Abstract:
Accurate medium-range weather forecasting is a longstanding challenge due to the chaotic nature of the atmosphere and the imperfect representation of physical processes. Classical Numerical Weather Prediction (NWP) systems are highly skillful but computationally intensive, limiting accessibility for many operational users. Recently, purely data-driven global AI models (e.g., FourCastNet, Pangu-Weather, GraphCast, FuXi) have emerged as efficient alternatives that produce forecasts rapidly on a single GPU; yet, like NWP, they exhibit persistent region- and variable-dependent systematic biases. This work develops and evaluates a machine-learning based multimodel superensemble (MMSE) that reduces those biases and enhances the forecast skill of modern AI weather models. The approach builds on the superensemble principle of bias-aware, performance-weighted blending (rather than equal-weight averaging), adapted here to AI forecasts with interpretable learning and spatiotemporal deep networks that preserve geophysical structure. Ground truth is provided by ERA5 and GPM IMERG for precipitation, with inputs harmonized at 0.25° and 6-hour intervals. To illustrate utility across distinct users and regimes, MMSE is configured for three applications: (i) January 10 m winds over Germany for renewable-energy operations (wind-ramp planning, day-ahead scheduling, grid balancing); (ii) May 2 m air temperature over India for early detection of heatwave conditions for public-health advisories or demand management; (iii) July rainfall over India for core monsoon operations (flood early warning, reservoir operations, agricultural decisions). We first implement a tabular, 2D XGBoost MMSE (XGB–MMSE), which performs strongly on absolute-error metrics (RMSE) but can dilute spatial-anomaly structure (ACC) when flattening fields. We therefore develop a 3D CNN MMSE (CNN–MMSE) that ingests full spatiotemporal forecast volumes, preserving coherent patterns and improving both RMSE and ACC at medium leads. Model interpretability is maintained via SHAP-based explanations that expose condition-dependent weighting and each contributor's relative strengths. Across all three configurations, the MMSE consistently outperforms its constituent AI models, demonstrating a forecast gain of ~1 day. Here, "forecast gain" denotes the difference in lead time at which the MMSE and a comparison model achieve the same accuracy (e.g., at a fixed RMSE/ACC threshold). This one-day head start enables earlier, better-targeted actions—issuing advisories, pre-positioning resources, and optimizing operations—thereby improving safety and reducing costs. Once trained, the AI-only super-ensemble blends multi-model forecasts to produce full 15-day products on a single GPU in a few minutes. The methodology is modular and readily retrainable for other regions, seasons, or variables, enabling broader adoption in renewable energy, disaster preparedness, agriculture