



“Gheorghe Asachi” Technical University of Iasi, Romania



## IMPUTING MISSING DATA IN A SWAT WATER QUALITY MODELLING STUDY USING STATISTICAL METHODS

Hülya Boyacıoğlu<sup>1,2\*</sup>, Meltem Kaya Uyar<sup>3</sup>, Hayal Boyacıoğlu<sup>4</sup>

<sup>1</sup>Dokuz Eylul University, Department of Environmental Engineering, Buca 35390, İzmir, Türkiye

<sup>2</sup>Dokuz Eylul University, Center for Environmental Research and Development (ÇEVMER), Buca 35390, İzmir, Türkiye

<sup>3</sup>Dokuz Eylul University Graduate School of Natural and Applied Sciences, Buca 35390, İzmir, Türkiye

<sup>4</sup>Ege University, Department of Statistics, Bornova 35100, İzmir, Türkiye

### Abstract

Large water-quality databases are useful in modeling studies to identify optimal measures for pollution mitigation and management of water basins. The objective of the study was to conduct statistical methods to impute missing data in the water quality simulation study in the Küçük Menderes River Basin, Türkiye, where missing data caused by a lack of periodic sampling is an important challenge. In the study, the Soil Water Assessment Tool (SWAT) was used to simulate nitrate-nitrogen concentrations (NO<sub>3</sub>-N). Water-quality data collected between 2001 and 2012 from the outlet of the basin was subjected to regression analysis-based imputation methods. In this scope, simple regression models were developed to estimate missing water quality data. Hence, a continuous data set was created, and then the SWAT water quality model was calibrated and validated. Since the calculated Nash–Sutcliffe model efficiency coefficient values were above 0.65, model simulations were judged "good". Furthermore, the Mann–Whitney U test was applied to test model performance by comparing continuous data generated by the SWAT model with the limited observed water quality data. It can be concluded that a simple regression model and non-parametric Mann–Whitney U tests can be performed to impute missing data and evaluate model performance in modeling studies of data shortage basins.

*Key words:* data imputation, regression analysis, SWAT model, water quality modeling

*Received:* August, 2022; *Revised final:* October, 2023; *Accepted:* January, 2024; *Published in final edited form:* March, 2024

\* Author to whom all correspondence should be addressed: e-mail: [hulya.boyacioglu@deu.edu.tr](mailto:hulya.boyacioglu@deu.edu.tr), [hulyaboyacioglu@gmail.com](mailto:hulyaboyacioglu@gmail.com)