

Tracking Drought Propagation Across Soil Moisture, Streamflow, and Vegetation in Illinois Using Machine Learning

Anika Raj¹, Kyla Wolski², Allison Goodwell¹

¹ Illinois State Water Survey, Prairie Research Institute, University of Illinois Urbana-Champaign

² Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign

Accepted abstract, August 2026 version. An extended abstract with updated results will follow before the meeting.

Drought impacts in the US Corn Belt propagate through the land surface: precipitation deficits reduce soil moisture, and sustained deficits can later appear in streamflow and vegetation conditions. However, these stages are observed by separate systems, complicating direct comparison of their timing and severity. In this study, we integrate observations from the 19-station Illinois Climate Network (ICN), U.S. Geological Survey stream gauges, and satellite-based vegetation indices to place these three types of drought impacts on a common scale and study their predictability.

We first predict the 5 cm and 50 cm soil-moisture anomaly one day ahead using linear, ridge, lasso, and random-forest models. To test how much predictive information is retained in longer-term antecedent soil conditions, we withhold the most recent 14 days of soil measurements; the model sees soil conditions from 14 days before the issue date, or 15 days before the prediction target. Because soil moisture is strongly autocorrelated, this is a more demanding test than initializing from the present state, which would draw substantial skill from persistence. Five predictor sets isolate the information sources: weather alone, the delayed soil readings alone, and weather plus the delayed 5 cm reading, the 50 cm reading, or both. Splits are chronological: training before 2023 (106,159 samples), model selection on 2023–2024, and 2025 held back entirely.

Using weather together with both delayed soil layers reduced RMSE by 26% at 5 cm ($R^2 = 0.70$) and 48% at 50 cm ($R^2 = 0.75$) relative to weather alone. It also outperformed the persistence baseline—the delayed reading carried forward unchanged—by 43% at 5 cm and 12% at 50 cm. This difference reflects the much stronger persistence of deeper soil moisture: the delayed reading alone explains 5% of anomaly variance at 5 cm but 68% at 50 cm, leaving substantially less room for improvement at depth. The information sources also reverse rank with depth. Using only the delayed soil readings, the best model explains 67% of variance at 50 cm compared with 6% for weather alone, essentially matching the persistence baseline; at 5 cm, this reverses to 44% for weather and 28% for soil. The best model class also changes with depth. Random forest is most accurate at 5 cm, consistent with more nonlinear near-surface soil-moisture dynamics. Meanwhile at 50 cm, which integrates forcing over longer periods, linear regression performs best.

During 2023–2024, the observed 50 cm soil-moisture anomaly averaged -0.30 relative to the pre-2023 seasonal climatology, indicating persistently drier-than-normal conditions at depth. Weather-only models instead predicted a mean anomaly near zero, producing a positive bias of 0.32 . Adding the delayed 50 cm soil observation reduced this bias to 0.10 , whereas adding the delayed 5 cm observation produced little change. This suggests that recent weather alone did not capture the persistent subsurface dryness, while the older deep-soil observation retained that information.

The streamflow and vegetation components are being incorporated using the same anomaly framework. Daily discharge at long-record Illinois U.S. Geological Survey gauges is converted to a 14-day rolling mean, log-transformed, and standardized seasonally, while satellite NDVI anomalies surrounding each station will represent vegetation response. Station–gauge linkage is determined by watershed containment rather than proximity, since a station can lie near a gauge while draining elsewhere; 36 pairs across 16 gauges currently qualify, with 18 overlapping the modeling period. We will present evaluation on the withheld 2025 period, station-level variation in prediction skill, soil–streamflow relationships at verified pairs, and progress on the vegetation component. This study contributes to the understanding and prediction of multiple facets of drought in the US Midwest.