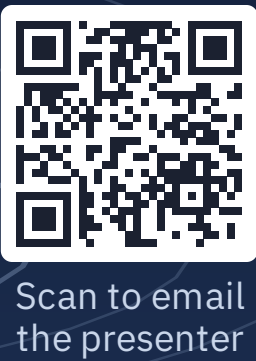


Hydro-Meteorological Drivers of Tomato Yield Variability Revealed through Explainable Machine Learning and Sensitivity Analytics



Pashupati Nath Singh^{1*}, Praval Upadhy¹, Prashant K. Srivastava¹

¹Remote Sensing Laboratory, Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi 221005, India ^{*}pashupati1110@bhu.ac.in

Warm nights, sunshine and surface pressure dominate tomato-yield variability across Madhya Pradesh, **mostly by acting together**. A regularised linear model predicts it with about one-tenth the error of the best deep network.

51 districts

10 hydro-met predictors

6 seasons, 2019–25

20 models, one protocol

306 district-seasons

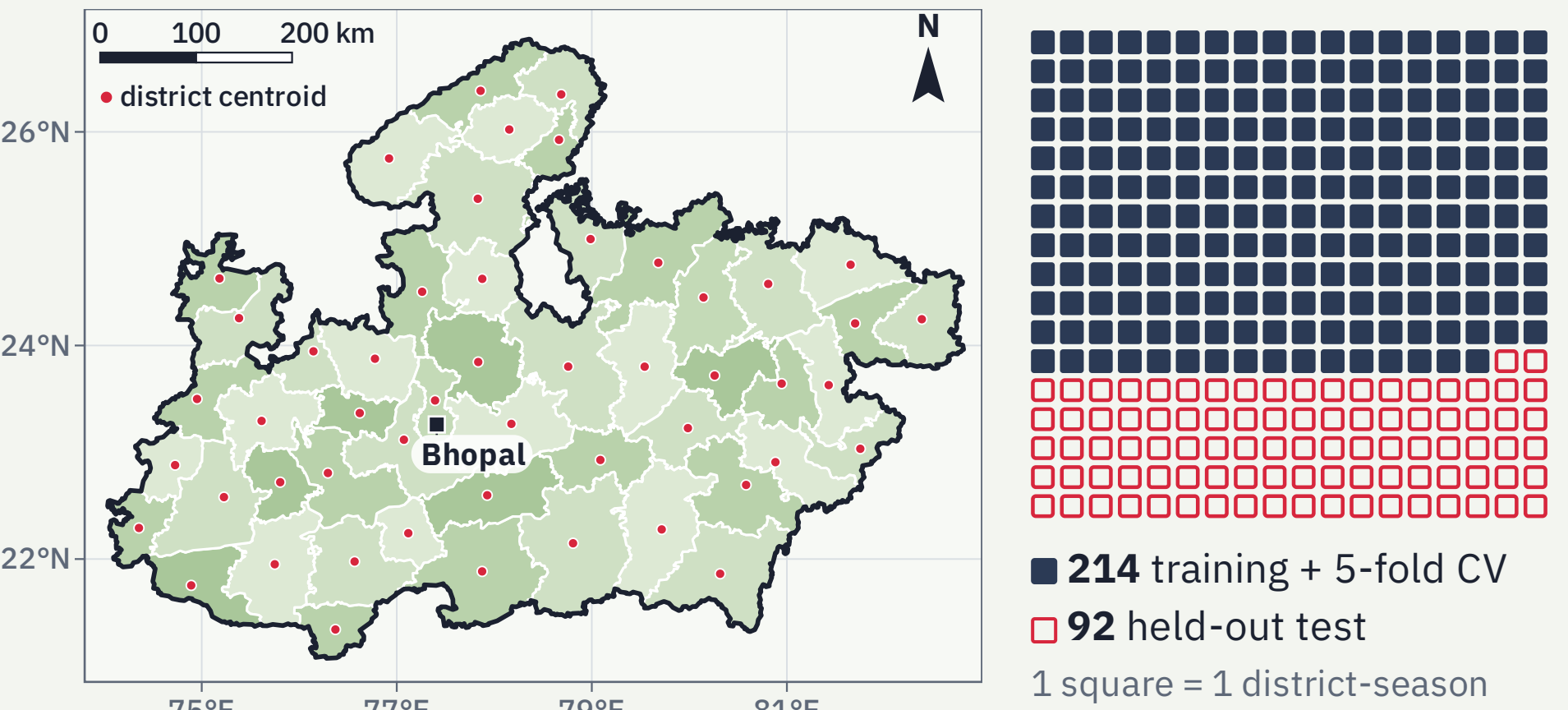
6 XAI and sensitivity tools

Why look up to forecast tomatoes?

Madhya Pradesh is one of India's leading tomato states (more than 3.5 million tonnes a year), grown mostly rain-fed in the kharif season. Monsoon timing, heat episodes and patchy soil moisture make yields swing between districts and years. Data-driven forecasts exist, yet machine learning (ML) and deep learning (DL) are seldom tested under one protocol, and few studies show *which atmospheric processes* a model actually relies on.

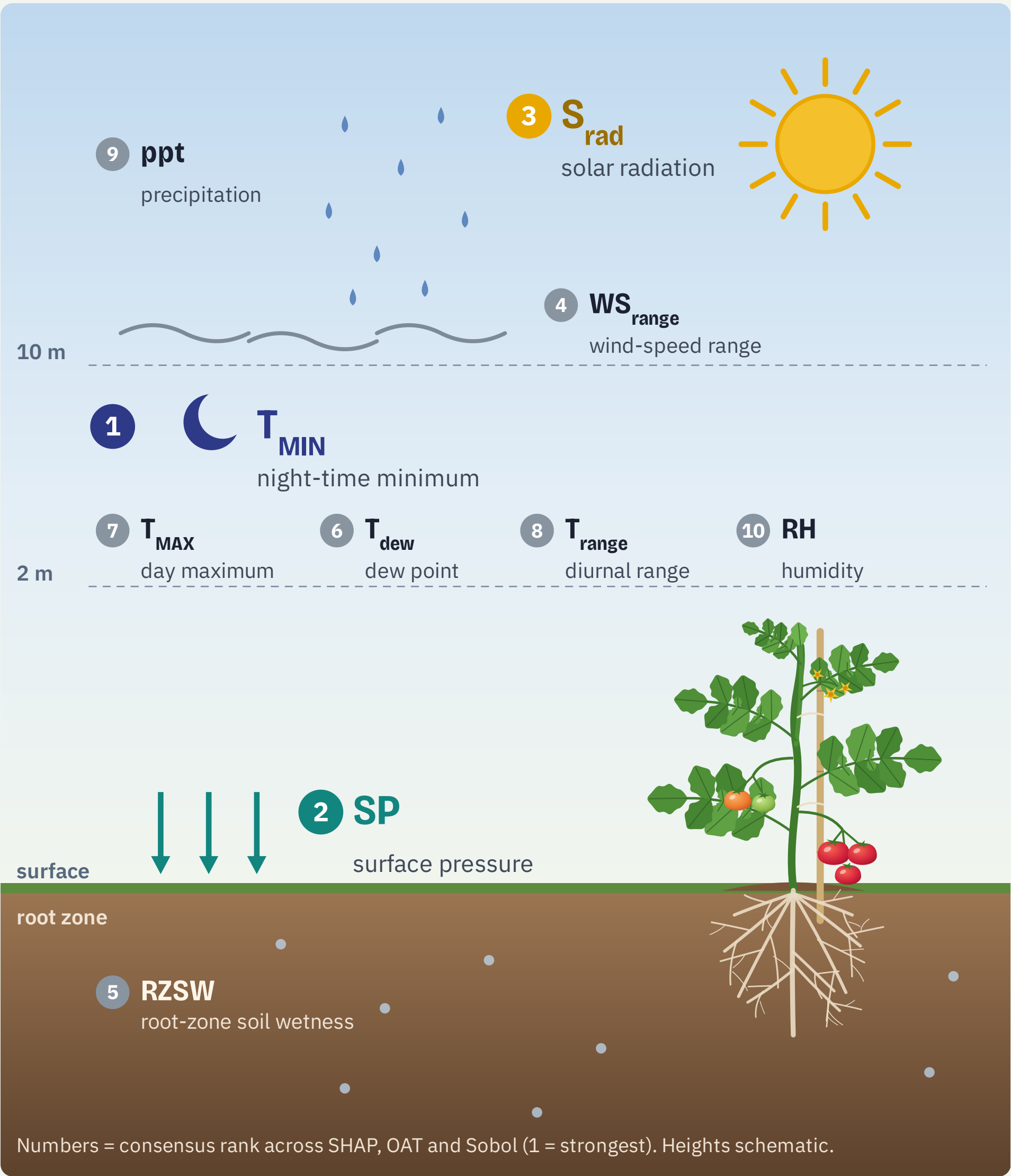
Aim: identify the hydro-meteorological controls on district-scale tomato yield and the model family that captures them best, with every prediction explained.

Study domain and data



The **51 districts** (dots) span 21.1–26.9°N and 74.0–82.8°E, from the semi-arid north-west (~750 mm rain a year) to the sub-humid south-east (>1,600 mm). **Climate:** NASA POWER / MERRA-2 reanalysis, daily values averaged per season. **Yield:** Directorate of Horticulture, MP, cross-checked with National Horticulture Board records for six seasons, 2019–20 to 2024–25; the test split is stratified by yield.

Ten predictors, sky to soil



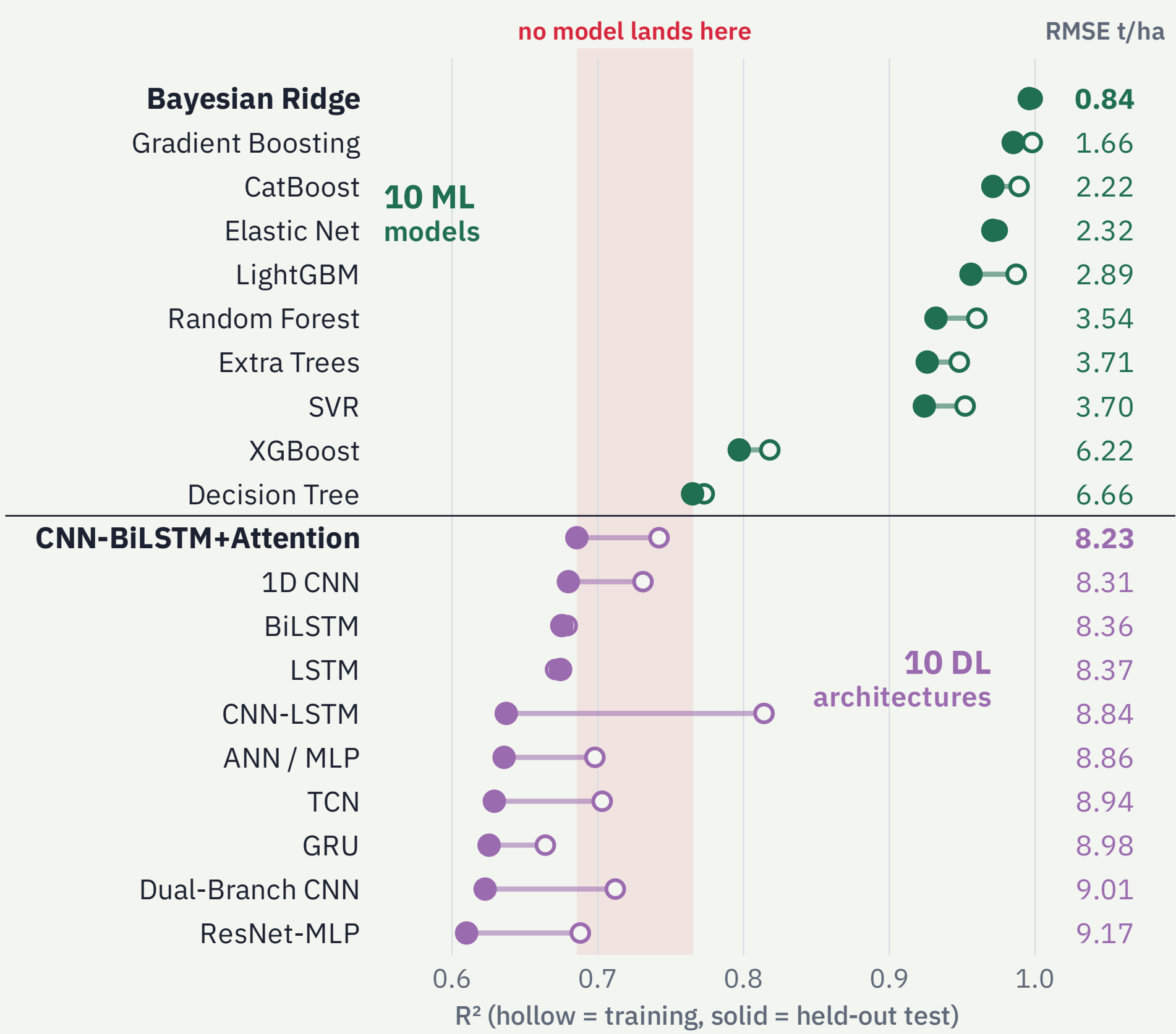
Plus latitude, longitude and an encoded district term: 13 base inputs per district-season.

Workflow

- Assemble** reanalysis and yield records into a gap-free 306 × 13 matrix.
- Prepare:** log-transform yield (skewness 1.43), encode districts, scale inputs, add three physics-based indices: heat stress $T_{MAX}(1 - RH/100)$, moisture supply $ppt \times RZSW$, and $S_{rad} \times T_{MIN}$.
- Compete:** 10 ML vs 10 DL models on the same stratified 70/30 split and 5-fold CV; Optuna tuning (100 trials) for ML, grid search with early stopping for DL.
- Score** with R^2 , RMSE, MAE, MAPE and the hydrological skill scores NSE and KGE.
- Explain and stress-test** the tree-based XGBoost learner with SHAP, LIME, OAT, Sobol, PDP/ICE and input noise.

Does a deeper network predict better?

No. Every one of the 10 ML models beat every one of the 10 DL architectures on held-out data.



Deep networks under-fit. They reach only R^2 0.66–0.81 even on their own training data, the signature of too few samples (214 training records) to constrain deep weights, not of over-fitting. Attention adds just $\Delta R^2 = 0.05$ over a plain MLP, and BiLSTM gains less than 0.001 over LSTM.

Best of each family

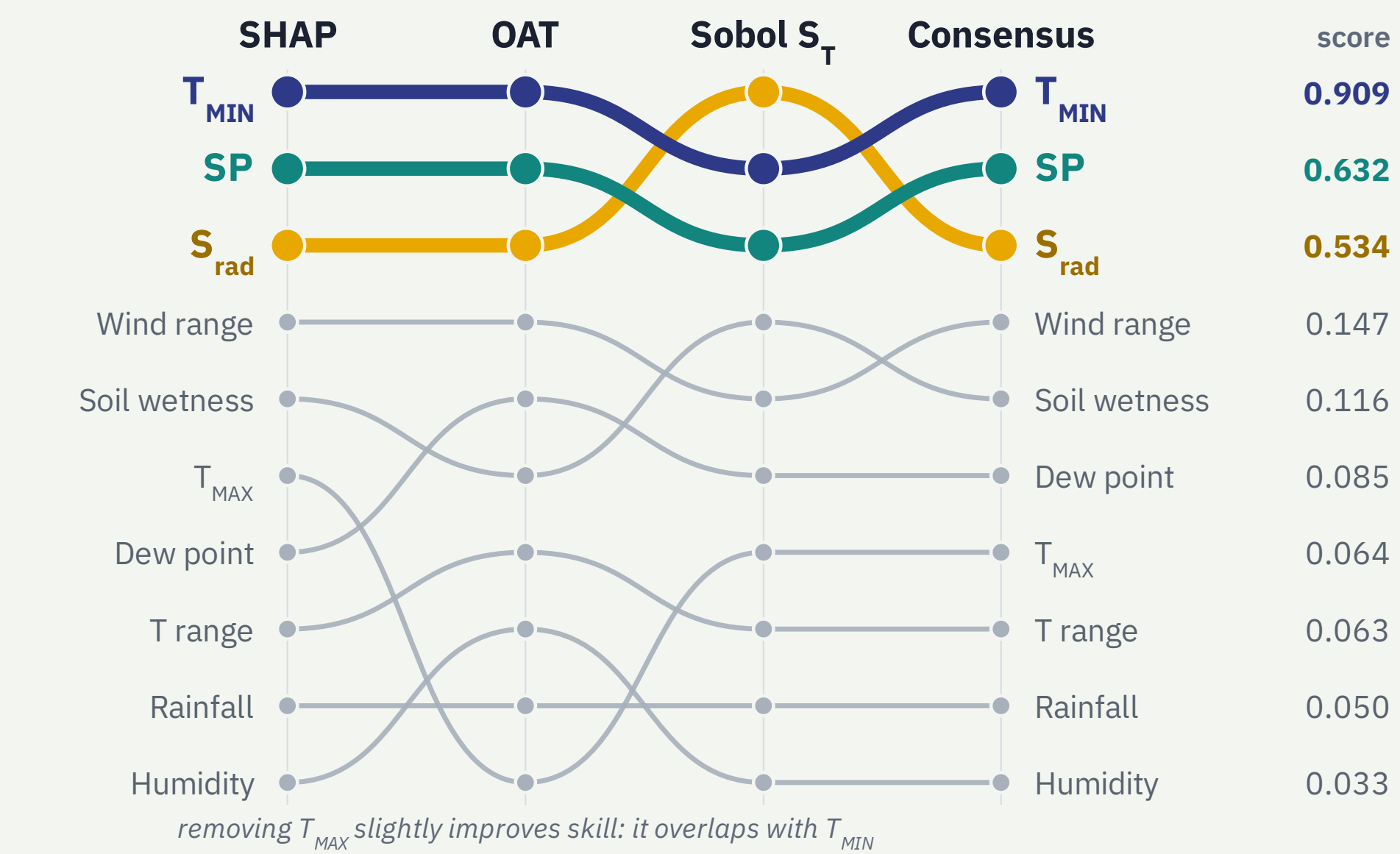
92 test district-seasons	Bayesian Ridge (ML)	CNN-BiLSTM + Attention (DL)
R^2	0.996	0.686
RMSE (t ha ⁻¹)	0.84	8.23
MAE (t ha ⁻¹)	0.65	6.26
MAPE (%)	2.3	23.6
Nash–Sutcliffe efficiency	0.996	0.686
Kling–Gupta efficiency	0.995	0.697
Mean bias (t ha ⁻¹)	+0.00	+0.41

Why a linear model wins here

- Log-transformed yield is close to Gaussian, which suits a probabilistic linear model.
- Bayesian priors shrink the strongly correlated temperature terms instead of fitting their noise.
- Each record is a single season, not a sequence, so recurrence and attention have little to work with.
- The physics-based indices hand the linear model the non-linearity it would otherwise lack.

Which part of the atmosphere sets the harvest?

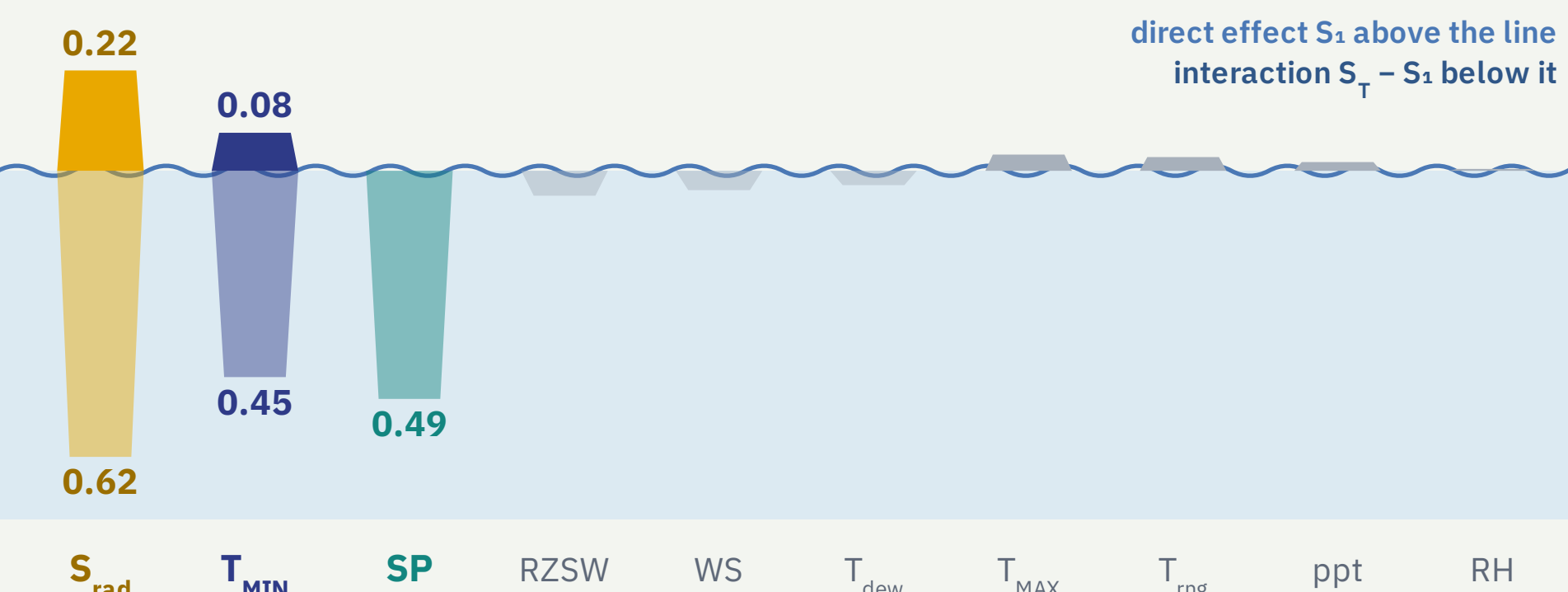
Night-time minimum temperature, surface pressure and solar radiation fill the top three places in every ranking.



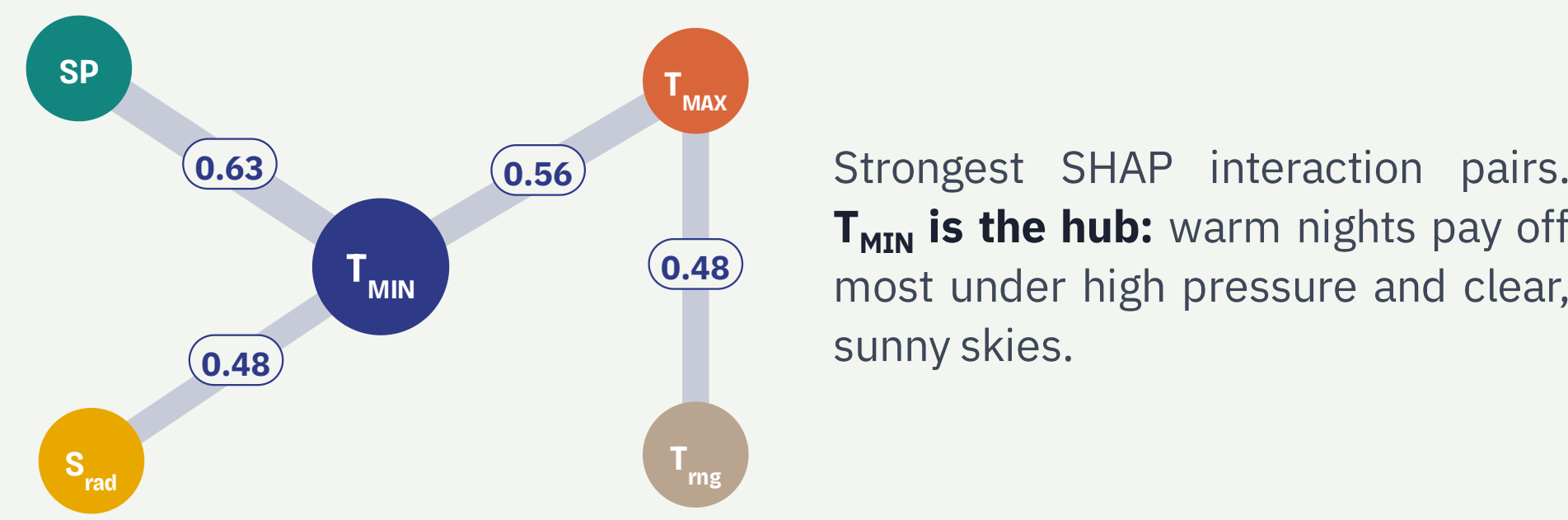
Rankings from the XGBoost learner; consensus score weights SHAP 40%, OAT 35% and Sobol S_T 25%. **Rainfall and humidity never rise above 8th:** soil wetness already carries the rain signal, and kharif rains usually meet crop demand, so heat and light are the binding limits.

Do the drivers act alone?

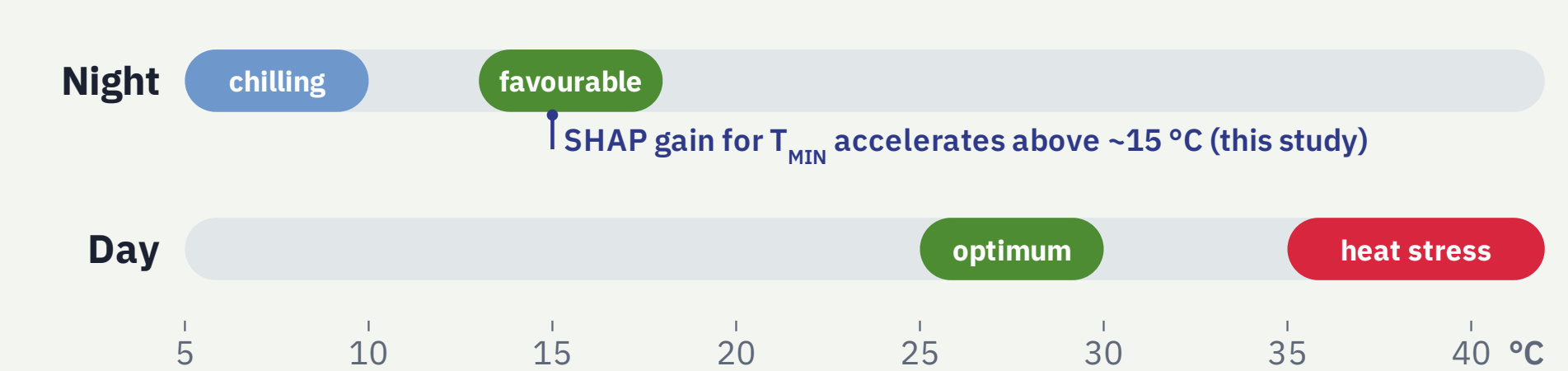
Rarely. Sobol variance decomposition shows most of their influence is shared.



S_{rad} has the largest total effect ($S_T = 0.84$) but a direct effect of only 0.22; surface pressure acts almost entirely through interactions. One-at-a-time tests would miss most of this.



Reading the drivers as crop physiology

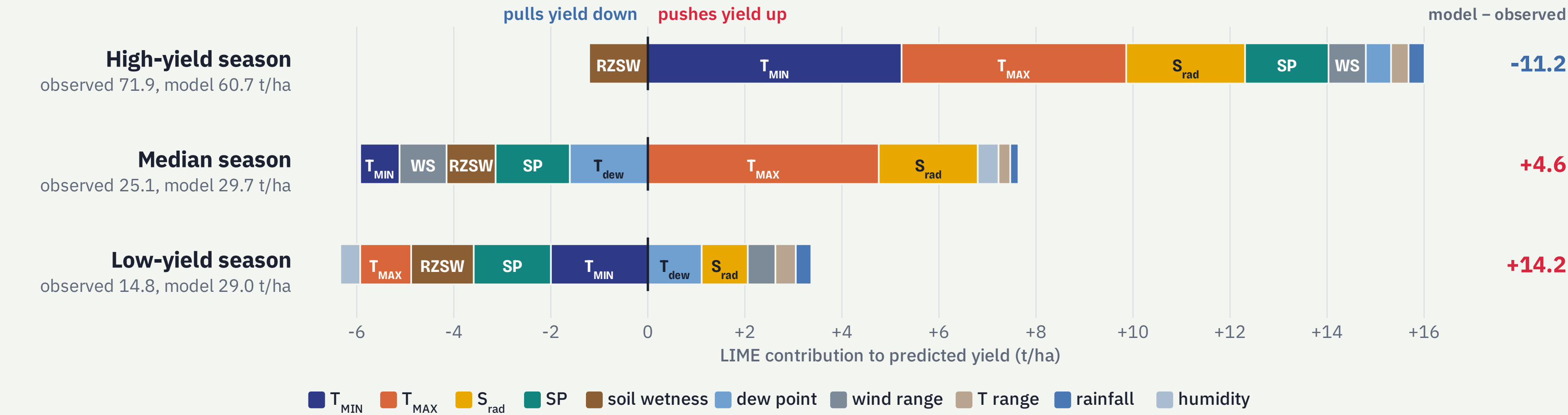


Temperature windows from tomato physiology (Peet & Wolfe, 2000). High surface pressure tracks clear-sky, high-radiation weather; low pressure brings cloud, humidity and disease pressure.

How stable are these answers? Adding ±5–10% Gaussian noise to the inputs lowered R^2 by only 0.02–0.05, within the typical uncertainty of reanalysis fields. SHAP, LIME, OAT and Sobol rest on different mathematics, yet point to the same drivers.

What does one district-season look like?

LIME splits three held-out XGBoost predictions into the local push of all ten predictors.



Night warmth, day temperature and sunshine lift the high-yield season; night temperature, surface pressure and soil wetness pull the low-yield one down. That season is still over-predicted by 14.2 t ha⁻¹, pointing to pest, disease or management losses that weather alone cannot see.

From atmosphere to action

- Measure what matters**
Networks in MP are built around rainfall; denser sub-district T_{MIN} and surface-pressure observations would sharpen forecasts most.
- Forecast lean**
Five predictors carry about 80% of SHAP importance, so a compact model fed by the NASA POWER API can run every season.
- Assess climate impacts with interactions**
Cloud- or aerosol-driven shifts in S_{rad} reach yield mainly through couplings with warmth and pressure.
- Match the model to the data**
With a few hundred tabular records, start with regularised ML before deep learning.

What is new, to our knowledge

- First head-to-head benchmark of 10 ML and 10 DL models for tomato yield.
- A four-method explanation consensus: SHAP, LIME, OAT and Sobol.
- Surface pressure revealed as a strong, previously unreported yield driver.

Limits and next steps

Six seasons cannot span ENSO or IOD extremes, district averages hide sub-district contrasts, and pests, soils and management are not yet inputs. Next: longer records, Sentinel-2 and MODIS indices, CMIP6-driven seasonal outlooks, and transfer to other South Asian tomato belts.

Key references

Lundberg & Lee (2017) *NeurIPS* 30 · Ribeiro et al. (2016) *KDD* 1135–1144 · Saltelli et al. (2019) *Environ. Model. Softw.* 114, 29–39 · Grinsztajn et al. (2022) *NeurIPS* 35 · Tipping (2001) *JMLR* 1, 211–244 · Peet & Wolfe (2000) *Climate Change and Global Crop Productivity*, 213–243

Acknowledgements

NASA POWER and GMAO MERRA-2 for open climate data; Directorate of Horticulture, Government of Madhya Pradesh, for district yield statistics.