



Regression equations of probability plot correlation coefficient test statistics using machine learning

Hyunjun Ahn, Sunghun Kim, Joohyung Lee, and Jun-Haeng Heo

Yonsei University, School of Civil and Environment Engineering, Seoul, Korea, Republic of (kamjakang@yonsei.ac.kr)

In the extremes hydrology field, it is essential to find the probability distribution model that is most appropriate for the sample data to estimate the reasonable probability quantile. Depending on the assumed probability distribution model, the probability quantile could be estimated with quite different values. The probability plot correlation coefficient (PPCC) test is one of the goodness-of-fit tests for finding suitable probability distributions for a given sample. The PPCC test determines whether assumed probability distributions are acceptable for the sample data using correlation coefficients between sample data and theoretical quantiles of assumed probability distributions. The critical values for identification are presented as a two-dimensional table, depending on the sample size and the shape parameters of models, for a three-parameter probability distribution. In this study, the applicability and utility of machine learning in the hydrology field were examined. For the usability of the PPCC test, a regression equation was derived using a machine learning algorithm with two variables: sample size and shape parameter.