사업 수자원의 지속적 확보기술개발 사업단.

정대일, 김영오 (2002). “앙상블 예측을 이용한 충주댐의 월 유입량 예측.” 대한토목학회 논문집, 대한토목학회, 제22권 제3-B호, pp. 321-331.

Brandon, D.G. (2005). “Using NWSRFS ESP for Making Early Outlooks of Seasonal Runoff Volumes into Lake Powell.” AMS Forum: Living with a Limited Water Supply Session 1. Probabilistic Hydrometeorological Forecasting Part II, San Diego, CA, U.S.A.

Croley II, T.E. (2000). Using Meteorology Probability Forecasts in Operational Hydrology. ASCE Press, Reston VA .

Day, G.N. (1985). “Extended Streamflow Forecasting Using NWSRFS.” ASCE J. Water Resour. Plng. & Mgmt., 111(2), pp. 157-170.

Hamlet, A.F., and D.P. Lettenmeier (1999). “Columbia River Streamflow Forecasting Based on ENSO and PDO Climate Signals.” ASCE J. Water Resour. Plng. & Mgmt., 125(6), pp. 333-341.

Leung, L. R., Hamlet, A. F., Lettenmaier, D.P., and Kumar, A. (1999). “Simulations of the ENSO hydroclimate signals in the Pacific Northwest Columbia River Basin.” Bull. Amer. Meteorol. Soc., 80, pp. 2313-2328.

National Weather Service (2003). Ensemble Streamflow Prediction Analysis and Display Program (ESPADP) User Manual. Riverside Technology, Inc.

Stedinger, J. and Kim Y. O. (2002). “Updating ensemble probabilities based on climate forecasts.” 2002 Conference on Water Resources Planning and Management. ASCE, CD.

Werner, K, D. Brandon, M. Clark and S. Gangopadhyay (2004). “Climate Index Weighting Schemes for NWS ESP-based Seasonal Volume Forecasts.” J. of Hydrometeorology, 5(6), pp. 1076-1090.

Wood, A. W., Maurer, E.P., Kumar, A., and Lettenmaier, D.P. (2002). “Long-range experimental hydrologic forecasting for the eastern United States.” J. Geophys. Res.-Atmos., 107, pp.4429, doi:10.1029/2001JD000659. ●