

다중 선형 회귀 모델을 이용한 8월 강수량의 예측

백희정 · 권원태

1. 서 론

우리나라 강수량은 여름철에 연강수량의 50% 이상이 집중되어 내리며 여름철 총강수량의 변동은 매우 커서 매년 변화가 심하다 (허창희와 강인식, 1988). 이러한 여름철 강수량의 예측은 농업, 생산업을 비롯한 전반적인 산업 분야와 국가적인 차원에서 에너지, 수자원, 기타 원료 수급 관리 등과 같은 장기 계획 수립에 기초자료로서 필수적이나 현재 통계적인 방법에 머물고 있는 실정이다. 이 연구에서도 북반구 500 hPa 고도와 태평양 해수면 온도 자료를 사용한 통계 모델로부터 8월 강수량을 2개월 앞서 예측할 수 있는 가능성을 제시하였다.

여름철 몬순 기간 동안 연강수량의 75%가 내리는 인도에서는 통계적 장기에보의 발전이 오래 전부터 있어왔다. 인도 남서 지역의 여름철 강수량과 지난해 겨울철 히말라야 눈의 양과의 상관관계를 근거로 인도와 버마 지역의 여름철 강수량을 장기에측한 Blanford (1884)을 시작으로 Walker(1924)에 이르러서는 인도의 강수량과 전세계 기압, 온도, 강수량 등과의 상관관계를 구하여 인도 강수량을 예측하기 위한 통계 모형을 구하였다. 인도 이외의 지역에서 통계적 모델을 사용한 장기에측에 대한 연구가 많이 수행되었다 (Bell 1977, Quinn and Burt, 1970) 우리나라에서는 강인식 등(1992)과 최정부 등(1993), 오재호 등(1994)은 한반도 여름철 강수량과 500 hPa 고도 및 태평양 해수면 온도와의 지연 상관관계를 분석하여 다중선형회귀모형을 만들어 한반도 여름철 강수량 예측을 시도하였다.

이 연구에서는 한반도 8월 강수량과 500 hPa 고도 및 태평양 해수면 온도 사이의 지연(lag) 상관관계를 구하여 지연에 따라 우리나라 8월 강수량에 큰 영향을 미치는 500 hPa 고도 및 태평양 해수면 온도 공간 변동의 구조를 구하였다. 이렇게 구한 변동구조를 지수화하고 이를 다중선형 회귀 모델의 변수로 채택하여 한반도 8월 강수량을 예측할 수 있는 최적 통계모형을 구하였다. 이 모델을 이용하여 예측을 시도하였으며, 예측된 강수량과 실제 강수량을 비교하여 모델의 숙련도를 검증하였다.

기상연구소 예보연구실

2. 자료 및 분석 방법

사용된 자료는 한반도 8월 강수량, 북반구 500 hPa 고도, 태평양 해수면 온도이다. 8월 강수량자료는 1953년부터 1993년까지 장기 자료를 가지고 있는 육지 12개 지점 자료를 사용하였으며, 각 지점에 대해서 1961년부터 1990년 30년간의 기후 평균과 표준편차로 정규화한 강수량 자료를 산술 평균하여 정규화된 한반도 8월 강수량으로 정의하였다. 500 hPa 고도값은 1952년 1월부터 1992년 12월까지 미국기상예보소(NMC)에서 분석된 월평균 자료로, 분석 영역은 위도 $20^{\circ}\text{N} \sim 90^{\circ}\text{N}$ 이며, 격자 간격은 위·경도 각각 5도이다. 해수면 온도는 일본 기상청에서 분석된 1952년 1월부터 1992년 12월까지 위도 $40^{\circ}\text{S} \sim 60^{\circ}\text{N}$, 경도 $100^{\circ}\text{E} \sim 80^{\circ}\text{W}$ 사이의 위·경도 4도의 자료이다. 이 두자료는 8월 강수량과 마찬가지로 1961년부터 1990년 30년간의 기후평균과 표준편차로 정규화하여 사용하였다.

강인식 등(1992)와 최정부 등(1993), 오재호 등(1994)은 여름철 강수량을 예측하기 위해서 일정한 기간(종속기간)에 대해서 통계모형을 구축하고, 통계모형 구축 기간이 아닌 기간(독립기간)에 대해서 예측을 시도하여 예측된 값의 정확도를 검증하였다. 그러나 이 연구에서는 1973년부터 1993년 21년 8월 강수량을 예측하기 위해서 예측년도마다 최근 20년 자료를 사용하여 모델을 새로이 구축하여 총 21개 모델을 생성하였다.

한반도 8월 강수량을 예측하기 위해서는 다중회귀모델의 예보인자를 우선 선정하여야 한다. 이를 위해서 한반도 8월 강수량과 500 hPa 고도 및 태평양 해수면 온도와의 지연상관 분석을 수행하여 지연상관도에서 99%($CC=0.65$)이상인 지역을 평균한 값을 1차적으로 예보인자로 선정한다. 이렇게 선정된 예보인자들의 안정도를 분석하여 안정하지 않는 예보인자는 제거하여 2차적으로 예보인자를 재선정하여 하였다. 마지막으로 2차적으로 선정된 예보인자들을 Stepwise 통계모형을 사용하여 유의수준 $\alpha=0.01$ 으로 각 변수들간의 부분검정을 거쳐 최종적으로 다중선형회귀모델을 최적화시키는 예보인자를 선정하였다.

위에서 최종적으로 선택된 예보인자를 사용하여 8월 강수량을 예측하기 위한 다음과 같은 다중선형회귀모델을 만들었으며, 다중회귀모델의 유의성을 검증하기 위해서 검정통계량 F와 결정계수(Reduction of error) R^2 을 구하였다. 검정통계량 F가 크고 결정계수 R^2 이 1에 가까울수록 회귀모델의 유의성은 커진다.

마지막으로 다중선형회귀모델을 사용하여 1973년부터 1992년에 대해서 한반도 8월 강수량을 예측하였으며, 예측의 정확도를 Barnett and Hasselmann (1979)이 사용한 숙련도 (Skill Score)를 구하여 평가하였다.

3. 결과

최종적으로 결정된 8월 강수량에 대한 다중회귀모델의 예보인자들은 1개 ~ 3개이다. Fig. 1에 8월 강수량에 대한 예보인자로 사용된 지역과 lag를 나타내었다. 그림에서 앞의 숫자는 지연(lag)을 나타내며, 뒤의 숫자는 모델 구축 기간을 나타낸다. 예를들어 6/5372는 1953년부터 1972년 20년에 대한 8월 강수량의 회귀모델에 사용되는 6개월전 예보인자를 의미하며, 실선은 8월 강수량과 양의 상관관계, 파선은 음의 상관관계를 의미한다.

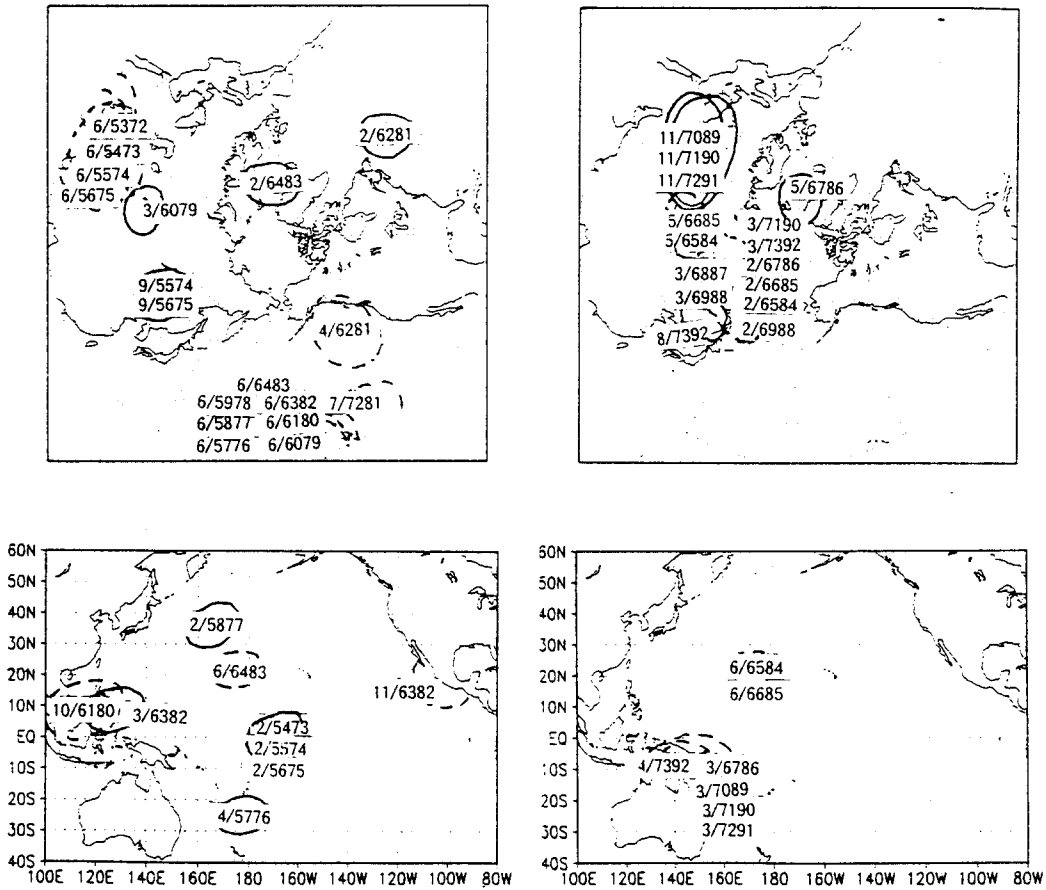


Fig. 1 Location of the predictors of the regression model for August precipitation. (a) and (b) show predictors of 500 hPa geopotential height and (c) and (d) show predictors of sea surface temperature.

500 hPa 고도 예보인자는 주로 만주, 지중해, 아열대 중태평양, 파미르 고원, 북미 서안, 북대서양, 그린랜드, 시베리아에서 나타났다. 특히 중앙아시아의 500 hPa 고도는 1953-1975년 동안 4개 회귀모델의 예보인자로 연달아 선정되었으며, 1964-1983년 이전 기간에 대해서는 8월 강수량과 아열대 중태평양 500 hPa 고도와의 상관관계가 매우 뚜렷하였으나 그 이후 기간에 대해서는 나타나지 않았고 오히려 서몽고 지역의 500 hPa 고도가 8월 강수량과 음의 상관성이 매우 뚜렷했다. 또한 8월 강수량과 지연 11개월에서 양의 상관성을 갖는 카스피해의 500 hPa 고도는 1970-1992년에 대한 회귀모델의 예보인자로 연속적으로 선정되었다. 그리고 태평양 해수면의 예보인자는 주로 적도를 중심으로 아열대 태평양에서 나타났으며, 1970년 이후 다중회귀 모델의 예보인자로 오스트레일리아 북부 해양의 해수면 온도가 사용되었다.

1953년부터 1992년 동안 총 21개의 다중회귀모델로부터 적합한 한반도 8월 강수량과 실제 관측값의 시계열(보이지는 않았음)에서 모델에 의한 값이 1950, 60년대에 적합되는 정도가 70년대 이후보다 더 낮으나 전반적으로 실제 8월 강수량에 잘 적합하고 있다. 이 모델들의 유의성을 알아보기 위해서 모형의 적합도를 나타내는 검정 통계량 F값과 결정계수 R^2 값을 Table 1에 나타내었다. 이 표에서 알 수 있듯이 1953-1972년의 회귀모델의 F값을 제외하고는 대부분 F값은 99% 신뢰도에서 유의하며 회귀모형에 의해서 이 모형들은 1953-1972년의 총변동의 41%를 설명하는 것을 제외하고는 총변동의 60% 이상, 최고 90%를 설명하고 있으며 결정계수값이 시간이 지남에 따라 높아지는 양상을 보이고 있다.

Table 1. R^2 and F values of multiple regression models for August precipitation.

reference period	number of predictors	F	R^2	reference period	number of predictors	F	R^2
1953-1972	1	14	0.42	1964-1983	3	40	0.87
1954-1973	2	19	0.68	1965-1984	3	31	0.84
1955-1974	3	27	0.83	1966-1985	3	45	0.89
1956-1975	3	28	0.83	1967-1986	3	46	0.89
1957-1976	2	18	0.67	1968-1987	1	30	0.61
1958-1977	2	24	0.72	1969-1988	3	19	0.68
1959-1978	1	29	0.60	1970-1989	2	50	0.90
1960-1979	2	29	0.76	1971-1990	3	64	0.90
1961-1980	2	25	0.73	1972-1991	2	17	0.88
1962-1981	3	41	0.88	1973-1992	3	17	0.65
1963-1982	3	30	0.84				

Fig. 2에 앞에서 구한 다중회귀모델들을 이용하여 그 다음 해인 1973년부터 1992년 8월 강수량을 예측한 값(예측된 값은 정규화된 값이므로, 이 값에 12개 지점의 30년 표준편차 118.5을 곱

하고 기후 평균값인 220.3을 더함)과 실제 8월 강수량을 나타내었다. 1993년 8월 강수량은 예보 인자로 선정된 1993년 500 hPa 고도자료의 부재로 구하지 못하였다. 그림에서 1980년, 1987년, 1990년과 같이 강수량이 많았던 해에는 실제 강수량보다 매우 적게 예측되었다. 그리고 대부분 예측 강수량이 실제 강수량보다 적었으나 그 값은 평균이내(표준편차의 25%이내)이며, 그 변화 경향은 매우 유사하다. 예측기간에 대한 모델의 숙련도는 0.92로 이 모델들은 실제 변동의 92%를 예측하고 있는 것으로 나타나 매우 높은 예측 정확도를 보였다. 이러한 결과는 강인식 등(1992)에 의해 구한 예보의 숙련도(0.97)보다는 작지만 예측되는 기간은 더 길므로 더 나은 결과를 보이는 것으로 사료된다. 또한 참고로 8월 강수량을 예측하기 위해서 생성된 20개 회귀모델에 대해서 다음해 1년에 대해서만 예측하지 않고, 각 모델을 사용하여 기간을 연장하여 예측을 시도한 결과 비록 예측기간에는 차이가 있으나 대부분 0.85~0.9의 높은 숙련도를 나타내었다.

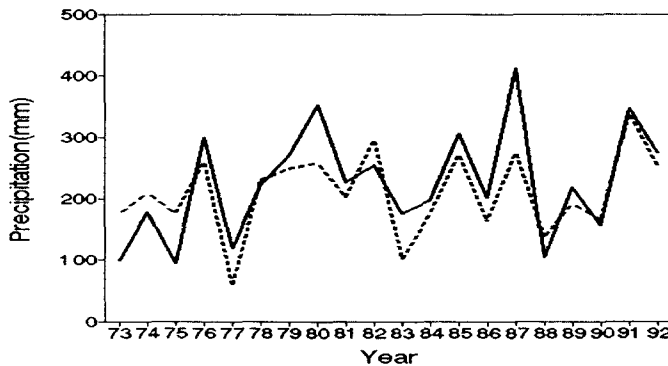


Fig. 2. The time series of August precipitation. Solid line indicates the observation and dotted line indicates predicted value by regression models.

4. 결론

이 연구에서는 한반도 8월 강수량을 예측하기 위해서 매년 최근 20년 자료를 사용하여 새로운 모델을 만들었다. 이를 위해서 8월 강수량과 북반구 500 hPa 고도 및 태평양 해수면 온도와의 상관분석을 수행하였으며, 회귀모델을 최적화하는 예보인자를 선정하여 다중회귀모델을 생성하여 예측을 시도하였다. 이러한 결과를 요약하면 다음과 같다.

- 8월 강수량을 예측하기 위한 다중회귀모델의 최종적인 500 hPa 예보인자는 만주, 지중해, 아열대 중태평양, 중앙아시아, 북미 서안, 북대서양, 그린랜드, 시베리아, 카스피해에서 나타

났다. 태평양 해수면의 예보인자는 주로 적도를 중심으로 아열대 태평양에서 나타났으며, 1970년 이후 다중회귀 모델의 예보인자로 오스트레일리아 북부 해양의 해수면 온도가 지속적으로 나타났다.

- 8월 강수량의 예측 결과 모델의 숙련도는 0.92이었으며, 예측 모델은 매우 유의하였다. 그리고 본 연구에서 구한 회귀모델들을 사용하여 회귀모델을 구축하는데 포함되지 않는 기간에 대해서 예측을 시도하였을 때, 비슷한 숙련도를 보였다.

이 연구를 통하여 한반도 8월 강수량을 예측하기 위해서 구한 다중회귀모델은 통계적으로 매우 유의하는 것으로 나타났으며, 앞으로 선정된 예보인자들의 지속적인 모니터링과 통계모델의 예보인자로 선택된 500 hPa 고도와 해수면 온도와의 한반도 기상요소와 물리적, 역학적 연관성을 계속적으로 연구해야 할 것이다. 또한 더 높은 예측 정확도를 갖는 통계모델과 역학모델의 개발이 진행되어야 할 것이다.

참고문헌

- 강인식, 허창희, 민경덕, 1992: 한반도 여름철 강수량의 장기예측. 한국기상학회지, 28, 283-292.
- 오재호, 권원태, 홍운, 박정규, 윤용훈, 나득균, 백희정, 1994: 통계적 방법을 이용한 장기예보 기법 개발 연구(II). 기상연구소 연구보고서, MR94F-002, 79pp.
- 최정부, 권원태, 홍운, 박정규, 윤용훈, 나득균, 백희정, 1993: 통계적 방법을 이용한 장기예보 기법 개발 연구(I). 기상연구소 연구보고서, MR93F-001, 122pp.
- 허창희, 강인식, 1988: 한국지역 강수의 변동성에 관한 연구. 한국기상학회지, 24, 38-48.
- Barnett, T. P. and Hasselmann, 1979: Techniques of linear prediction, with application to ocean and atmospheric fields in the Tropical Pacific. *Rev. Geophys. Space Phys.*, 17, 949-968.
- Bell, G. J., 1976: Seasonal forecasts of Hong Kong summer rainfall. *Weather*, 31, 208-212.
- Blanford, H. F., 1884: On the connection of the Himalayan snowfall with dry winds and seasons of droughts in India. *Proc. Roy. Soc. London*, 37, 3-22.
- Quinn, W. H. and W. V. Burt, 1970: Prediction of abnormally heavy precipitation over the equatorial Pacific dry zone. *J. Appl. Met.*, 9, 20-28.
- Walker, G. T., 1924: Correlation in seasonal variation of weather. India Met. Dept. *Memoirs.*, XXIV, 275-345.