



Assessment of predictability of categorical probabilistic long-term forecasts and its quantification for efficient water resources management

Son, Chanyoung^a · Jeong, Yerim^a · Han, Soohee^a · Cho, Younghyun^{b*}

^aHydrometeorological Cooperation Center

^bKorea Water Resources Corporation (K-water)

Paper number: 17-047

Received: 23 May 2017; Revised: 30 June 2017; Accepted: 30 June 2017

Abstract

As the uncertainty of precipitation increases due to climate change, seasonal forecasting and the use of weather forecasts become essential for efficient water resources management. In this study, the categorical probabilistic long-term forecasts implemented by KMA (Korea Meteorological Administration) since June 2014 was evaluated using assessment indicators of Hit Rate, Reliability Diagram, and Relative Operating Curve (ROC) and a technique for obtaining quantitative precipitation estimates based on probabilistic forecasts was proposed. The probabilistic long-term forecasts showed its maximum predictability of 48% and the quantified precipitation estimates were closely matched with actual observations; maximum correlation coefficient (R) in predictability evaluation for 100% accurate and actual weather forecasts were 0.98 and 0.71, respectively. A precipitation quantification approach utilizing probabilistic forecasts proposed in this study is expected to enable water management considering the uncertainty of precipitation. This method is also expected to be a useful tool for supporting decision-making in the long-term planning for water resources management and reservoir operations.

Keywords: Categorical probabilistic long-term forecasts, Quantification, Precipitation

효율적인 수자원관리를 위한 범주형 확률장기예보의 예측력 평가 및 정량화

손찬영^a · 정예림^a · 한수희^a · 조영현^{b*}

^a수문기상협력센터, ^b한국수자원공사(K-water)

요 지

기후변화로 인해 강수의 불확실성이 증가하는 현 시점에서 효율적인 물 관리를 위한 계절예측 및 기상 예보의 활용은 필수적이다. 본 연구에서는 기상청에서 2014년 6월부터 시행하고 있는 범주형 확률장기예보를 Hit Rate, Reliability Diagram, Relative Operating Curve (ROC)의 평가지표를 활용하여 예측력을 검증하였고, 추가적으로 확률예보를 활용하여 정량적인 예측 강수량을 생산하는 기법을 제안하였다. 확률장기예보의 예측력 검증결과 최대 48%의 예측력을 갖는 것을 확인할 수 있었다. 확률예보를 활용하여 예측 강수량을 추정한 결과, 정량적으로 관측 자료와 유사하게 모의되는 것을 확인할 수 있었으며 예측 적합도 평가결과 100%의 정확도를 가진 예보의 경우 최대 0.98, 실제 예보의 경우 최대 0.71의 상관계수를 보였다. 본 연구에서 제안하는 확률예보를 활용한 예측 강수량 추출기법은 강수의 불확실성을 고려한 물 관리를 가능하게 해줄 것으로 판단되며 효율적인 수자원 장기 이수계획 및 저수지 운영의 의사결정지원 등에 활용 가능할 것으로 기대된다.

핵심용어: 범주형 확률장기예보, 정량화, 강수량

*Corresponding Author. Tel: +82-2-2150-0560
E-mail: yhcho@kwwater.or.kr (Y. Cho)

1. 서론

최근 이상기후로 인해 강수량의 불확실성이 가중되고 홍수, 가뭄 등 수재해의 발생빈도 및 강도가 증가함에 따라 안정적인 용수공급 등 수자원 관리에 어려움을 겪고 있어 장기 기상예측기반의 수자원관리가 요구되고 있는 실정이다. 장기 기상예측정보는 수자원의 효율적인 이용 및 증가되는 수요와 불확실한 공급에 대한 의사결정을 지원해주며 가뭄 대비를 위한 물 공급 제한 등에 활용될 수 있는 매우 중요한 정보이다. 현재 수자원 분야에서 장기 수자원의 효율적인 공급 및 운영을 위해 매년 10월부터 다음 해 6월까지 댐 저수지 수위 운영 및 홍수기 대비 운영 목표 수위 계획을 수립 및 운영하고 있지만, 장기예보 자료의 불확실성 및 수문해석 모형 입력 자료로의 활용 어려움, 정성적인 자료 제공, 예측 자료의 검증 미흡 등으로 기상청에서 제공하는 장기예보를 참고할 뿐 실제로는 과거 관측 자료에 기반을 둔 빈도분석 결과를 활용하여 댐 운영 계획을 수립하고 있는 실정이다.

장기예보의 경우 기후시스템의 복잡한 상호작용에 대한 충분한 연구 및 기술력이 부족함에 따라 단기·중기예보에 비해 상대적으로 정확성은 낮고 불확실성은 큰 것이 현실이다. 보다 효율적인 의사결정을 위하여 국내·외에서는 불확실성에 대한 정보를 확률론적으로 제공하는 방법을 운영하고 있다. 현재 기상청에서는 국내 처음으로 2014년 6월부터 기존 단정예보에서 범주형 확률장기예보(1·3개월 전망)로 변경하여 매주 혹은 매달 제공하고 있다. 범주형 확률장기예보는 기상청 전지구 계절예측시스템의 장기예측 모델 자료와 전세계 장기예측자료생산센터(Global Producing Centre, GPC) 자료, WMO 장기예보 다중모델 앙상블 자료를 종합적으로 분석 및 토의하여 30년 평균강우량을 기준으로 평년 이하(Below normal, BN), 평년과 유사(Near normal, NN), 평년 이상(Above normal, AN) 등 3가지 범주의 확률을 주별 혹은 월별로 예보하고 있다(KMA, 2014, 2016). 이러한 범주형 확률장기예보는 이미 미국, 영국, 캐나다, 일본, 중국, 러시아 등 많은 국가에서 운영하고 있으며 공통적으로 주요 예보인자는 기온과 강수량, 범주는 우리나라와 마찬가지로 3가지 혹은 극치사상(extreme event)을 포함한 5가지로 분류하여 예보하고 있다.

물 관리 계획을 위해 국·내외에서는 과거 자료를 활용하여 강수량 및 유량을 예측하는 방법과 장기예보 자료에 기반을 둔 방법을 주로 활용하고 있다. 이와 관련된 연구동향을 살펴보면 Simpson et al. (1993)은 예측 가능한 태평양 지역의 해수면온도 자료를 이용하여 하천의 연 유출량을 예측하고 설정된

범주(BN, NN, AN)에 대해 연간 하천 유출량의 확률들을 추정하는 기법을 제안한 바 있다. 또한 Croley (2000)는 범주형 확률강수예보를 활용하여 표본시나리오에 가중 값을 부여하는 방법을 제시한 바 있으며, Stedinger and Kim (2010)은 확률밀도함수를 이용하여 표본시나리오에 가중 값을 부여하는 PDF-ratio 방법을 제안한 바 있다. Gouweleeuw et al. (2005)는 범주형 중장기 기상예보자료와 강우-유출모형을 조합하여 1995년과 1997년의 홍수사상에 대하여 분석을 수행한 바 있으며, Kang et al. (2005)은 중장기 수자원 운영 및 관리에 활용하고자 수치예보와 월 강수 전망을 조합하여 Ensemble Streamflow Prediction (ESP) 기법을 활용하는 방안을 제시한 바 있다. 추가적으로 Kim et al. (2011)은 금강유역을 대상으로 구성된 바 있는 RRFS (Rainfall-Runoff Forecasting System)-ESP 시스템에 PDF-ratio 방법을 기반으로 한 사전 처리기능을 추가하여 보다 효율적인 중장기 예측시스템으로의 확장을 시도한 바 있으며, Kang (2013)은 해수면온도(Sea Surface Temperature, SST)와 북반구 500 hPa 지위고도(Geopotential Height, GPH) 자료로부터 선정된 예측인자를 이용하여 안동댐 계절별 유입량에 대한 카테고리 확률예측과 조건부 확률밀도함수에 의한 예측을 수행한 바 있다. Yu et al. (2014)은 강수량과 기온에 대한 3개월 예보자료를 이용하여 각 월별 앙상블을 생성하고 HBV (Hydrologiska Byrans Vattenbalansavdelning) 물수지 모형에 입력하여 물 부족을 확률적으로 전망하는 방법을 제시한 바 있다.

과거 기상자료의 평균을 기준으로 한 예보는 ‘적음’과 ‘많음’에 대한 기준이 모호하여 정량적 정보를 획득하기 위한 정보로 이용하기 어려웠다. 하지만, 2014년 6월부터 범주형의 형태로 예보를 수행하고 있어 이를 활용하여 보다 효율적인 이수 측면의 물 관리가 가능할 것으로 기대된다. 수자원 분야에 1개월 및 3개월 범주형 확률장기예보를 활용하기 위해서는 우선적으로 확률장기예보에 대한 예측력 검증이 선행되어야 하며 추가적으로 정성적인 확률장기예보자료를 활용하여 불확실성을 포함한 강수량으로 정량화되어야 한다. 따라서 본 연구에서는 범주형 확률장기예보를 Hit rate 등 Anders (2011)와 KMA (2014)에서 제안하는 다양한 검증기법을 통해 예측력을 판단하였고, 범주형 확률장기예보를 저수지 운영 계획 등에 활용하고자 불확실성을 포함한 정량적 강수량 산출기법을 제안하고자 하였다. 본 연구의 결과는 미래 불확실성을 고려한 이수 측면의 수자원 운영 및 관리에 활용 가능할 것으로 판단된다.

2. 이론적 배경

2.1 확률예보 평가방법

2.1.1 Hit Rate

범주형 확률장기 예보의 평가를 위해 관측과 예측의 각 경우에 대해 예측이 맞힐 때를 1, 틀렸을 때는 0으로 하여 경우의 수를 Table 1과 같이 정리하였다. Hit Rate은 예보의 적중률을 의미하므로 전체 경우의 수 중에 관측과 예측이 맞은 경우의 비율을 나타내며, Eq. (1)을 통해 상대적으로 쉽게 산정할 수 있다. Hit Rate는 0에서 1 사이의 값을 갖고 예보의 적중률이 높을수록 1에 가까운 값을 갖는다.

$$Hit\ Rate = \frac{a+e+i}{N} \quad (1)$$

여기서, a, e, i 및 N 은 Table 1에 제시된 바와 같다.

2.1.2 Reliability Diagram

Reliability Diagram은 각 확률구간에 따라 예측한 확률과 관측 빈도의 유사한 정도를 평가하는 방법이다. Reliability Diagram에는 Perfect Reliability를 나타내는 대각선이 존재하고, No skill line을 의미하는 경우에 따라 다른 기울기를 가지는 대각선이 존재한다(Fig. 3). 완벽한 예보일수록 Perfect Reliability선 위에 위치하고, 데이터가 Perfect Reliability선 위에 위치하면 과소모의, Perfect Reliability선 아래에 위치하면 과대모의 하는 것을 의미한다. No skill line은 Perfect Reliability와 평균값의 중앙값을 의미하며, Perfect Reliability을 갖더라도 No skill line에 위치하면 통계적으로 의미를 갖지 못한다. Reliability Diagram은 확률예보의 편향정도를 그래픽으로 확인할 수 있지만, 정성적인 평가만 가능하고 정량적인 평가는 어렵다는 단점이 있다.

2.1.3 Relative Operating Curve (ROC)

ROC는 예측한 경우의 비율(Probability of Detection, POD)

과 예측하지 못한 경우의 비율(False Alarm Ration, FAR)의 대립되는 두 결과를 구분하는 예보 능력을 평가하는 방법이다. 각 확률구간별로 두 경우를 계산하여 그래프로 나타낸다. 0과 1 사이의 값을 갖고, 1에 가까울수록 그래프는 대각선 좌측 위쪽에 위치하고 예측성능이 우수함을 의미한다. 그러나 ROC는 카테고리 별 확률의 오차에 대한 정보를 제공하지 못하고 예보 편차에 민감하지 않기 때문에 신뢰도를 평가할 수 없다는 단점이 있다.

2.2 범주형 확률장기예보를 활용한 강수량 산출기법

범주형 확률장기예보는 30년(1981~2010년) 월평균 강수량(μ_p)을 기준으로 평년보다 적을 확률(P_{BN}), 평년과 유사할 확률(P_{NN}), 평년보다 많을 확률(P_{AN})을 월별로 제공하고 있다. 추가적으로 평년 비슷 범위(r_{mm})를 제공하는데, 예로 7월의 평년 강수량이 280 mm라 할 때 r_{mm} 이 $\pm 10\%$ 라 하면 252 mm (280×0.9)에서 308 mm (280×1.1)는 평년과 정량적으로 유사한 것으로 판단하게 된다. 본 연구는 이들 범주형 확률장기 예보의 정보를 활용하여 미래 예측 강수량 및 불확실성을 아래 절차로 산정하였다. 첫 번째로 1981~2010년 동안 발생한 각 월별 일 강수량(y)을 각 월의 일 개수(N)만큼 무작위로 추출하고 i 번 모의 발생시킨다($Y_i = \{y_{1,i}, y_{2,i}, \dots, y_{j,i}, \dots, y_{N,i}\}$). 여기서, Y_i 는 i 번째 무작위로 모의 발생한 일 강수량의 조합, N 은 각 월의 일수(예, 31일)를 의미한다. 두 번째로 Eq. (2)와 같이 강우발생 확률(P_o)을 각 월별로 산정하고 강우(P_o) 및 무강우($1-P_o$) 발생확률로 0 (무강우) 또는 1 (강우)을 N 만큼 무작위로 추출하고 i 번 모의 발생시킨다($R_i = \{r_{1,i}, r_{2,i}, \dots, r_{j,i}, \dots, r_{N,i}\}$).

$$P_o = \frac{1}{N_{yr}} \sum_{t=1}^{N_{yr}} \frac{n_t}{N_t} \quad (2)$$

여기서, N_{yr} 은 30개년(1981~2010년), n 은 강우 발생 일수이다.

세 번째로 모의발생한 일강수량(Y_i)과 강우발생유무(R_i)의 조합을 곱하여 강수일수를 고려하여 모의된 일 강수량 조합(S_i)을 구성한다(Eq. (3)).

$$S_i = R_i \cdot Y_i = \{r_{1,i} \times y_{1,i}, r_{2,i} \times y_{2,i}, \dots, r_{j,i} \times y_{j,i}, \dots, r_{n,i} \times y_{n,i}\} \quad (3)$$

네 번째로 30년 월평균 강수량을 기준으로 각 산정된 i 번째 모의된 일 강수량의 합($\sum S_i$)이 작을 경우($\sum S_i < \mu_p \times (1-r_{mm})$), S_{BN} 으로 분류하고, 비슷할 경우($\mu_p \times (1-r_{mm}) \leq \sum S_i \leq \mu_p \times (1+r_{mm})$), S_{NN} , 클 경우($\sum S_i > \mu_p \times (1+r_{mm})$), S_{AN} 으로 각각 분류

Table 1. Definition of the 3×3 contingency table for categorical forecast

Forecast \ Observation	BN	NN	AN	SUM
BN	a	b	c	A
NN	d	e	f	B
AN	g	h	i	C
SUM	X	Y	Z	N

하고 추출하고자 하는 모의 강수량의 총 개수(k)를 범주형 확률장기예보의 예보확률(P_{BN}, P_{NN}, P_{AN}) 비율만큼 각각 추출하여 최종적인 모의 일 강수량 조합($\hat{S}_i, i = 1, 2, \dots, k$)을 구성한다. 예를 들어 $k = 10,000$ 이고 P_{BN}, P_{NN}, P_{AN} 의 예보확률이 각각 50%, 30%, 20%라면 S_{BN} 은 5,000개($k \times P_{BN}$), S_{NN} 은 3,000개($k \times P_{NN}$), S_{AN} 은 2,000개($k \times P_{AN}$)을 추출하여 \hat{S}_i 을 구성한다.

2.3 적합도 평가방법

예측 강수량 결과의 적합도 평가를 위해 본 연구에서는 평균절대오차(Mean absolute error, MAE), 평균제곱근오차(Root mean square error, $RMSE$), 효율성계수(Coefficient of efficiency, CoE), 일치도지수(index of agreement, IoA), 상관계수(Pearson correlation coefficient, R), 결정계수(Coefficient of determination, R^2) 등 총 6가지 기법을 이용하여 분석을 수행하였다. 여기서, 평균절대오차와 평균제곱근오차는 작을수록, 효율성계수, 일치도지수, 상관계수, 결정계수는 1에 가까울수록 적합도가 높은 것을 의미한다. 적합도 평가방법의 주요 식은 Eqs. (4)~(7)과 같다.

$$MAE = \frac{1}{N} \sum_{i=1}^N |S_i - O_i| \tag{4}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - O_i)^2} \tag{5}$$

$$CoE = 1 - \frac{\sum_{i=1}^N (S_i - O_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \tag{6}$$

$$IoA = 1 - \frac{\sum_{i=1}^N (S_i - O_i)^2}{\sum_{i=1}^N (|S_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{7}$$

여기서, N 은 자료 개수, O 및 \bar{O} 는 관측 값 및 관측 평균 값, S 는 모의 값을 의미한다.

3. 대상유역 및 자료

본 연구에서는 범주형 확률장기예보를 활용한 예측 강수

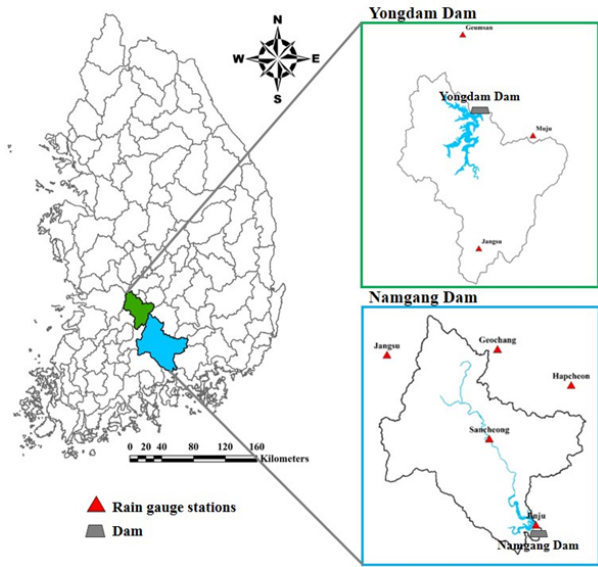


Fig. 1. Location of study areas (watersheds) and rain gauge stations

량을 추정하고 이를 수자원 운영 및 관리에 활용 가능한지를 평가하고자 우리나라 이·치수관리에 큰 역할을 하고 있는 남강댐 및 용담댐 유역을 분석 대상으로 선정하였다(Fig. 1). 관측 강수량은 Thiessen 가중법을 통해 산정하였으며 강우관측소는 남강댐 유역의 경우 진주, 거창, 합천, 산청, 장수지점, 용담댐 유역의 경우 금산, 무주, 장수지점을 선정하였으나 이중 장수지점의 경우 1988년부터 강우관측이 이루어져 평년기간(1981~2010년)을 포함하지 못함에 따라 제외하였다.

범주형 확률장기예보는 기상변수(강수량, 온도)를 평년대비 높음, 비슷, 적음으로 단정하여 예보하는 방식이 아닌 평년대비 많은 강수가 발생할 확률, 비슷할 확률, 적을 확률과 같이 발생 가능성에 대하여 확률론적으로 예보하는 방식이다. 확률론적 예보는 미래의 불확실성이 포함된 예측정보를 정량적인 확률 값으로 제공함에 따라 이상기후에 대한 대응정책 수립 및 수자원 분야의 장기계획 수립 등에 다양한 의사결정을 할 수 있는 장점이 있다. 기후변화로 인해 우리가 경험하지 못한 극치사상(홍수 및 가뭄 등)의 빈도 및 강도가 증가함에 따라 장기예보의 정확도 향상 및 실용성 등이 요구되고 있는 상황에서 과거 운영된 단정적인 예보보다는 확률적인 예보가 탄력적 대응을 가능하게 할 것이다. 이에 기상청은 확률론적 예보를 2014년 6월부터 매주 1개월 예보와 매달 3개월 예보를 전국 단위 혹은 시·도 단위로 예보하고 있다. 확률장기예보에 포함되는 주요 정보로는 각 시·도별 30년(1981~2010년)간의 기후 평균 값, 평년 비슷 범위, 3분위(높음, 비슷, 적음) 확률 값을 포함한다.

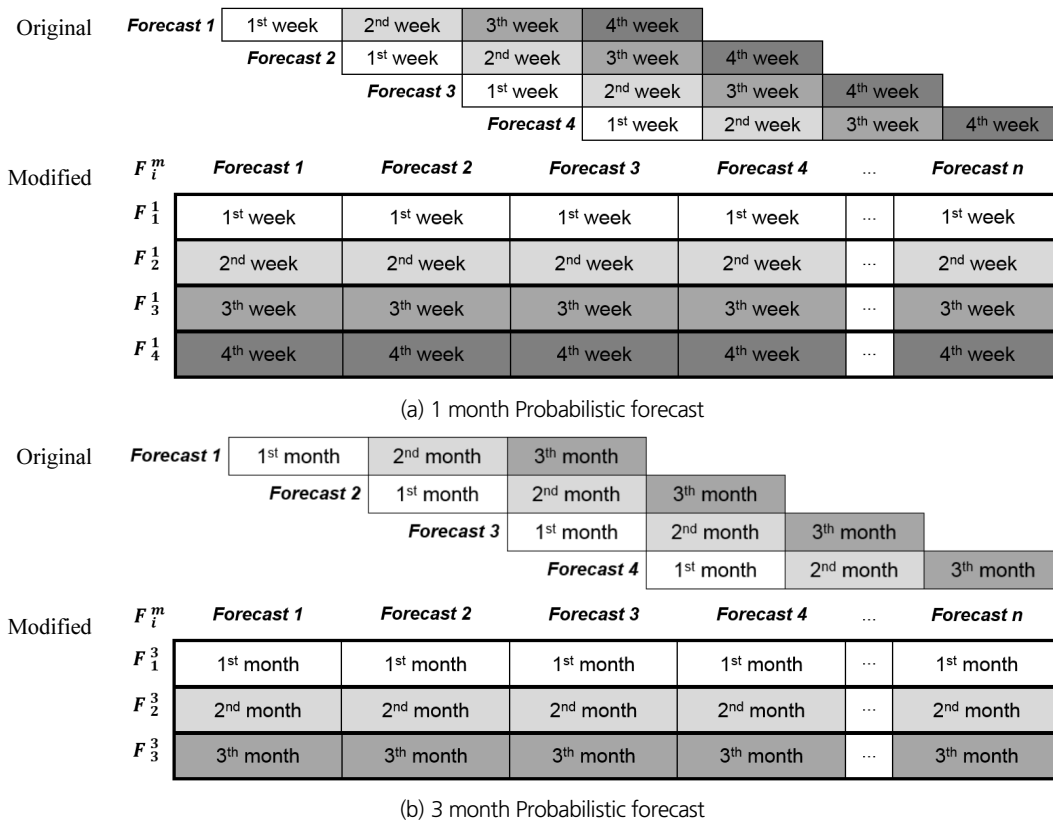


Fig. 2. Time series of original and modified probabilistic forecasts for (a) 1-month and (b) 3-month

본 연구에서는 범주형 확률장기예보의 예측력 검증을 위해 1개월 및 3개월 확률장기예보자료를 모두 검증하였으며, 월별 예측 강수량 추정의 경우는 3개월 확률장기예보 자료를 활용하여 분석하였다. Lead time별 확률예보의 예측력 검증을 위해 Fig. 2와 같이 동일한 lead time을 갖는 자료를 구분하여 시계열을 재구성하였다. Fig. 2의 F_i^m 는 예보확률을 의미하고, m 은 예보문의 종류를 의미하는 것으로 1개월 확률예보($m=1$), 3개월 확률예보($m=3$)를 나타낸다. i 는 lead time을 나타내는 것으로 1개월 확률예보일 때는 주 단위($i \leq 4$)이고, 3개월 확률예보일 때는 월 단위($i \leq 3$)를 갖는다.

4. 결과

4.1 범주형 확률장기예보의 예측력 검증

용담댐 유역을 포함하는 대전·세종·충청남도지역 1개월 확률예보의 F_1^1 (1개월 확률예보의 첫 번째 lead time)부터 F_4^1 (1개월 확률예보의 네 번째 lead time)의 Hit Rate를 산정한 결과, F_1^1 에서 최대 38%의 예측력을 보였으며 이후 F_2^1 는

33%, F_3^1 은 30%, F_4^1 는 32%의 적중률을 보여 전반적으로 lead time이 커질수록 Hite Rate가 작아지는 특성을 보였다.

확률예보의 편향 정도를 분석한 Reliability Diagram 결과는 Fig. 3과 같다. Fig. 3의 막대그래프는 확률구간별 예측된 경우의 수를 백분율로 표현한 것이며, Perfect Reliability(적색 대각선)상에 데이터가 존재하면 예측력이 완벽하다는 것을 나타내고, no skill line(적색 대각선)상에 존재하면 예측력이 없음을 의미한다. 대전·세종·충청남도 1개월 확률예보에서는 Perfect Reliability인 대각선에 가장 가까운 AN의 예측 성능(Figs. 3(c), 3(f), 3(i), and 3(l))이 가장 우수한 것으로 분석되었다. 범주별 특징을 살펴보면, BN은 Perfect Reliability 선 위에 데이터들이 존재하므로 과소예측하고 있음을 알 수 있고(Figs. 3(a), 3(d), 3(g), and 3(j)), NN의 경우 대부분 Perfect Reliability선 아래에 데이터가 존재하므로 과대예측(Figs. 3(b), 3(e), 3(h), and 3(k))하고 있다.

Fig. 4는 대전·세종·충청남도지역 1개월 확률예보의 ROC 검증 결과를 나타내고 있다. 왼쪽 상단 쪽으로 갈수록 예측성능이 우수한 것으로 판단하므로 ROC 검증 방법을 바탕으로 해석하면 Figs. 4(a) and 4(d)와 같이 F_1^1 과 F_2^1 는 BN의

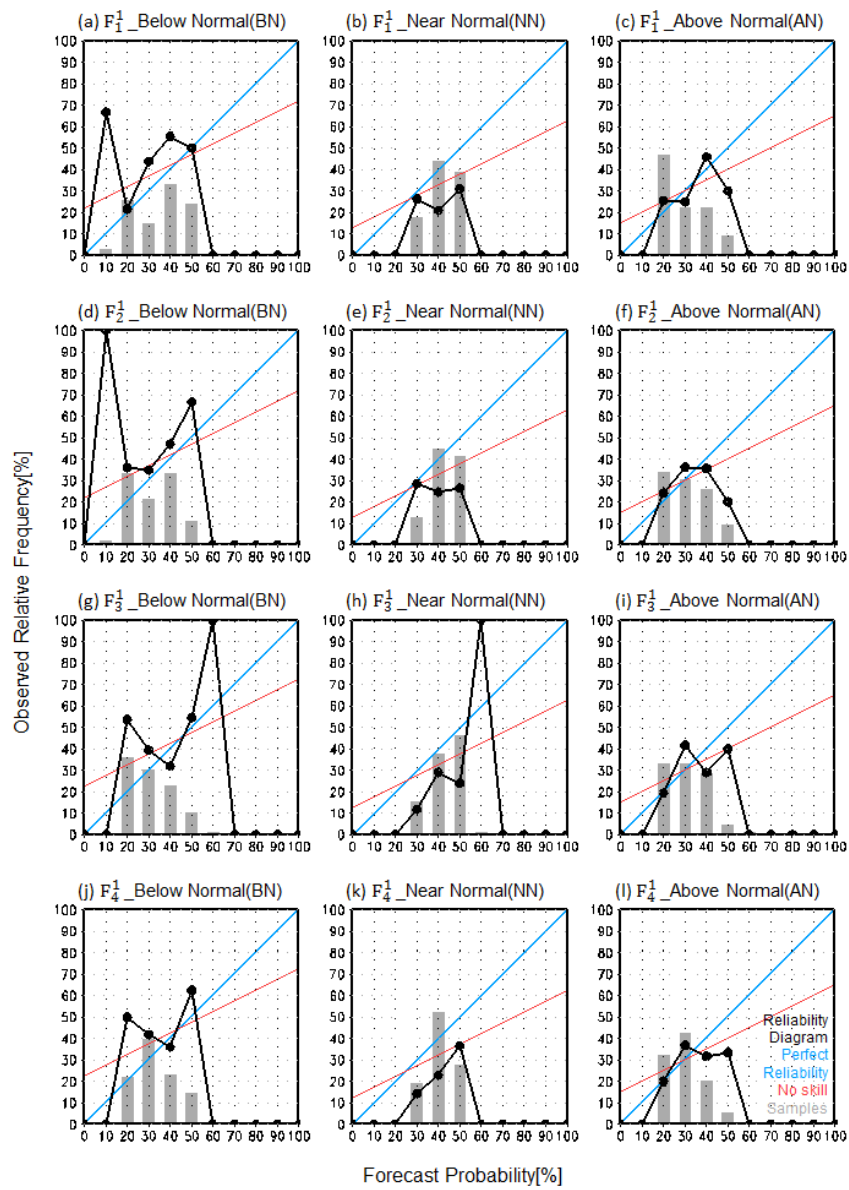


Fig. 3. Reliability diagrams of 1-month probabilistic forecasts for Daejeon·Sejong·Chungcheongman-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3, (j)~(l) leadtime = 4)

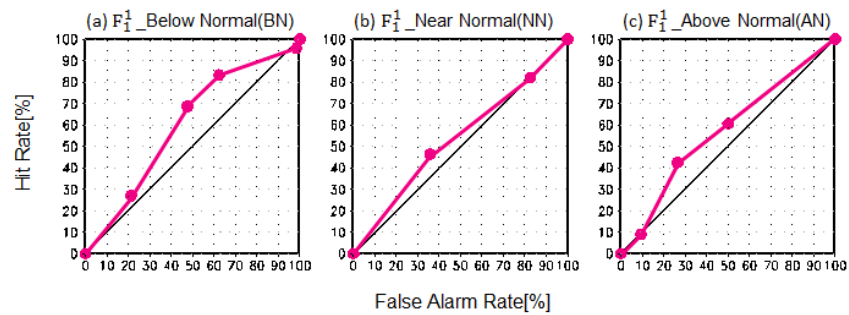


Fig. 4. Relative operating curves of 1-month probabilistic forecasts for Daejeon·Sejong·Chungcheongman-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3, (j)~(l) leadtime = 4)

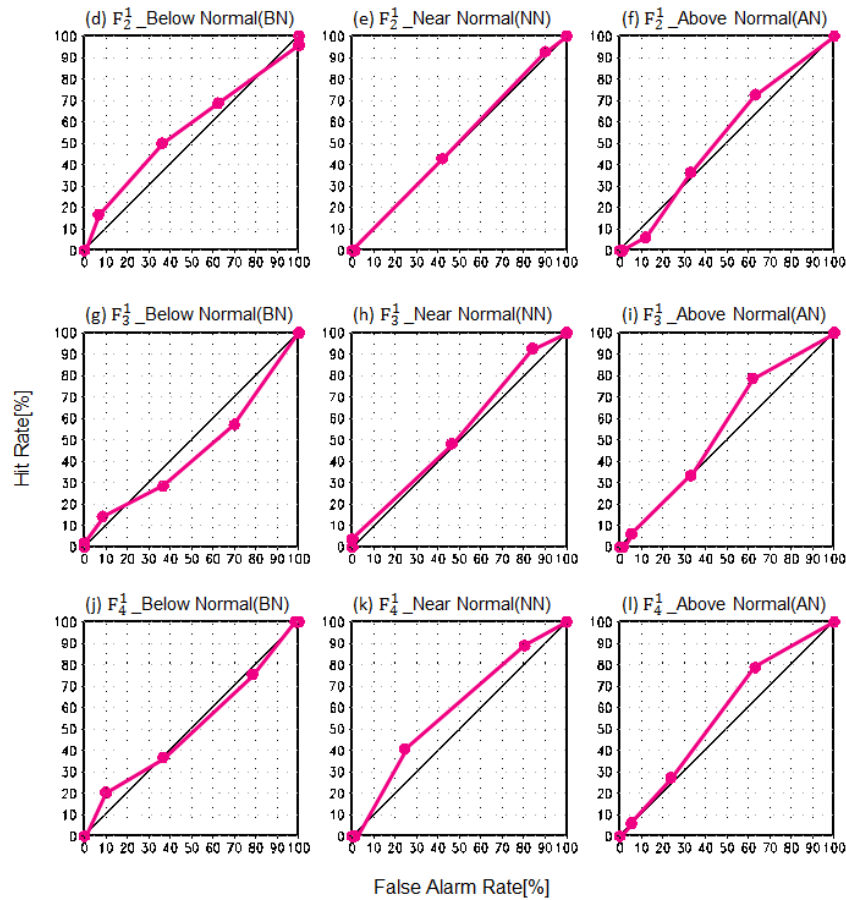


Fig. 4. Relative operating curves of 1-month probabilistic forecasts for Daejeon · Sejong · Chungcheongman-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3, (j)~(l) leadtime = 4) (Continue)

예측성능이 가장 우수하였으나, F_3^1 과 F_4^1 는 AN과 NN의 성능이 BN보다 우수함을 알 수 있다. 또한, Fig. 4를 보면, lead time이 커질수록 예측성능이 저하되는 것으로 분석되었다.

대전 · 세종 · 충청남도지역 3개월 확률예보의 F_1^3 부터 F_3^3 의 Hit Rate를 구한 결과, F_1^3 에서 최대 48%의 예측력을 보였으며 F_2^3 와 F_3^3 는 36%의 예측력을 보였다. 이처럼 3개월 확률예보 역시 lead time이 짧을수록 비교적 높은 적중률을 나타냈으며 1개월 확률예보와 비교하여 높은 Hit Rate를 갖는 것을 확인할 수 있었다.

Reliability Diagram 분석결과(Fig. 5), F_1^3 의 경우 AN이 가장 우수한 예측력을 보였으나 lead time이 길어질수록 BN이 가장 우수한 예측성능을 보였다. 각 범주별 특징을 살펴보면 1개월 확률예보와 마찬가지로 BN은 관측보다 과소예측(Figs. 5(a), 5(d), and 5(g))하고 있고, NN은 대부분 과대예측(Figs. 5(b), 5(e), and 5(h))하고 있으며, AN은 과소예측(Figs. 5(c), 5(f), and 5(i))하는 특징을 보였다. ROC 검증결과(Fig. 6),

Fig. 6(b)와 같이 F_1^3 은 NN의 예측성능이 가장 우수한 것으로 분석된 반면, F_2^3 와 F_3^3 에서는 AN이 가장 높은 예측력(Figs. 6(f) and 6(i))을 보였다.

남강댐 유역을 포함한 부산 · 울산 · 경상남도지역 1개월 확률예보의 F_1^1 부터 F_4^1 의 Hit Rate를 산정한 결과, F_1^1 은 41%, F_2^1 는 31%, F_3^1 은 33%, F_4^1 는 31%로 약 30~40%의 예측력을 보였으며 대전 · 세종 · 충청남도지역 결과들과 마찬가지로 lead time이 길수록 Hit Rate가 작아지는 것으로 분석되었다. Fig. 7은 Reliability Diagram 결과로 Perfect Reliability와 가장 유사한 AN의 예측성능(Figs. 7(c), 7(f), 7(i), and 7(l))이 가장 우수하게 나타났다. ROC 검증결과 대전 · 세종 · 충청남도지역의 결과와 같이 F_1^1 과 F_2^1 는 BN의 예측성능(Figs. 8(a) and 8(d))이 가장 우수한 반면 F_3^1 과 F_4^1 의 경우 AN의 예측성능(Figs. 8(i) and 8(l))이 우수한 것으로 분석되었다. 특히 대부분 기간에 대해 NN의 예측 성능은 50% 미만으로 BN과 AN에 비해 예측성이 낮은 것으로 분석되었다.

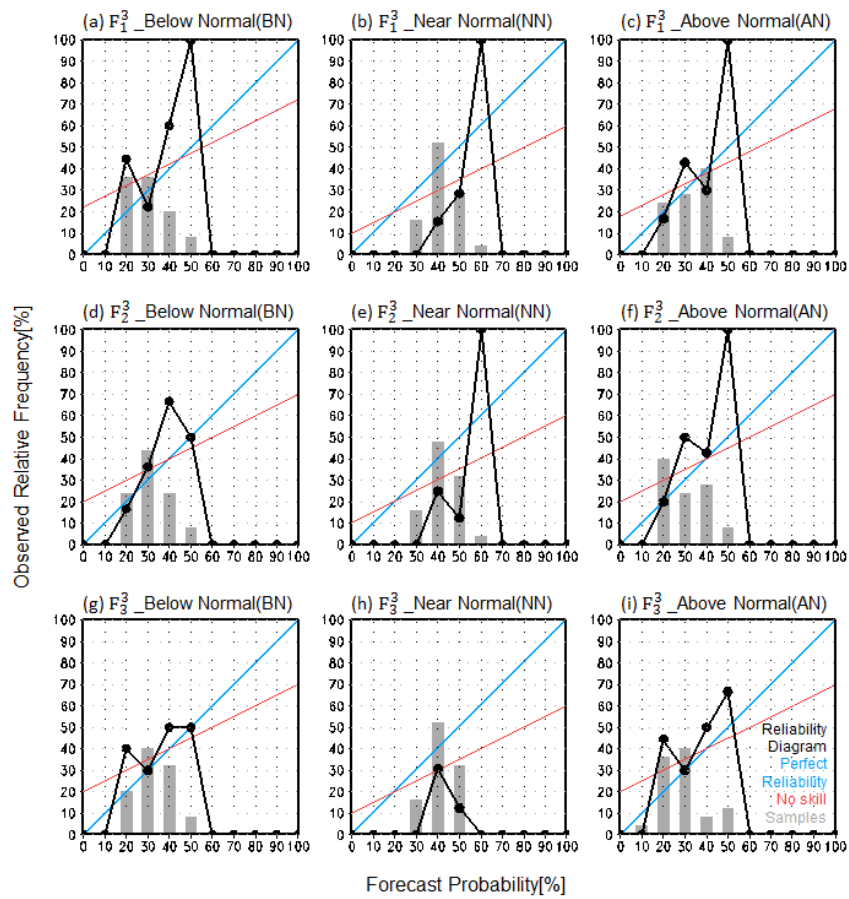


Fig. 5. Reliability diagrams of 3-month probabilistic forecasts for Daejeon·Sejong·Chungcheongman-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3)

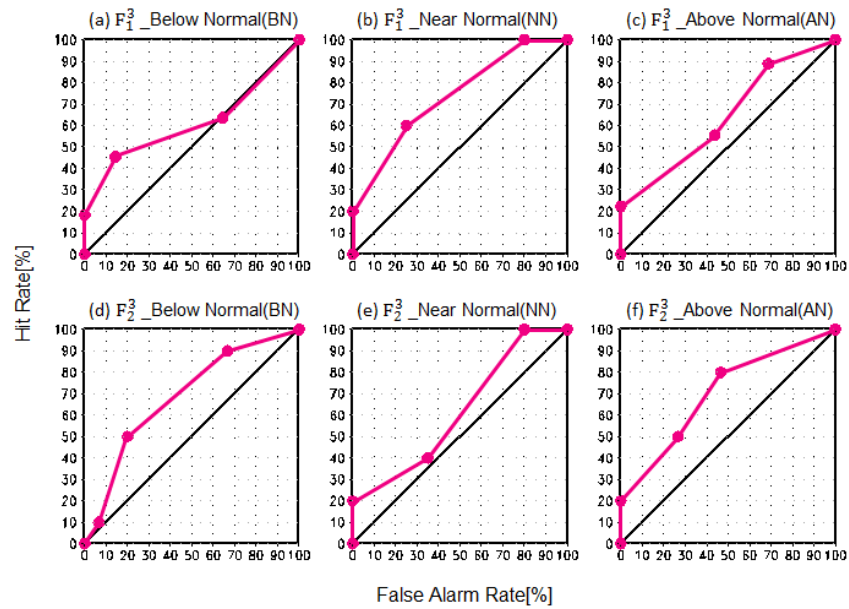


Fig. 6. Relative operating curves of 3-month probabilistic forecasts for Daejeon·Sejong·Chungcheongman-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3)

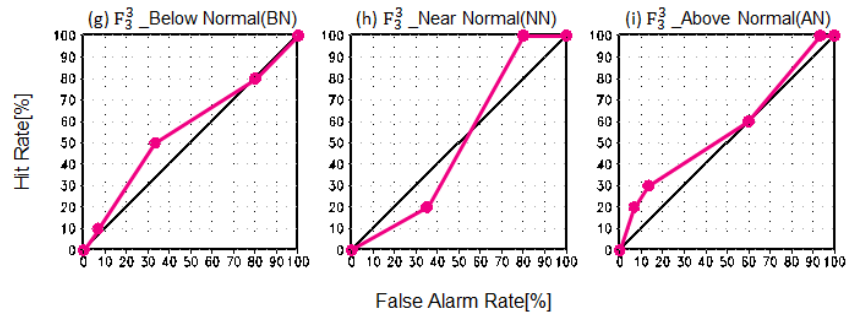


Fig. 6. Relative operating curves of 3-month probabilistic forecasts for Daejeon · Sejong · Chungcheongman-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3) (Continue)

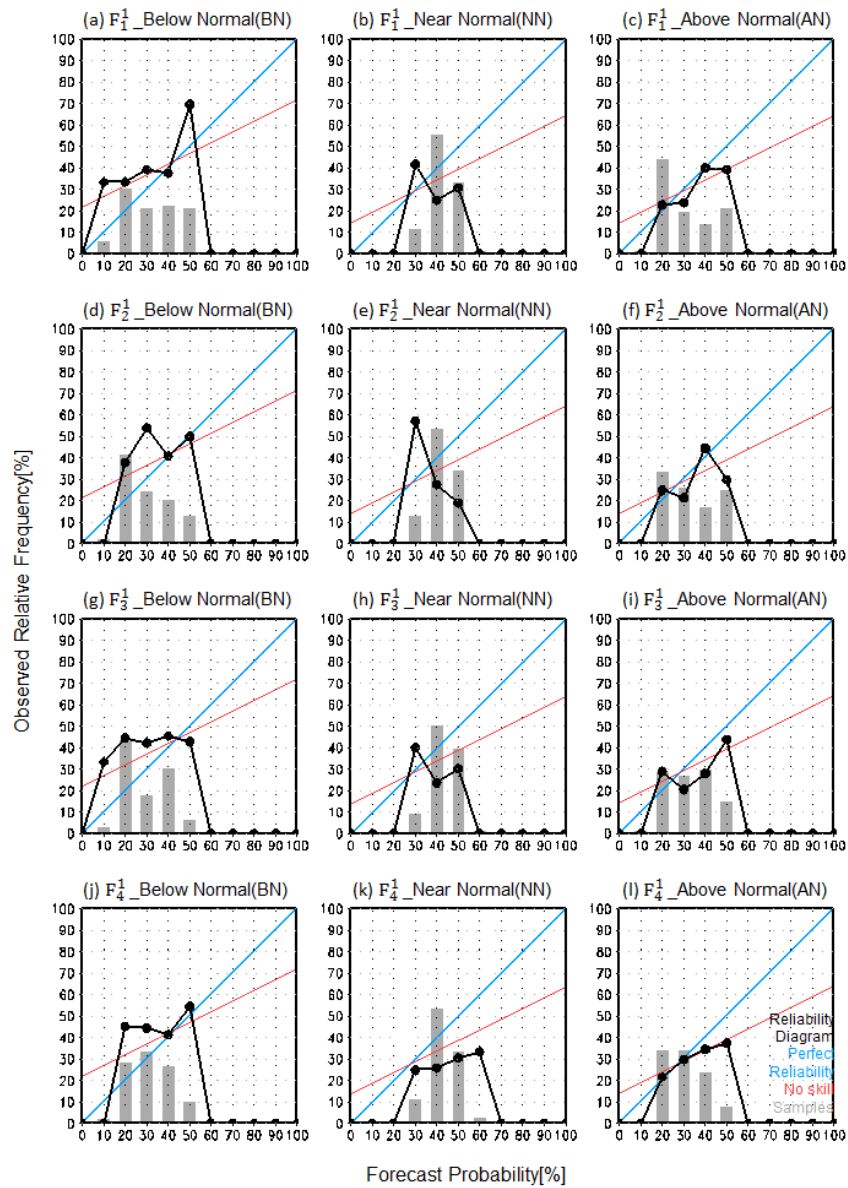


Fig. 7. Reliability diagrams of 1-month probabilistic forecasts for Busan · Ulsan · Gyeongsangnam-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3, (j)~(l) leadtime = 4)

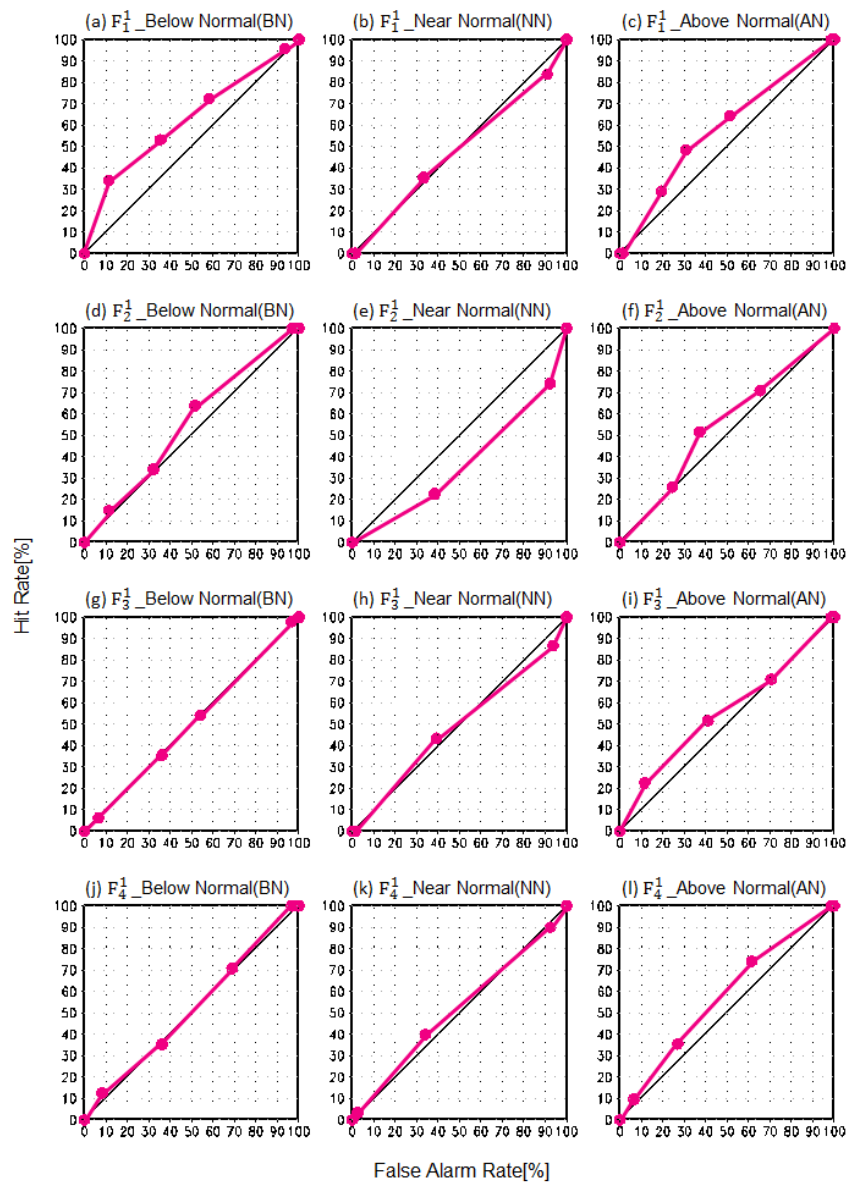


Fig. 8. Relative operating curves of 1-month probabilistic forecasts for Busan · Ulsan · Gyeongsangnam-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3, (j)~(l) leadtime = 4)

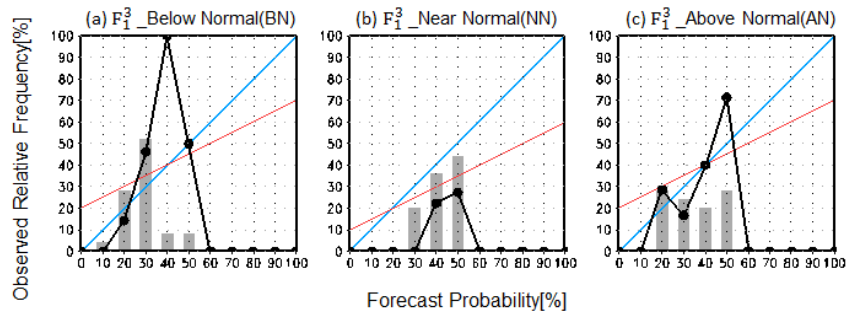


Fig. 9. Reliability diagrams of 3-month probabilistic forecasts for Busan · Ulsan · Gyeongsangnam-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3)

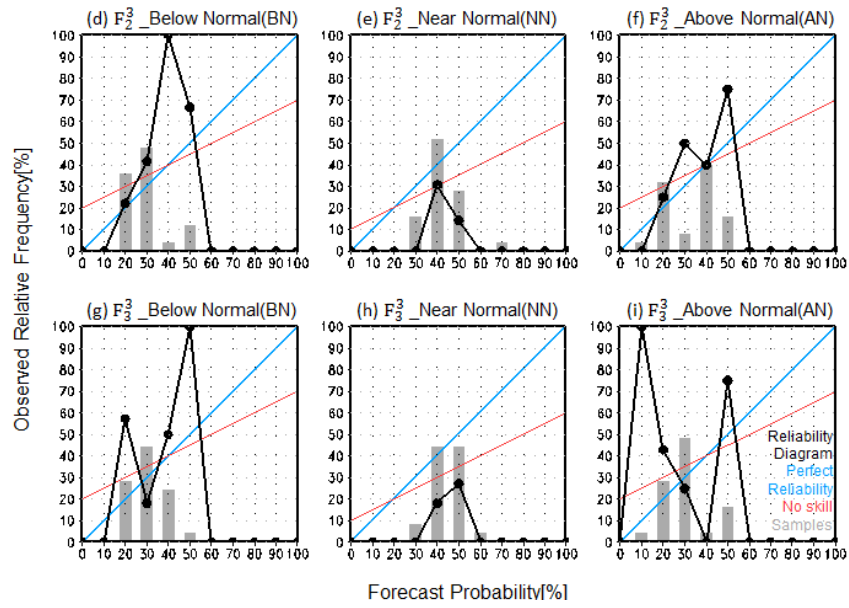


Fig. 9. Reliability diagrams of 3-month probabilistic forecasts for Busan · Ulsan · Gyeongsangnam-do ((a)~(c) leadtime = 1, (d)~(f) eadtime = 2, (g)~(i) leadtime = 3) (Continue)

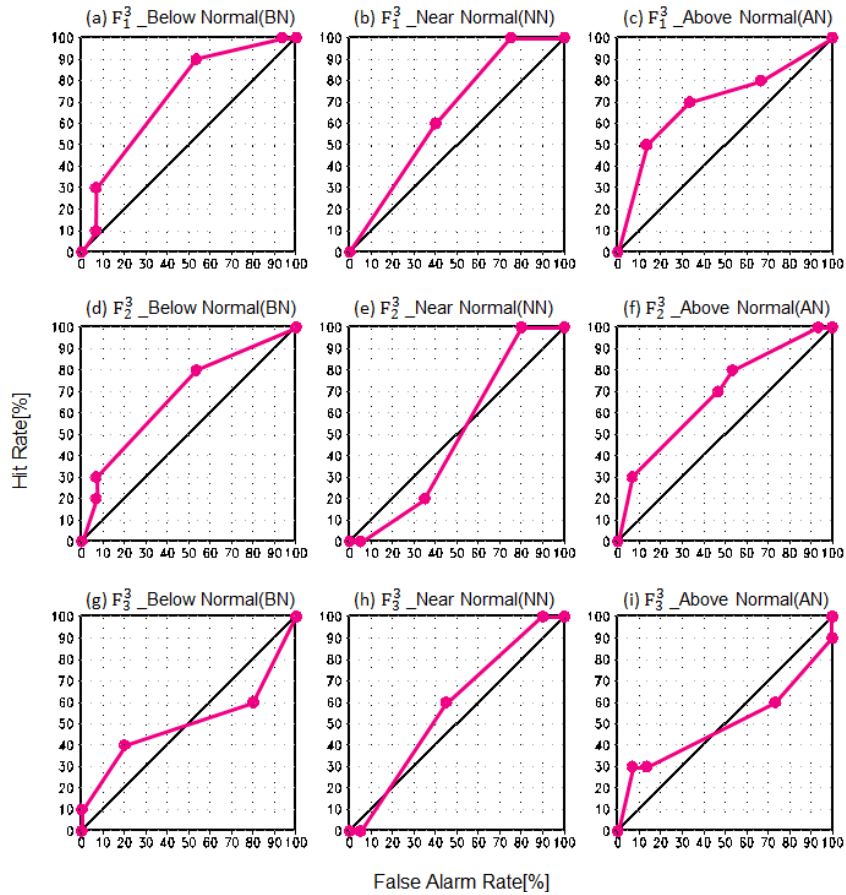


Fig. 10. Relative operating curves of 3-month probabilistic forecasts for Busan · Ulsan · Gyeongsangnam-do ((a)~(c) leadtime = 1, (d)~(f) leadtime = 2, (g)~(i) leadtime = 3)

부산·울산·경상남도지역 3개월 확률예보의 Hit Rate를 구한 결과, F_1^3 은 40%, F_2^3 는 44%, F_3^3 은 44%로 1개월 확률예보 대비 높은 Hit Rate를 갖는 것으로 분석되었다. Fig. 9는 Reliability Diagram 분석 결과를 정리한 것으로 Fig. 9(c)와 (f)처럼 F_1^3 과 F_2^3 의 AN이 가장 높은 예측력을 보였으며 F_3^3 은 NN의 예측력이 가장 높게 나타났다. ROC 검증결과(Fig. 10), F_1^3 과 F_2^3 는 BN의 예측성능(Figs. 10(a) and 10(d))이 가장 높게 나타난 반면 F_3^3 의 경우 NN의 예측성능(Fig. 10(h))이 가장 우수한 것으로 분석되었다.

4.2 범주형 확률장기예보를 활용한 예측 강수량 산정

본 연구에서 제안한 범주형 확률장기예보를 활용한 예측 강수량 산정기법이 정량적으로 예측 강수량을 적절히 추출할

수 있는지 검토하기 위해 2014년 6월부터 2016년 6월까지 각 월별 관측된 강수의 범주(높음, 비슷, 낮음)를 모두 정확하게 예보했다고 가정하였다. 예를 들어 2014년 6월의 관측된 실제 강우량이 6월의 평년 강우량 대비 높았다면 각 범주별 확률 P_{BN} , P_{NN} , P_{AN} 은 각각 0%, 0%, 100%로 가정되어 예측 강수량 추출을 위한 범주별 예보 확률로 활용하게 된다.

Fig. 11은 2014년 6월부터 2016년 6월까지 100%로 가정된 확률 값을 적용하여 남강댐 유역 인근 강우관측소(Figs. 11(a)~11(d))와 용담댐 유역 인근 강우관측소(Figs. 11(e)~11(f))의 예측 강수량 평균 및 신뢰구간(95%, 99%)을 추정한 결과와 실제 관측된 월 강수량을 비교한 그림으로 분석결과 모든 강우 관측소에서 실제 관측된 강수가 대부분 모의된 예측 강수량 상사블의 신뢰구간에 포함되고 강수의 전반적인 계절 패턴을 잘 유지하고 있음을 확인할 수 있으며 정량적으로 관측

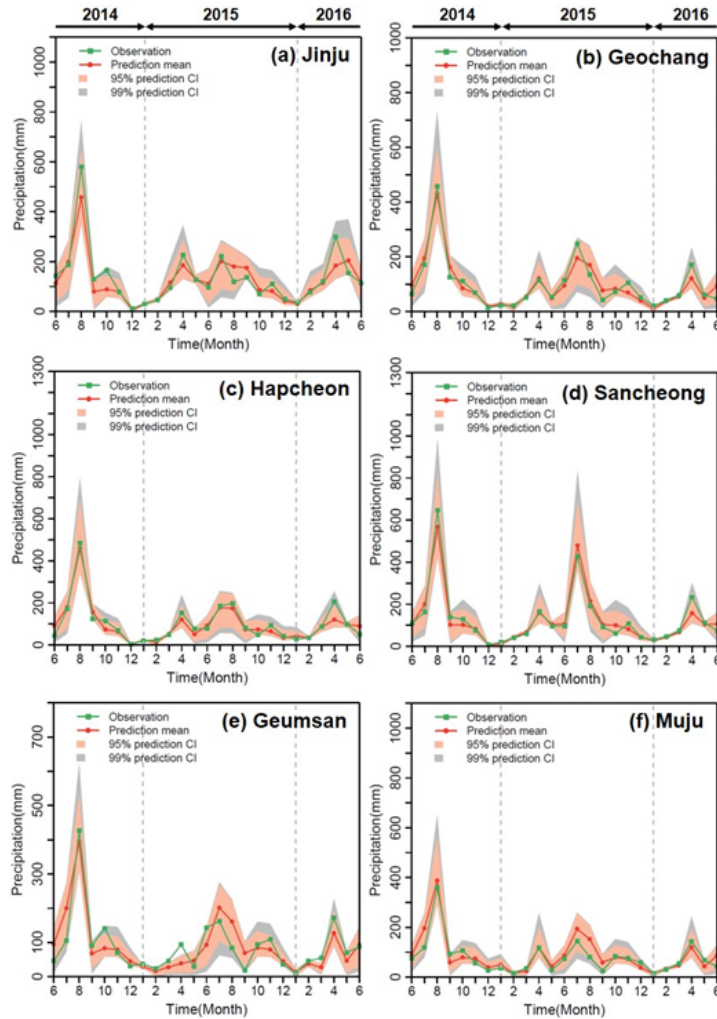


Fig. 11. Observed and simulated (under perfect forecast) time series of monthly precipitation with 95% and 99% confidence intervals for two study watersheds (Namgang Dam: (a)~(d), Yongdam Dam: (e) and (f)) from June 2014 to June 2016

강수량과 예측 강수량의 평균이 유사한 결과를 보이고 있었다. 특히, 여름철(6~8월)의 경우 상대적으로 관측 월 강수량 대비 예측이 많은 차이를 보였으며 불확실성 역시 큰 것으로 나타났다.

적합도 평가를 위해 평균절대오차(MAE), 평균제곱근오차(RMSE), 효율성계수(CoE), 일치도지수(IoA), 상관계수(R), 결정계수(R²) 등 총 6가지 기법을 이용하였다. 평균 예측 강수량과 관측값과의 적합도 평가 결과(Table 2), 남강댐

Table 2. Statistical results comparing observed and simulated (ensemble mean under perfect forecast) monthly precipitation for two study watersheds from June 2014 to June 2016

Dam	Namgang				Yongdam	
	Jinju	Geochang	Hapcheon	Sancheong	Geumsan	Muju
Efficiency measure						
Mean absolute error (MAE, mm)	29.59	20.26	19.59	20.71	30.58	22.06
Root mean square error (RMSE, mm)	44.69	25.99	27.46	30.11	38.65	29.79
Coefficient of efficiency (CoE)	0.75	0.91	0.9	0.94	0.77	0.85
Index of agreement (IoA)	0.95	0.98	0.98	0.99	0.94	0.96
Pearson Correlation coefficient (R)	0.93	0.96	0.96	0.98	0.89	0.93
Coefficient of determination (R ²)	0.87	0.92	0.92	0.95	0.79	0.87

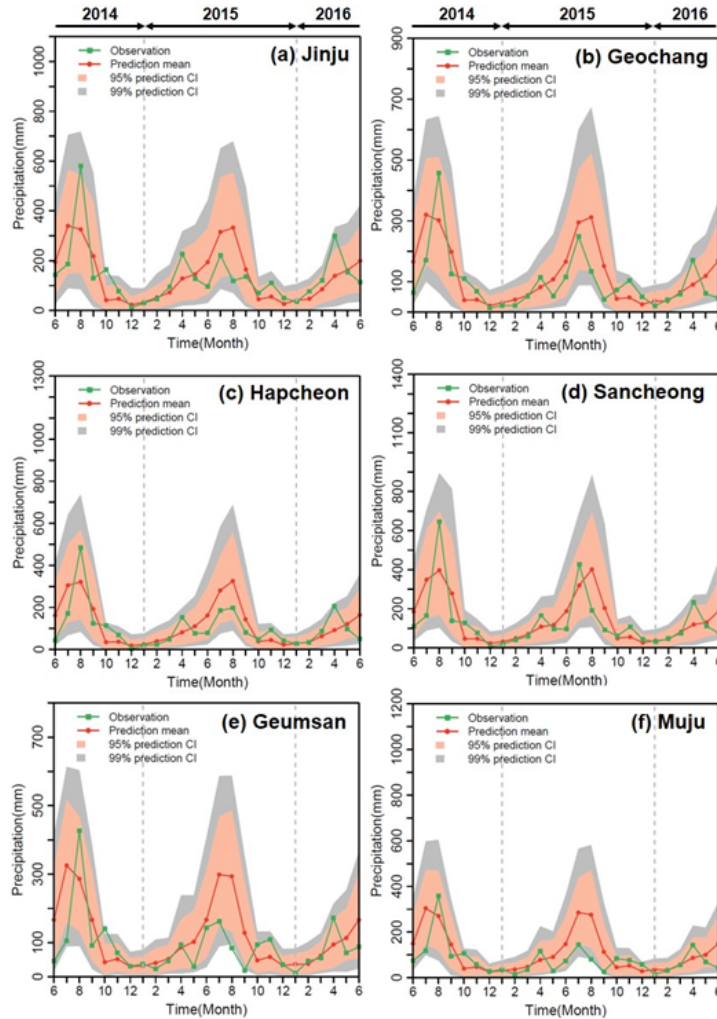


Fig. 12. Observed and simulated (using actual forecast probabilities) time series of monthly precipitation with 95% and 99% confidence intervals for two study watersheds (Namgang Dam: (a)~(d), Yongdam Dam: (e) and (f)) from June 2014 to June 2016

Table 3. Statistical results comparing observed and simulated (ensemble mean using actual forecast probabilities) monthly precipitation for two study watersheds from June 2014 to June 2016

Dam	Namgang				Yongdam	
	Jinju	Geochang	Hapcheon	Sancheong	Geumsan	Muju
Efficiency measure						
Mean absolute error (<i>MAE</i> , mm)	68.90	59.46	58.73	70.12	64.07	57.10
Root mean square error (<i>RMSE</i> , mm)	95.99	77.62	75.95	97.99	89.07	77.67
Coefficient of efficiency (<i>CoE</i>)	0.13	0.36	0.40	0.31	0.07	0.20
Index of agreement (<i>IoA</i>)	0.76	0.81	0.83	0.83	0.71	0.72
Pearson Correlation coefficient (<i>R</i>)	0.60	0.69	0.71	0.71	0.53	0.59
Coefficient of determination (<i>R</i> ²)	0.36	0.47	0.51	0.51	0.28	0.35

유역의 경우 평균절대오차는 합천지점에서 가장 낮게 산정되었으며 진주지점에서 가장 높게 산정되었다. 평균제곱근오차의 경우 거창지점에서 가장 낮게 산정되었으며 진주지점에서 가장 높게 산정되었다. 효율성계수, 결정계수의 경우 진주지점을 제외하고는 0.9 이상으로 산정되어 상당히 높은 적합도를 보였으며 일치도지수와 상관계수는 모든 지점에서 0.9 이상으로 분석되어 상대적으로 관측 자료와 유사한 결과도 출됨을 확인할 수 있었다. 용담댐 유역의 경우 무주지점에서 가장 낮은 평균절대오차와 평균제곱근오차를 보였으며 효율성계수, 일치도지수, 상관계수, 결정계수 역시 무주지점에서 가장 높게 산정되었다. 따라서, 본 연구에서 제안하는 범주형 확률장기예보를 활용한 예측 강수량 산출기법은 전체적으로 관측 자료와 매우 높은 적합도를 보임에 따라 충분히 활용 가능할 것으로 판단된다.

범주형 확률장기예보에서 제공하는 실제 예보 확률 값과 평년기간의 강수량을 활용하여 예측 강수량을 추정하였다. Fig. 12는 2014년 6월부터 2016년 6월까지 실제 예보된 확률 값을 활용하여 남강댐(Figs. 12(a)~12(d))과 용담댐(Figs. 12(e)~12(f))유역의 예측 강수량의 평균 및 신뢰구간(95%, 99%)을 추정한 결과를 실제 관측된 월 강우와 비교한 그림이다. 분석결과, 실제 관측된 월강수량과 추정된 예측 강수량의 양상별 평균값과 정량적인 차이를 보이고 있으나 전반적인 계절 강수패턴을 비교적 잘 유지하는 것을 확인할 수 있었다. 이 정량적인 차이는 확률장기예보의 실패로 인해 실제 관측된 강우의 범주(예, 높음)에 해당하는 양상별을 적게 추출함에 따라 과소 및 과대 산정을 야기한 것으로 판단된다. 특히, 여름철(6~8월)의 경우 상대적으로 관측 월강수량 대비 많은 차이를 보였으며 불확실성 역시 큰 것으로 나타났다. 실제 2014~2015년 극심한 가뭄이 발생한 해였음에도 불구하고 관측 강수량이 본 연구에서 제안하는 예측 강수량의 범위 내에 대부분 포함되고 있어 충분히 불확실성을 고려한 수자원 장기 이 수계획 등에 활용 가능할 것으로 보인다.

실제 예보된 확률을 이용하여 모의된 평균 예측 강수량 양상별과 관측 치와의 적합도 평가 결과(Table 3), 남강댐 유역의 경우 합천지점, 용담댐 유역의 경우 무주지점에서 평균절대오차와 평균제곱근오차가 가장 낮게 산정되었으며 효율성계수, 일치도지수, 상관계수, 결정계수는 가장 높게 산정되었다. 산정지점의 경우 상대적으로 큰 평균절대오차와 평균제곱근오차를 보이지만 효율성계수, 일치도지수, 상관계수, 결정계수는 상대적으로 크게 산정되어 관측 자료의 패턴을 비교적 잘 유지하는 것을 확인할 수 있었다.

5. 요약 및 결론

본 연구는 기상청에서 공식적으로 예보하고 있는 1개월 및 3개월 범주형 확률장기예보에 대하여 예측력을 검증하고 범주형 확률장기예보를 댐 저수지 운영계획 등 수자원 운영 및 관리 분야에 활용하기 위하여 범주형 확률장기예보에서 제시하는 각 범주별 예보확률을 이용한 강수량 추정기법을 제안하였다. 용담댐 및 남강댐유역에 대한 확률장기예보의 예측력 검증결과, 전체적으로 약 30~40%의 예측력을 보였는데 특히, 용담댐유역 대전·세종·충청남도지역의 3개월 예보의 경우는 최대 48%의 예측력을 보였다. 또한, 전반적으로 lead time이 길어질수록 예측력이 떨어짐을 확인할 수 있었다.

본 연구에서 제안한 범주형 확률장기예보를 활용한 정량적 강수량 추정기법에 대한 적합성 검토를 위해 강수 범주를 정확하게 예보했다고 가정하여 추출한 결과 모든 관측소에서 실제 관측된 강수가 대부분 모의된 예측 강수량 양상별의 신뢰구간에 포함되고 0.9 이상의 상관계수를 보임에 따라 본 연구에서 제안된 예측 강수량 추정기법은 충분히 활용 가능할 것으로 판단하였다. 또한, 실제 예보 확률을 이용하여 강수량을 추정한 결과, 관측 월강수량과 모의된 예측 강수량의 평균값이 정량적인 차이를 보이나 전반적인 계절 강수패턴을 비교적

잘 유지하는 것을 확인할 수 있었다. 정량적인 차이는 실제 관측된 강우의 범주에 해당하는 양상분을 적게 추출함에 따라 과소 및 과대 산정을 야기한 것으로 판단된다. 그러나 2014~2015년 극심한 가뭄이 발생한 해임을 감안할 때, 본 연구에서 산출한 예측 강수량의 범위 내에 실측이 대부분 포함되어 있는 것으로 보아 산출방법이 불확실성을 충분히 고려하고 있음을 알 수 있다.

범주형 확률장기예보는 최대 3개월까지 제공됨에 따라 실제 장기이수계획(10월~6월)에 활용하기에는 현실적으로 부족한 자료일 수 있으나 3개월 이하의 물 관리는 확률예보 자료 기반, 3개월 이상은 다수의 계절예측 모델을 활용한 MME (multi-model ensemble) 결과 혹은 기존 물 관리에 활용하고 있는 갈수빈도(20년빈도) 유입량을 이용하는 것도 하나의 방안이 될 수 있을 것으로 판단된다. 향후 확률예보의 정확도 및 예측기간의 개선이 이루어진다면 장기 예측정보를 활용한 수자원의 효율적인 이용과 불확실한 강수량 자료에 대한 의사결정을 지원할 수 있게 함으로써 이수 및 치수측면의 수자원 운영 및 관리에 크게 활용될 수 있을 것으로 기대된다.

감사의 글

본 연구는 한국기상산업진흥원의 “기상See-AT기술개발” 사업의 연구비지원으로 수행한 ‘확률장기예보 물관리 활용 기술개발 및 환류 체계 구축(KMIPA 2015-2120)’ 과제에 의해 수행되었습니다.

References

- Croley II, T. E. (2000). *Using meteorology probability forecasts in operational hydrology*. ASCE Press, p. 206.
- Gouweleeuw, B. T., Thielen, J., Franchello, G., De Roo, A. P. J., and Buizza, R. (2005). “Flood forecasting using medium-range probabilistic weather prediction.” *Hydrology and Earth System Sciences*, Vol. 9, No. 4, pp. 309-368.
- Kang, B., Rleu, S. Y., and Ko, I. H. (2005). “Long-term streamflow prediction for integrated real-time water management system.”, *Proceedings of the Korea Water Resources Association Conference 2005*, KWRA
- Kang, J. W. (2013). “Probabilistic forecasting of seasonal inflow to reservoir.” *Journal of Environmental Science International*, Vol. 22, No. 8, pp. 965-977.
- Kim, J. C., Kim, J., and Lee, S. J. (2011). “Improvement of mid/long-term ESP scheme using probabilistic weather forecasting.” *Journal of Korea Water Resources Association*, Vol. 44, No. 10, pp. 843-851.
- KMA (Korea Meteorological Administration) (2014). *The improvement of regional long-term forecast system*. KMA Publication No. 11-136000-001095-14, KMA, pp. 65-73.
- KMA (Korea Meteorological Administration) (2016). *A forecast characteristics analysis technical note of GloSea5*. KMA, pp. 3-14.
- Persson, A., Andersson, E., and Tsonevsky, I. (2015). *User guide to ECMWF forecast products*. Ver.1.2, ECWMF, pp. 101-113.
- Simpson, H. J., Cane, M. A., Herczeg, A. L., Zebiak, S. E., and Simpson, J. H. (1993). “Annual river discharge in Southeastern Australia related to El Niño-southern oscillation forecasts of sea surface temperatures.” *Water Resources Research*, Vol. 34, No. 11, pp. 3035-3044.
- Stedinger, J. R., and Kim, Y. O. (2010). “Probabilities for ensemble forecast reflecting climate information.” *Journal of Hydrology*, Vol. 391, pp. 9-23.
- Yu, P. S., Yang, T. C., Kuo, C. M., and Wang, Y. T. (2014). “A stochastic approach for seasonal water-shortage probability forecasting based on seasonal weather outlook.” *Water Resources Management*, Vol. 28, No. 12, pp. 3905-3920.