Erdös Institute Data Science Bootcamp May 2022

Predicting Fertilizer Input for Rice Cultivation in India

Team Barrel:

Akshay Suresh

Arman Darbinyan

Dmitry Shcherbakov

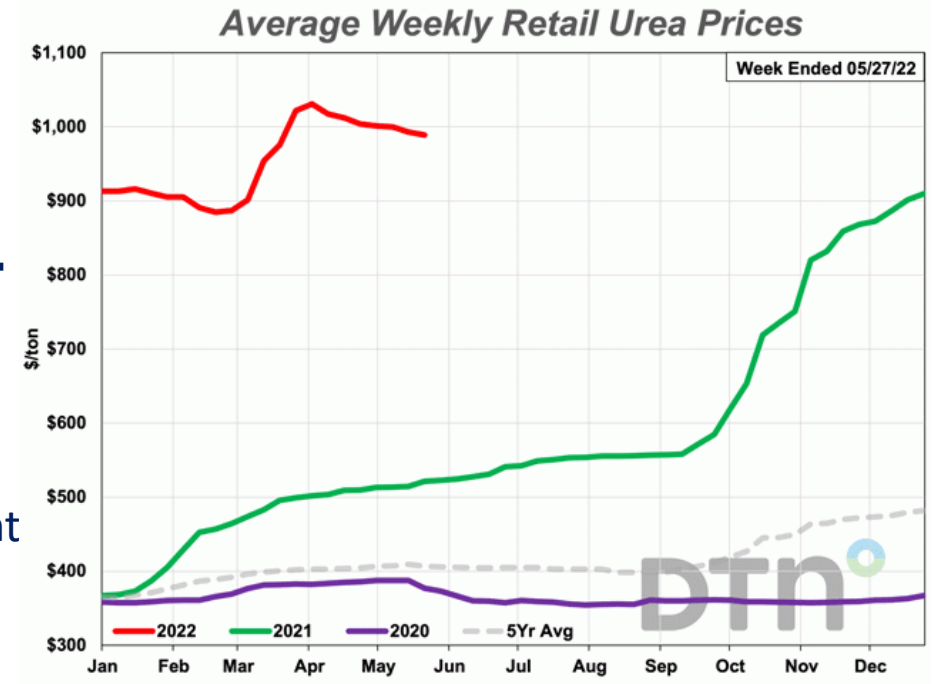
Emilio Codecido

Leonardo Santana

Sustaining India’s Food Self-sufficiency

Resource-intensive farming

Soaring fertilizer prices



Provision of consistent subsidies to farmers

India’s dependence on imports:
Urea – 33 %, Phosphate – 90%, Potash – 100%

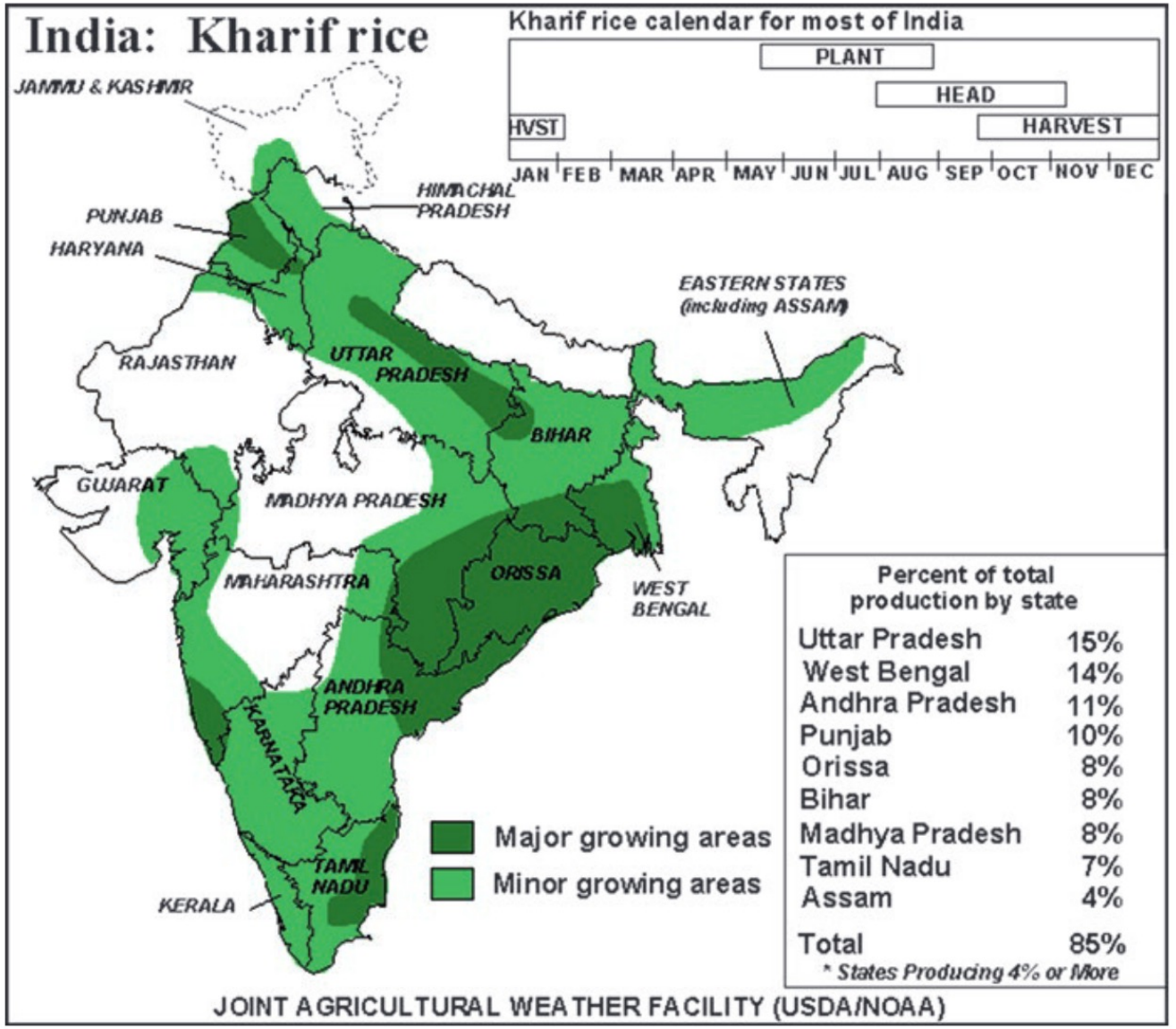
Precise fertilizer budgeting required by policy-makers

Our Goal

Help policy-makers by

"predicting expected NPK fertilizer input (kg/ha) to obtain a desired rice yield (kg/ha)"

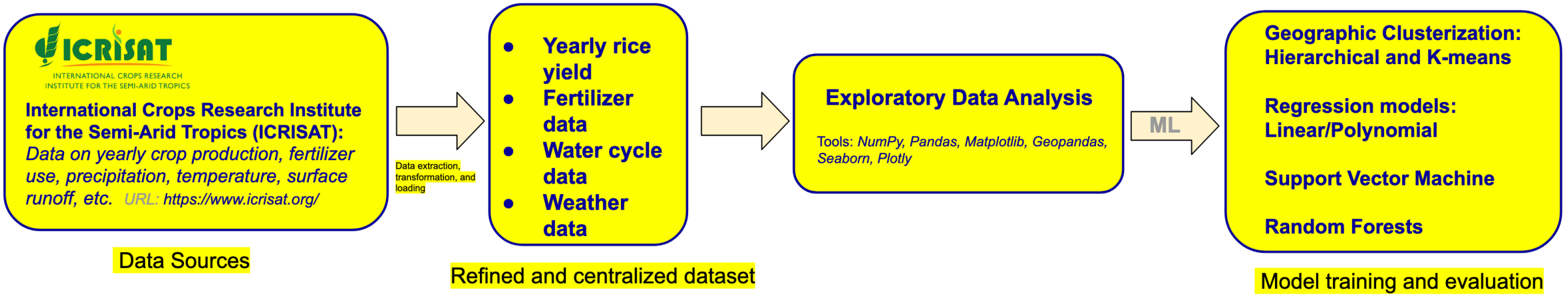
accounting for diverse rice cultivation environments.

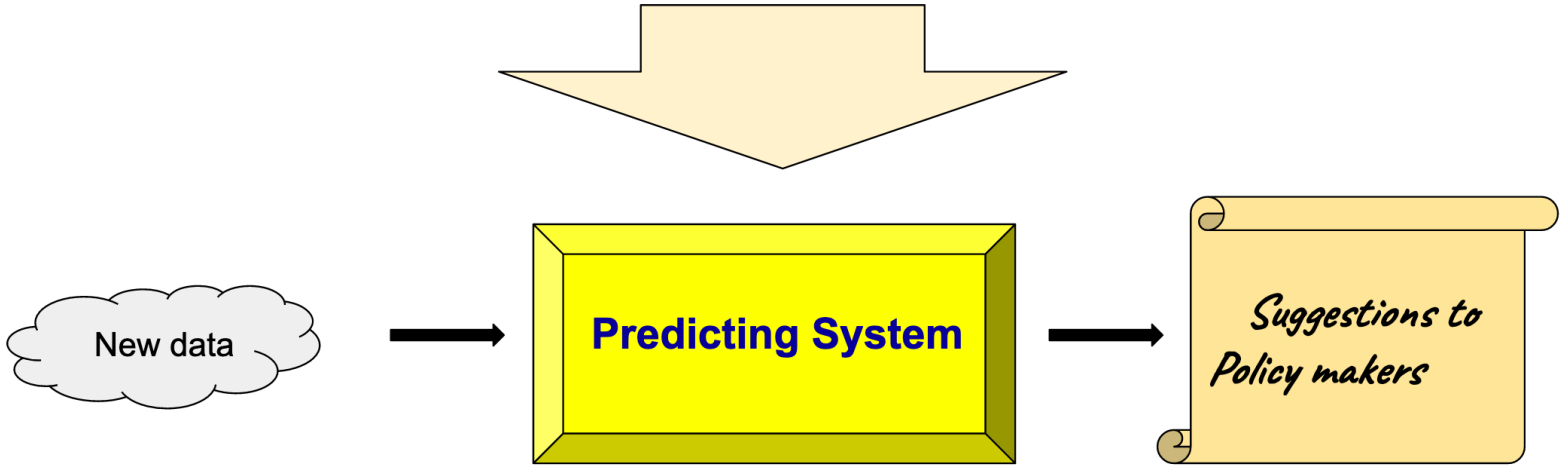


Mahajan, G., Kumar, V., Chauhan, B.S. (2017)

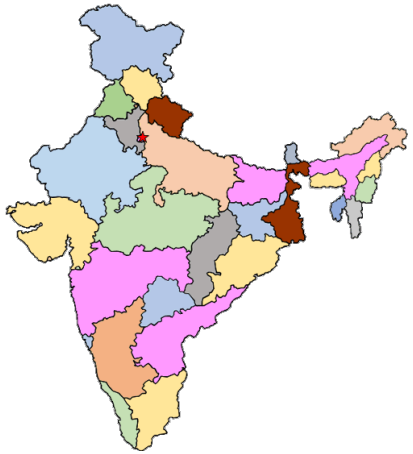
Workflow

(Akshay, Emilio) (Leonardo, Arman, Dmitry) (all team members)



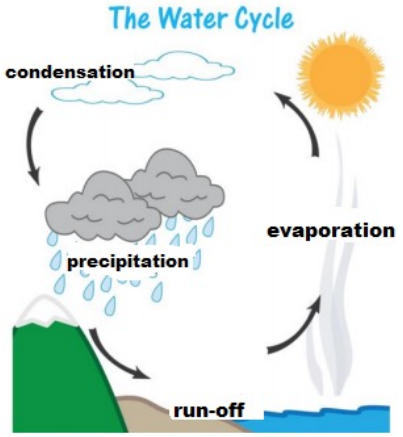


Data Features



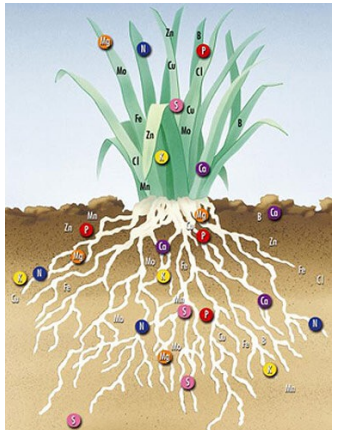
Geography

Seasonality and temperature



Precipitation, runoff, evapotranspiration

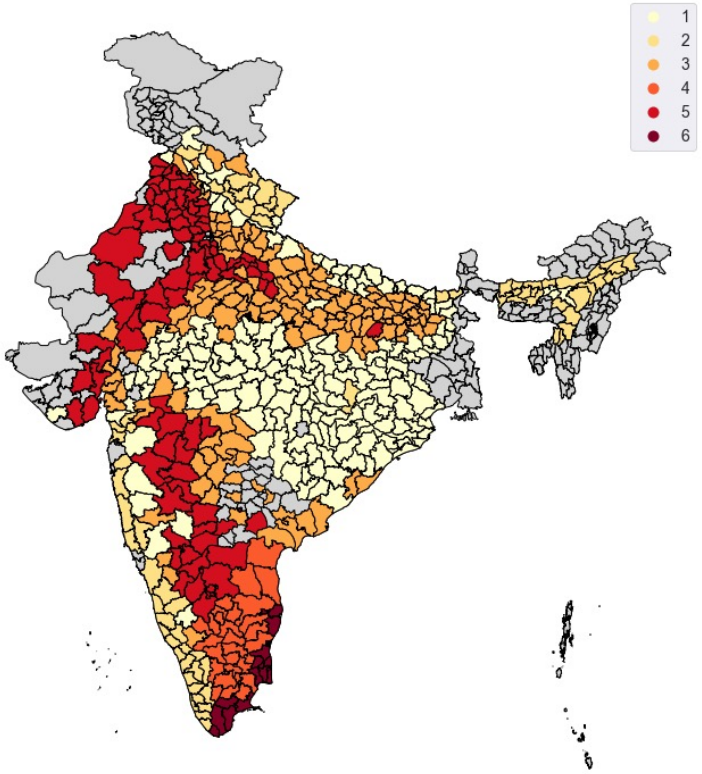
Wind speed



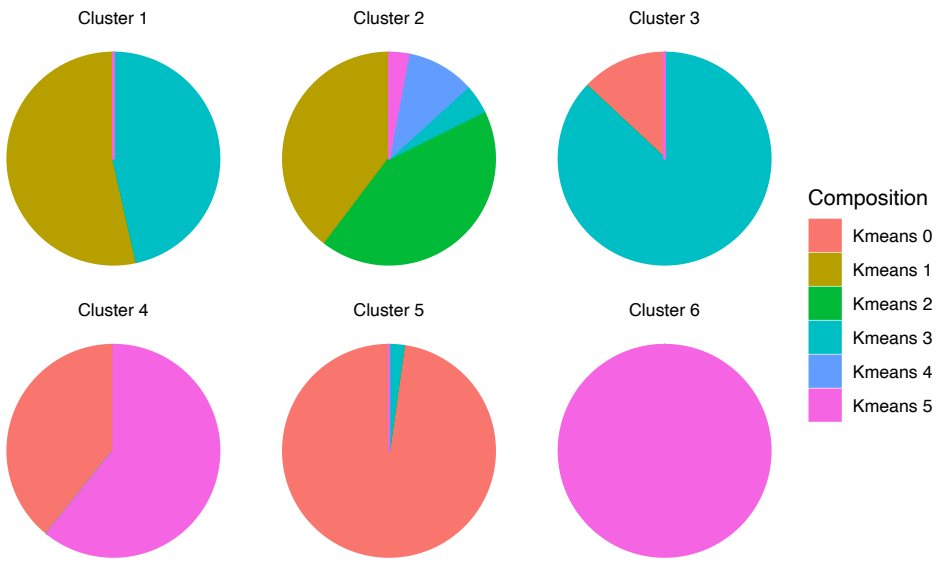
Fertilizer input

Clustering

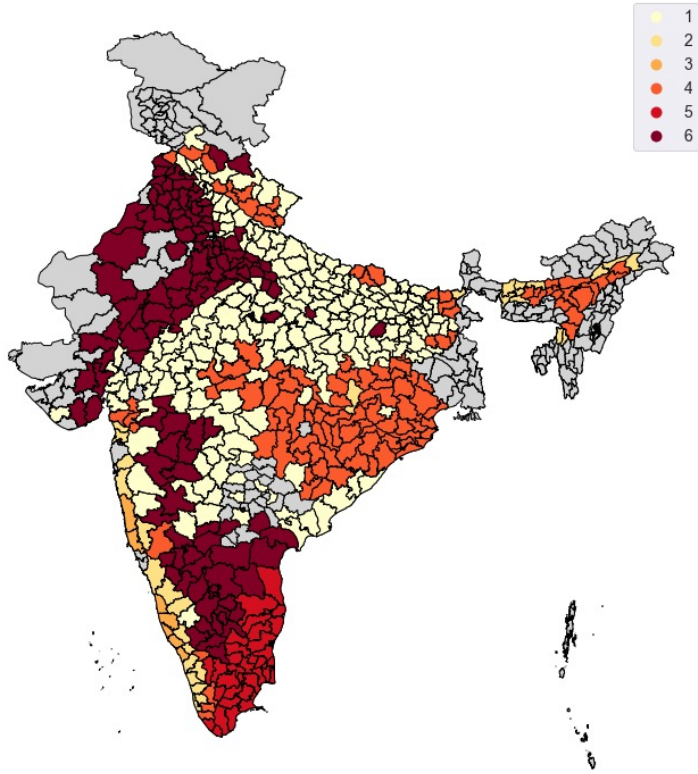
Method 1: Correlation-based Hierarchical Clustering on 12 environmental parameters



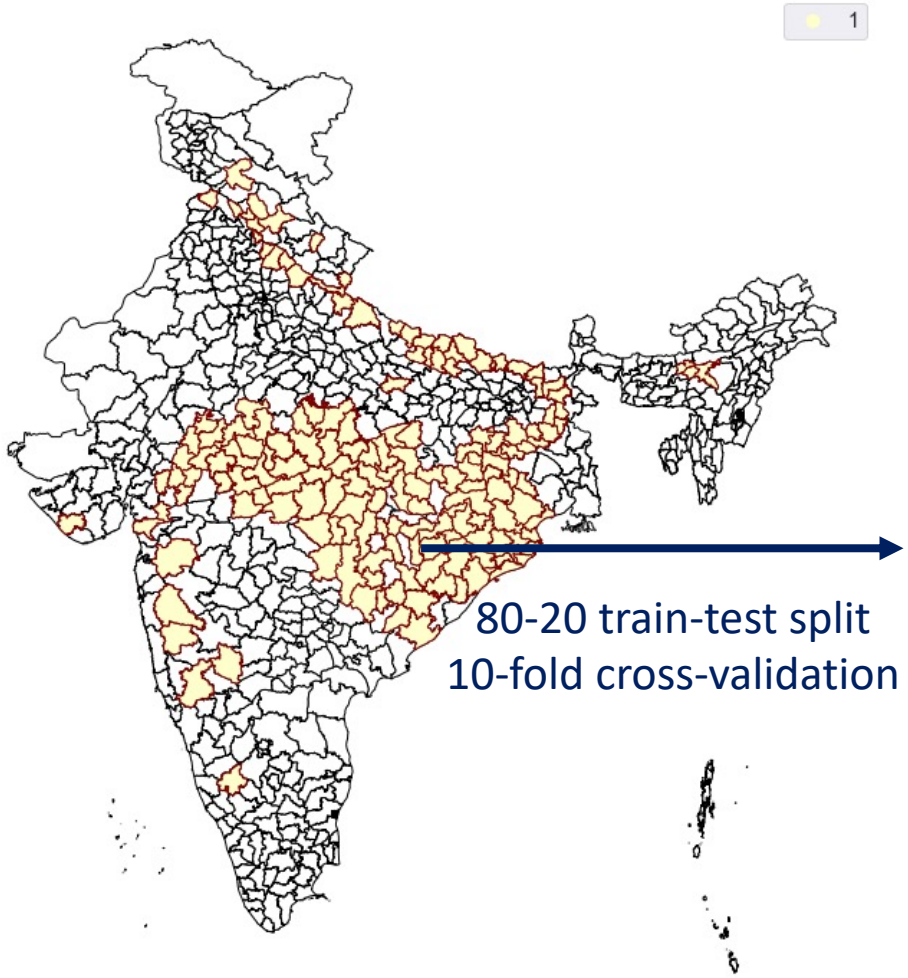
Composition of clusters in Method 1 by the classification of districts from Method 2:



Method 2: K-means clustering (validation)

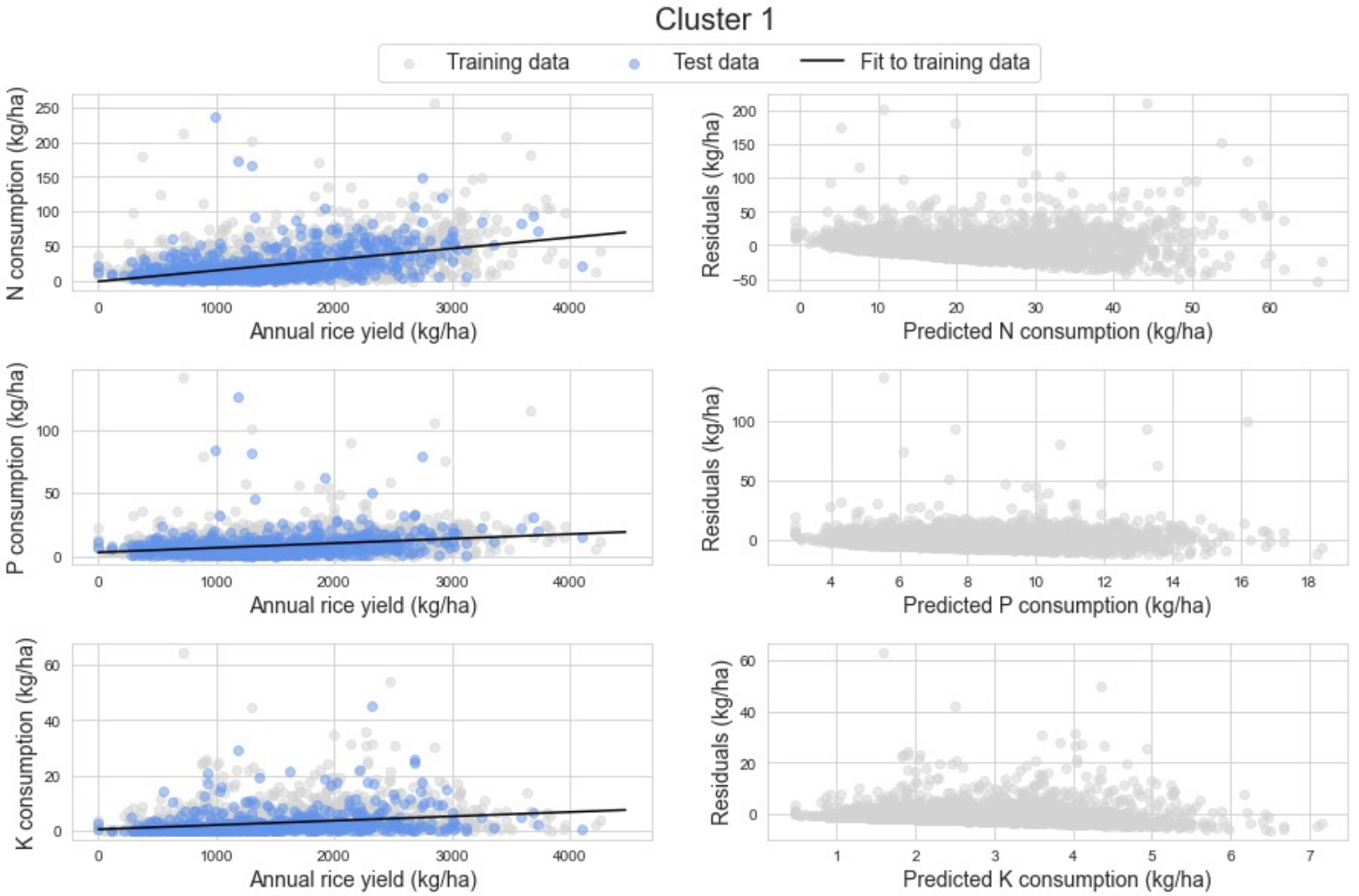


Cluster-level Modeling



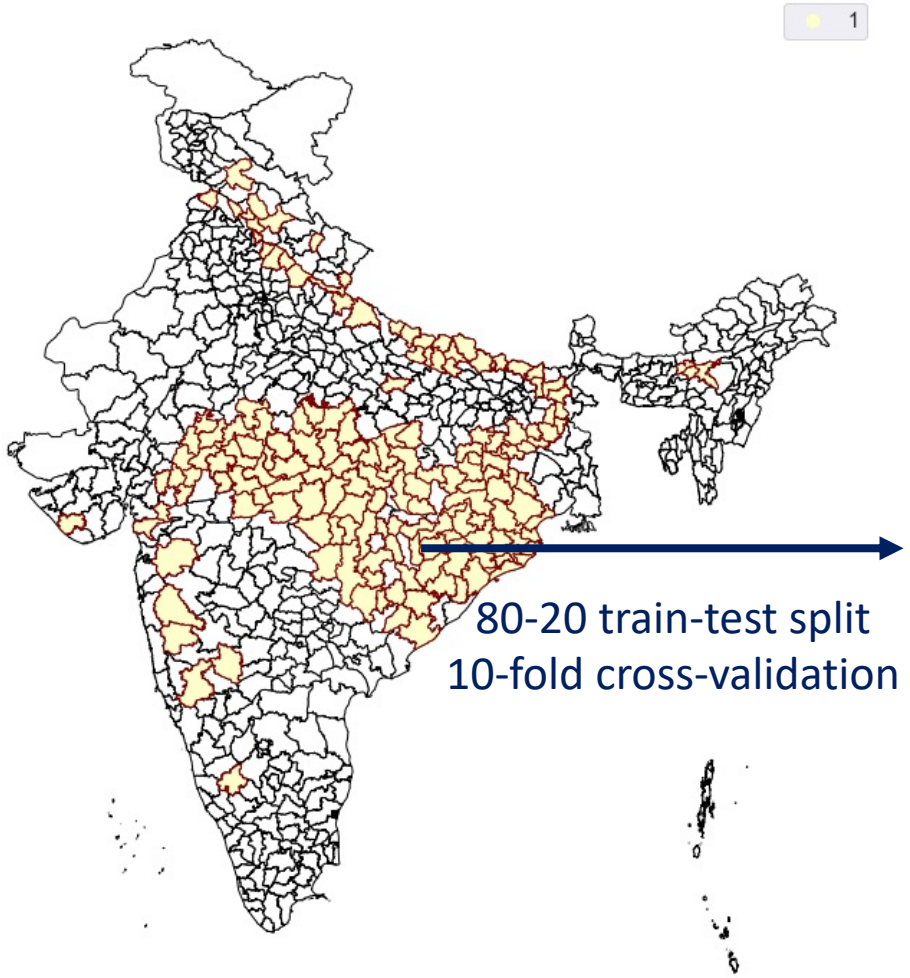
80-20 train-test split
10-fold cross-validation

Simple linear regression

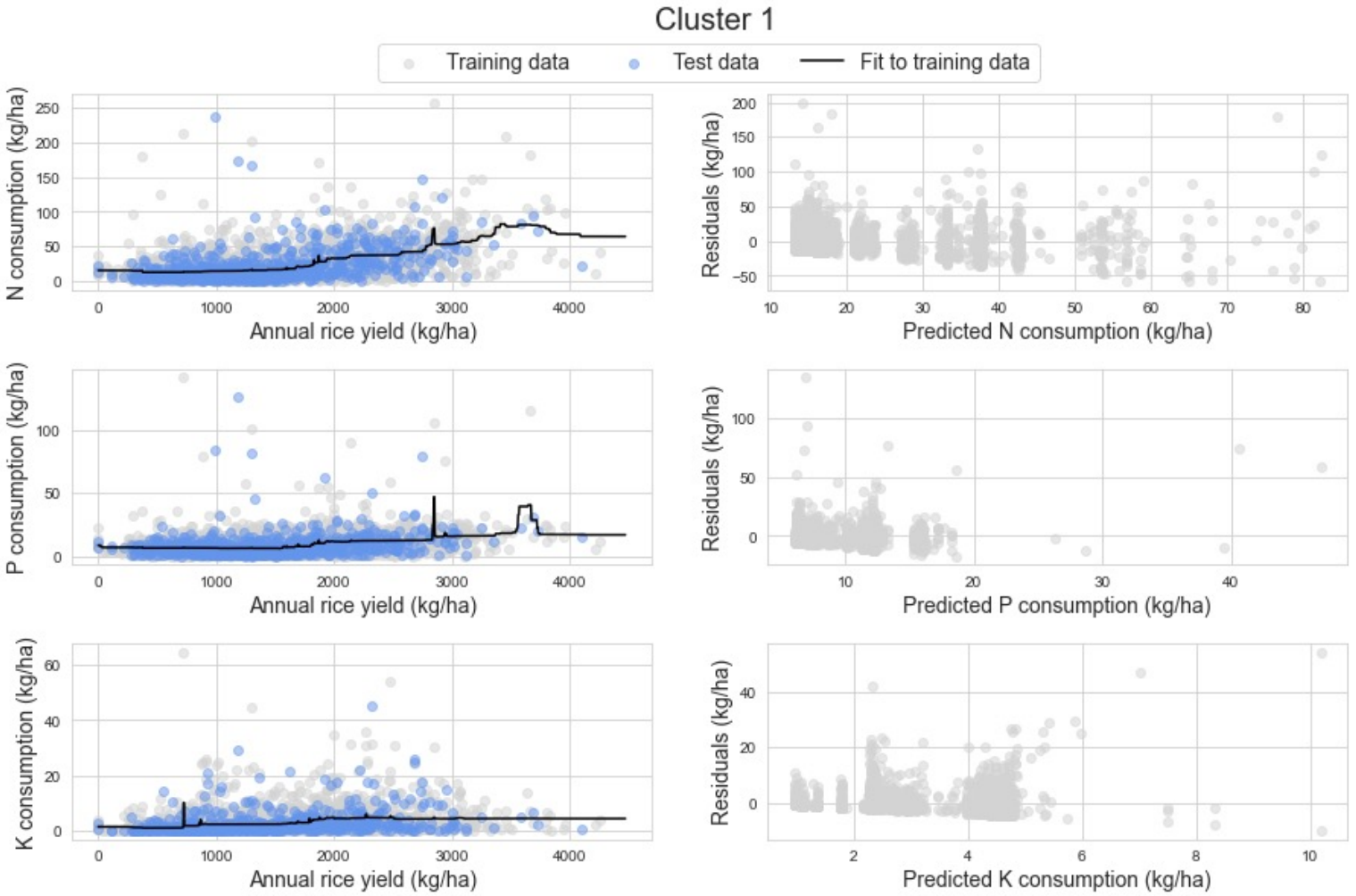


Cluster-level Modeling

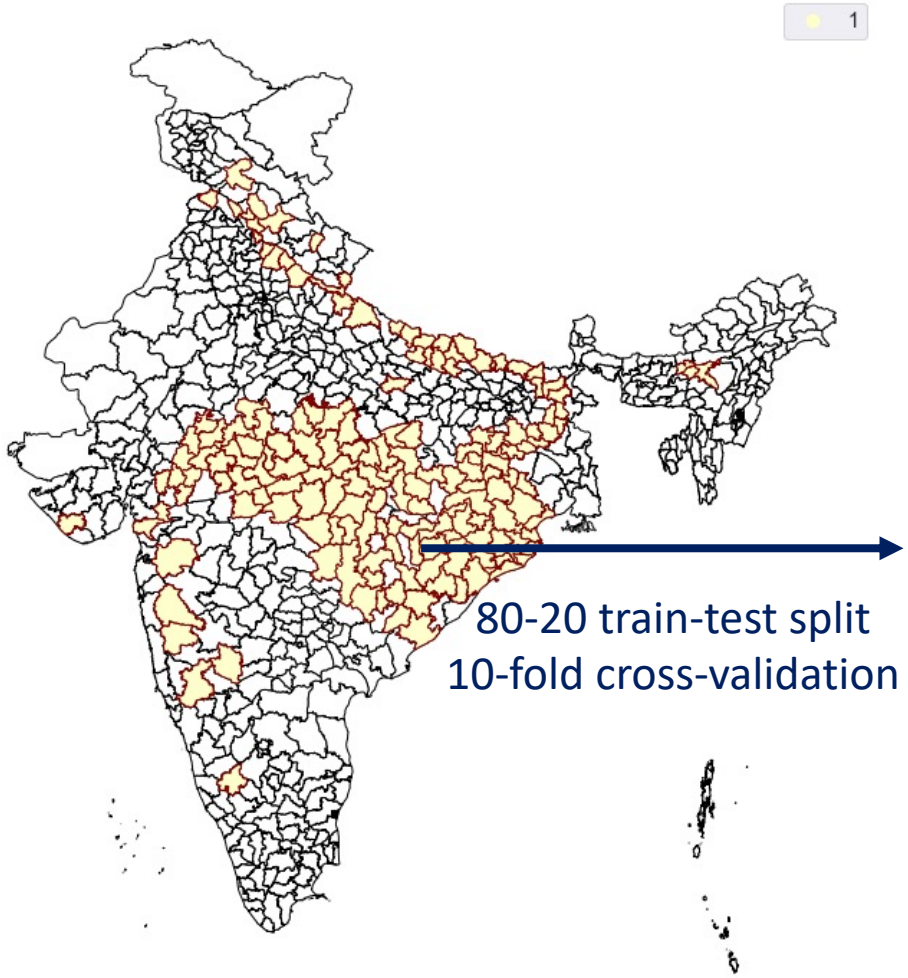
Random forest regression



80-20 train-test split
10-fold cross-validation



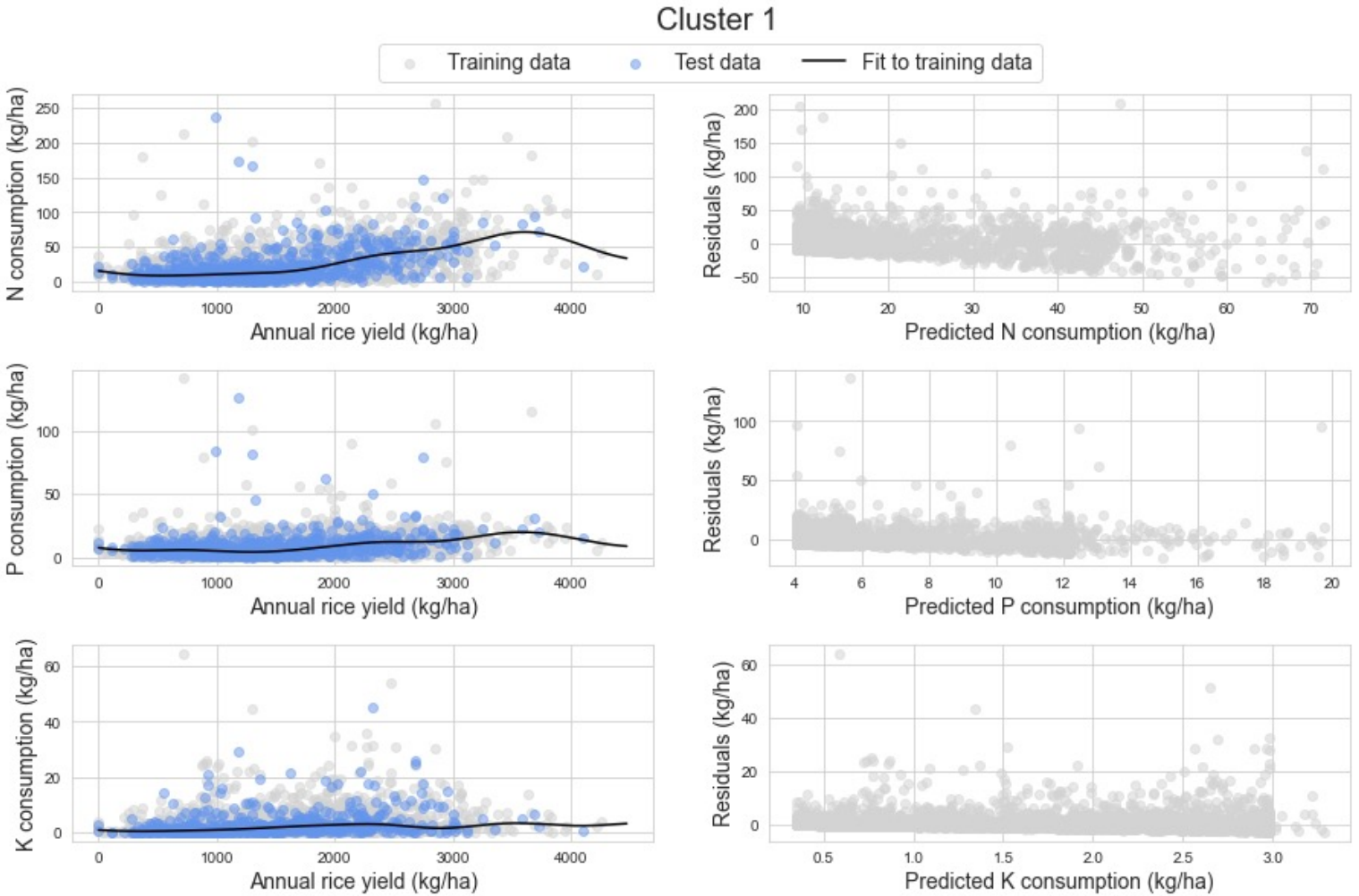
Cluster-level Modeling



80-20 train-test split
10-fold cross-validation

Radial basis function kernel

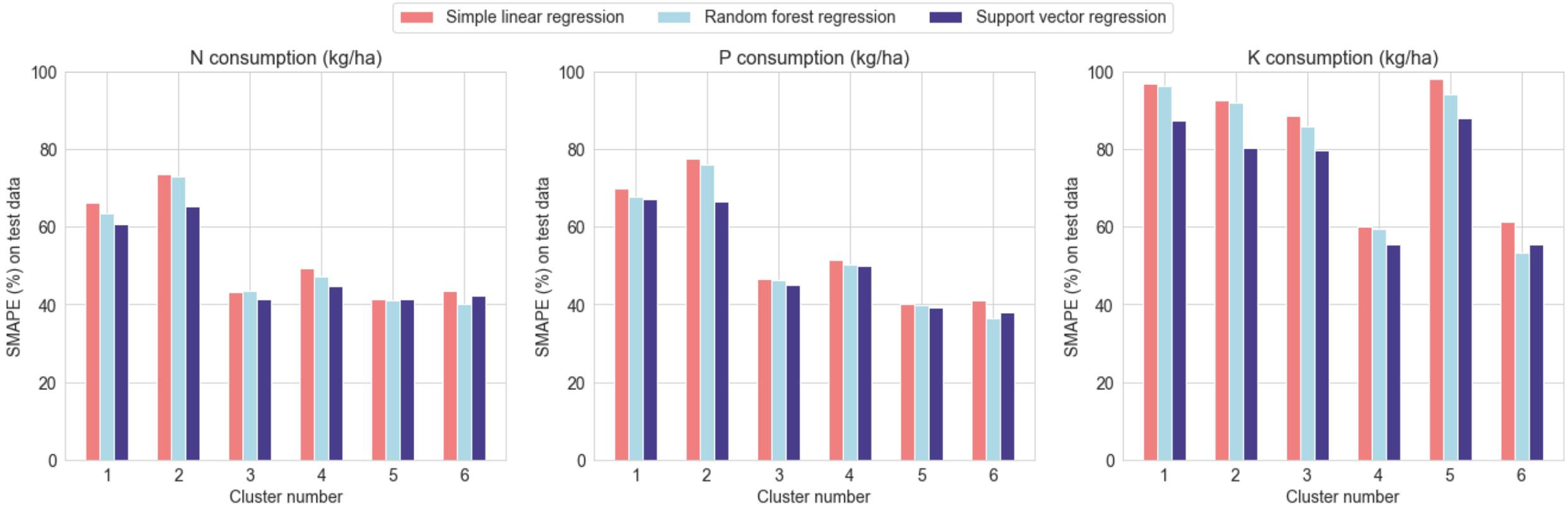
Support vector regression



Model Comparison

Symmetric mean absolute percent error:

$$\text{SMAPE} = \left(\frac{100\%}{N_{\text{observations}}}\right) \sum_{\text{observations}} \left(\frac{|\text{True value} - \text{Predicted value}|}{(|\text{True value}| + |\text{Predicted value}|)/2}\right)$$



Summary and Future Work

- Built cluster-level models predicting fertilizer input for a given rice yield.

<u>Future model extensions:</u>

- New variables:

solar irradiance

soil nutrients

cloud cover

- Impact of crop rotation and off-season farming practices.

Stakeholders: Agriculture policy-makers

<u>Acknowledgements:</u>
Special mention to James Bramante for mentoring us through our project.