

서울지역 연 강수량자료에 대한 분석과 추계학적 모형 적용 및 예측 가능성 검토

유철상¹⁾, 문장원²⁾, 김중훈³⁾

1. 서론

수문 분야에서 가장 중요한 문제는 홍수와 가뭄이며 이 두 가지 문제가 일어나는 근본 원인은 강수이다. 강수의 많고 적음에 따라 홍수와 가뭄이라는 극단적인 문제가 나타나게 되며 이러한 강수에 대한 중요성 때문에 기상과 관계된 많은 요인 중 가장 긴 기간의 관측기록을 가지고 있는 것이 강수량에 대한 자료이다.

전통적 농업국가였던 우리나라에서는 강수에 대한 정보의 필요에 의해 조선 세종 때 측우기가 발명(1441)된 후로 현재에 이르기까지 관측기기로 강수를 정량적으로 측정하고 있다. 그 동안 자료의 많은 손실에도 불구하고 220여 년이란 세계에서 가장 긴 강수량 측정자료와 1626년부터의 강수일수 기록을 가지고 있다(조신섭등, 1996).

이와 같이 긴 관측기간을 갖는 서울지역의 강수량자료를 이용하여 지금까지 많은 학자들에 의해 연구가 이루어져 왔다. 임규호와 정현숙(1992)은 간단한 통계분석과 시계열분석을 서울지방의 연 강수량자료에 적용하여 총 220년의 조사기간 중 1900년대를 전후로 약 25년 동안 연 강수량이 적게 나타나는 기간을 건조시기로, 나머지를 시간 순으로 습윤시기 1과 습윤시기 2로 분류하였다. 또한 이들 기간동안에 관측된 자료의 통계적인 특성을 연구하여 서울지방의 최근 연 강수량이 측우기 관측치로 대표되는 서울의 과거 연 강수량과 동일한 수준임을 밝히려 하였다. 또한 정현숙과 임규호(1994)는 서울지역 강수량과 강수일수 시계열을 이용하여 각각의 자료가 갖는 특성을 분석하고 상호관계를 검토하여 고대 측우기 관측자료의 신뢰성을 검토하였다. 조신섭등(1996)은 서울지역의 연 강수량자료를 간섭모형을 이용하여 적합해보고자 하였으며, 전종갑과 문병권(1997)은 서울지역 강수량 관측자료의 분석을 통해 측우기 관측자료와 현대 관측자료사이의 연 강수량 경년변화를 비교하였고 우리나라 강수주기의 경우 약 5~6년 주기가 현저하게 나타난다고 하였으며 이를 태양흑점자료와 비교하였다.

본 연구에서는 서울지역 연 강수량자료를 임규호와 정현숙(1992)의 연구결과와 같이 1778년부터 1883년까지의 습윤시기 1, 1884년부터 1910년까지의 건조시기, 그리고 1911년부터 1998년까지의 습윤시기 2로 나누어 각각의 자료에 대해 분석해보고 그 특성을 파악해보고자 한다. 또한 자료 구간을 좀더 세분화하여 적절한 ARIMA 모형을 결정해보고 선정된 추계학적 모형의 형태에 따라 전체 자료에 대해 일관된 모형의 적용 및 예측 가능여부를 판단해보고자 한다.

2. 대상 자료

본 연구에서 사용할 자료는 서울지역 연 강수량자료이며 총 자료기간은 1778년부터 1998년까지 총 221년이다. 다음 그림 1은 대상 자료인 서울지역 연 강수량자료를 도시한 결과이다.

-
- 1) 고려대학교 환경공학과 조교수
 - 2) 고려대학교 토목환경공학과 석사과정
 - 3) 고려대학교 토목환경공학과 부교수

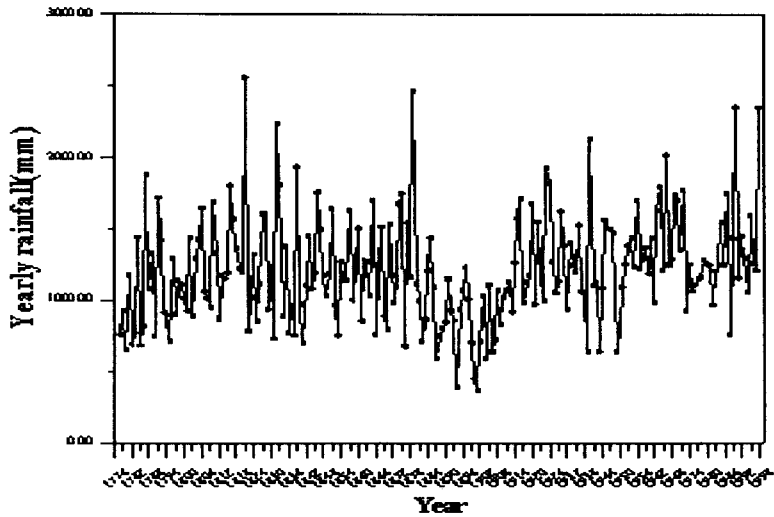


그림 1. 서울지역 연 강수량자료(1778-1998)

그림 1에서 1778년부터 1907년까지의 자료는 조선시대 측우기에 의해 관측된 자료를 현대의 단위로 환산한 결과이며 1908년부터 1998년까지의 자료는 근대 우량계로 관측된 자료이다. 그림 1을 이용하여 자료의 추세를 판단해보면 특별한 증가나 감소 추세를 보여주고 있지 않으며 분산 또한 일정하게 유지된다고 볼 수 있으므로 자료에 대한 특별한 처리는 필요 없을 것으로 판단된다.

3. 자료 분석

3.1 기초 통계분석

각각의 자료에 대한 특성을 보기 위해 먼저 각 자료 기간에 따른 평균, 표준편차, 왜곡도 계수를 구하여 비교하였으며 표 1은 그 결과를 나타낸 것이다.

표 1. 각 자료 기간에 따른 기초 통계량

구 분	평균 (mm)	표준편차 (mm)	왜곡도 계수
전체 자료 (1778-1998)	1207.499	378.9018	0.724
습윤기간1 (1778-1883)	1201.094	385.7187	1.017
건조 기간 (1884-1910)	871.961	268.8626	-1.093
습윤기간2 (1911-1998)	1318.165	338.3781	0.731

먼저 평균을 살펴보면 전 기간의 경우 1207.499 mm, 습윤기간 1은 1201.094 mm, 습윤기간 2는 1318.165 mm이며 건조기간은 871.961 mm로 산정 되었다. 표준편차의 경우 습윤기간 1과 습윤기간 2는 각각 385.7187 mm와 338.3781 mm로 유사하게 나타나고 있으나 건조기간의 경우에는 268.8626 mm로 차이를 보여주고 있음을 알 수 있다. 왜곡도 계수 역시 습윤기간 1, 2의 경우 양의 값으로 나타나 오른쪽으로 왜곡된 분포를 보여주고 있으나 건조기간의 경우에는 이와 달리 왼쪽으로 왜곡된 분포를 보여주고 있다. 따라서 습윤기간 1, 2는 자료의 특성이 유사하게 나타나고 있으나 건조기간은 이와 다른 자료특성을 갖는다고 할 수 있으며 서울 지역 연 강수량자료 분석 시 전체 자료에 대한 일괄적인 분석보다는 자료 특성이 다른 각 기간별로 분석을 수행하는 것이 보다 적절함을 나타내주는 결과라 할 수 있을 것이다. 그림 2는 각 자료 기간 별로 산정된 평

균값을 자료의 시도표에 함께 도시한 결과이다. 이를 통해 자료 기간 별로 평균이 다르게 나타남을 보다 확실하게 알 수 있으며 습윤 기간 1, 2와 건조 기간으로 자료를 구분하는 것이 어느 정도 타당함을 나타내 준다고 할 수 있다. 여기서 습윤 기간 1과 2에서 평균이 차이가 나는 이유로는 조선시대 강우 관측방법을 생각해 볼 수 있다. 측우기의 특성상 2mm이하의 강우와 동절기의 강설량을 포함하고 있지 않으므로 이와 같은 차이가 나고 있다고 볼 수 있으며 따라서 습윤 기간 1과 2의 평균 역시 유사하게 나타나고 있다고 볼 수 있다.

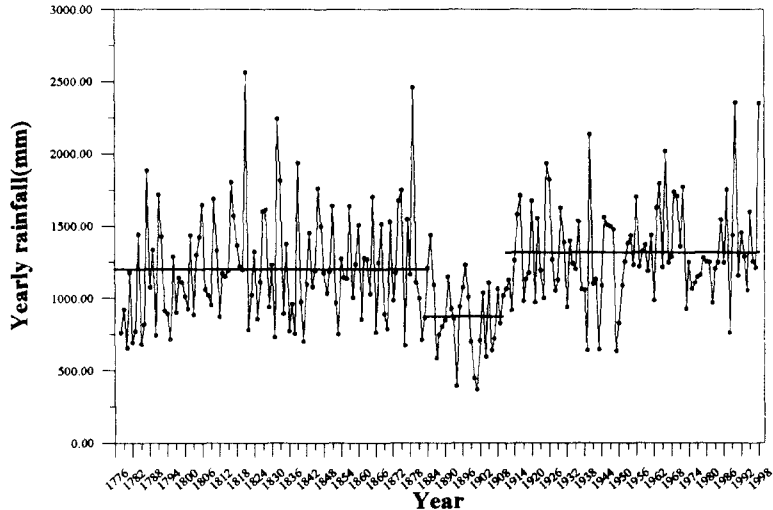


그림 2. 자료 기간에 따른 평균의 변화

3. 2 스펙트럼 분석

스펙트럼 분석은 빈도영역(frequency domain)에서 시계열 분석을 수행하는 것이다. 빈도영역의 시계열 분석이란 주어진 시계열을 싸인과 코싸인의 합성 함수로 표현하고 이 때 이들 주기 함수의 빈도(frequency) - 단위 시간내의 이동각 - 를 이용하여 자료 분석하는 것을 말한다. 스펙트럼 분석의 1차 목적은 계열내 잠재된 주기를 찾아내는 데 있으며 2차 목적은 스펙트럼 밀도의 추정 그 자체에 있다(박유성, 허명희, 1996).

따라서 서울지역 연 강수량자료에 잠재된 주기를 확인하기 위해 스펙트럼 분석을 실시하였으며 이 분석 역시 전체자료기간, 습윤기간 1, 습윤기간 2, 그리고 건조기간으로 나누어 그 결과를 검토하였다. 다음 그림 3은 각 자료 구간에 따라 스펙트럼 분석 결과 얻어진 주기를 나타낸다.

그림 3을 통해 살펴보면 서울지역 연 강수량자료의 전체자료의 경우 큰 Peak가 대략 3지점에서 나타나고 있음을 알 수 있으며 이 때의 주기는 110.50년, 55.25년, 3.62년으로 대략 110년, 55년, 그리고 3~4년의 주기가 크게 나타나고 있음을 나타낸다. 습윤기간 1의 경우 우세한 주기는 15.14년, 5.89년, 3.66년으로 과거 측우기에 의해 관측된 자료의 경우 15년, 그리고 약 4~6년의 주기가 우세하게 나타나고 있음을 알 수 있다. 또한 건조기간에 대한 주기를 살펴보면 대략 6~7년의 주기가 우세하게 나타나고 있음을 알 수 있으며 습윤기간 2의 경우 고빈도 영역에서 우세한 주기가 나타나는 것으로 보인다. 이 때의 주기는 대략 2~6년 사이의 값을 나타내고 있으며 따라서 근대 관측이 시작된 이후 서울지역 연 강수량자료는 2~6년 사이의 주기가 가장 우세하게 나타나고 있음을 알 수 있다.

지금까지 스펙트럼 분석을 통해 나타난 결과를 종합하여 비교해보면, 1778년부터 1998년까지의 전체 자료를 이용한 경우에는 110년, 55년, 그리고 약 3~4년의 주기가 가장 우세한 것으로 판단되었으며 측우기 관측자료인 습윤기간 1의 경우 3~6년의 주기가 가장 우세하게 나타났고 또한 15년의 주기도 우세하게 나타나고 있었다. 습윤기간 1, 2의 중간에 위치하고 있는 건조기간의 경우 약 6~7년의 주기가 크게 나타나고 있었

고 마지막으로 최근 관측자료인 습윤기간 2의 경우 2~6년의 우세한 주기를 판단할 수 있었다. 최근 자료일 수록 우세한 주기가 점점 짧아지고 있음을 알 수 있으며 모든 기간에 대해서 3~6년의 주기는 계속적으로 나타나고 있었다. 따라서 서울지역 연 강수량자료의 경우 대체적으로 3~6년의 주기성분이 우세한 것으로 판단할 수 있으며 이는 과거 전종갑과 문병권(1997)의 연구결과인 5~6년 주기와 어느 정도 일치한다고 볼 수 있다.

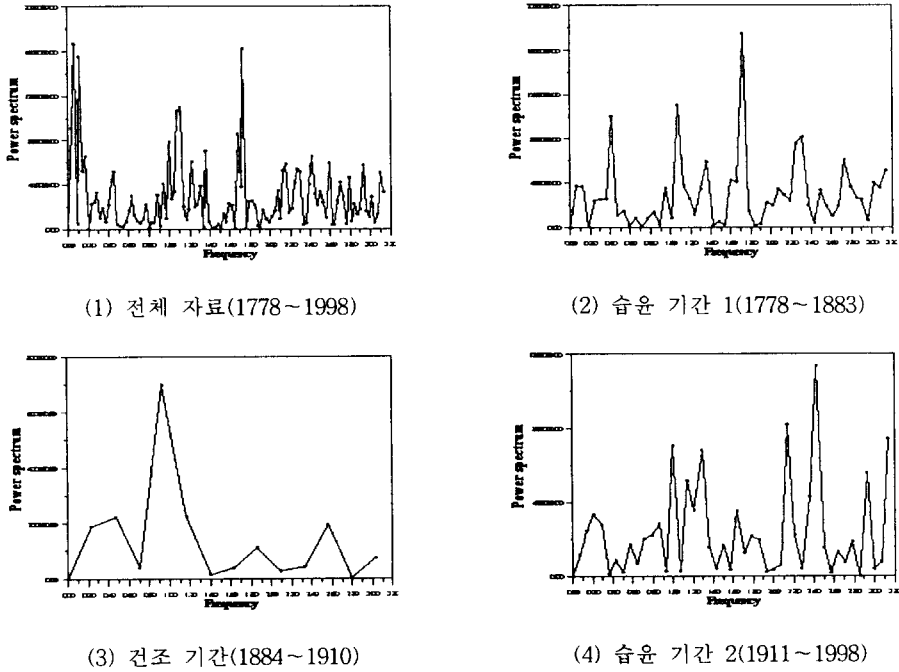


그림 3. 각 자료 구간별 주기도

3. 3 적절한 추계학적 모형 결정

시계열 자료분석의 최종적인 목적은 과거 자료를 이용하여 미래에 나타날 결과에 대한 예측을 실시하는 것이다. 일반적으로 이러한 예측을 위해 주로 사용되는 방법은 자료특성에 따라 적절한 추계학적 모형을 결정한 후 이를 이용하여 미래에 대해 예측을 수행하는 것이다. 따라서 본 연구에서는 전체 자료 기간과 습윤 기간 1, 건조 기간, 습윤 기간 2에 대해 적절한 추계학적 모형을 산정하고 비교하기로 하였다. 또한 상대적으로 자료 기간이 길다고 볼 수 있는 습윤 기간 1, 2를 각각 1-a, 1-b, 2-a, 2-b로 세분하여 동일한 분석을 수행하였으며 이를 통해 서울지역 연 강수량 자료에 대한 추계학적 모형 적용 시 전 자료에 대해 일관된 모형이 적용 가능한 지를 판단하고자 하였다.

자료에 대해 적절한 추계학적 모형을 결정하는 일반적인 방법은 먼저 자료의 시도표를 통해 자료의 비정상성, 즉 추세와 계절성을 나타내고 있는 지를 판단한 후 만약 추세와 계절성을 나타내고 있다면 이러한 요소의 제거를 위해 차분을 실시하게 된다. 대상 자료 계열에서 추세와 계절성이 제거되어 정상성 조건을 만족하였다면 해당 자료의 ACF와 PACF를 관찰하여 모형의 차수를 결정하게 된다.

서울지역 연 강수량자료의 경우 특정한 추세가 나타나고 있지 않다고 판단할 수 있으므로 추세 제거를 위한 차분이 필요하지 않다고 판단할 수 있다. 따라서 자료의 ACF와 PACF를 이용하여 후보 모형의 차수를 결정한 후 결정된 후보 모형의 AIC값을 비교하여 가장 적절한 모형을 선정하였다. 각 자료 구간별로 ACF와 PACF를 도시하여본 결과 시차에 따른 상관성은 매우 낮은 것으로 나타나고 있었으며 자료 기간 중 전체 자

료의 경우와 습윤 기간 2-b의 경우 무작위한 자료 특성을 보이고 있었다. 나머지의 경우 습윤 기간 1-a와 1-b, 그리고 습윤 기간 2-a는 AR모형으로 판단할 수 있었으며 건조 기간의 경우에는 MA모형으로 결정할 수 있었다. 결정된 후보 모형을 이용하여 적정 추계학적 모형을 산정하기 위해 AIC 기초 통계량을 이용하였으며 후보 모형 중 가장 작은 AIC값을 나타내는 모형으로 적정 모형을 결정할 수 있다. 다음 표 2는 각 자료 기간별 결정된 적정 추계학적 모형을 나타낸 것이다.

표 2. 각 자료 기간별 적정 모형 산정 결과

구 분	전체 자료 (1778~1998)	습윤기간 1-a (1778~1830)	습윤기간 1-b (1831~1883)	건조기간 (1884~1910)	습윤기간 2-a (1911~1954)	습윤기간 2-b (1955~1998)
적정모형	백색 잡음	AR(1)	AR(1)	MA(1)	AR(3)	백색 잡음

표 2의 산정 결과를 보면 각각의 적용 자료 기간에 따라 모형이 다르게 산정 되고 있음을 알 수 있으며 자기 회귀 모형(AR), 이동 평균 모형(MA), 그리고 백색 잡음등으로 그 형태 또한 다양한 결과를 보이고 있다. 따라서 같은 서울지역에서 관측된 자료라 할 지라도 전체를 하나의 일관된 특성을 갖는 자료로 보고 분석을 수행하는 것보다는 자료의 기간에 따라 먼저 자료 특성을 파악한 후 그 결과에 따라 적절한 분석을 수행하는 것이 바람직할 것으로 판단된다. 습윤기간 1과 2의 결과가 차이를 보이는 이유는 관측계기가 다르고 또한 습윤기간 1의 경우 관측자료 자체가 2mm이하의 소량의 강수와 강설을 포함하지 않고 있다는 자료의 특성으로 인해 근대 관측이 수행된 이후의 자료와 다른 특성을 나타내는 것으로 생각할 수 있다. 그리고 관측계기가 같은 습윤 기간 2-a와 2-b의 경우에는 자기 회귀 과정과 백색 잡음 과정으로 판이하게 다른 결과를 보여 주고 있으며 이러한 결과를 주는 원인으로는 최근 들어 엘니뇨나 라니냐와 같은 이상현상으로 인해 우리나라 강수 특성이 영향을 받은 것이 하나의 원인이라고 볼 수 있을 것이다.

4. 결 론

과거 서울 지역의 강수 경향을 정확히 이해하고 이를 이용하여 앞으로의 예측에 이용하고자 한다면 고대의 강수기록을 면밀히 검토해야 할 필요가 있다. 이러한 점에서 조선 시대에서부터 지금까지 유지되어 오고 있는 서울 지역 강수 관측 자료는 매우 소중한 자료라고 할 수 있으며 대상 자료에 대한 특성 파악은 무엇보다 우선되어야 한다.

이러한 관점에서 본 연구에서는 서울 지역 연 강수량 자료에 대해 과거 연구자들에 의해 제안된 자료 기간 구분을 이용하여 각 기간별 자료의 통계적 특성 분석 및 적절한 추계학적 모형의 구축을 통해 그 결과를 비교하여 보았다. 먼저 자료 기간을 습윤 기간 1, 건조 기간, 습윤 기간 2로 나누어 기초적인 통계량을 산정하였으며 각각에 대해 스펙트럼 분석을 수행하여 자료에 나타나고 있는 우세 주기를 판단하여 보았다. 마지막으로 습윤 기간 1과 2를 보다 세분하여 추계학적 모형을 적합하여 보았으며 결정된 추계학적 모형을 각 자료 기간별로 비교하여 보았다.

그 결과, 각 자료 기간 별 기초 통계량 산정 결과로부터 습윤 기간 1, 2는 전체 자료의 특성과 유사하게 나타나고 있음을 알 수 있었으나 건조 기간의 경우 다른 자료 특성을 얻을 수 있었다. 따라서 과거 연구자에 의해 제안된 습윤 기간 1, 건조 기간, 습윤 기간 2와 같은 자료의 구분은 타당한 것으로 생각할 수 있다. 또한 각 자료 기간에 대해 스펙트럼 분석을 수행한 결과 각 자료 기간별로 우세한 주기는 3~6년 주기가 공통적으로 나타나고 있음을 알 수 있었다. 하지만 최근 들어 우세 주기는 조금씩 짧아지고 있음을 알 수 있었으나 이러한 결과가 보편적인 경향으로 나타나고 있는 지는 추가 연구를 통해 확인할 필요가 있을 것으로 생각된다. 최근 들어 우세 주기가 짧아지고 있는 원인을 생각해보면 엘니뇨나 라니냐와 같은 이상 기후 현상으로 인해 강수 형태가 변하고 있으며 이로 인해 강수 특성 자체가 영향을 받아 나타나는 것으로 생각된다. 마지막으로 각 자료 기간별로 적절한 추계학적 모형을 적합해 본 결과 각 기간별로 선정된 모형에는 일관성이 없이 다양하게 결정되고 있었다. 이를 통해 서울 지역 연 강수량 자료의 경우에는 시간에 따라 자료의 특성

이 변경되고 있다고 판단할 수 있으며 분석하고자 하는 자료 기간별로 적절한 분석이 수행되어야 함을 나타내 준다고 할 수 있다. 즉, 서울 지역 연 강수량 자료의 경우 전체 자료 기간에 대해 일관된 한가지 모형이 적용되기 어렵다고 할 수 있다. 또한 이러한 자료의 특성을 갖는 서울 지역 연 강수량 자료를 추계학적 모형으로 예측을 수행하기 위해서는 각 자료 기간별로 구분되어 수행되어야 하며 강수 자료는 시차에 따른 상관관계가 매우 낮고 자료 자체에 비 정상성을 포함하고 있으므로 이러한 자료의 특성을 고려해줄 수 있는 방법의 도입이 필요하다고 할 수 있다. 즉, Kalman Filter와 같이 비 정상성을 갖는 자료를 고려해줄 수 있는 방법을 이용하여 예측을 시행한다면 기존의 방법을 이용한 결과보다 좋은 결과를 얻을 수 있을 것으로 생각된다. 특히, Kalman Filter 이론 중에서 모형의 매개변수 변화를 고려할 수 있는 방법을 이용하여 실제 강수 자료에 적용 및 예측을 수행하는 것이 보다 적절할 것으로 사료된다.

6. 참고 문헌

- 1) 박유성, 허명희(1996), 시계열 자료 분석, 자유 아카데미
- 2) 임규호, 정현숙(1992), “서울지방 연강수량의 경년 변동, 1771-1990”, 한국기상학회지, 제 28권 제 2호, pp.125~132
- 3) 전종갑, 문병권(1997), “측우기 강우량 자료의 복원과 분석”, 한국기상학회지, 제 33권 제 4호, pp.691~707
- 4) 정현숙, 임규호(1994), “서울 지역 월 강수량과 강수일수, 1770-1907”, 한국기상학회지, 제 30권 제 4호, pp.487~505
- 5) 조신섭, 이정형, 김병수(1996), “시계열 모형을 이용한 측우기 자료의 분석”, 응용통계연구, 제 9권 제2호, pp.25~43