University of Reading

Building and interpreting a data-driven energy demand model for Indian states

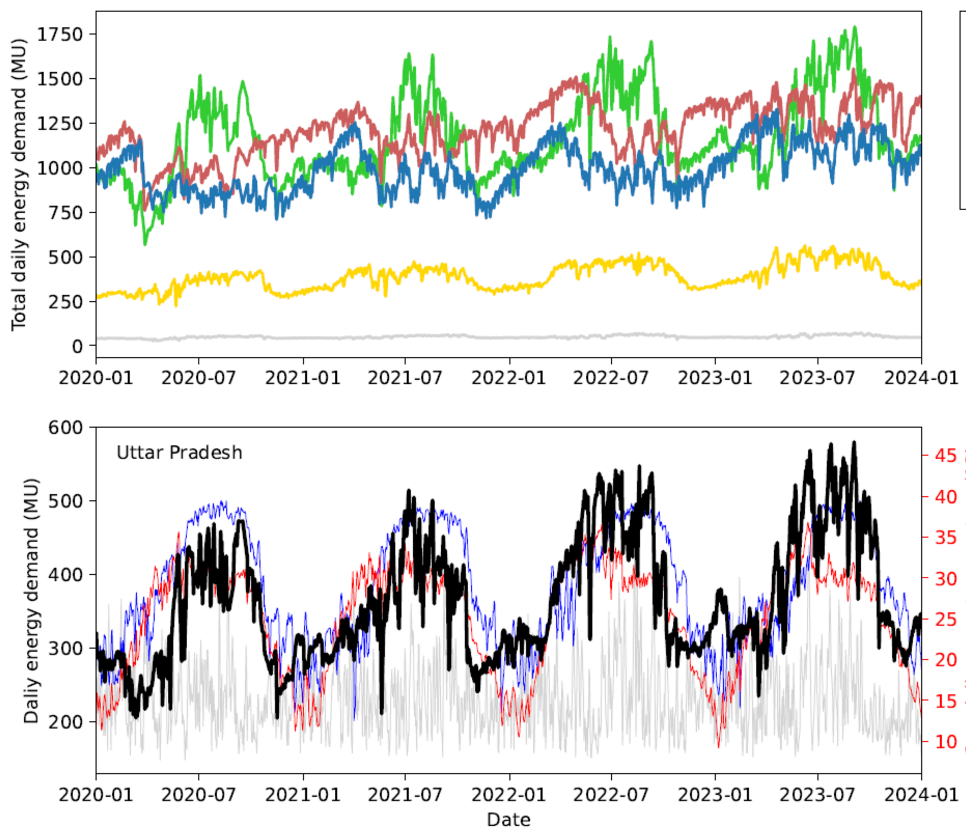
Kieran M R Hunt (University of Reading) and Hannah Bloomfield (University of Newcastle)

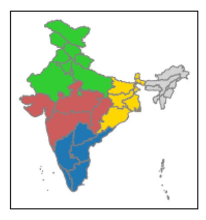
With India rapidly decarbonising its energy grid, reliable forecasts of potential energy deficit, and hence energy demand, are required. Here, we build an accurate and interpretable data-driven modelling pipeline that accurately predicts weather-based demand in Indian states.

1. Methods

XGBoost is a scalable and efficient machine learning algorithm that builds an ensemble of additive decision trees, with each new tree attempting to correct the errors made by the previous trees. We build one XGBoost model for each state, using weather variables as predictors, and energy demand as the predictand.

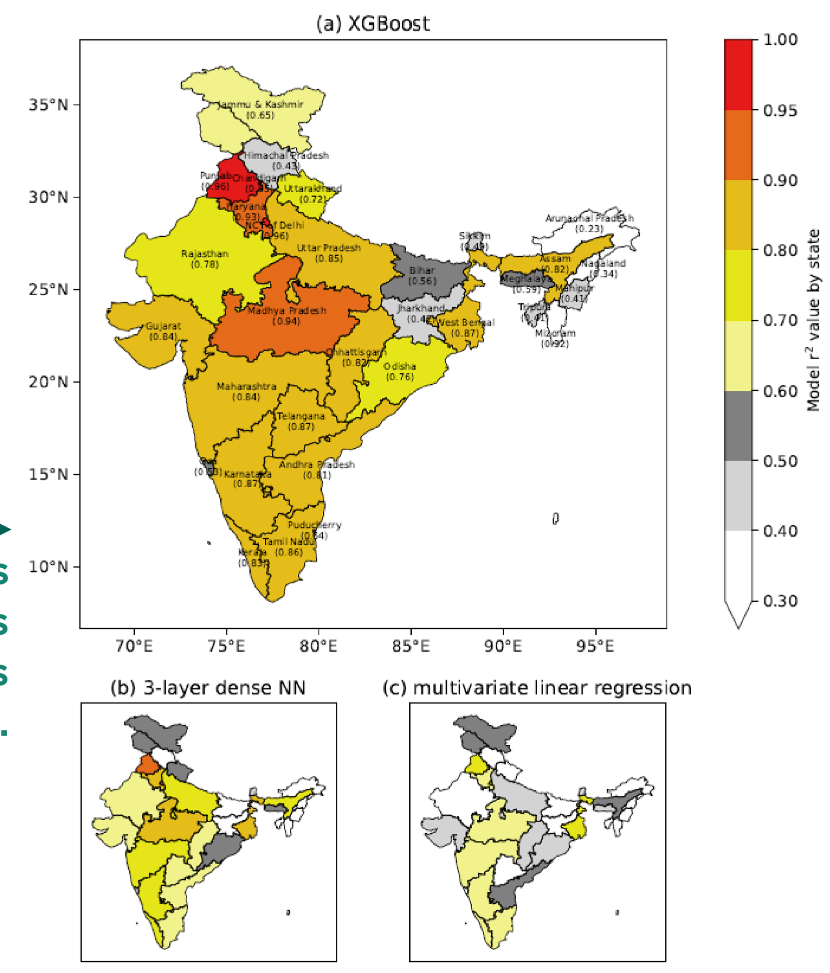
Demand data are scraped from daily reports released by Grid India. These are quality controlled and detrended, and the variance is calibrated to 2022 economic conditions. Weather data are from ERA5, weighted by population and averaged over each state, including daily minima, means, and maxes, as well as 7-day and 30-day means. We also include days of the week and holidays.





Energy demand varies by day, season, and region, but is predominantly weather-driven. However, we also need to correct for trends in the mean and variance.

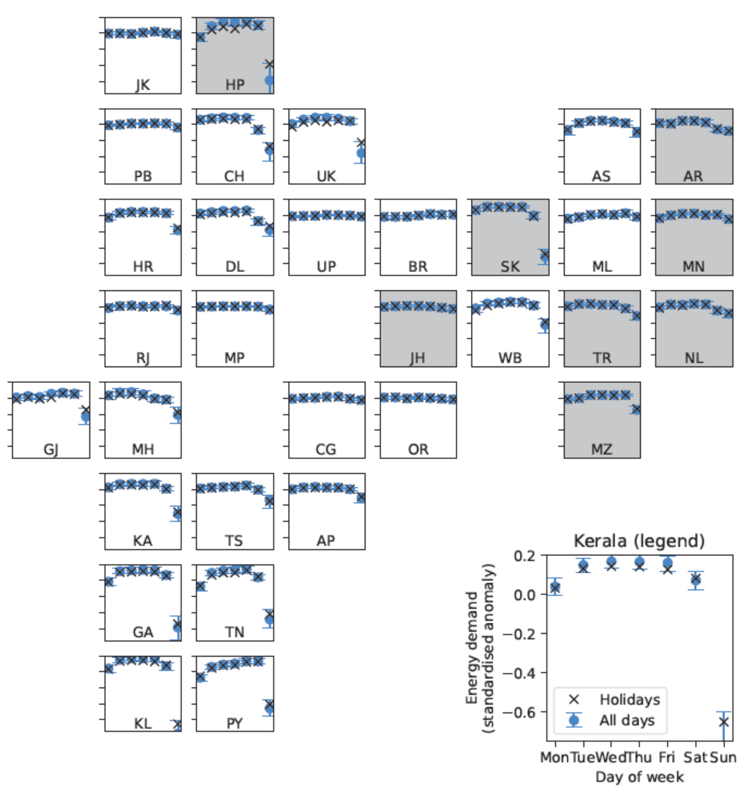
Statewise XGBoost models outperform both deep neural nets and traditional regression. This is useful for explainability.



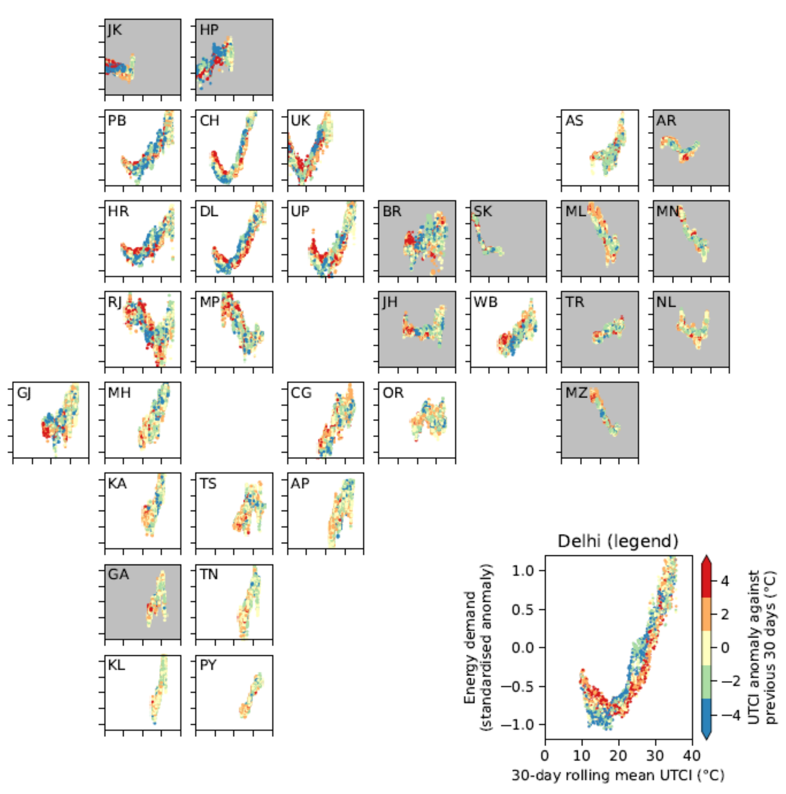
2. Explainability

Explainable models have several benefits. They allow stakeholders to understand and trust the models, but also give us interpretable insights into the complex relationships between predictors and predictions. For this, we use Shapley analysis, which works by distributing the total contribution to a prediction among the predictors in a model based on their marginal contributions. This gives one value per predictor per prediction, quantifying its importance.

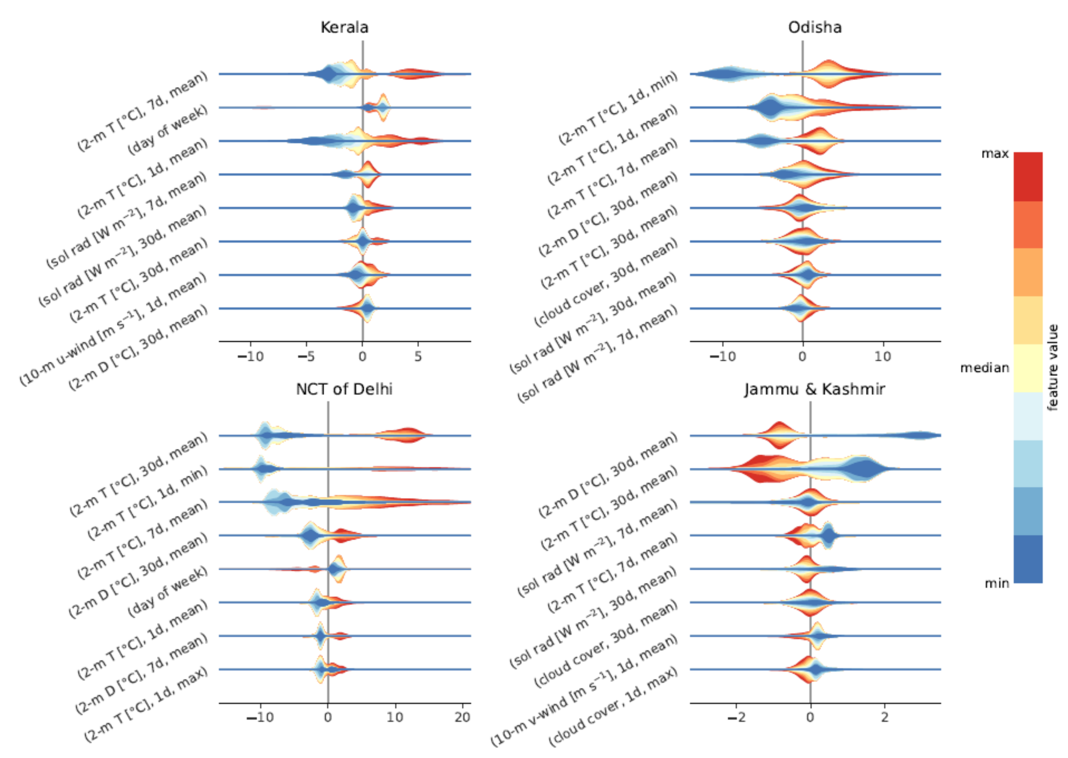
Shapley values can be used to understand nonlinear responses such as the increased demand at both low temperatures (heating) and high temperatures (air conditioning). They can also be used to identify key predictors and their relationships with each other.



Energy demand is sensitive to days of the week and state/national holidays – especially in the south.



Demand has expected U-shape relationship with thermal stress. However, point of inflection varies by state.

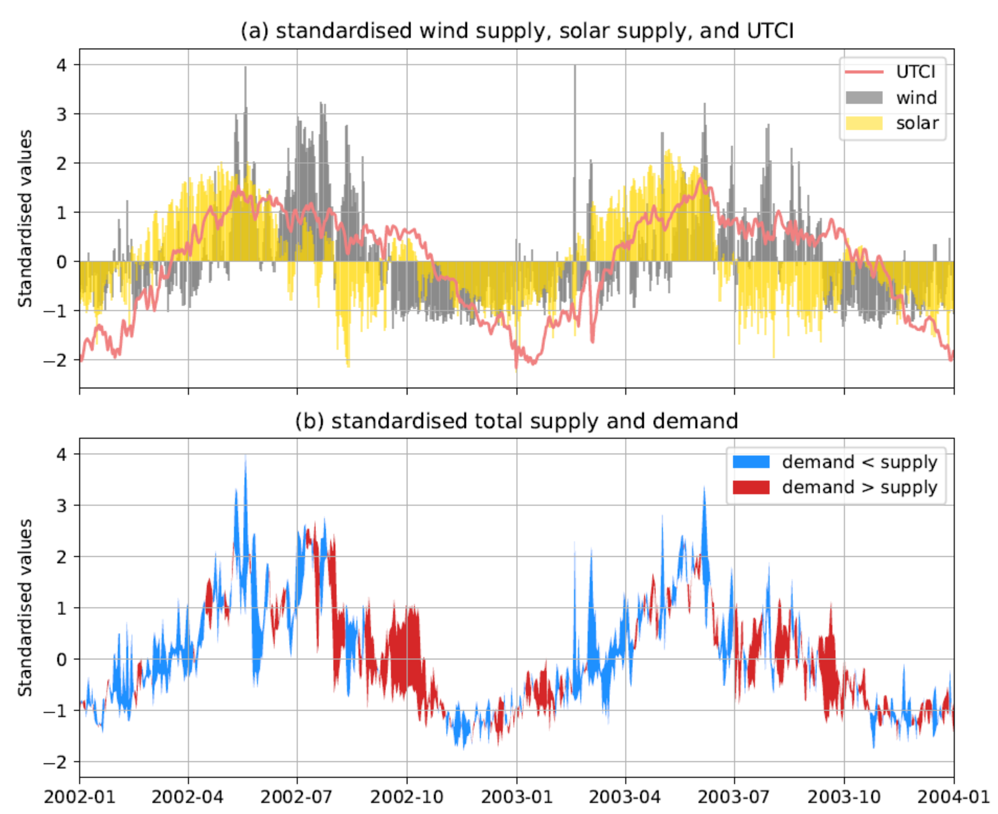


Distributions of Shapley values show a range of linear and nonlinear responses. Typically, demand is more sensitive to seasonal variables rather than daily ones, and to temperature rather than dewpoint.

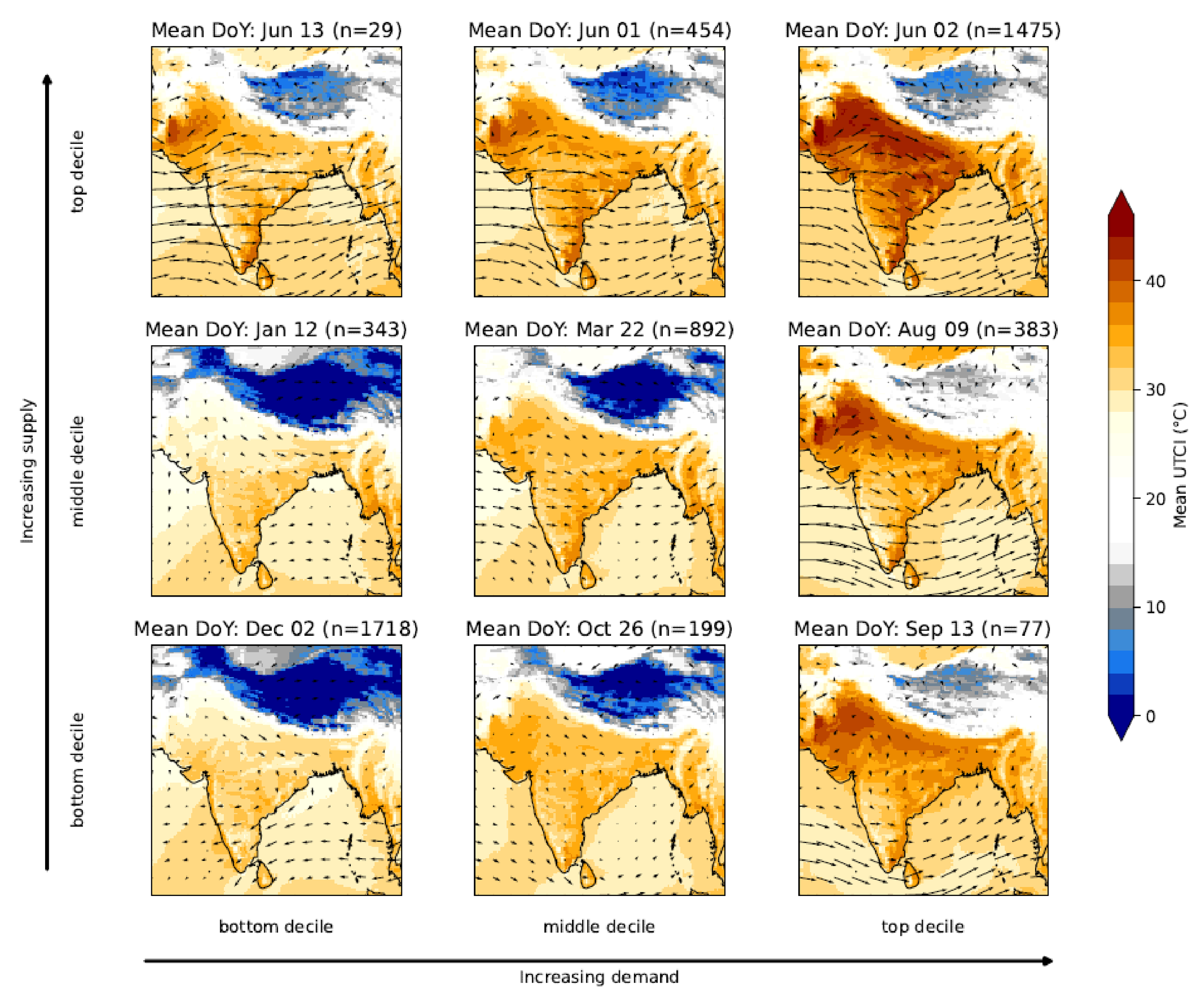
3. Impacts

By driving the demand models with historic reanalysis data, we can estimate what energy demand (either for a given state, or by adding them up, all India) would have been on each day from 1979 onwards, assuming 2022 socioeconomic conditions. We can then combine this dataset with a similar one constructed for wind and solar supply over India (Hunt and Bloomfield, 2024), to identify potential energy deficits, caused either by weak generation or surplus demand.

On seasonal timescales, production typically follows demand, both peaking in the summer months and reaching minima in the winter. However, they do not match perfectly, leading to risky deficits in September. This motivates the use of weather regimes, whose predictability over India usefully extends the time available for mitigation (Dijkstra et al, 2024). In this example, we see an early monsoon withdrawal is the cause of both high demand and low generation.



Modelled all-India demand and supply, calibrated to 2022. Wind and solar have high complementarity (top), but both are weak during the monsoon withdrawal, leading to potentially large deficits (bottom).



Weather regimes associated with high demand and low supply (bottom right) are typical of the monsoon withdrawal: weakening winds, high temperatures, and high humidity.

References and resources

Chen and Guestrin (2016). XGBoost: a scalable tree boosting system. *KDD '16*
Dijkstra, Bloomfield, and Hunt (2024). Identifying weather patterns responsible for renewable energy production droughts over India. *Adv Geo (submitted).*
Hunt and Bloomfield (2024). Quantifying renewable energy potential and realized capacity in India: Opportunities and challenges. *Met Apps.*
Lundberg and Lee (2017). A unified approach to interpreting model predictions. *Adv. Neur. Inf. Proc. Sys.*

Hunt and Bloomfield (2024)

Energy capacity and production data