Title: Key Variables for Discriminating Groundwater Recharge Mechanisms in a Semi-Arid Context: The Case of Burkina Faso

Younoussa GAMSORE[1]*, Mahamadou KOITA[1], Jean-Michel VOUILLAMOZ[2], Roland YONABA[1], Babacar DIENG[3], Nestor Y. Fiacre COMPAORE[4], Elie Serge Gaetan SAURET[5], Fabrice LAWSON[6], Moussa Diagne FAYE[1], Japhet OUEDRAOGO[3]

[1]*Laboratoire Eaux Hydrosystèmes et Agriculture (LEHSA), Institut 2iE, Ouagadougou, Burkina Faso*
[2]*IRD/CNRS, Université Grenoble Alpes, Grenoble, France*
[3]*Programme D'Approvisionnement en Eau et Assainissement (PAEA), Ouagadougou, Burkina Faso*
[4]*Laboratoire Géosciences et Environnement (LaGE), Université Joseph KI-ZERBO, Ouagadougou, Burkina Faso*
[5]*INERA/CNRST, Bobo-Dioulasso, Burkina Faso*
[6]*IRD, Cotonou, Benin*
**Corresponding author: younoussa.gamsore@2ie-edu.org*

ABSTRACT

In semi-arid sub-Saharan Africa, particularly in Burkina Faso, groundwater is essential for drinking water supply and socio-economic development. However, its long-term sustainability is threatened by overexploitation, pollution, and often unfavorable hydrogeological conditions. In this context, sustainable groundwater management requires a better understanding of aquifer recharge, especially its underlying mechanisms. Although the aridity index is commonly used to discriminate recharge mechanisms at large scales, it does not adequately capture the variability of the environmental factors controlling recharge. This study was designed to identify the key environmental covariates involved in discriminating recharge mechanisms in a semi-arid setting. To this end, a dataset was compiled from documented sites, including identified recharge mechanisms and a range of environmental variables derived from a literature review and expected to influence them: mean annual precipitation (P_moy), actual evapotranspiration (ETR), potential evapotranspiration (ETP), aridity index (Ind_Ar), land use and land cover (LULC), readily available water for plants (RFU), saturated hydraulic conductivity (Ksat), slope, elevation, distance to streams, HAND, TWI, Aspect, TPI, lithology, weathering depth, and static water level. Parsimonious variable selection was first performed using Sequential Forward Selection, followed by consensus nested cross-validation (cnCV) within a spatially validated Random Forest pipeline. The resulting core set of variables was then used for final Random Forest modeling under repeated spatial cross-validation. Model performance was assessed using balanced accuracy (Bacc) and AUC-ROC, while model interpretation relied on permutation importance, SHAP values, and the combined use of partial dependence plots (PDP) and individual conditional expectation (ICE) curves. The final variable set included P_moy, Ksat, RFU, LULC, and Aspect. Despite the limited sample size ($n = 238$), the model achieved satisfactory performance, with a Bacc of 0.64 ± 0.06 and an AUC-ROC of 0.67 ± 0.08. Permutation importance and SHAP provided consistent results, identifying P_moy as the dominant predictor, followed by Ksat and Aspect. RFU and LULC showed more moderate but still meaningful contributions. PDPs revealed non-linear relationships between the target and the selected variables, whereas ICE curves indicated that model responses arise from combined effects and interactions among covariates. Overall, the results suggest that the model captures physically meaningful aspects of groundwater recharge and highlight the major role of rainfall, soil physical properties, and slope orientation in discriminating recharge mechanisms at large scale. This study confirms the potential of machine learning approaches to address the complexity of environmental processes in data-scarce settings and opens the way for more reliable spatial mapping of recharge mechanisms and their improved integration into regional hydrogeological models.

Keywords: Groundwater recharge mechanisms; semi-arid environment; Random Forest; spatial cross-validation; Burkina Faso.