

Disentangling Climate and Socioeconomic Contributions to Water Shortage in the Western U.S.

Yifan Luo^{1*}, Danielle S Grogan², Shan Zuidema², Jiameng Zheng³, Matthew D Lisk⁴, Karen Fisher-Vanden⁴, Sheila 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

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

Key Question

How can **machine learning** and **explainable AI (SHAP)** improve the prediction and understanding of **unmet water demand** across the Western U.S. to support **sustainable water resource management**?

Problem Background

The Western U.S. faces growing water shortages driven by climate change, prolonged droughts, and increasing socioeconomic pressures. Since 2000, a historic megadrought has highlighted the strain on water resources, affecting ecosystems, agriculture, and urban areas. Understanding how climate variability and socioeconomic factors contribute to these shortages is critical for developing effective water management strategies. This study uses data-driven methods to disentangle these drivers and provide insights for sustainable resource allocation.

Data Sources

The primary data source is the **Water Balance Model (WBM)** output (Grogan et al., 2022), referenced in "Bringing Hydrologic Realism to Water Markets" (Grogan et al., under review). It tracks **hydrologic dynamics** in the Western U.S., with **unmet water demand** (difference between water demand and allocation) as a key measure of shortages across Water Management Areas (WMAs) from 2010 to 2019. This study also integrates ERA5 reanalysis data for **temperature** at 2m and **precipitation**, **irrigation efficiency** data derived from WBM gross irrigation output, calibrated to U.S. Geological Survey (USGS) water use statistics (2010). **Domestic and industrial water use per capita** data come from the USGS (2016), and **population density** is sourced from United Nations World Population Prospects-Gridded Population of the World (UNWPP GPW) Version 4, adjusted to 2015 UNWPP country totals.

Method Flowchart

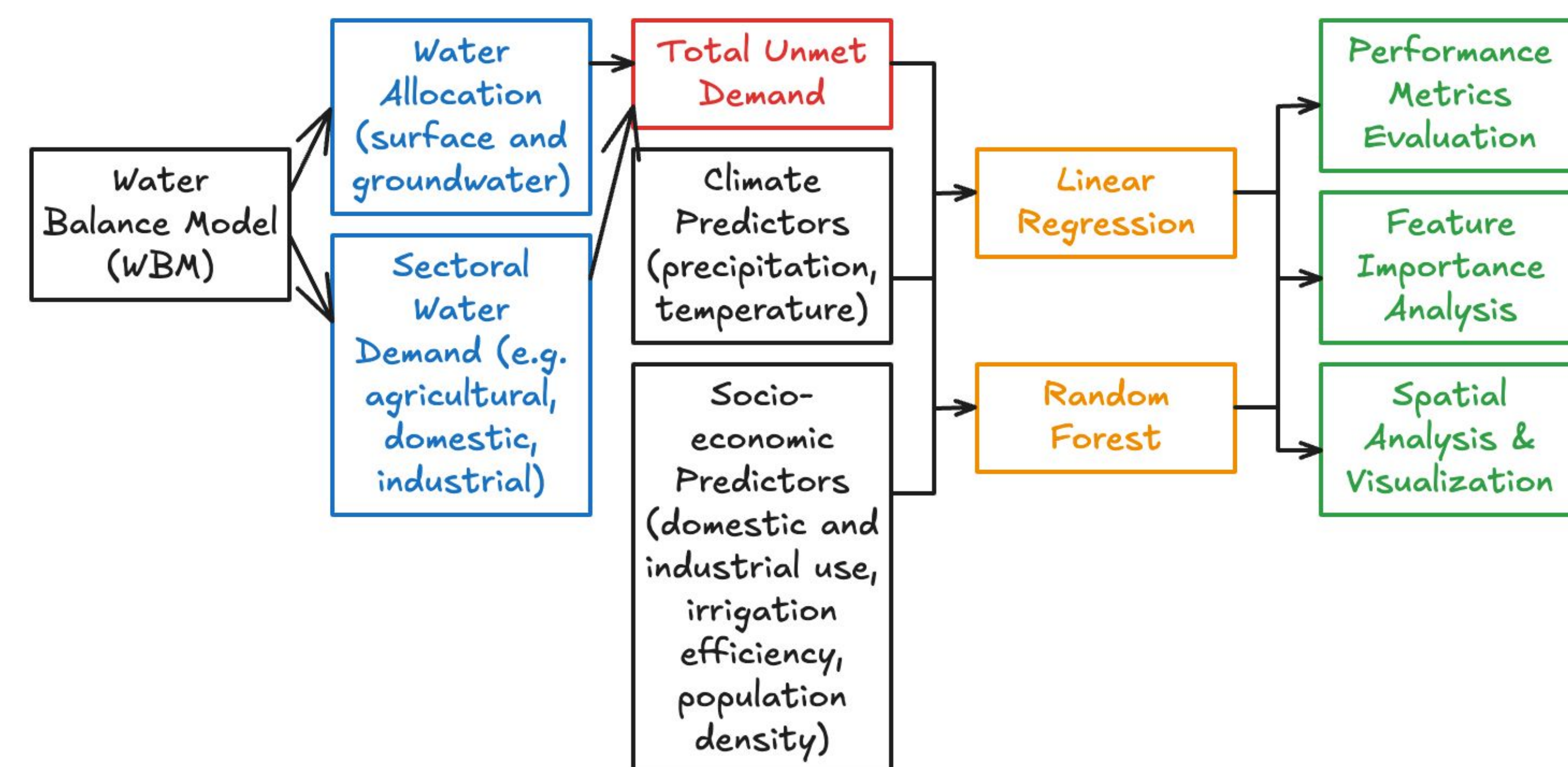


Figure 1. Flowchart illustrating the key methods.

Research Methods

- **Predictive Modeling:** Applied Random Forest and Linear Regression models to predict unmet water demand across WMAs, using cross-validation to ensure every WMA was included in test sets.
- **Feature Importance Analysis:** Used SHAP (SHapley Additive Explanations) to quantify the contribution of each predictor to unmet demand and provide local explanation at the WMA level.
- **Evaluation & Insights:** Compared model performance using Root Mean Square Error (RMSE) and R^2 scores, then identified key drivers of unmet demand and their spatial distribution for policy-relevant insights.

Spatial Distribution of Unmet Water Demand

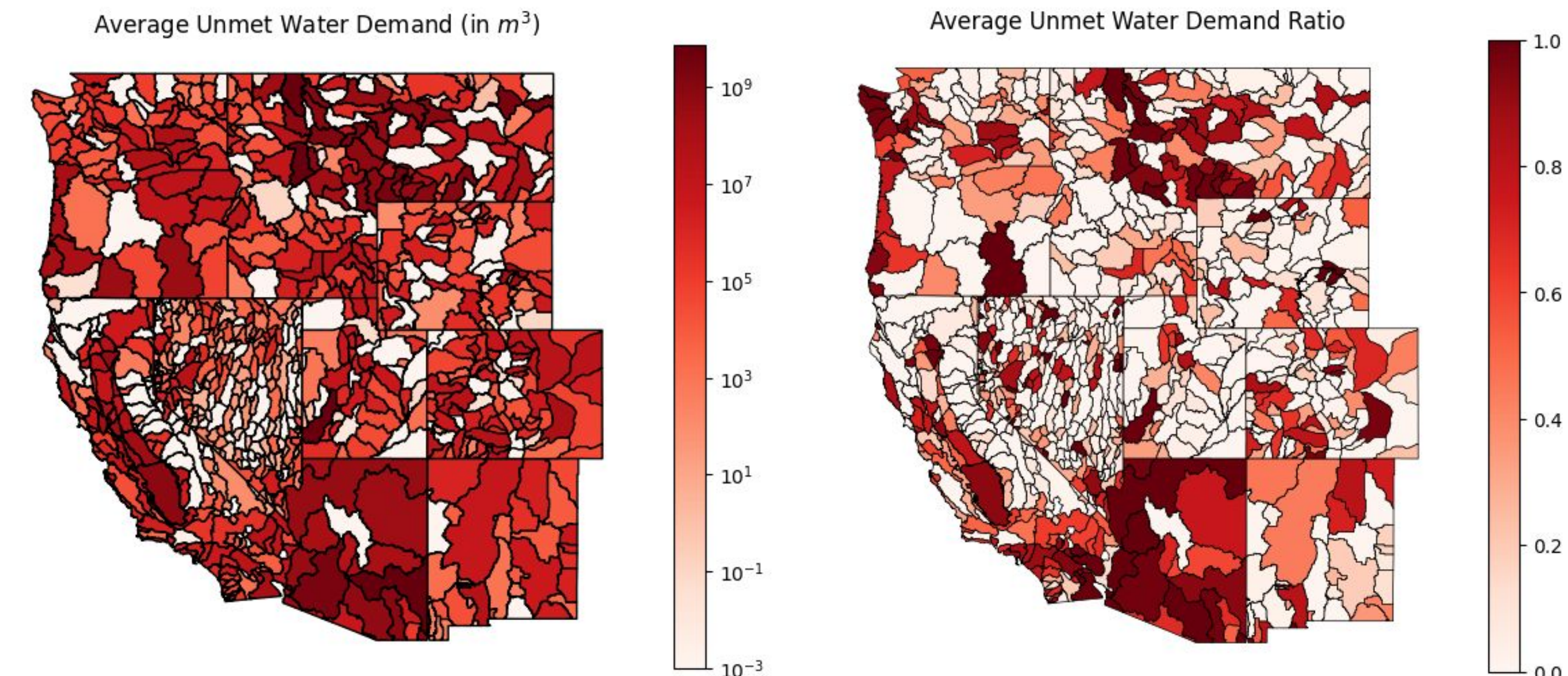


Figure 2. Spatial Distribution of average unmet water demand and its ratio across WMAs. The left panel illustrates the total average unmet water demand (in cubic meters) for each WMA, using a logarithmic scale to capture the wide range of values. Darker regions indicate higher unmet demand. The right panel presents the unmet demand ratio, calculated as the fraction of total water demand that remains unmet. Higher ratios (dark red) highlight areas experiencing severe water shortages, while lower values (lighter shades) indicate relatively better water supply conditions.

SHAP Analysis

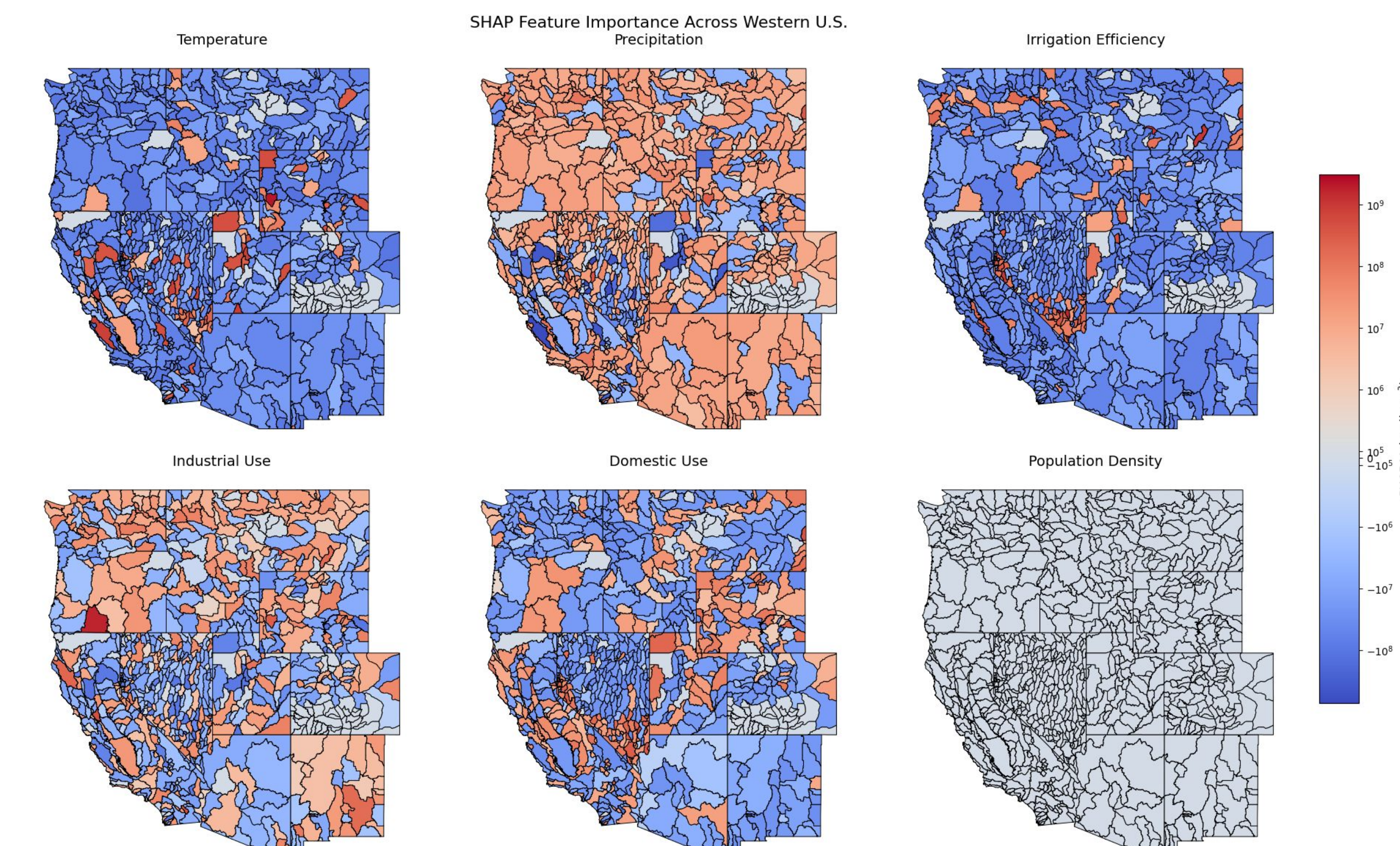


Figure 3. SHAP Feature Importance Maps. A spatial representation of SHAP values across different WMAs, showing the influence of various features such as temperature, precipitation, irrigation efficiency, domestic use, industrial use, and population density on the model's predictions.

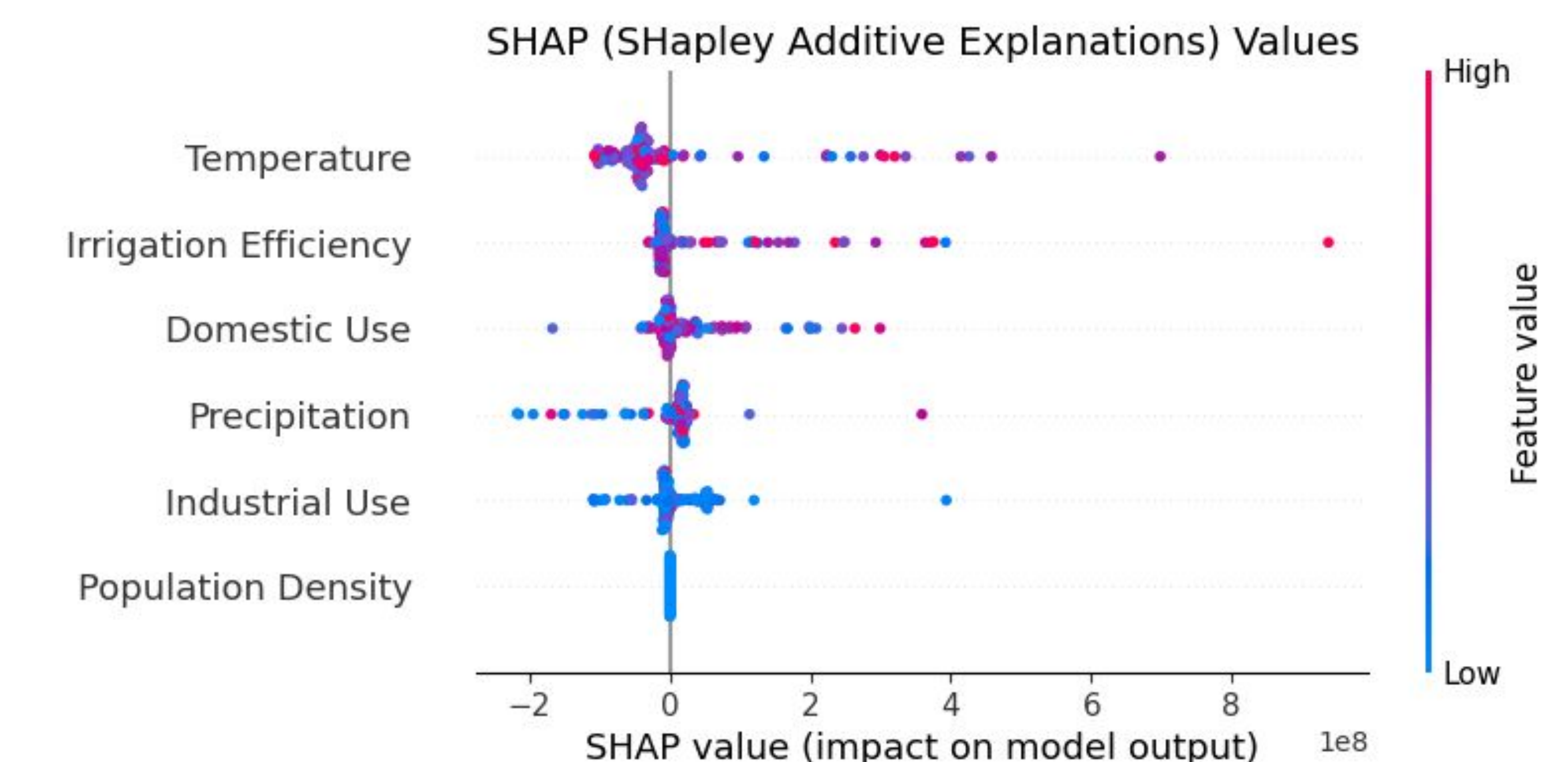


Figure 4. SHAP Summary Plot across the features. This plot visualizes the SHAP values for different features, showing their impact on the model's output. Red points indicate higher feature values, while blue points represent lower values. The X-axis (SHAP value) represents each feature's contribution to the model's output, where positive values increase unmet demand, and negative values reduce it. Each dot represents a WMA's feature impact, with a wider spread indicating variable influence (temperature, irrigation efficiency) and a concentrated spread suggesting a more consistent effect (industrial use, population density).

Feature Importance Analysis

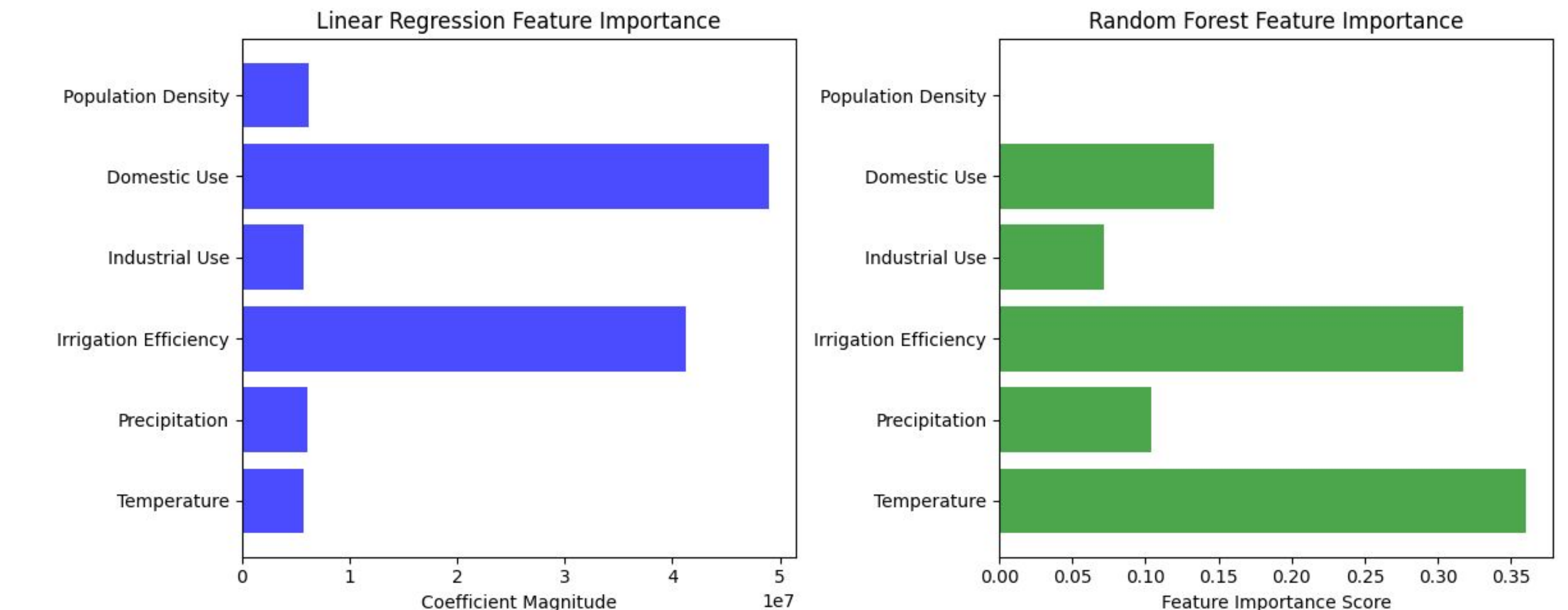


Figure 5. Feature Importance Comparison (Linear Regression vs. Random Forest). The left bar chart displays the coefficient magnitudes for linear regression, indicating the relative importance of each feature, while the right bar chart shows feature importance scores from a Random Forest model, highlighting key factors influencing predictions.

Key Findings

- **Temperature and Irrigation Efficiency:** These emerge as the most influential variables, with predominantly negative impacts and significant spatial variation across WMAs.
- **Precipitation:** Unlike temperature, precipitation has a predominantly positive impact, increasing model predictions across most of the west.
- **Domestic and Industrial Use:** Domestic water use generally has stronger and more widespread positive impacts on predictions than industrial use, though both show notable spatial variability across the Western U.S.
- **Population:** Population density has minimal direct impact on unmet demand.
- **Regional Unmet Demand Variability:** Certain regions face extreme water deficits, highlighting the critical need for tailored management strategies.

Conclusion and Discussion

- **Policy and Management Strategies:** Findings suggest that improving irrigation efficiency could significantly alleviate unmet demand. Additionally, implementing adaptive water management strategies tailored to regional climate conditions is crucial.
- **Modeling Improvements:** The contrast between linear regression and random forest feature importance underscores the need for robust, non-linear modeling techniques to capture complex water demand dynamics.
- **Climate Change Considerations:** Given the strong influence of temperature and precipitation, future work should incorporate climate projections to assess long-term water security risks.

References

1. Grogan, D., Lisk, M., Zuidema, S., Zheng, J., Fisher-Vanden, K., Lammers, R., Olmstead, S., Fowler, L., & Prusevich, A. (under review). Bringing Hydrologic Realism to Water Markets. MSD-LIVE Data Repository. <https://doi.org/10.57931/2283495>
2. Grogan, D. S., Zuidema, S., Prusevich, A., Wollheim, W. M., Glidden, S., & Lammers, R. B. (2022). Water Balance Model (WBM) v.1.0.0: A Scalable Gridded Global Hydrologic Model with Water-Tracking Functionality. Geoscientific Model Development, 15(19), 7287–7323. <https://doi.org/10.5194/gmd-15-7287-2022>

Acknowledgement

This work was supported by the U.S. Department of Energy, Office of Science, Biological and Environmental Research Program, Earth and Environmental Systems Modeling, MultiSector Dynamics under Cooperative Agreement DE-SC0022141.