

국내 강우의 이상치검정 방법의 비교 연구

A Comparative Study on Methods for Outlier Test of Rainfall in Korea

이정식*, 신창동**
Jung Sik Lee, Chang Dong Shin

요 지

이상치는 표본자료에서 크게 어긋나 다른 자료들로부터 떨어져 표시되는 자료로써, 실제로 발생할 확률이 매우 낮은 자료로 정의되고 있다. 설계홍수량을 산정하기 위하여 적용하고 있는 극치계열의 연최대치 강우자료에는 기계오작동 및 엔지니어의 표독오류가 발생하고 있으며, 기후변화에 따른 거대태풍 및 국지적인 집중호우 발생 등으로 인한 극치값 등에서 이상치가 관측되고 있다. 통상 이상치들은 통계분석시 자료 본연의 특성을 왜곡시켜 편향된 결과를 산정할 수 있으므로 빈도해석시 이상치해석 절차를 수행하여 자료의 적정성을 확인하여야 한다. 현재 실무에서는 설계홍수량 산정요령과 하천설계기준·해설 등에서 관련 내용을 기술하고 있지만, 국내 강우자료의 기록연수의 부족으로 인하여 빈도해석시 이상치 해석이 미수행되고 있어 이상치에 따른 자료편의가 발생하면 결과물인 확률강우량이 왜곡되게 산정될 수 있다.

따라서, 본 연구에서는 국내 주요 도시의 강우자료를 이용하여 이상치검정을 수행하였다. 대상지점으로는 서울, 부산, 대전, 대구, 인천, 광주, 울산 등의 비교적 긴 관측년수를 보유하고 있는 광역시를 선정하였으며, 지속기간은 10분, 1~24시간의 25개 강우자료를 적용하였다. 이상치검정 방법으로는 타 방법에 비하여 이상치 검정력이 뛰어난 것으로 알려진 2가지 방법을 채택하였으며, 표본자료의 평균과 표준편차로 표준화된 z -값을 이용하여 상·하 한계선을 초과하는 값을 확인하는 z -Score 방법중 향상된 중위수 절대편차(MAD)에 의한 수정 z -Score 방법(Hoaglin, 1993)과 Box-Plot 방법(Tukey, 1969)을 적용하였다. Box-Plot 방법(Tukey, 1969)은 전체 자료를 25%씩 사분위로 구분하는 방법으로 정렬된 자료계열을 중앙값, 박스, 수염(whiskers), 이상치로 구분한다. 정렬된 25~75% 값들을 박스로 포함하여 외곽의 수염값들을 이상치로 분류하며, 특히 사분위수의 도식화로 데이터의 분포를 파악하기 좋으며, 이상치들의 위치와 자료의 비대칭 여부를 쉽게 파악할 수 있다.

본 연구의 수행으로 수정 z -Score 방법의 경우에는 서울과 대구지점에는 이상치가 없으며, 부산지점에는 13개, 대전지점 7개, 인천지점 5개, 광주지점 32개, 울산지점 26개가 나타났다. Box-Plot 방법으로는 서울지점 35개, 부산지점 39개, 대전지점 32개, 대구지점 38개, 인천지점 51개, 광주지점 61개, 울산지점 65개의 이상치가 분석되었다. 연구를 수행한 결과, 수정 z -Score 방법에 비하여 Box-Plot 방법에 의한 이상치가 더 많이 발생하였으며, 각각의 방법으로 지속기간 및 연도별 이상치 발생자료를 확인하였다. 방법별 이상치 발생현황 등을 분석하여 지점별 발생횟수를 분석하였으며, 추후 지점 및 자료의 보완이 수행되면 활용성을 증대시킬 수 있을 것으로 판단된다.

핵심용어 : 이상치검정, 수정 z -Score 방법, Box-Plot 방법

* 정회원 · 금오공과대학교 토목환경공학부 교수 · E-mail : jungsik@kumoh.ac.kr

** 정회원 · 금오공과대학교 토목공학과 시간강사 · E-mail : sichado88@kumoh.ac.kr