# Exploring Explainability (XAI)

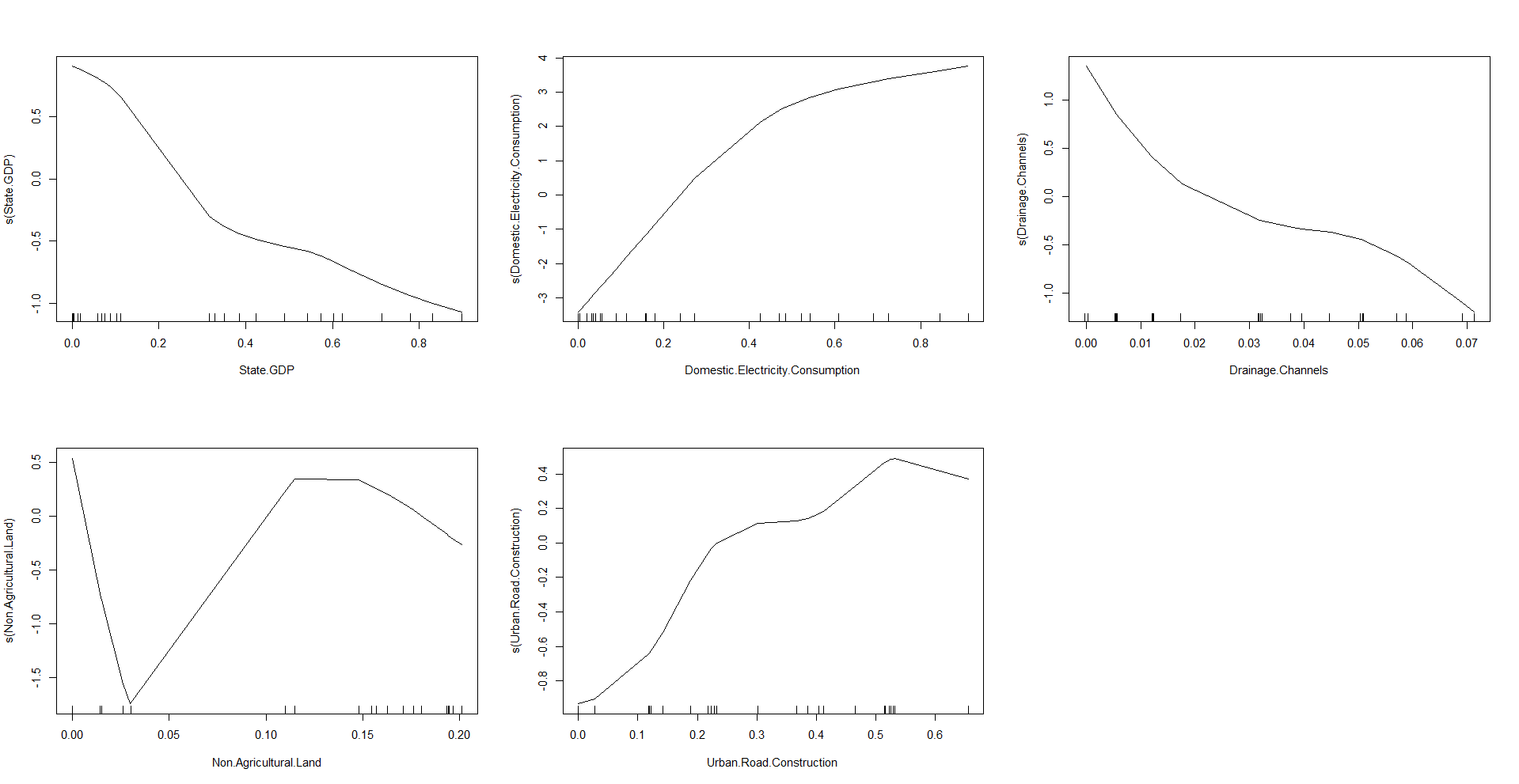
**Note on SHAP-Based Interpretability:**  
The SHAP analysis presented in this report is based on a relatively small dataset comprising 25 annual observations. While SHAP values offer valuable insights into feature-level contributions to individual predictions, caution must be exercised in interpreting global patterns and rankings. In small samples, SHAP values may be sensitive to outliers, limited permutations, and potential multicollinearity, which can lead to inflated or unstable attributions. The visualizations and explanations provided here should therefore be viewed as **exploratory aids** rather than definitive indicators of causal relationships or feature importance.

## Generalized Additive Models (GAM)

### Inherent Interpretability via Smooth Plots

GAMs are inherently explainable due to their additive structure:

* Where,
* Each is a non-parametric smooth function estimated from data using splines.
* is the linear predictor on log-scale due to the canonical link for the NB family.

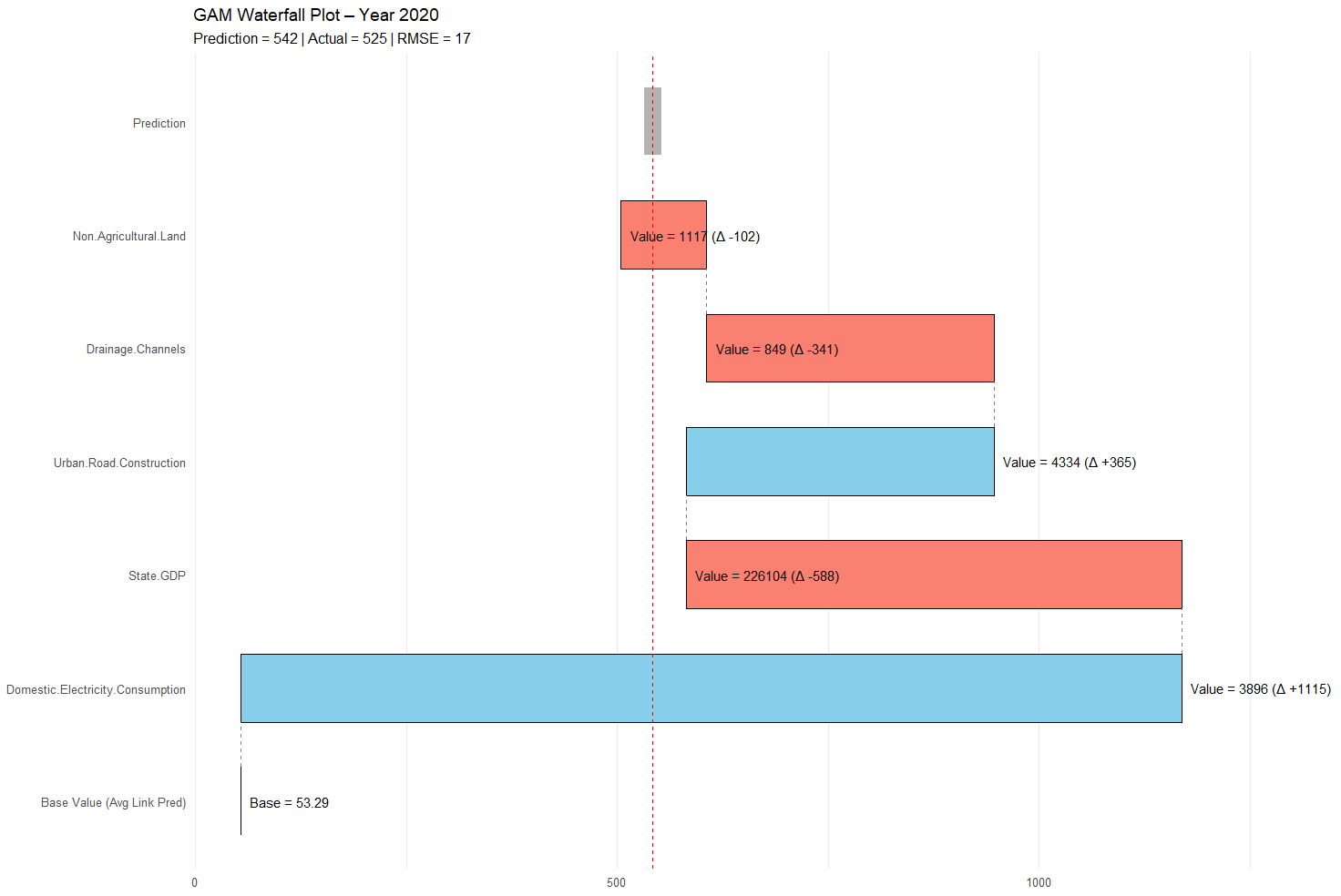


* The **x-axis** has normalized values of predictors, **y-axis** has centred smooth effect of the predictor. These are additive contributions to the log of expected flood counts. The **rug marks** on the x-axis indicate data density — where actual observations lie.
* **State GDP:** Decreasing effect → Higher GDP = Lower flood count? This is counter-intuitive, it may reflect non-linear effects or dominance of other correlated predictors.
* **Domestic Electricity Consumption:** The smooth is sharply increasing, especially for normalized values between 0.2 to 0.6. Suggests that higher electricity consumption (a proxy for urbanization) is associated with a higher flood risk, possibly due to increased impermeable surfaces.
* **Drainage Channels:** Declining curve → Better drainage may reduce floods. Makes physical sense, though the effect is subtle and not statistically significant.
* **Non-Agricultural Land:** Nonlinear "hill-shaped" relationship. Moderate values (0.05 to 0.15) increase flood risk, but both very low and very high values reduce it. Possible interpretation: urban fringe zones (moderately converted land) are more flood-prone than either fully rural or fully urbanized areas.
* **Urban Road Construction:** Increasing till a peak → Moderate infrastructure worsens flooding, but levels off. Again, reasonable — roads can worsen water runoff, but after a threshold, the marginal impact diminishes.

### GAM Waterfall Breakdown

A white paper with black text

AI-generated content may be incorrect.



* We start with the model’s baseline prediction of 53 floods.
* Domestic Electricity Consumption is very large and positive (∆ +1115).

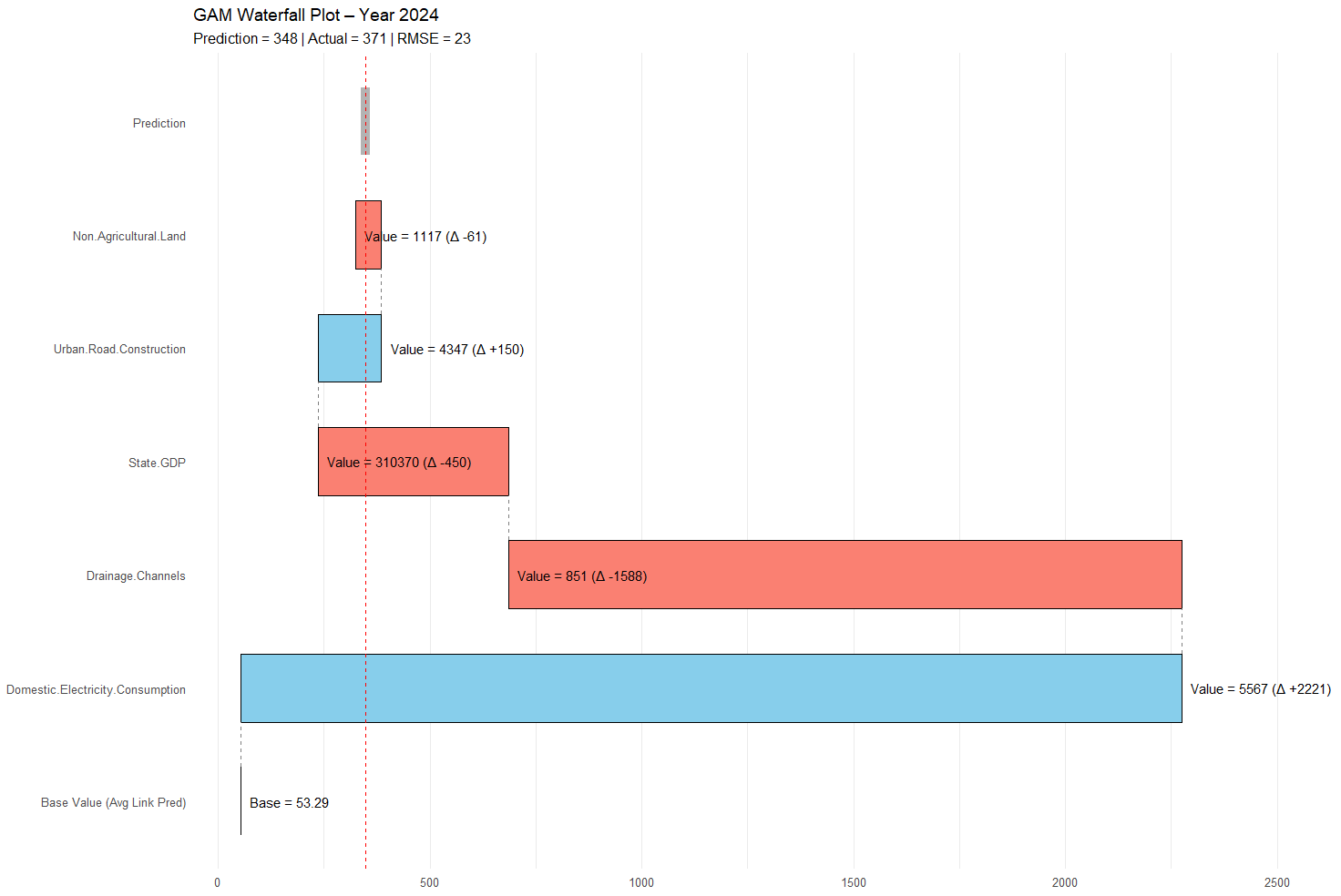
– In 2020 electricity usage was high; the model interprets this as dense urban growth ⇒ more impervious area ⇒ a big upward push in expected floods.

* Impact of State GDP is strongly negative (∆ ‑588).

– High GDP dampens flood risk in the model (perhaps reflecting better infrastructure spending).

* Urban Road Construction adds +365, reinforcing flood risk.
* Drainage Channels remove ‑341 because better drainage mitigates flooding.
* Non‑Agricultural Land removes ‑102 (a modest buffer effect).

Urban pressure (electricity + roads) drove floods up; GDP and drainage moderated but not enough to offset the surge, so the net outcome was still well above the baseline.



* Base value again ≈ 53 floods.
* Domestic Electricity Consumption: still high but Δ +2221 – even larger than 2020 (electricity growth accelerates).
* Drainage Channels turn strongly negative (∆ ‑1588), indicating major drainage improvements in 2024 that counteract part of the urban pressure.
* Impact of State GDP is still negative (∆ ‑450) – prosperity again buffers risk.
* Urban Road Construction is a small positive (∆ +150).
* Non‑Agricultural Land slightly negative (∆ ‑61).

Electricity growth alone would have exploded flood numbers, but big investments in drainage (and GDP‑related resilience) pulled the prediction back down. The model therefore attributes 2024’s moderate flood outcome to successful mitigation despite heavier urban demand.

**Take‑aways from the two years**

|  |  |  |
| --- | --- | --- |
|  | **2020** | **2024** |
| **Urban demand (Electricity)** | +1115 | +2221 |
| **Mitigation (Drainage + GDP)** | −929 combined | −2038 combined |
| **Net prediction** | 542 (close to real 525) | 348 (close to real 371) |

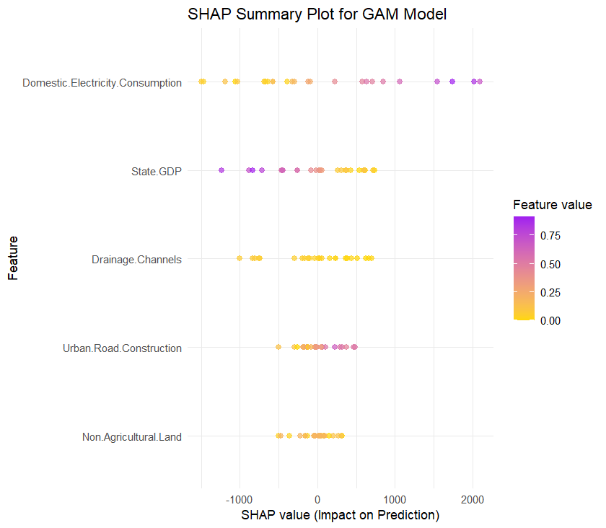
* In **2020** the upward push from electricity + roads was partly offset by GDP and modest drainage gains; flood pressure and mitigation were of similar order (≈ +1 480 vs − 1 030).
* In **2024** electricity pressure is much larger, but a **dramatic enhancement in drainage capacity** (− 1588) almost cancels it, leaving a lower net prediction—despite the bigger urban load.

Hence the chief discrepancies are:

1. **Electricity impact grows sharply.**
2. **Drainage flips from a moderate to a dominant suppressor.**
3. **Road‑construction and GDP cushioning effects both fade somewhat.**

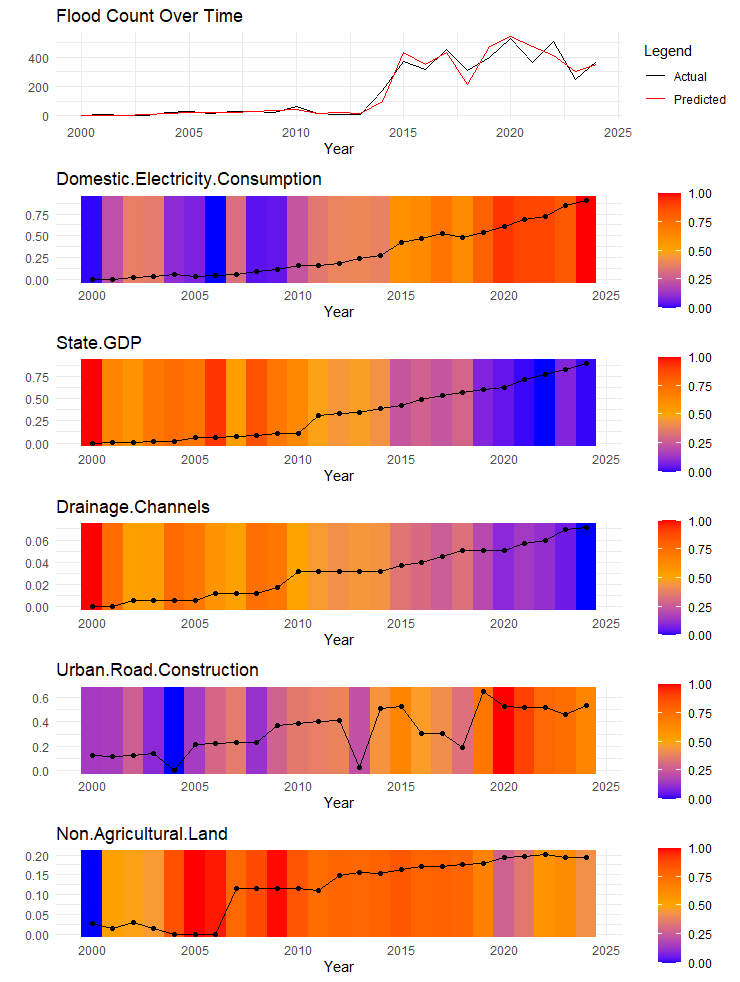
These shifts explain why the model moves from predicting a high flood count (~542) in 2020 to a much lower one (~348) in 2024 even though the urban‑demand signal is stronger—the mitigation variables behave very differently between the two years.

### SHAP Summary Plot



* Domestic electricity consumption and State GDP consistently have the strongest impact on model predictions.
* SHAP values for Domestic Electricity Consumption show a wider range, indicating a high level of influence in both directions depending on the feature value. Drainage Channels and Non-Agricultural Land have a narrower spread and smaller impact overall.

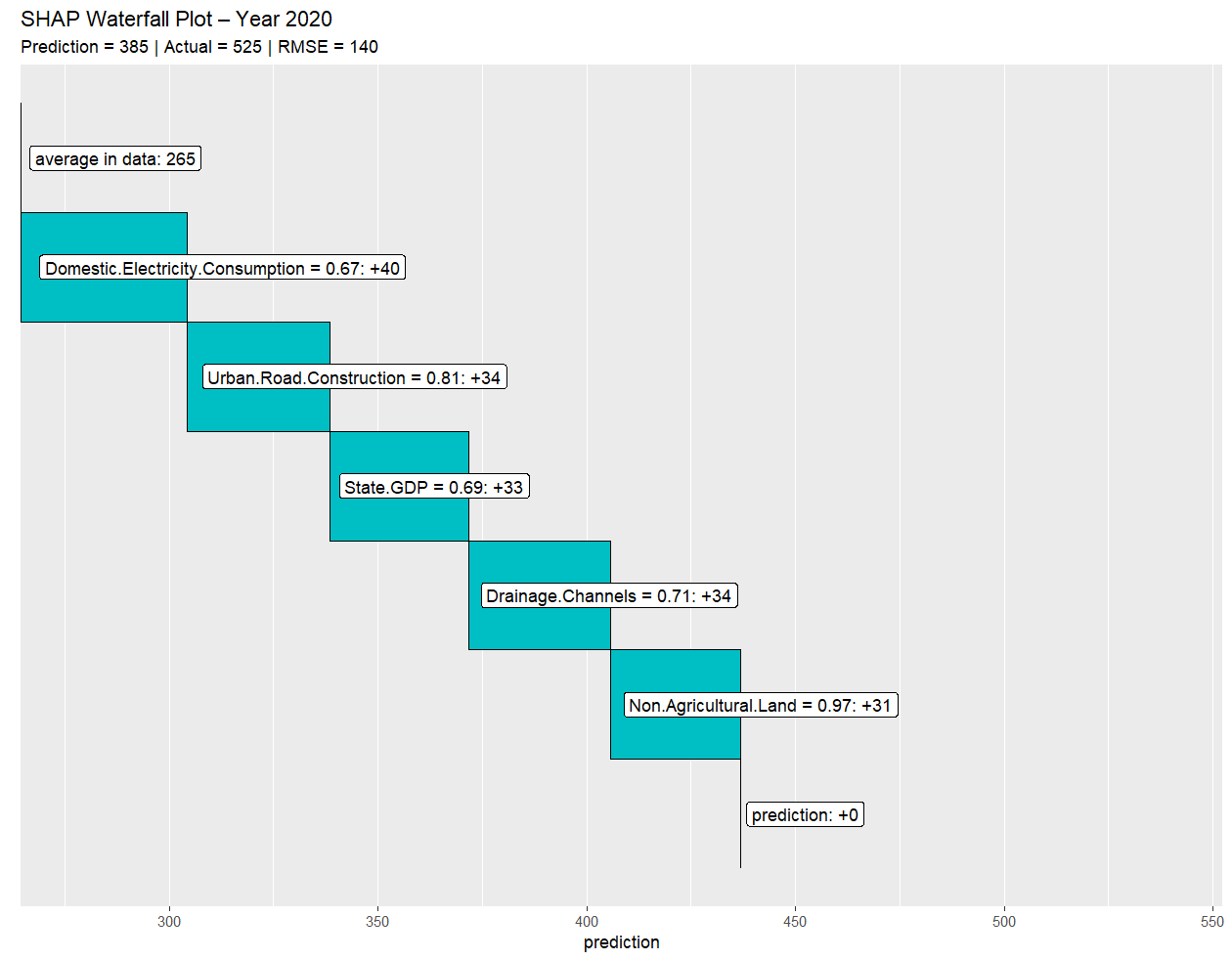
### SHAP Over Time



* In years like 2014–2020, where floods increase sharply, SHAP values (colour intensity) also spike for Domestic Electricity Consumption, Urban Road Construction, and State GDP.
* Drainage Channels become less impactful over time (despite rising in value), suggesting diminishing marginal explanatory power. Drainage channels increase in quantity, but their importance in predicting floods diminishes. This could reflect ineffectiveness or poor maintenance, which wasn’t clear from EDA alone.
* Non-Agricultural Land shows relatively stable SHAP intensity—despite a rising trend in the feature, its impact plateaus.

## Random Forest

### SHAP Waterfall Breakdown

A graph with blue squares

AI-generated content may be incorrect.

### SHAP Summary Plot



### SHAP Over Time

