

Quantifying Explanation Uncertainty in Water Shortage Attribution: An Ensemble eXplainable Artificial Intelligence (XAI) Approach to Deep Learning Emulators

Yifan Luo^{1*}, Danielle S. Grogan², Shan Zuidema², Richard B. Lammers², Jiameng Zheng³, Matthew D. Lisk⁴, Karen Fisher-Vanden⁵, Sheila M. Olmstead¹ and Vivek Srikrishnan¹
 (1) Cornell University, Ithaca, NY (2) University of New Hampshire, Durham, NH (3) Louisiana State University, Baton Rouge, LA (4) Pennsylvania State University, University Park, PA (5) Arizona State University, Tempe, AZ

*Corresponding author: Yifan Luo (yl3699@cornell.edu)

Research Question

How can a Deep Learning (DL) emulator of a process-based hydrologic model, combined with an ensemble of eXplainable Artificial Intelligence (XAI) methods, be used to produce spatially explicit, uncertainty-aware attribution of Western U.S. water shortages to meteorological variability versus human water use and efficiency?

Problem Background

- **Complex Drivers:** Water scarcity results from tangled interactions between meteorological variability (supply) and human use (demand), complicating attribution.
- **Limitation of Process-Based Models:** Process-based hydrologic models, such as the UNH Water Balance Model (WBM), represent these physics in detail. But they are too computationally expensive for large-scale sensitivity and attribution experiments.
- **Opportunity and Risk with Deep Learning:** DL emulators can approximate WBM outputs with orders-of-magnitude speedup, but they are often treated as “black boxes” whose predictions are difficult to interpret.
- **The Explainable AI Challenge:** Individual XAI methods can give conflicting explanations because they rely on different mathematical assumptions and are sensitive to different aspects of the model and data.
- **Consensus Approach:** We apply multi-method XAI ensemble to map the dominant drivers and quantify explanation consistency and uncertainty.

Data & Study Area

- **Study Area:** 11 conterminous Western U.S. states
- **Spatial Unit:** Around 900 Water Management Areas (WMAs)
- **Temporal Scope:** 2010–2019 (monthly resolution)
- **Input Data:**
 - **Meteorological Predictors (Monthly):** Temperature and Precipitation
 - **Socioeconomic Indicators (Static):**
 - Population Density
 - Cropland Fraction
 - Irrigation Efficiency
 - Domestic and industrial water-use intensity
- **Target Variable [0-1]:** Unmet water demand ratio = (demand – allocation) / demand from WBM (0 = all demand met; 1 = complete shortage)

Key Takeaways

- **DL Emulator Enables Attribution:** ConvLSTM matches WBM behavior efficiently, enabling fast, large-scale attribution experiments.
- **Regional Driver Patterns:** Precipitation dominates in Pacific Northwest, temperature in Southwest, agriculture in Central Valley.
- **Consensus as a Confidence Map:** Disagreement between XAI methods is not a failure but a diagnostic tool. Using an ensemble of XAI methods provides a necessary “confidence map” for attribution.

Acknowledgement

We would like to acknowledge high-performance computing support from the Derecho and Casper system provided by the National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation (NSF). This research was conducted with support from the Cornell University Center for Advanced Computing, which receives funding from Cornell University, the National Science Foundation (NSF), and members of its Partner Program. This work was supported by the U.S. Department of Energy (DOE), Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics under Cooperative Agreement DE-SC0022141.

Methods: Three-Stage Framework

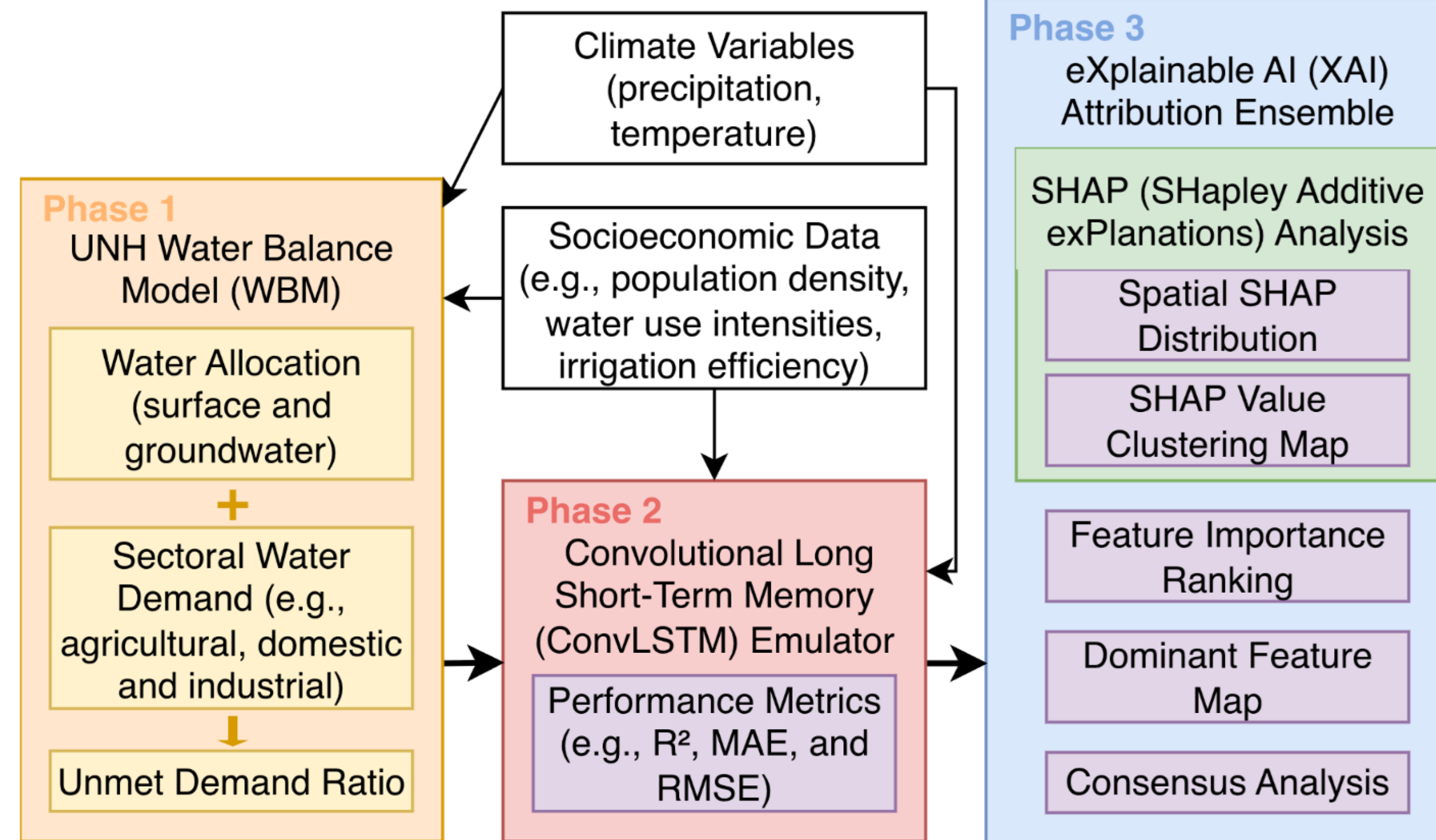


Figure 1. Three-phase framework methods workflow. Meteorological and socioeconomic inputs feed the UNH Water Balance Model (Phase 1) to simulate unmet demand ratio, which trains a ConvLSTM emulator and is evaluated with performance metrics (Phase 2). An attribution ensemble (Phase 3) for spatial attribution, ranking, dominant-feature maps, and consensus analysis.

Phase 1: Process-Based Hydrologic Model Simulation (WBM)

Use the UNH Water Balance Model (WBM) to simulate hydrology, infrastructure operations, and water rights, representing surface and groundwater processes, upstream–downstream propagation, and sector-specific withdrawals to compute monthly unmet water demand.

Phase 2: Deep Learning Emulator (ConvLSTM)

Train a ConvLSTM network that combines convolutional layers (spatial patterns) with LSTM units (temporal dynamics) to approximate the WBM input–output mapping, enabling fast emulation for attribution analysis.

Phase 3: XAI Attribution Ensemble & Consensus Analysis

Apply six complementary XAI methods to quantify driver importance, then use a consensus framework to identify the plurality-vote dominant driver in each WMA and measure consensus strength as the proportion of methods that agree.

Method	Description
R ² -Based Leave-One-Covariate-Out (LOCO)	Drop in explained variance (R ²) by setting a feature's values to zero to simulate its removal from the model
PFI (Permutation Feature Importance)	Increase in prediction error when feature values are randomly shuffled across samples
SHAP (SHapley Additive exPlanations)	Game-theoretic Shapley value contribution of each feature across all possible feature coalitions
Connection Weights (Garson's Algorithm)	Contribution based on absolute magnitudes of learned parameters in early convolutional layers
Input Gradients (Vanilla Gradients)	Sensitivity of output to input changes, computed as average magnitude of gradients
LIME (Local Interpretable Model-agnostic Explanations)	Local influence by fitting simple, interpretable linear surrogate models around individual predictions to explain behavior in that specific neighborhood

Table 1. Six XAI methods for attributing water shortage predictions to input drivers.

Results: Emulator Performance

Using 12-month input windows, the ConvLSTM emulator successfully reproduces WBM simulations with high accuracy ($R^2 = 0.865$, $MAE = 0.072$, $RMSE = 0.148$), that is comparable to training results.

Results: SHAP Analysis

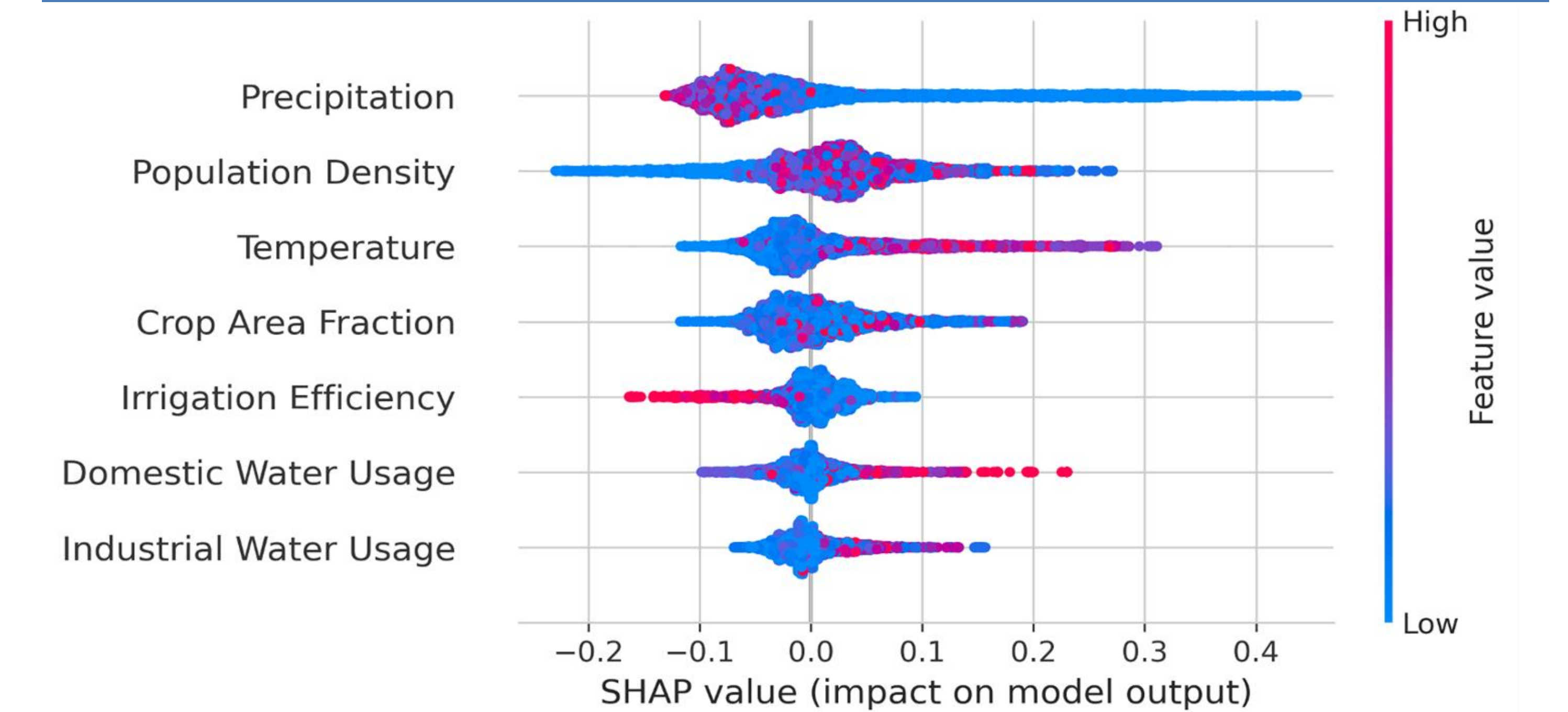


Figure 2. SHAP beeswarm plot showing the impact of key drivers on water unmet demand.

Results: Multi-Method Attribution

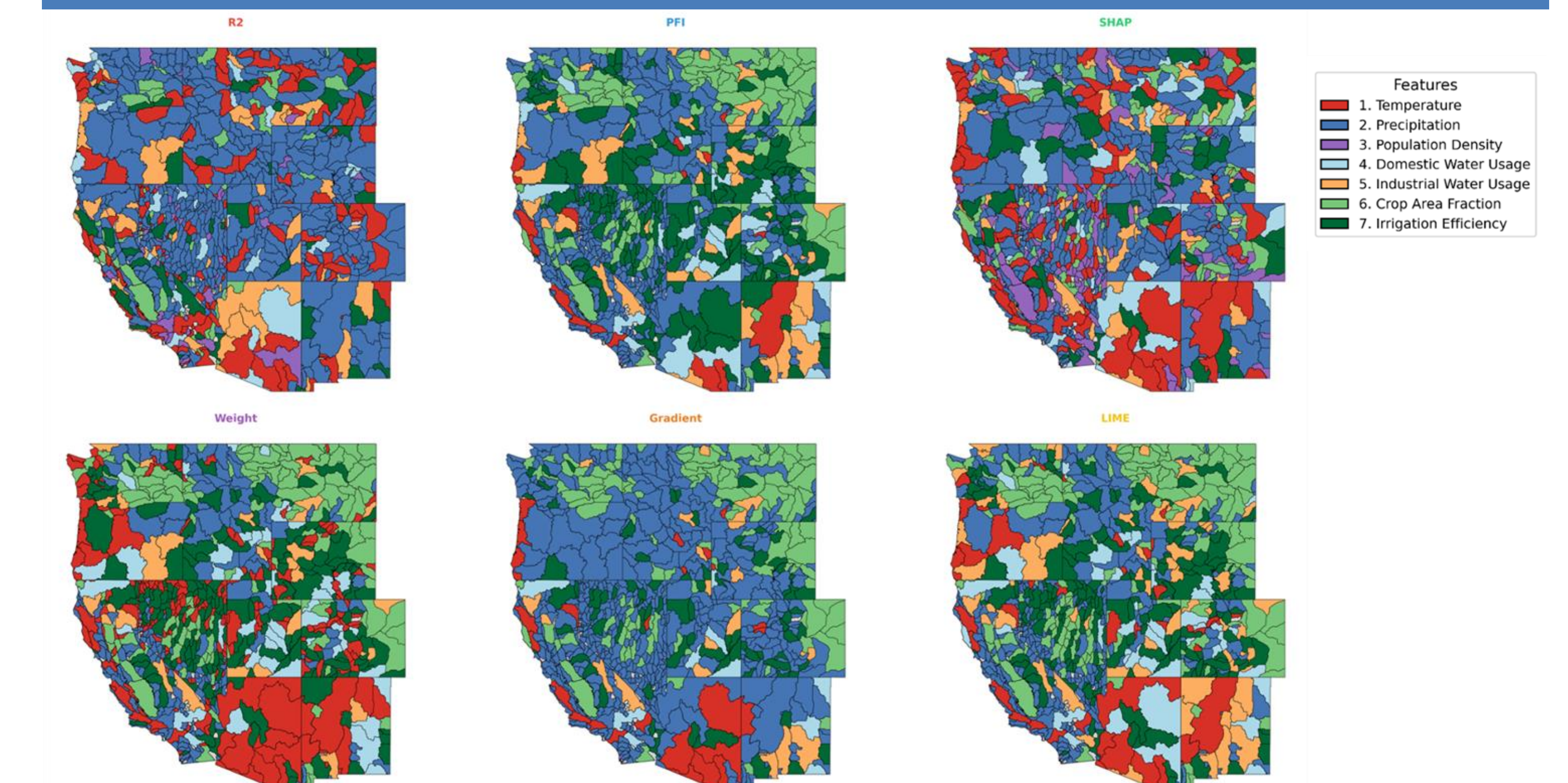


Figure 3. Spatial distribution of the dominant feature by WMA for each XAI method, with each region colored by the most influential driver on unmet water demand.

Results: Consensus Analysis

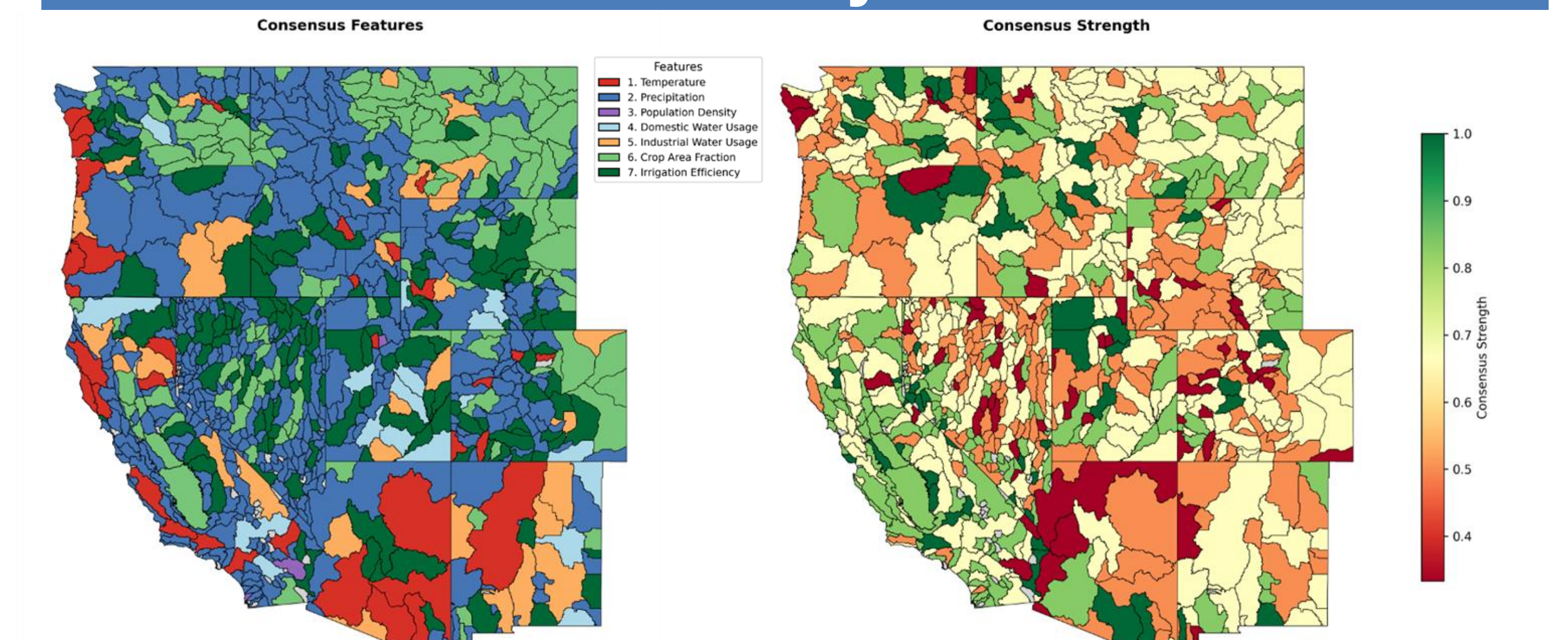


Figure 4. Spatial analysis of feature importance across Western U.S. watersheds. The Consensus Features map (left) identifies the single most dominant driver in each region. The Consensus Strength map (right) quantifies the confidence in that finding, from low (dark red) to high (dark green).