



"Gheorghe Asachi" Technical University of Iasi, Romania



COMPARATIVE EVALUATION OF MULTIPLE IMPUTATION METHODS FOR RAINFALL DATA GAPS

Sathish Gokenapally¹, Banjul Bhattacharyya^{2*}, Vivek Menon Sreekumar²

¹Department of Agricultural Statistics, PGIHS, Sri Konda Laxman Telangana State Horticultural University, Hyderabad, Telangana, India

²Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, West Bengal, India

Abstract

Missing rainfall data remains a significant challenge in agrometeorological research and agricultural decision-making. This study compares Normal (NORM), Predictive Mean Matching (PMM), and Random Forest (RF) methods for handling gaps in daily rainfall series that are subsequently aggregated to monthly mean daily rainfall. Using daily rainfall data from the Nadia and Murshidabad districts over 1970-2025 as a case study, we computed imputation efficiency and related diagnostics. While all three methods demonstrated acceptable performance with relative efficiencies consistently exceeding 0.98, our findings reveal limitations in the Normal method, particularly the generation of implausible imputations for highly positively skewed rainfall distributions. In contrast, PMM and RF methods showed comparable efficiency with superior practical applicability for positively skewed and zero-bounded rainfall. PMM, which is directly influenced by the empirical distribution of observed values, emerged as the preferred method for imputing missing values in the New Alluvial Zone, demonstrating enhanced reliability in maintaining distributional characteristics while avoiding physically implausible imputations. RF methods, capable of capturing complex nonlinear relationships between rainfall variables, showed equivalent performance with marginally superior variance preservation in high-variability months. Overall, this research indicates that while multiple imputation methods are broadly efficient, PMM offers an optimal balance between accuracy, computational efficiency, and theoretical soundness for rainfall applications in the New Alluvial agroclimatic zone.

Key words: missing rainfall data, MICE, multiple imputation, new alluvial zone, predictive mean matching, random forest

Received: March, 2026; Revised final: July, 2026; Accepted: August, 2026

* Author to whom all correspondence should be addressed: E-mail: bhattacharya.banjul@bckv.edu.in