

IDENTIFYING RAINFALL PATTERNS USING ROBUST PRINCIPAL COMPONENT ANALYSIS COUPLED WITH BICLUSTER ANALYSIS

by Dr. Sri Andayani, S.Si., M.Kom, Kismiantini, S.Si, M.Si, Ph.D, Nikenasih Binatari, S.Si., M.Si

ABSTRACT

Rainfall data are the most significant values in hydrology and climatology modelling. However, the datasets are prone to missing values due to various issues. This study aspires to impute the rainfall missing values by using various imputation methods such as Replacing by Mmean (RM), Nearest Neighbor (NN), Random Forest (RF), Non-linear Interactive Partial Least-Square (NIPALS) and Markov Chain Monte Carlo (MCMC). Monthly rainfall datasets from 24 rainfall stations in Yogyakarta, Indonesia were used in this study. The datasets were then used for bootstrapping to obtain an estimate of the within-imputation standard errors for each imputed dataset. The performances of five methods were evaluated using root mean square method (RMSE). The experimental results showed that the RF-Bootstrap (RF-B) approach was attained as the most satisfying fitting for missing rainfall data in Yogyakarta, Indonesia.

Kata Kunci: *MCMC, Missing value, nearest neighbor, NIPALS, random forest, replace by mean, bootstrap*