ABSTRACT

Understanding the rainfall–runoff process is fundamental for managing water movement through natural and built environments, with direct implications for flood risk reduction, water resource planning, and environmental sustainability. A key metric in this process is the event-based runoff coefficient (ERC), a dimensionless parameter expressing the proportion of precipitation converted into direct runoff. ERC provides critical insight into watershed response, yet its controlling factors vary widely across regions due to differences in climate, hydrology, and landscape characteristics. Despite its importance, these controls remain only partially understood, and predictive methods—particularly at the event scale—have yet to be fully developed and tested using large-scale datasets.

To address these gaps, this study investigates the variability, controls, and prediction of ERC through three integrated objectives: (1) to analyze the short- and long-term controls on ERC variability and assess trends in ERC and its drivers; (2) to evaluate the performance of Support Vector Regression (SVR) in predicting ERC and determine how temporal and spatial controls affect model skill; and (3) to build a predictive model for ERC across the contiguous United States (CONUS), enhancing generalizability through the integration of deep learning methods.

The first objective focuses on the Ohio region, utilizing data from the North American Land Data Assimilation System phase-2 (NLDAS-2) Mosaic Land Surface Model spanning 2000 to 2020. Short-term controls, including climatic factors (rainfall intensity, amount, duration), hydrological variables (antecedent soil moisture, drainage density, curve number), and topographic features (land use, slope, elevation, watershed shape), were analyzed alongside long-term controls such as land use change and climatic trends. Findings reveal that ERC increases with antecedent soil moisture and rainfall intensity, while higher elevations, often characterized

by forested land, exhibit lower ERC under high soil moisture and rainfall intensity condition. Larger watersheds demonstrate lower ERC under low rainfall intensity but higher values during high-intensity events. Long-term analysis identifies soil moisture as the primary control, with land cover changes as a secondary influence.

The second objective evaluates the performance of Support Vector Regression (SVR) for predicting ERC in both temporal and spatial scales. The study leverages data from the Ohio region (study area) and the Mid-Atlantic region (evaluation area), including streamflow data from the United States Geological Survey (USGS), rainfall data from National Oceanic and Atmospheric Administration (NOAA), and forcing data from NLDAS-2. Results show that SVR achieves strong predictive accuracy, though performance varies depending on scale and watershed characteristics. Temporally, watersheds dominated by climatic conditions pose greater challenges for prediction compared to those influenced by pre-event factors. Spatially, urbanized watersheds and regions with higher elevations demonstrate lower predictive accuracy due to complex runoff dynamics. These findings highlighted the limitations of applying uniform learning algorithms across heterogeneous landscapes, pointing to the necessity of spatially adaptive modeling strategies.

The third objective addresses these limitations by establishing a regime-based predictive model across the contiguous United States (CONUS). Moving beyond global parameterization, this phase integrated unsupervised K-Means clustering with regime-specific Extreme Gradient Boosting (XGBoost) models to explicitly disentangle continental hydrological heterogeneity. By stratifying 3,515 watersheds into six physically coherent regimes, the model demonstrated that predictive skill is strictly governed by landscape homogeneity, achieving the highest reliability in the response-dominated Agricultural Plains and Appalachian Forests. Physically, the analysis revealed a fundamental divergence in runoff mechanics: storage-buffered high-relief systems governed by deep geological retention in headwaters, versus transmissivity-limited low-gradient landscapes controlled by the efficiency of subsurface networks. This research ultimately provides a robust, physically interpretable methodology for estimating ERC in ungauged basins, underscoring the necessity of resolving regime-specific physical constraints rather than applying uniform parameterizations across macro-scales.

Overall, this research advances understanding of ERC variability and its controls, demonstrates the potential of machine learning for event-scale prediction, and lays the foundation for a scalable, transferable predictive model to inform flood risk management, water resource planning, and climate resilience strategies across diverse U.S. landscapes.