International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Comprehensive Method for Medium-Term Analysis and Forecast of Agricultural Commodity Prices

CMAF

Rotem Zelingher

Postdoctoral Research Fellow, Advancing Systems Analysis Program, IIASA

Advancing Systems Analysis (ASA)

Program research goals

- Identify, develop and deploy **new systems-analytical methods, tools, and data**
- Address the most pressing global sustainability challenges
- Find **solutions** to those challenges that are **both realistic and appropriate**

ASA Research Groups

Group Leader:
Sebastian Poledna

EM

Exploratory Modeling of Human-Natural Systems

SYRR

Systemic Risk and Resilience

Group Leader:
Reinhard Mechler

NODES

Novel Data Ecosystems for Sustainability

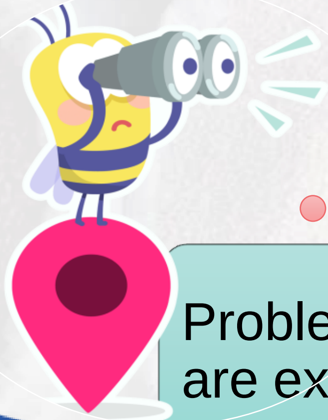
Group Leader: Ian McCallum

CAT

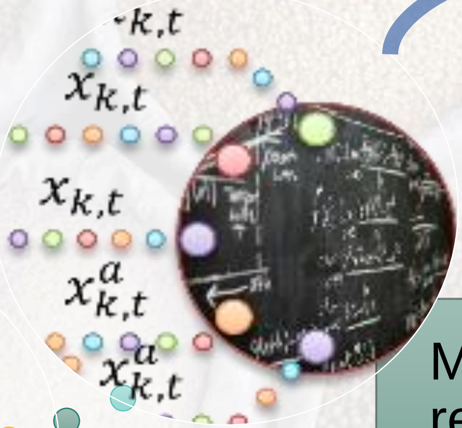
Cooperation and Transformative Governance

Group Leader:
Nadezhda Komendantova

CMAF



Problem: Food prices are extremely volatile



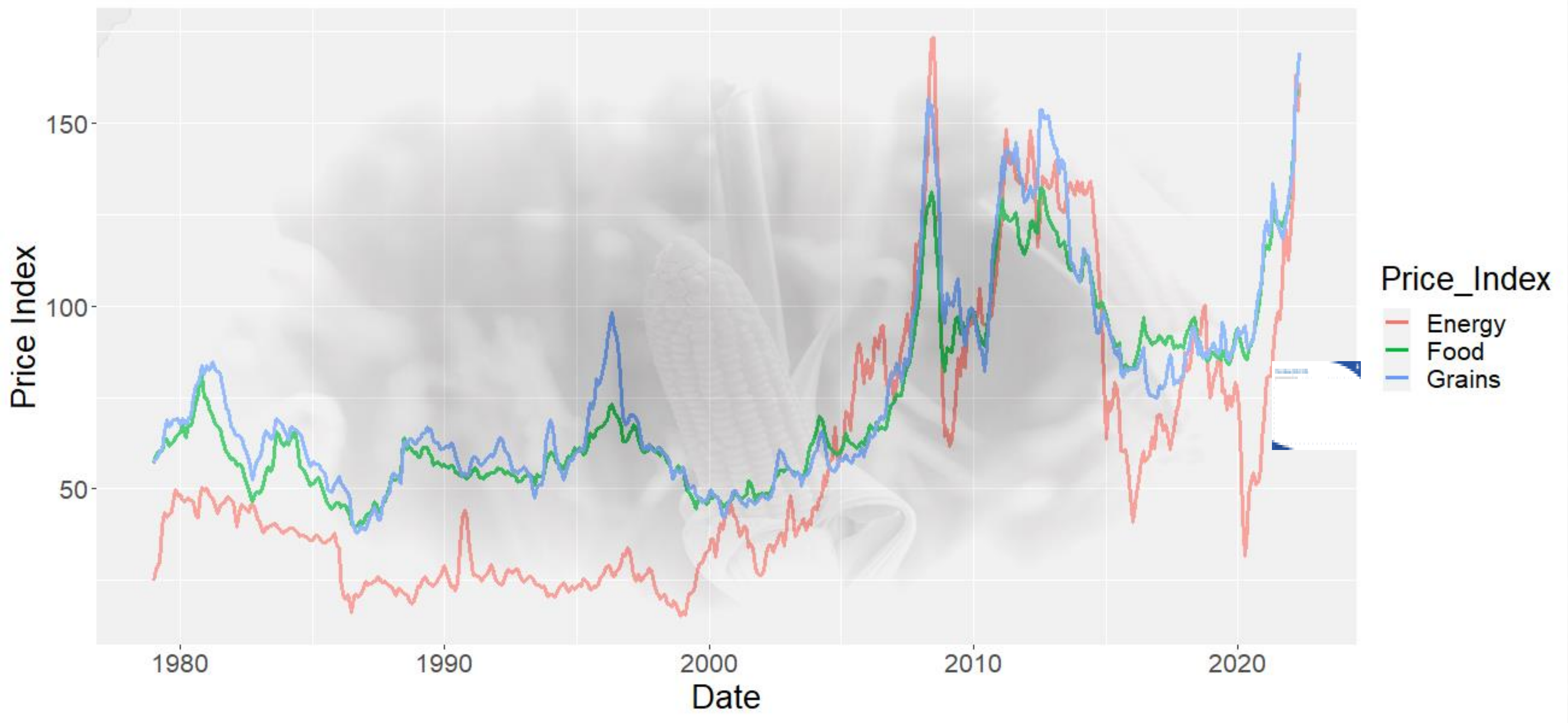
Many possible reasons/Data to use

Studying the problem using econometric & ML models

Model agnostic

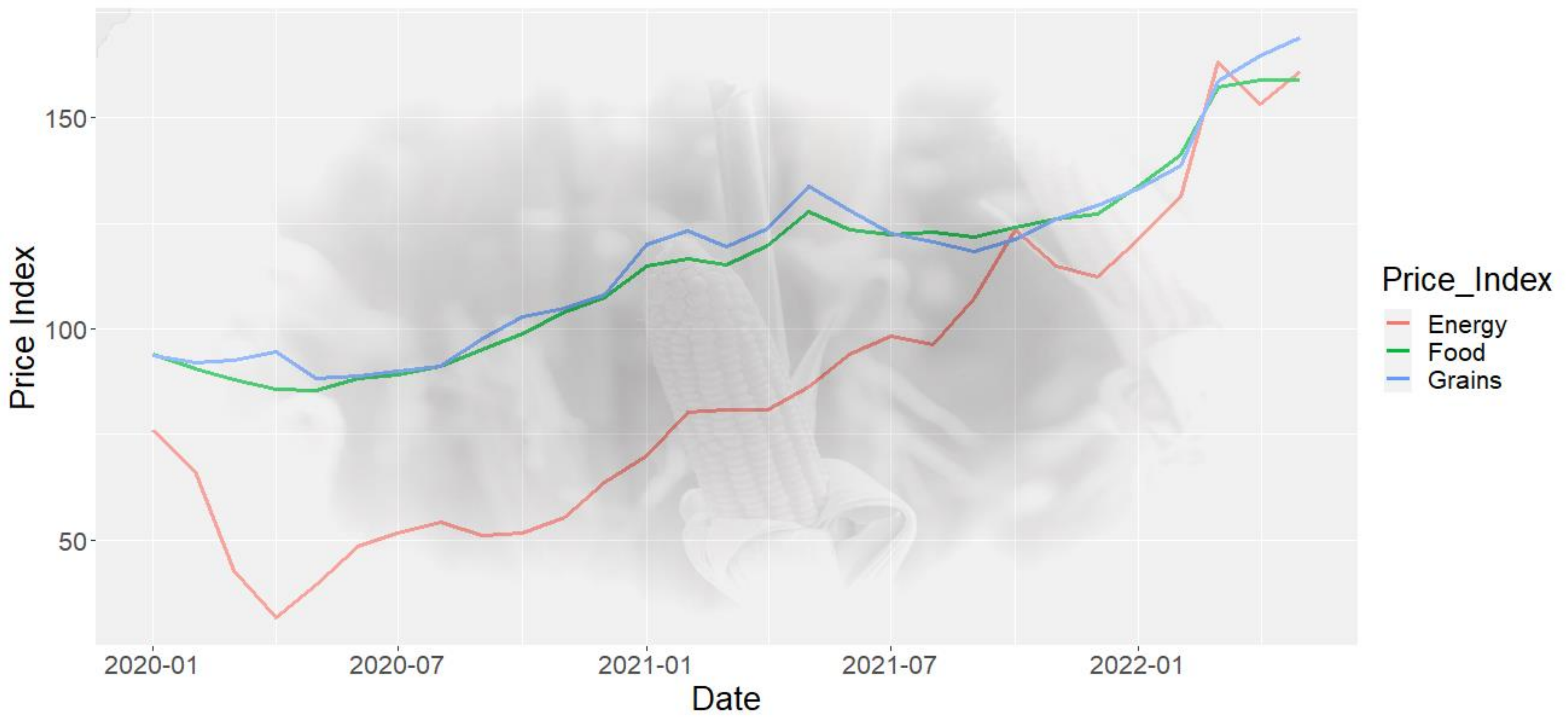
Everyone can understand

Price indices – High correlation



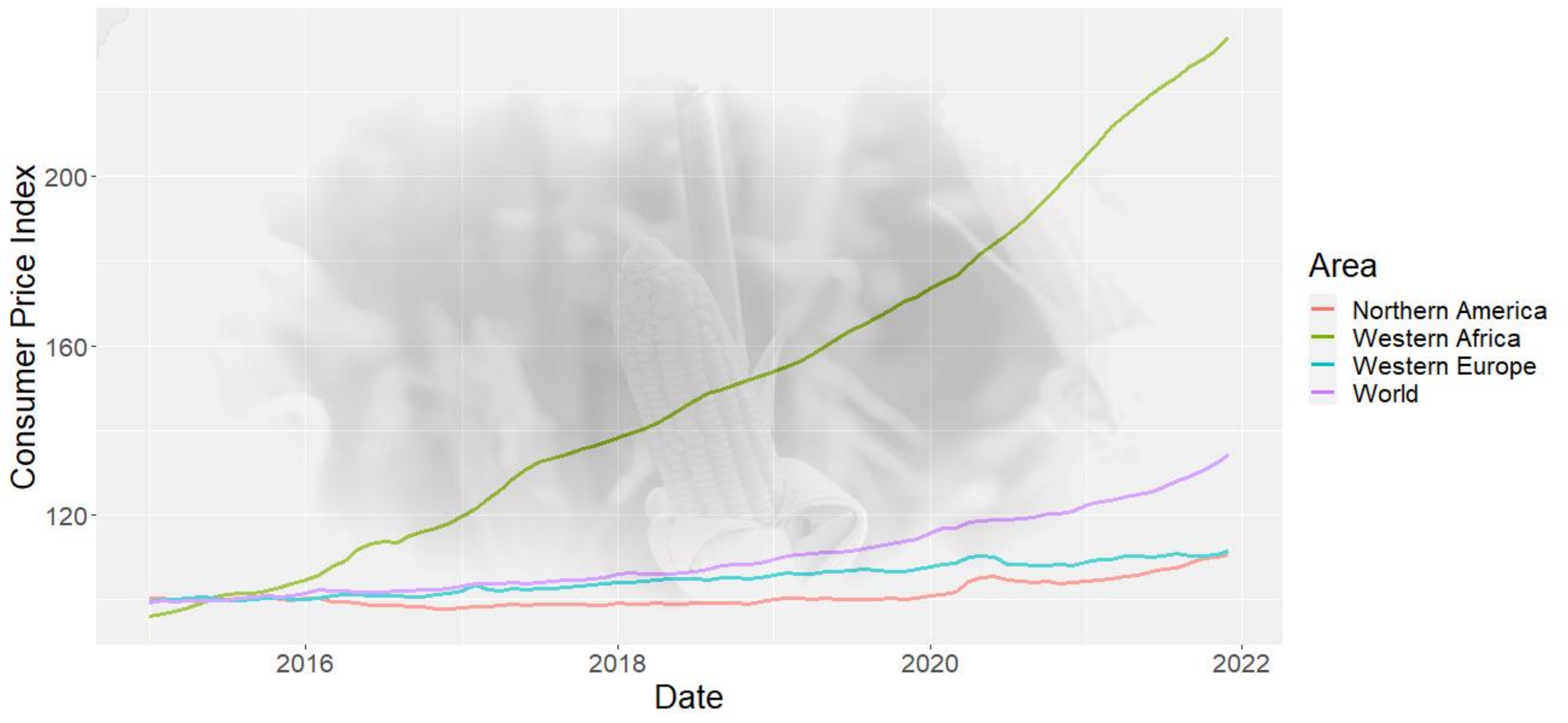
Source: the World Bank

Price indices (2010=100)

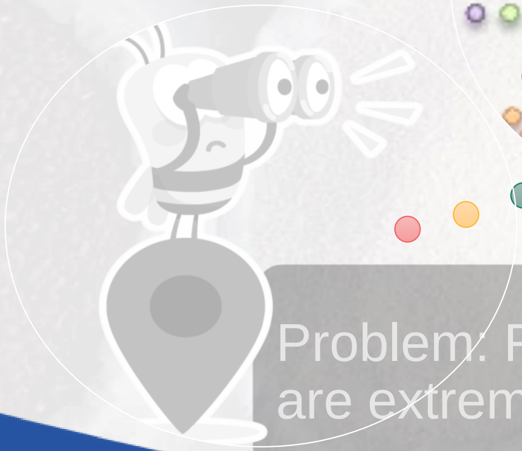


Source: the World Bank

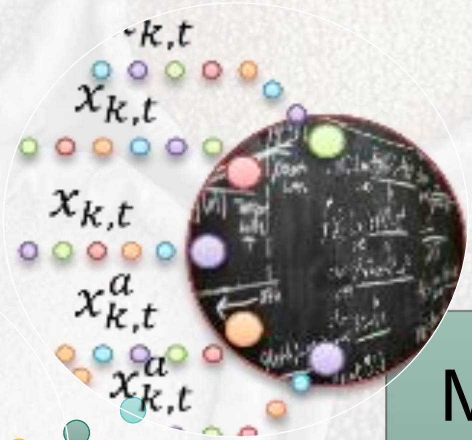
Consumer Food Price index (2015=100)



Source: FAO Stat



Problem: Food prices are extremely volatile



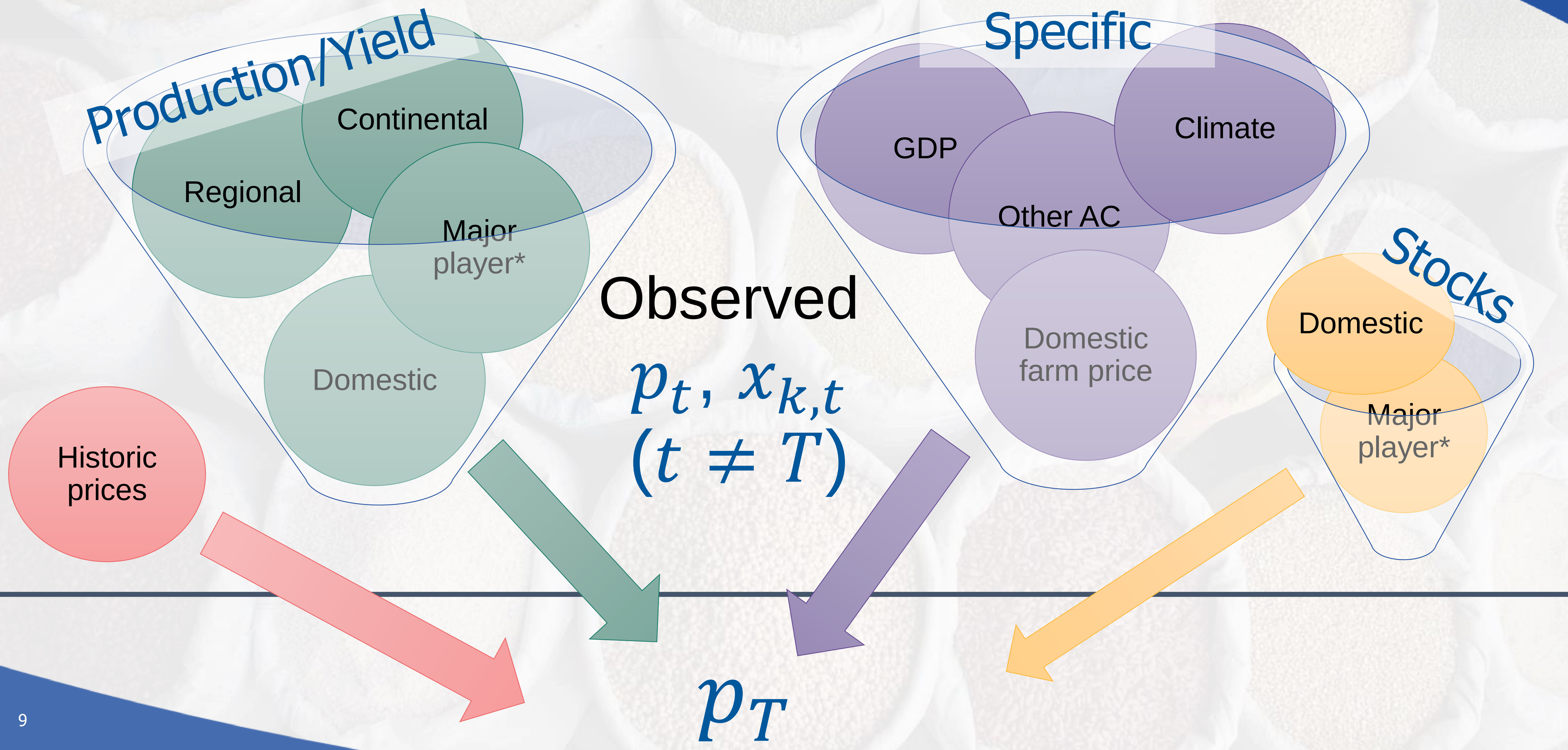
Many possible reasons/Data to use

Studying the problem using econometric & ML models

Model agnostic

Everyone can understand

Model overview

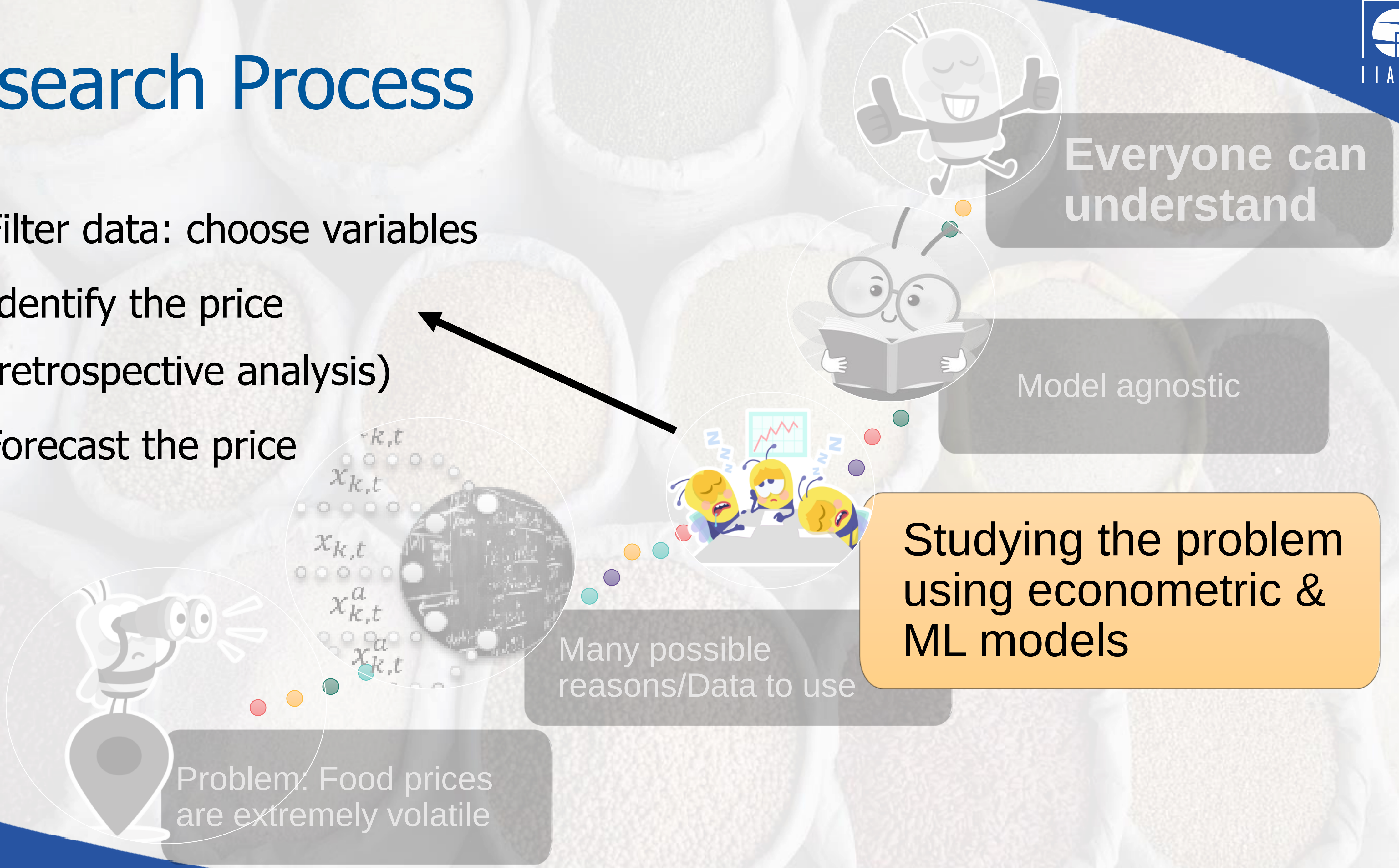


Research Process

I. Filter data: choose variables

II. Identify the price (retrospective analysis)

III. Forecast the price



Model specification & application

- Two types of $p_{m,Y}$ **An**alysis & **For**ecasting
  - **Reg**ression → Relative change
  - **Clas**sification → Increase/decrease
- Approach: **Eco**nometric/**M**achine **L**earning
- Model accuracy assessment**:
  - Regression: $RMSE$ 
  - Classification: Area under curve (AUC)

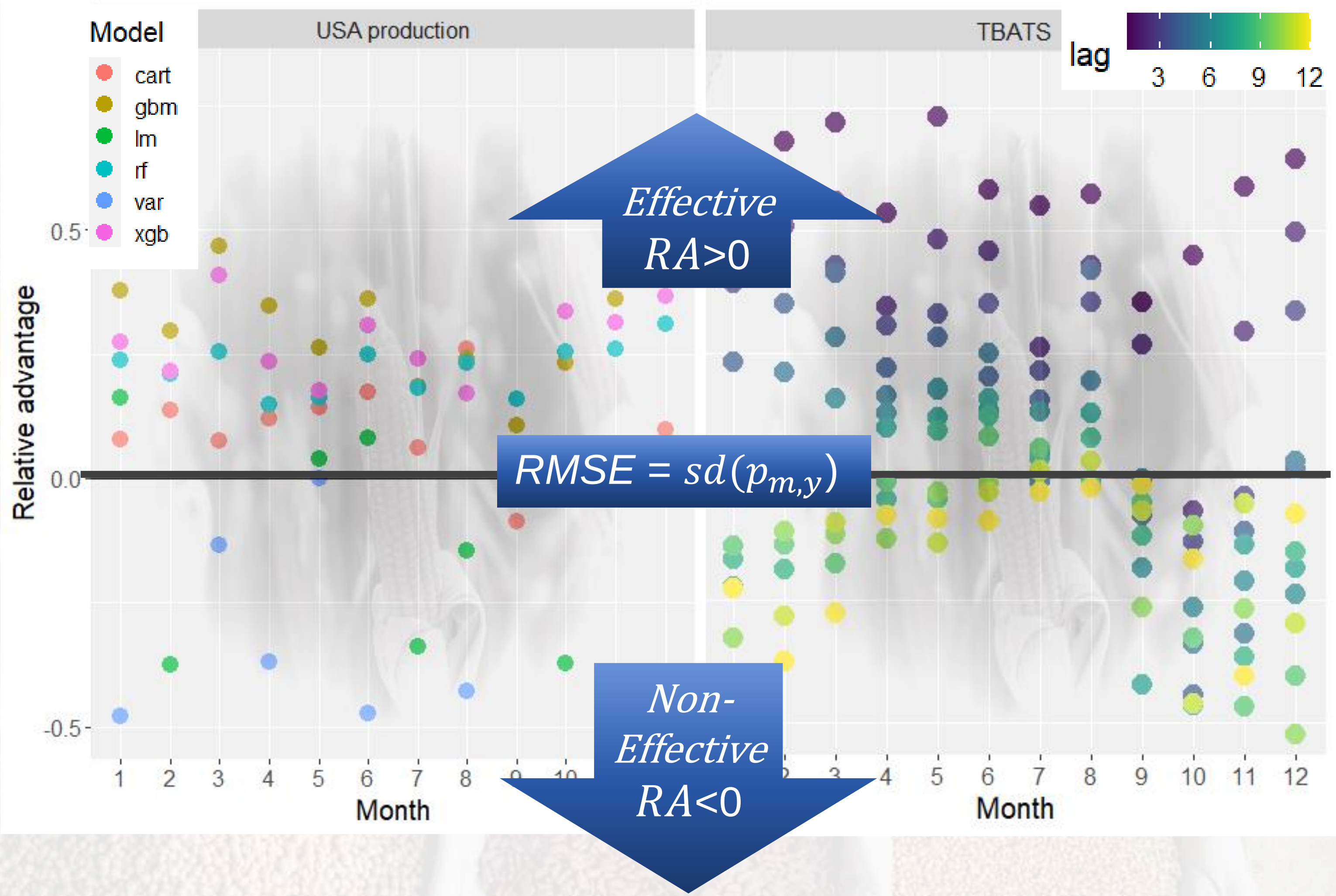
#	Algorithm	An	For	Reg	Clas	Model Type
1	LM, GLM		✓	✓	✓	Eco
2	VAR		✓	✓		Eco
3	CART	✓	✓	✓	✓	ML
4	RF	✓	✓	✓	✓	ML
5	GBM	✓	✓	✓	✓	ML
6	XGBoost	✓	✓	✓	✓	ML
7	KNN				✓	ML
8	TBATS		✓	✓		ML

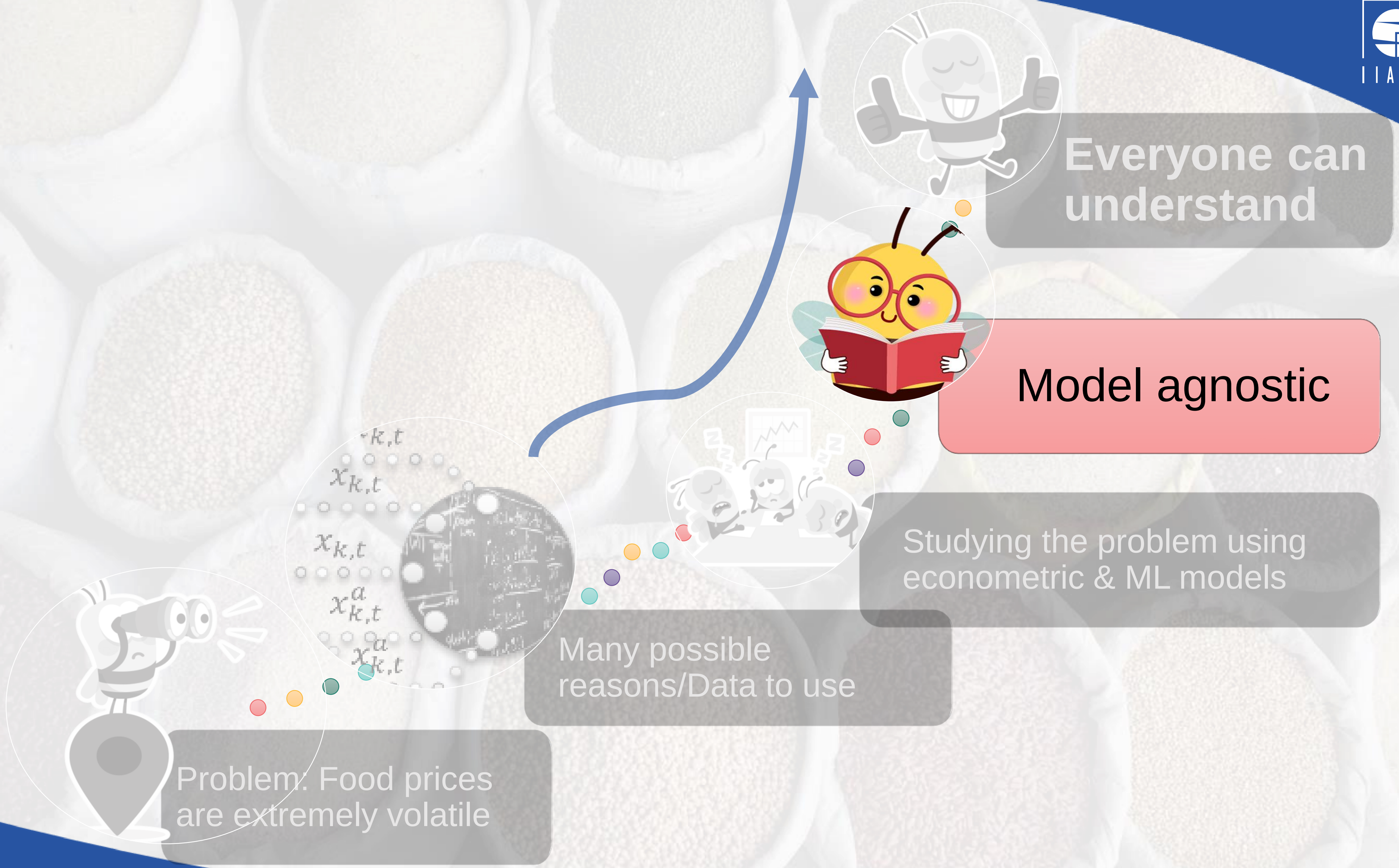
Relative advantage

$$\mathrm{RA} = 1 - \frac{RMSE}{sd(p_{m,y})}$$

Higher = Better performance

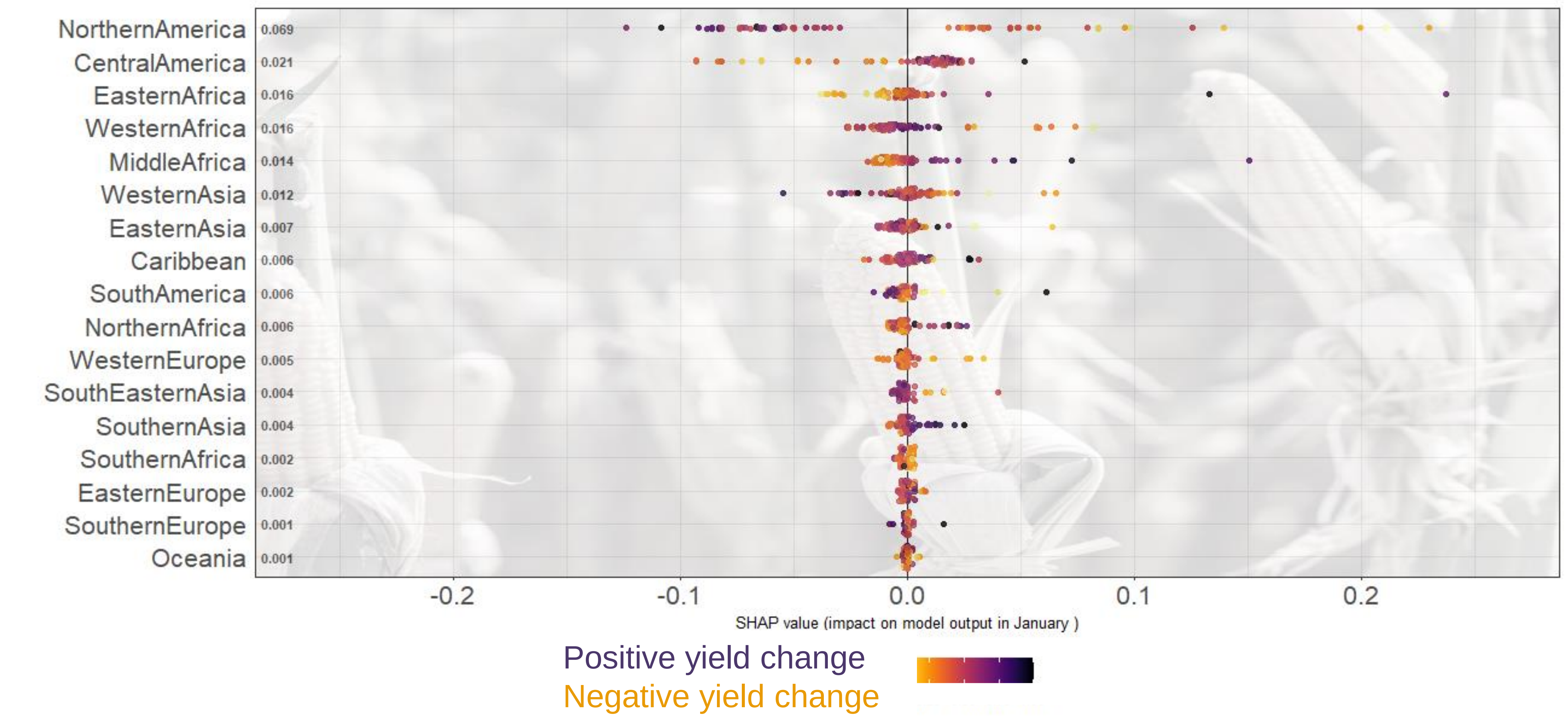
Lower = Worse performance





Everyone can understand
Model agnostic
Studying the problem using econometric & ML models
Many possible reasons/Data to use
Problem: Food prices are extremely volatile

Explore Relative Importance (SHAP)

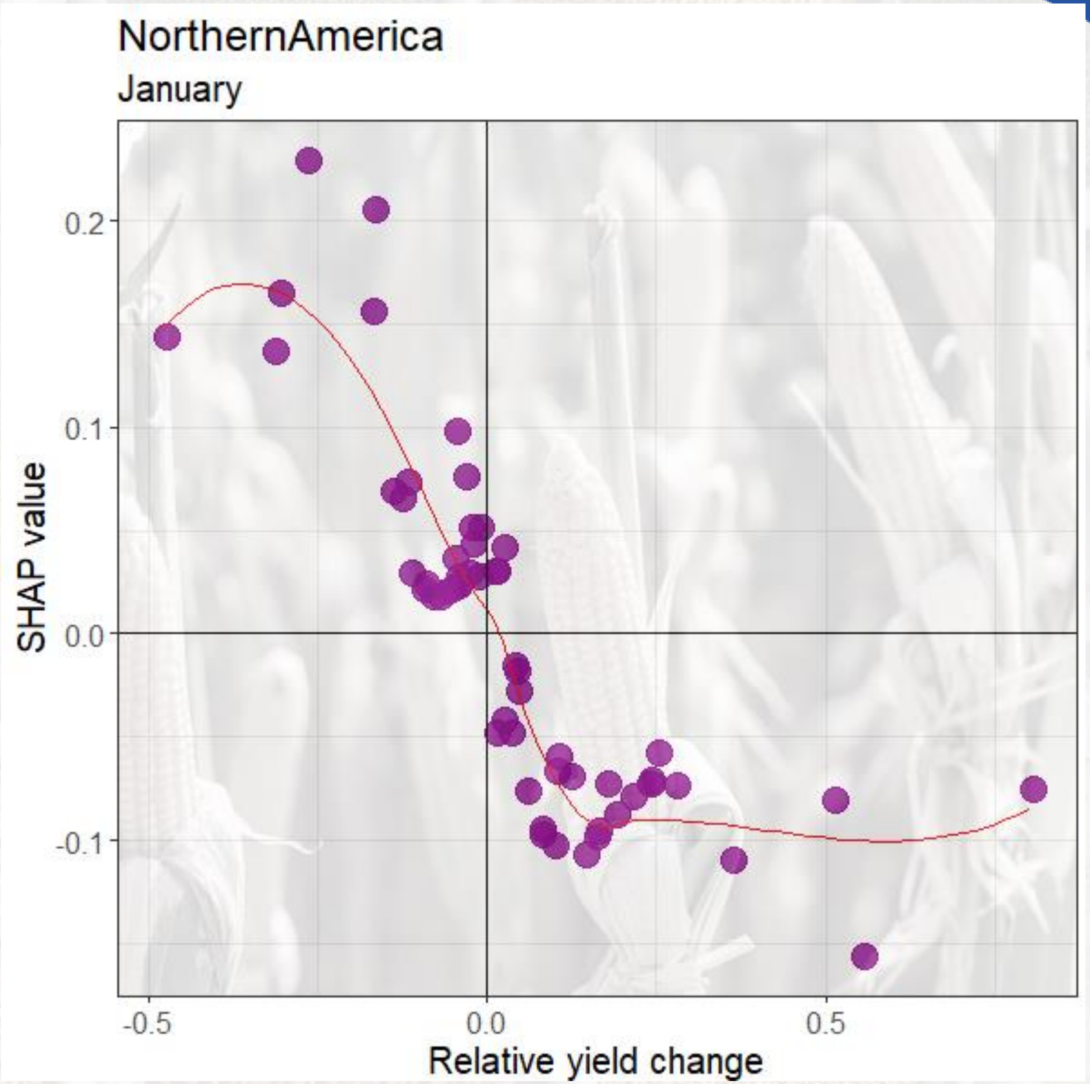


Explore Partial Dependence

- Average response of $p_{m,y}$ to variations of maize yield Northern America
- PDPs show negative correlation

$$x_{k,y} \downarrow \; \Rightarrow \; p_{m,y} \uparrow$$

- Strong negative influence of Northern America yield over global maize price
- $x_{k,y} < 0$ affects more than $x_{k,y} > 0$



Where does this project head to?

Make price forecasting and analysing a **Social Good**

- The first tool to provide medium-term price analysis and forecasting of AC in a way that would be useful for all humans.
- Promote understanding of the global food trade to enhance food security and social equity
- Publicly available as an open online platform

Everyone can understand

My research

Statistical modelling & Applied research

Using statistical and machine learning methods

Agriculture	Food security	Environment and ecology	Trade	And more

International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

THANK YOU

Rotem ZELINGHER

רתם זלינגר

zelingher@iiasa.ac.at

Bibliography

- Barrett, C. B., & Headey, D. (2014). A proposal for measuring resilience in a risky world.
- Costinot, A., & Rodríguez-Clare, A. (2018). The US gains from trade: Valuation using the demand for foreign factor services. *Journal of Economic Perspectives*, *32*(2), 3-24.
- De Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, *106*(496), 1513-1527.
- Molnar, C. (2022). Interpretable machine learning, Lulu. com. *Com: Morrisville, NC, USA*.
- Piot-Lepetit, I., & M'Barek, R. (2011). Methods to analyse agricultural commodity price volatility. *Methods to analyse agricultural commodity price volatility*, 1-11.
- Schaub & Finger (2020) Environmental Research Letters, 15, 034014. doi:10.1088/1748-9326/ab68ab

Appendix

Data sources

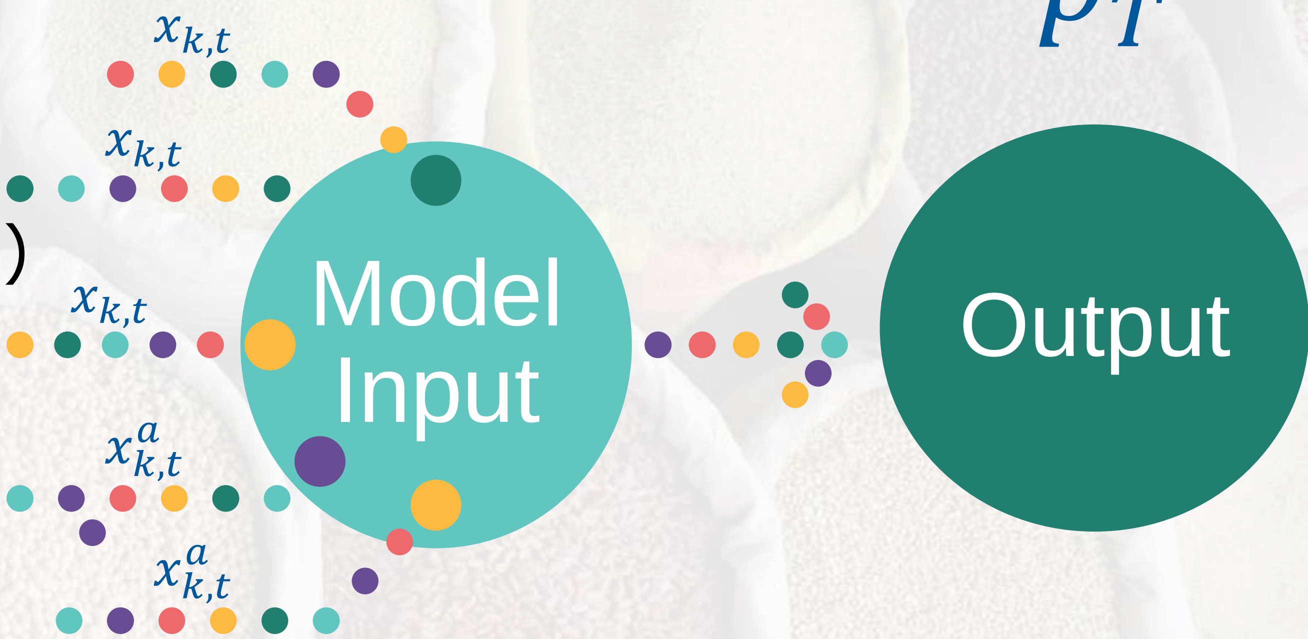
- Production, yield, stocks ($x_{k,t}^{d}$)
  - Four geographic scales (k)
  - Observed annually or monthly
  - Source: FAO-STAT, USDA*
- Additional & Case specific variables ($x_{k,t}^{a}$)
- Global prices, 1960-2022 (p_t)
  - Source: World Bank*
- Units: relative annual change

Observed data

$p_t, x_{k,t}$
$(t \neq T)$

Price to produce

p_T



Practical Motivation:
Forecasting global prices for stabilising local prices

If the decision-maker can:

- Spot unusual price shifts
- Foresee them
- Long enough for adaptation

Research Goals (II)

Enhance social equality

Promote understanding and equality in the global AC trade markets

- Data and knowledge are no longer resource-dependent

Assist low-income decision-makers

- Market analysis and price forecasting
- Recognise threats and opportunities
- Strategic design – sell/buy, consume/stock, substitutions

Random Forests (RF, Bagging)

- Random K sub-samples of the training set
  - Many trees $\rightarrow$ multiple independent training-data
- Result: Many trees lead to a preferable final result
  - Reduced correlation effect
  - Each prediction = average of K (up to 500) “weaker” trees

Boosting algorithm – GBM

- Creates multiple trees
- Each training set learns from the tree before
- High performance → Frequent appearance
- The result: Powerful final decision tree

Part I – Choose variables

Adding new variables

Are they potential regressors?

- Test for stationarity
- Correlation with price
- Lagged effects

How do they behave together?

- Correlation between features
- Regression Price ~ remained variables
  - Random Forests, GBM, XGB, CART
- Importance ranking
- Remove variables with negative impact on accuracy

New dataset

Part II – Retrospective analysis

1. **Split** data: Train (i years), Test (1 year) sets
2. **Train** an algorithm using the training set
   - Test different hyperparameters (in ML)
3. Use the Train set to **Identify**

$$p_{m,Y} = f\left(x^d_{1,y}, \dots, x^d_{k,y}, x^a_{k,m,y}\right)$$

4. **Test** of the algorithms based on test dataset
5. **Assess** detection accuracy using LOOCV
6. **Rank** features by their contribution level
7. **Filter** data - leave only those improving accuracy

	y=1	y=2	…	…	y=Y-1	$y = \mathrm{Y}$
y=1	Test	Train				
y=2	Train	Test	Train			
…	Train		Test	Train		
…	Train			Test	Train	
y=Y-1	Train				Test	Train
$y = \mathrm{Y}$	Train					Test

Part III – Price forecast

1. **Split** data: Train (45≤ i years), Test (1 year) sets
2. **Train** an algorithm using the training set
   - Tune ML algorithms
3. Use the Train set to **Forecast**

$$p_{m,Y} = f\left(x^{d}_{1,y}, \dots, x^{d}_{k,y}, x^{a}_{k,m,y}\right); Y = y_{max} + 1$$

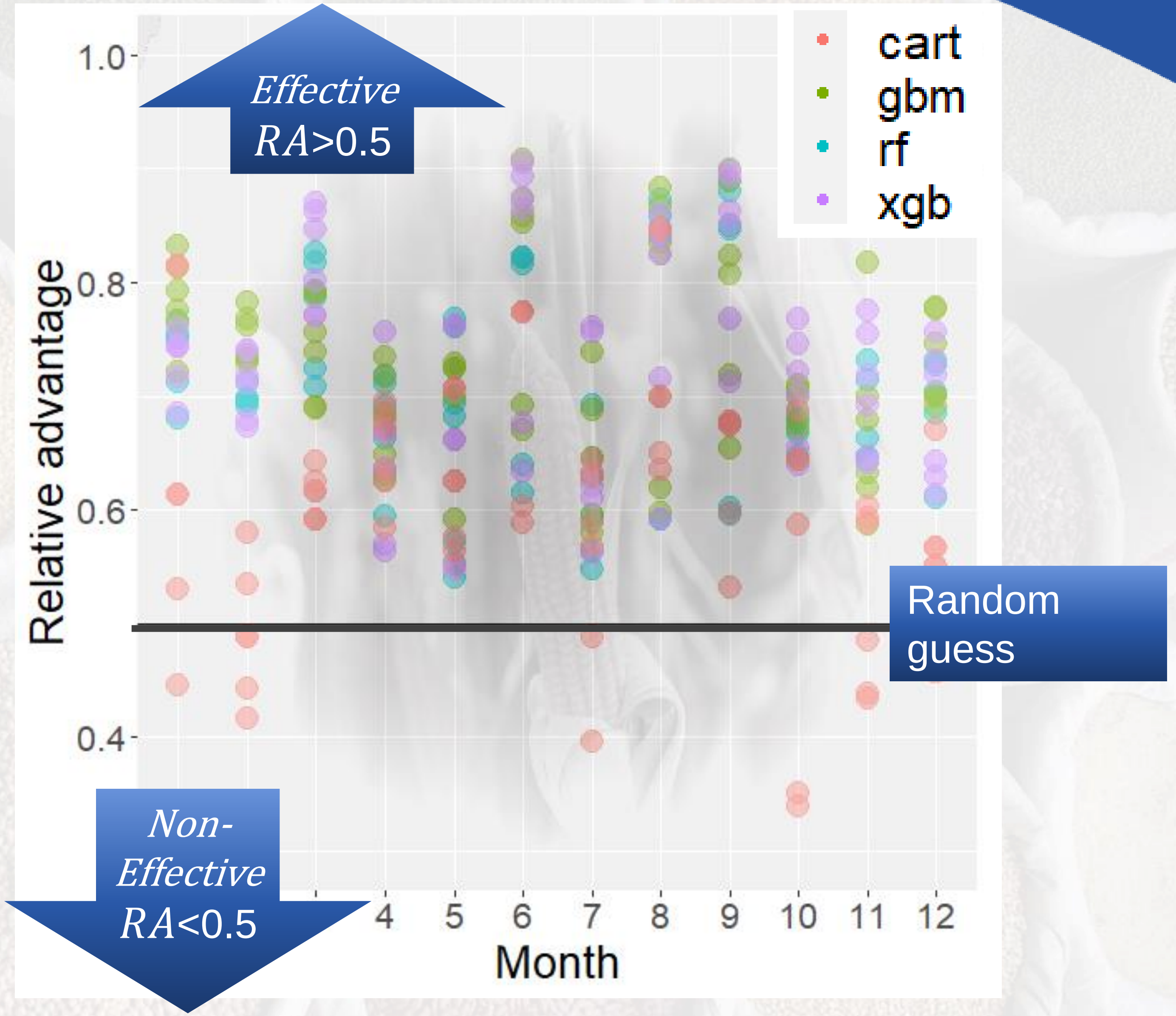
4. **Test** of the algorithms based on test dataset
5. **Assess** forecasting accuracy using Rolling CV
6. **Rank** features by their contribution level

<table>
<tr><th></th><th>y<44</th><th>…</th><th>y=44</th><th>…</th><th>y=Y-1</th><th>y = Y</th></tr>
<tr><td>y<44</td><td colspan="6"></td></tr>
<tr><td>…</td><td colspan="6"></td></tr>
<tr><td>y=44</td><td colspan="2">Train</td><td>Test</td><td colspan="3"></td></tr>
<tr><td>…</td><td colspan="3">Train</td><td>Test</td><td colspan="2"></td></tr>
<tr><td>y=Y-1</td><td colspan="4">Train</td><td>Test</td><td></td></tr>
<tr><td>y = Y</td><td colspan="5">Train</td><td>Test</td></tr>
</table>

Area Under ROC Curve (AUC)

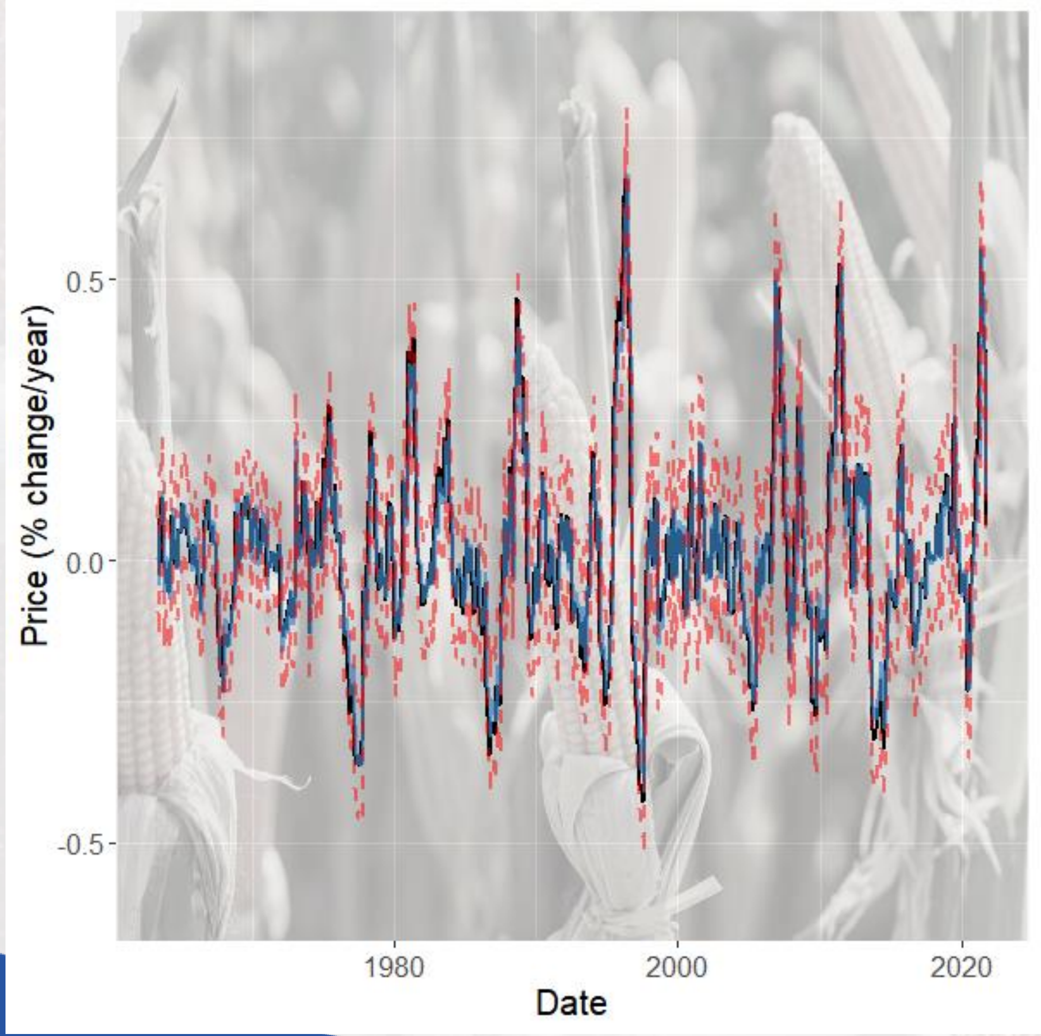
Higher = Better performance

Lower = Worse performance

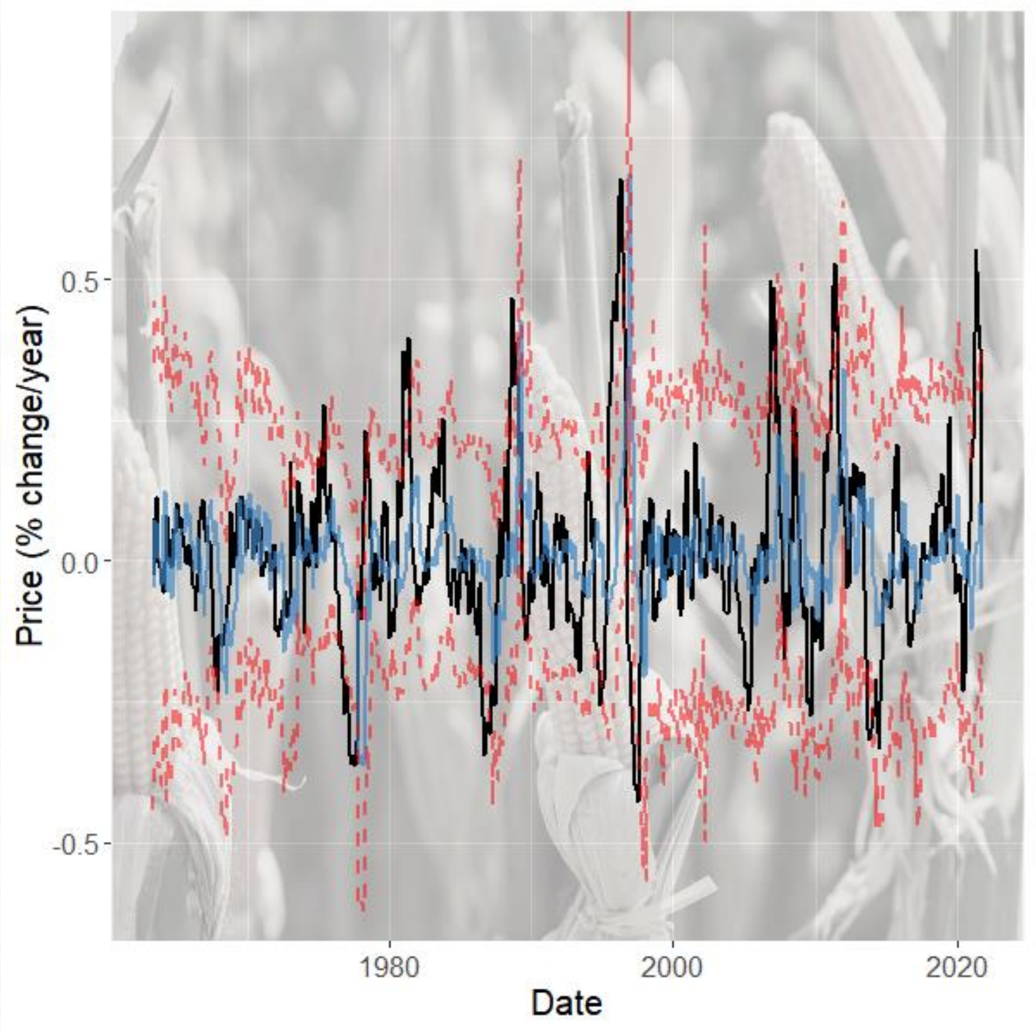


Time-Series Model - TBATS

Lag = 1 month



Lag = 12 months

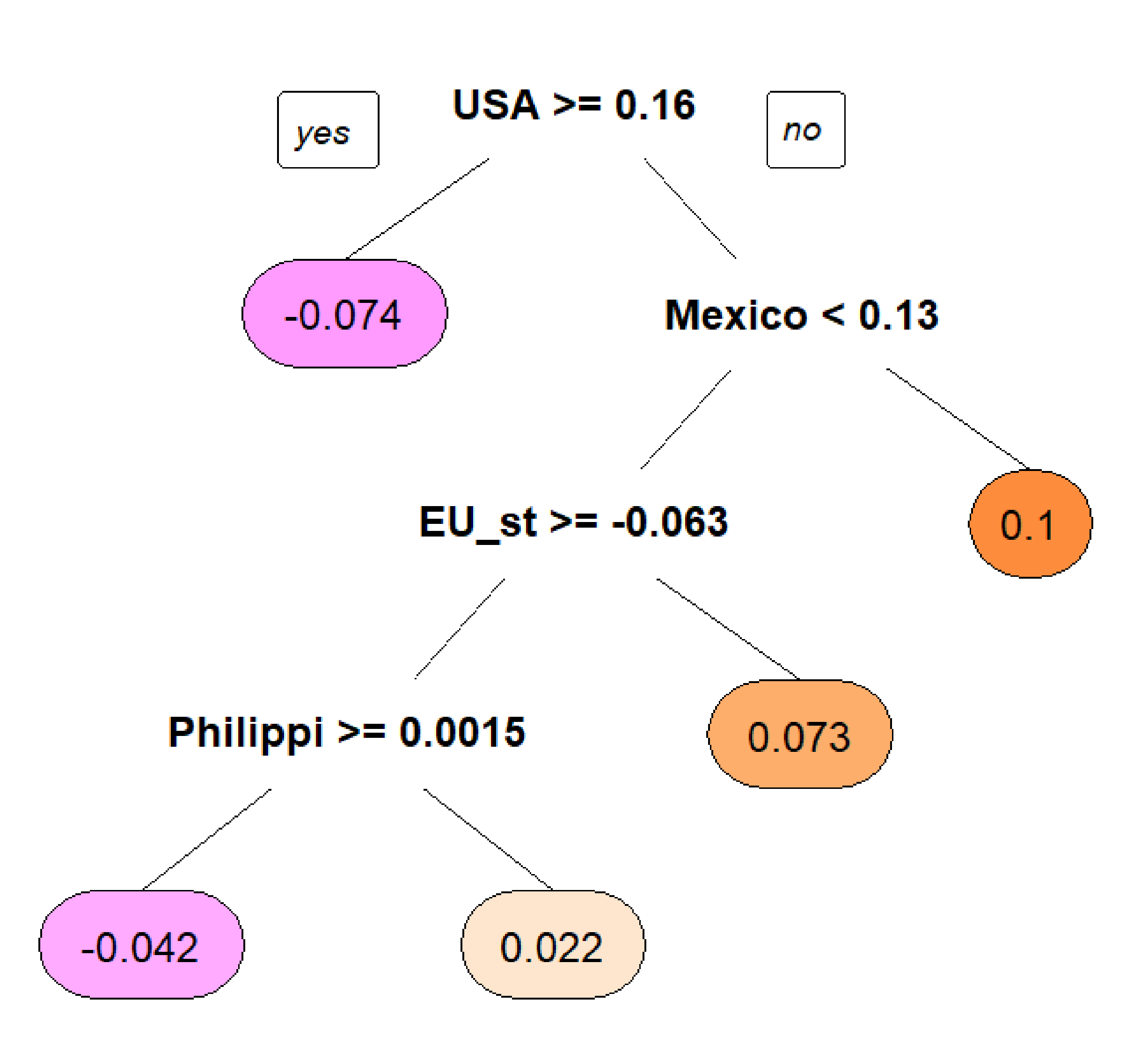


What drove the model's decision?

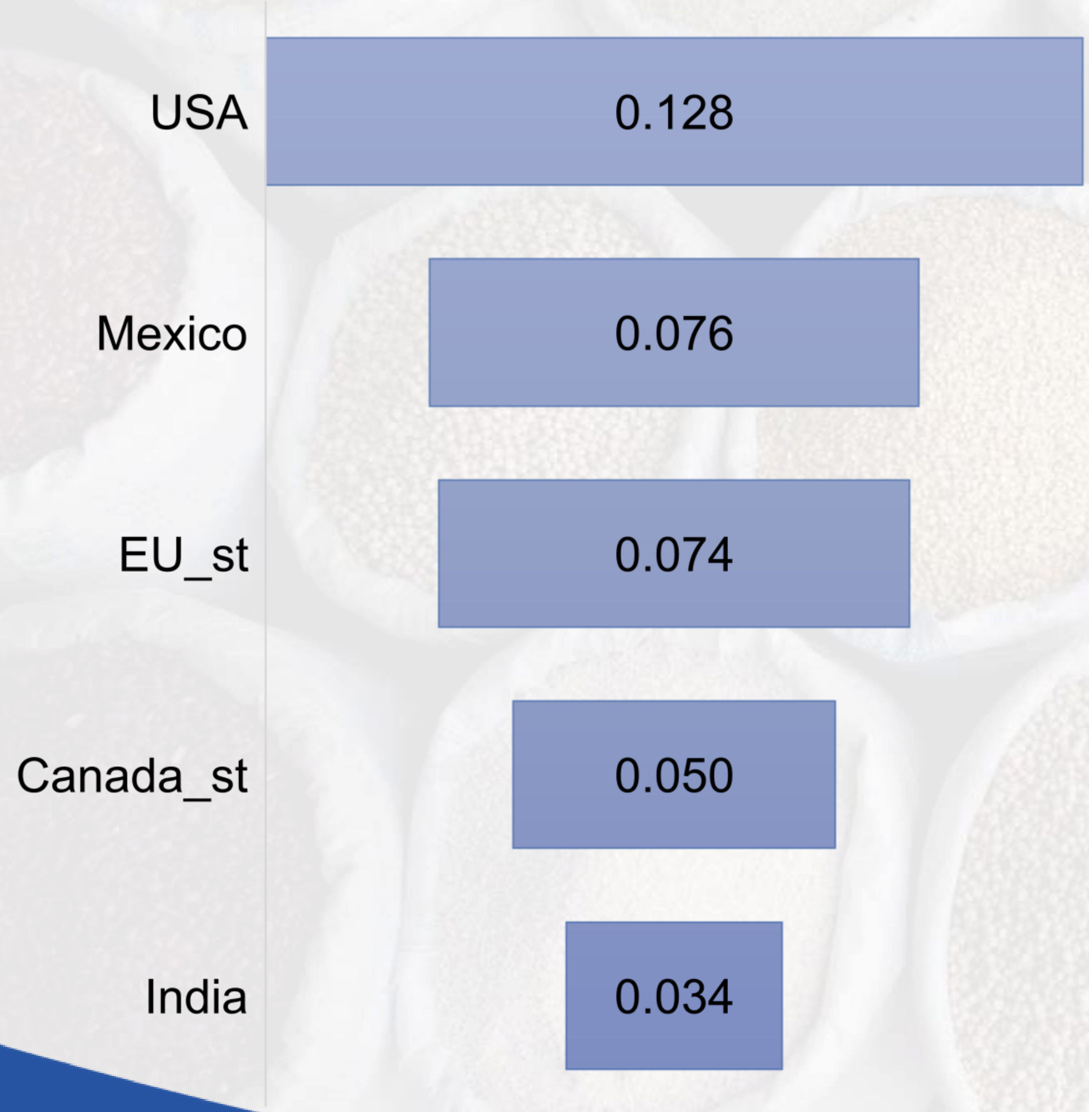
CART

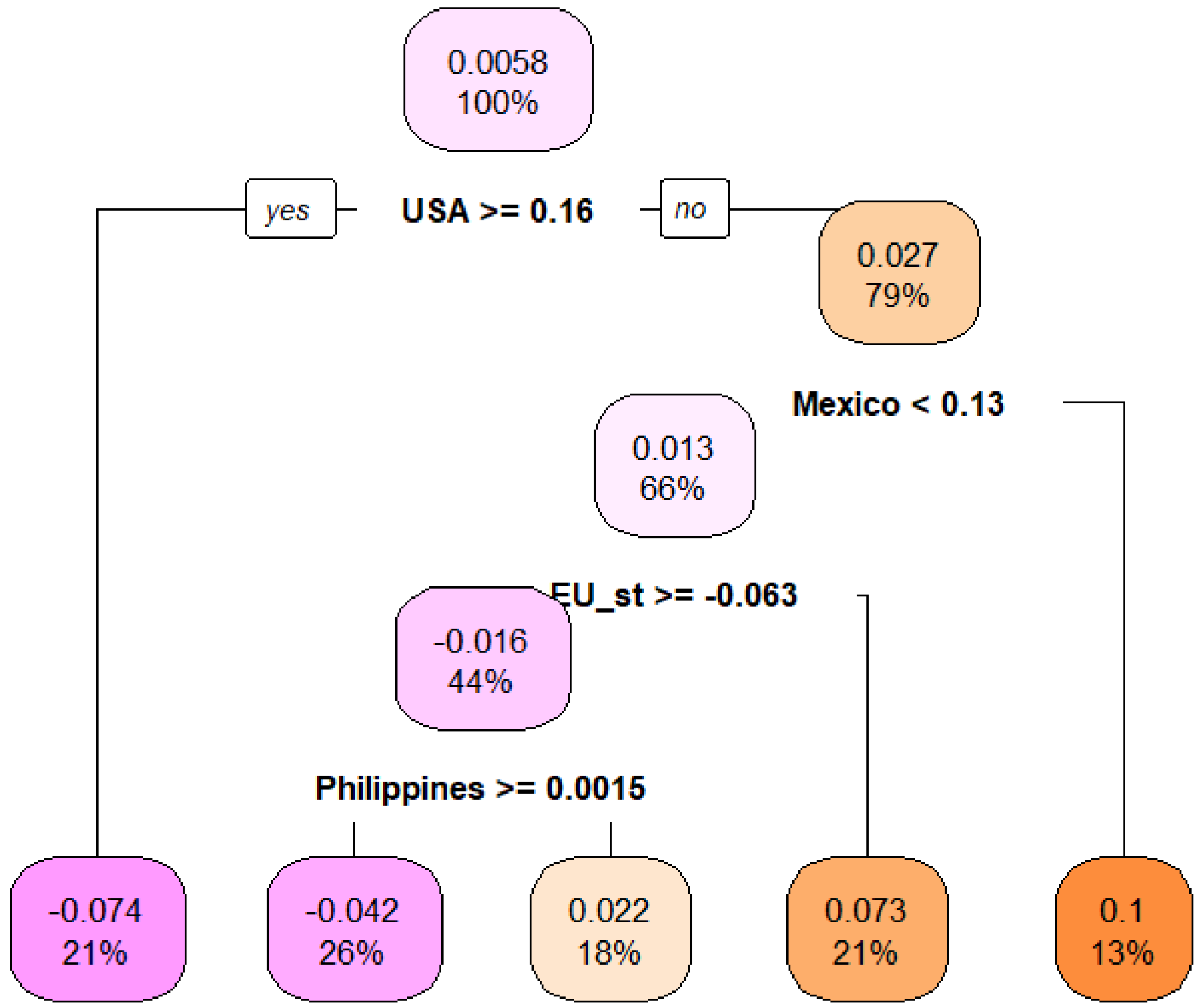
Classification **A**nd **R**egression **T**ree

- Split = partition into 2 subsets
- Simply visualised
- Easy to interpret
- Applied on black-box models

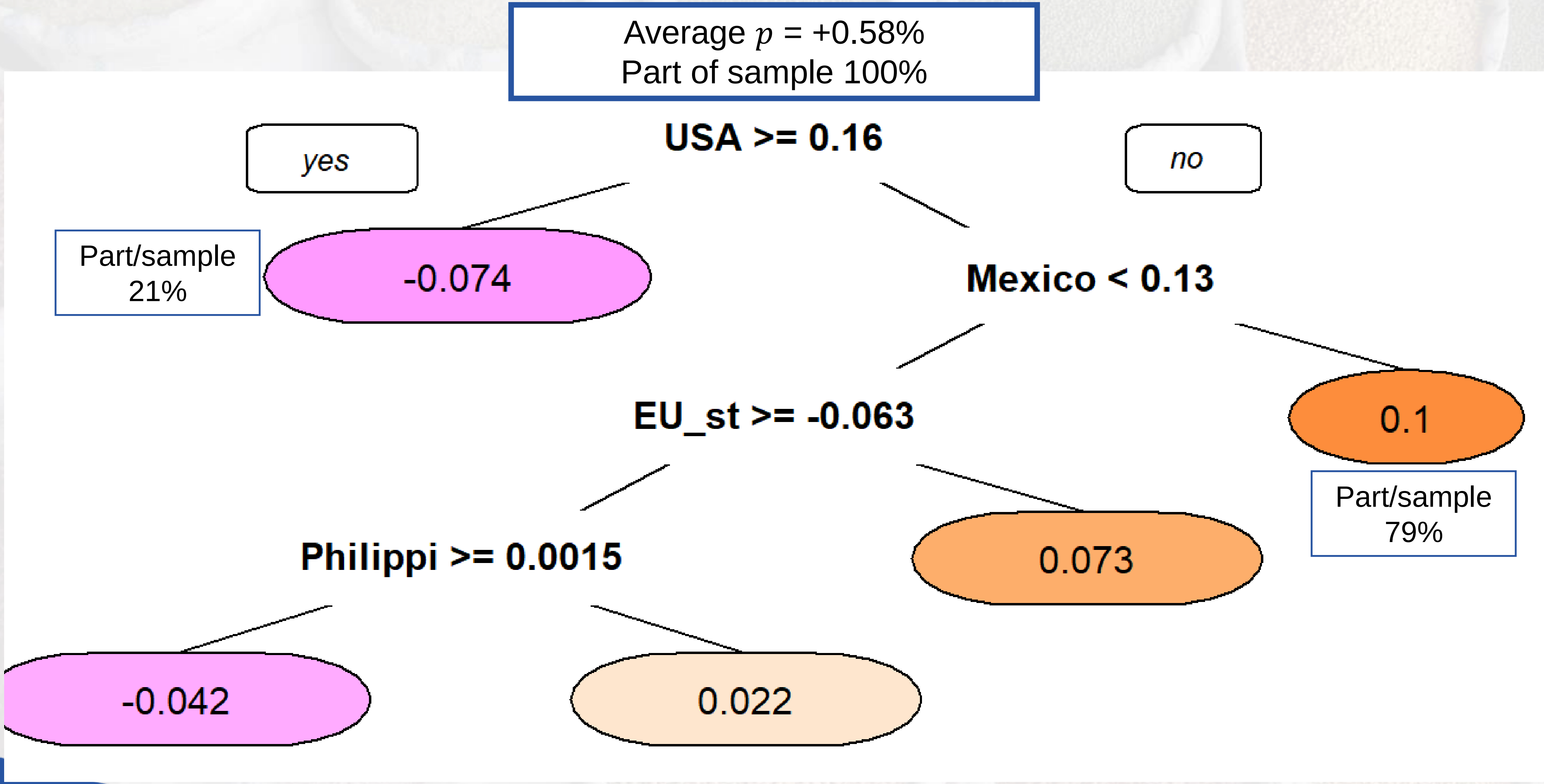


Variable Importance (October price)

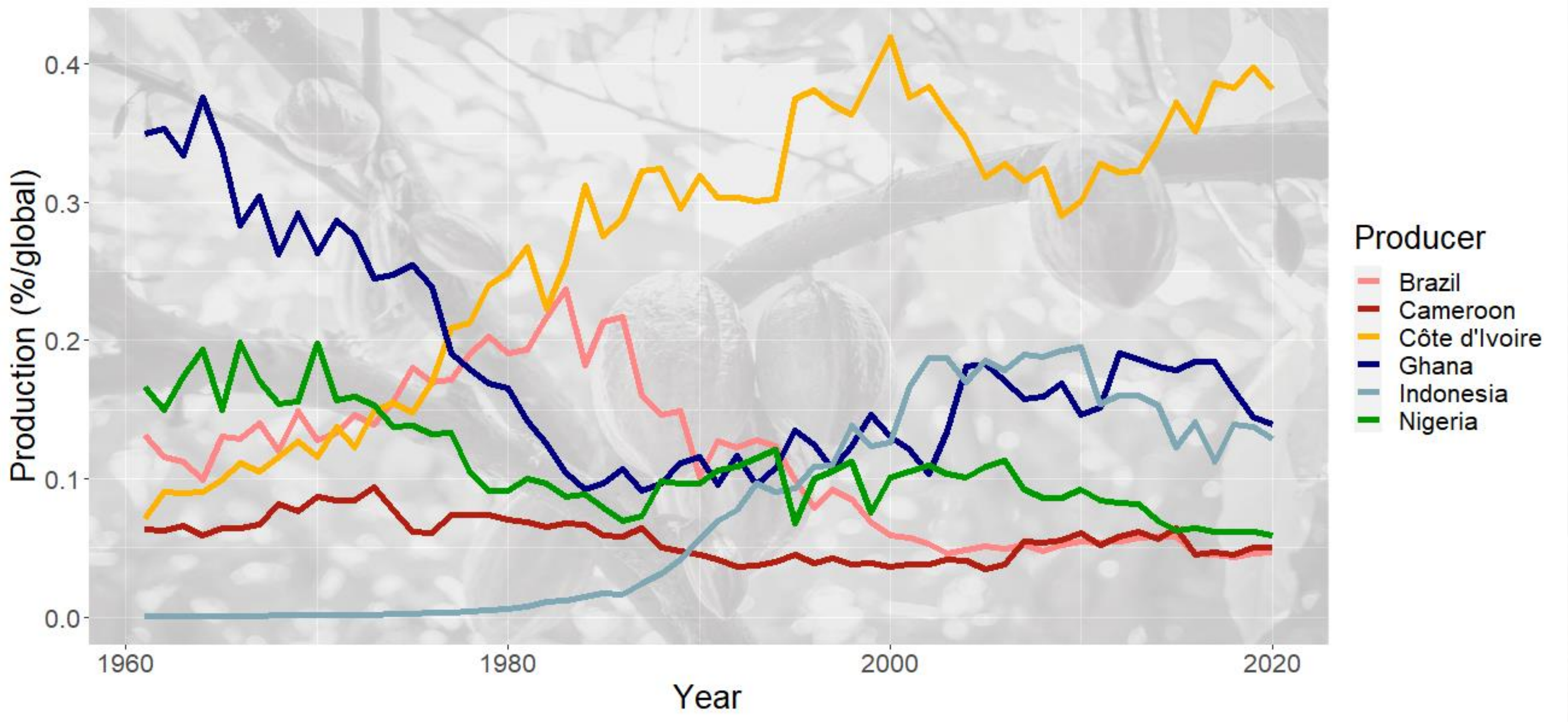




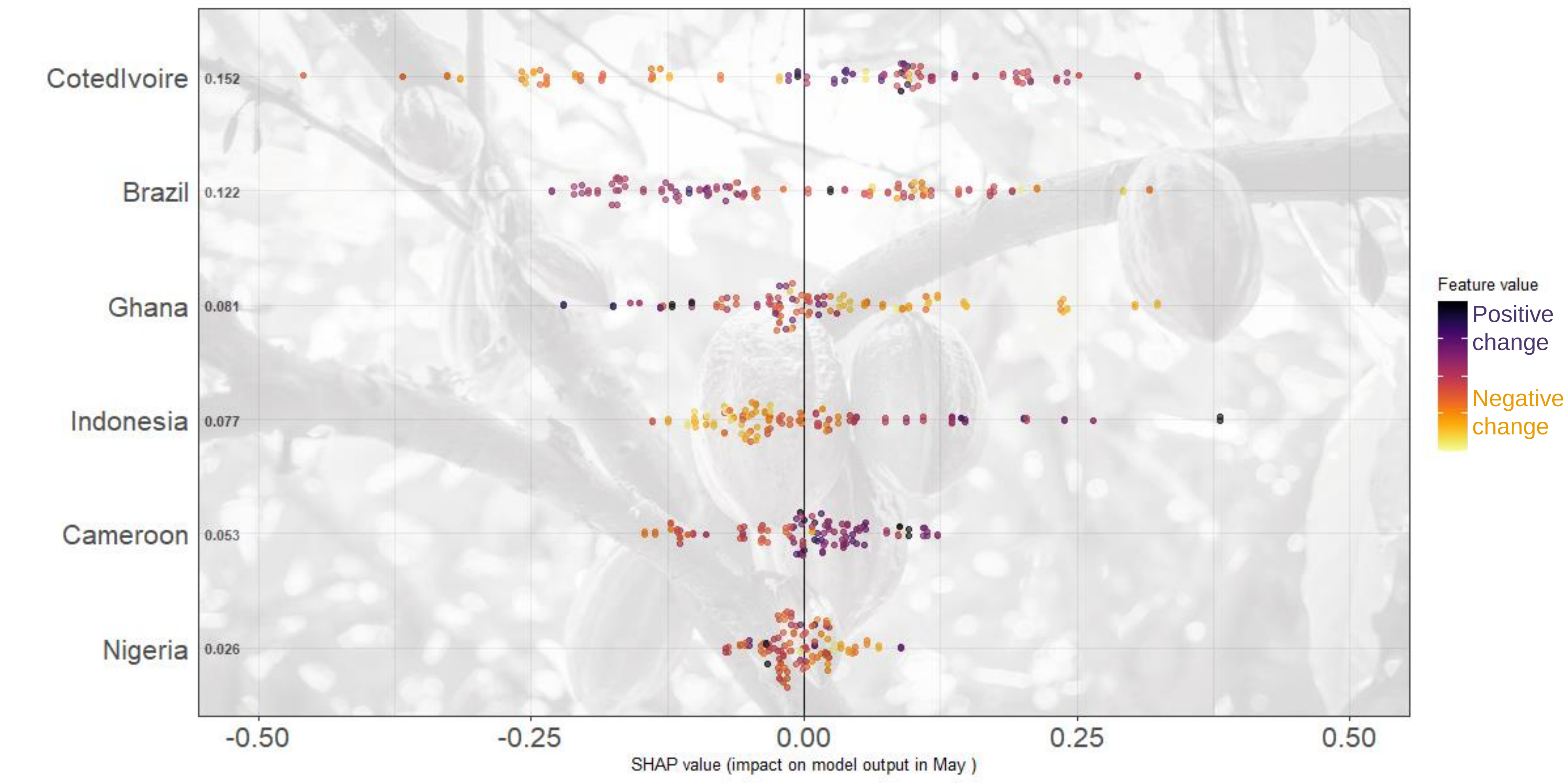
What happened in the model (October price)



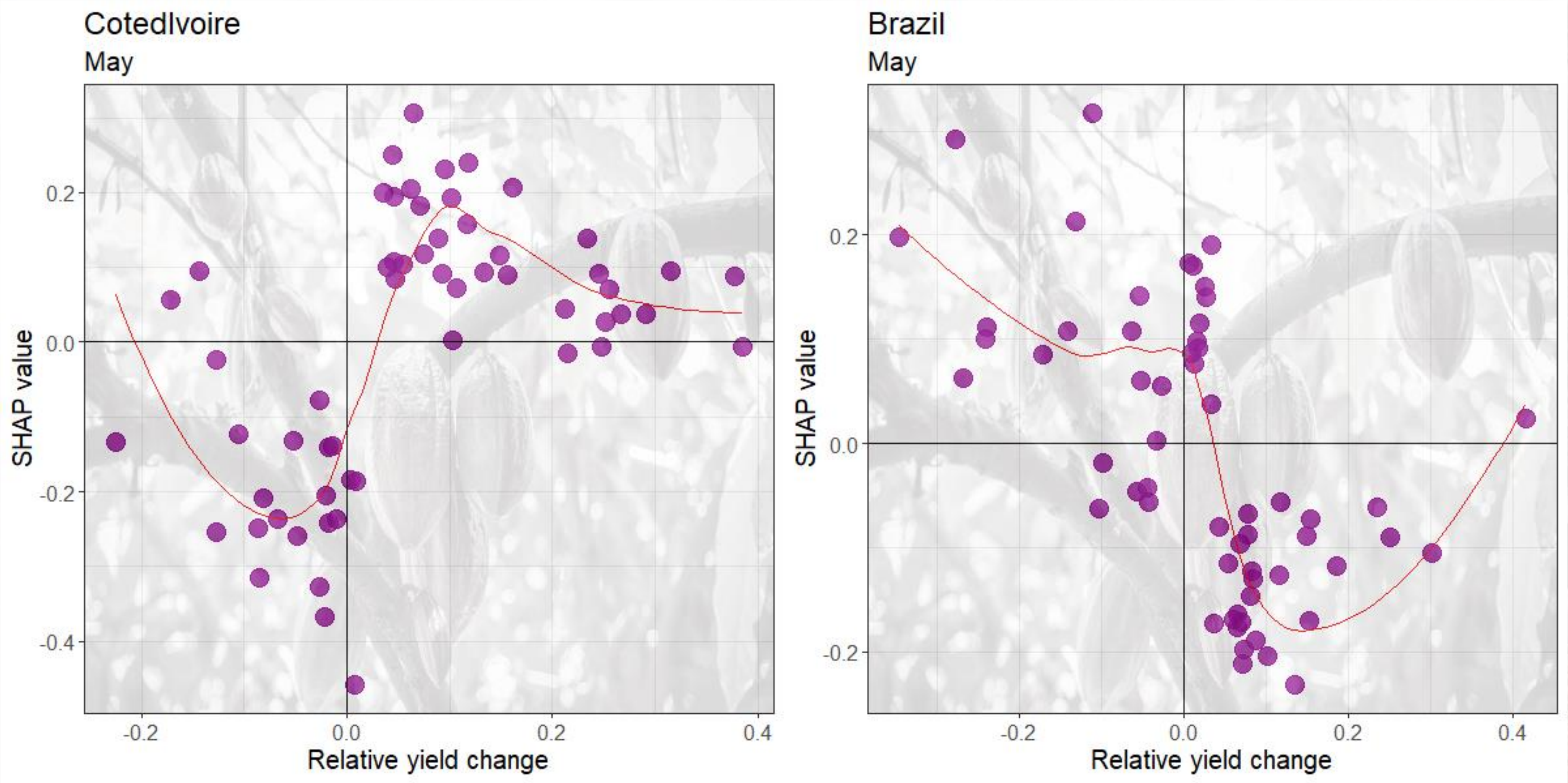
Cocoa top producing markets – Radical market chages



Relative Importance (SHAP) Cocoa



Partial Dependence - Cocoa



What did we achieve?

1. Understand how changes in local features affect AC global prices ✓
2. Forecast AC world prices from local production ✓
   - Applying comprehensive ML methods
   - Use public, regularly updated data
   - Achieve accurate forecasting results
   - In optimal time frame for adaptation (up to a year ahead)
3. Provide policy makers a transparent, ready-to-use model ✗