INSEAD The Business School for the World®

Not just random forests

Predicting forest loss through economic indicators

Data Science for Business – Final Project

SECTION AB GROUP 07

Erik UROSA | Martin MILENOVSKY | Maura FARRELL | Hill PRUKSANANONT | Mia BAO | Ian PARK

Agenda

A **Executive Summary**

B **Problem & Dataset**

C **Model definition**

D **Model interpretation & real world implications**

A

Executive Summary

Global environmental watchdogs and non-profits should focus their attention to a subset of ~20 countries which are in greatest danger of losing significant forested areas

Net loss of 1 Mn Ha between 2016 and 2024

- Brazil, Russia, and Indonesia expected to have greatest loss
- Net gain in countries like China does not offset overall decrease trend

GDP growth, GDP levels, and Agricultural % of part of GDP

- Countries with high % of agriculture are expected to experience the greatest loss to accommodate and grow crops
- High GDP growth countries are expected to see a net gain in forest land as they develop other industries and services

Most net loss in South America, Most net gain in Continental Asia

- South America's Amazon forest is most at risk; this is led by Brazil but 5 other countries in the top 15
- Africa and South East Asia are high loss regions as well
- China's overwhelming forest gain should be verified

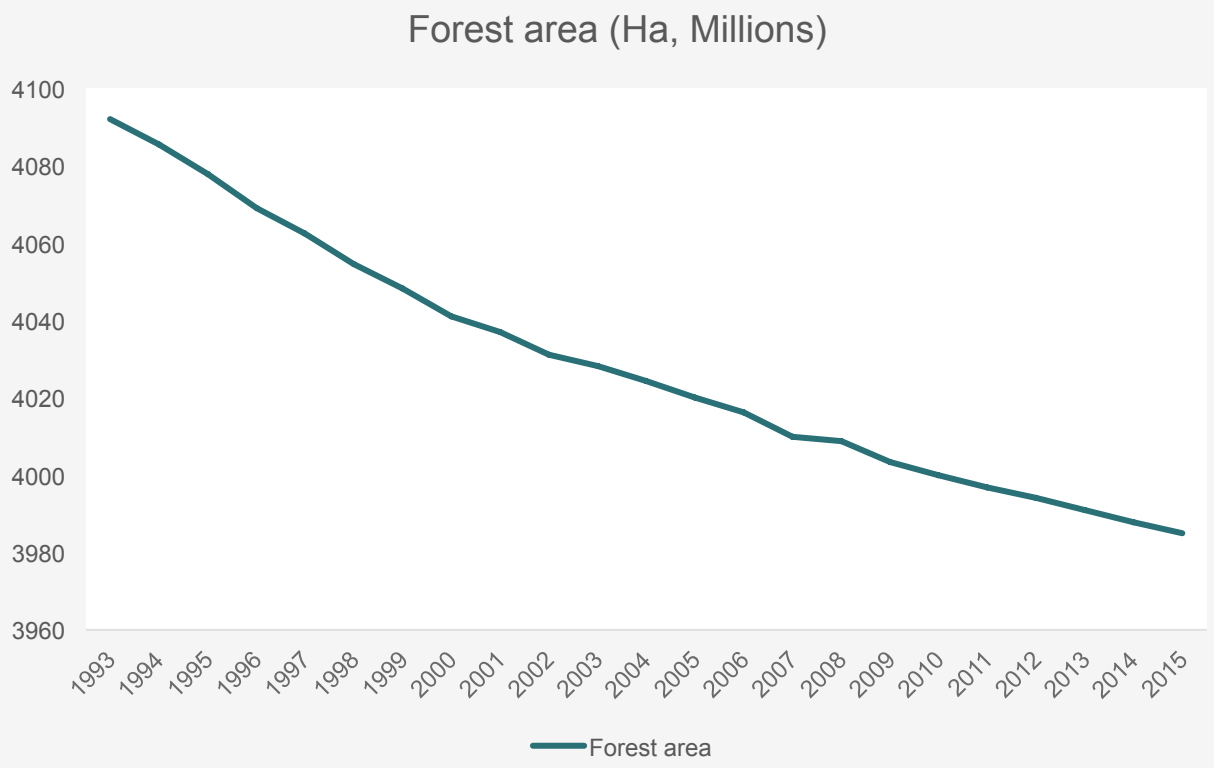
B

The world has lost 90M Ha of forest in 25 years

Problem

- Global **forested area declined steadily** between 1990 and 2015
- **Most extreme loss experienced in fast-growth developing countries** (e.g. Brazil 10% or Indonesia 20%)
- In the meantime other, mostly **developed countries, experienced net growth** in the net forest area (e.g. Sweden, France)
- Most **common reasons for de- or re-forestation** include:
  - Economic growth and development
  - Population growth
  - Agriculture activity
  - Changing commodity prices

Forest area (% of land surface) – 1990 - 2015



Is the de-forestation trend going to continue or will the trend turn around as the countries get more industrialized?

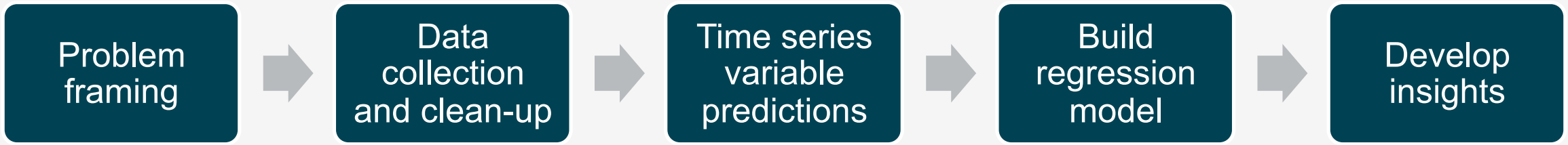
B

We are seeking to understand how economic growth affects the health of our forests through data science

Key questions

- Will the forest loss continue? How **much of forest area will be lost by 2024?**
- What are the **economic factors that are related** to forest area loss - Can we prevent further forest loss by addressing these underlying financial incentives?
- Which specific countries or regions of the world are **most vulnerable** to future forest loss?
- Are there any regions where the forest loss **trend will revert?**

Process



B

Data collection and clean up was a time intensive process given various country differences

Field Meaning	Description
Country Code	Unique 3-letter country codes assigned by ISO
Country Name	Generally accepted 1- or 2-word names for UN-recognized countries
Year	Standard Gregorian Calendar Year
Forest Area	Surface area of country which has been designated as a “forest” by international standards
Forest Area y-o-y change	Change in forest area vs. previous year
Population	Specific country population per previous census or estimates
Area	Total country land area
Population density	No. of people per sq. km
Agriculture / GDP	Share of agricultural production as % of total GDP (nominal, US$)
Gross Domestic Product per Capita (PPP)	Economic productivity by country, per capita, considering “purchasing power parity”, a measure of standard of living
Gini coefficient	Measure of statistical dispersion intended to represent the income or wealth distribution of a nation's residents
Price of Hard Logs	Soft Logs, Average Export price from the U.S. for Douglas Fir
Price of Soft Logs	Hard Logs, Best quality Malaysian meranti, import price Japan
Price of Palm Oil	Malaysia Palm Oil Futures (first contract forward) 4-5 percent FFA
Shock (soft log)	Has there been a month in which soft log price increased by more than 20% vis-à-vis 12-month rolling average?
Shock (hard log)	Has there been a month in which soft log price increased by more than 20% vis-à-vis 12-month rolling average?
Shock (palm oil)	Has there been a month in which soft log price increased by more than 20% vis-à-vis 12-month rolling average?
Forested country	Yes if total forest area in 1990 represented more than 50% of total country area

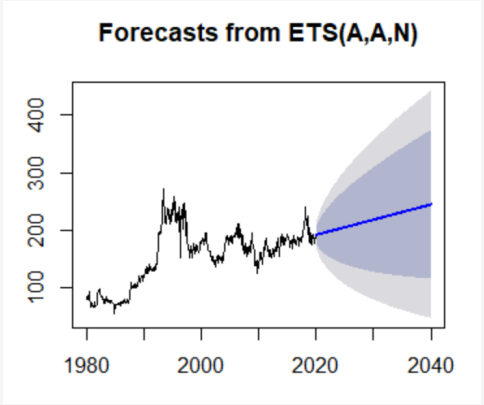
C

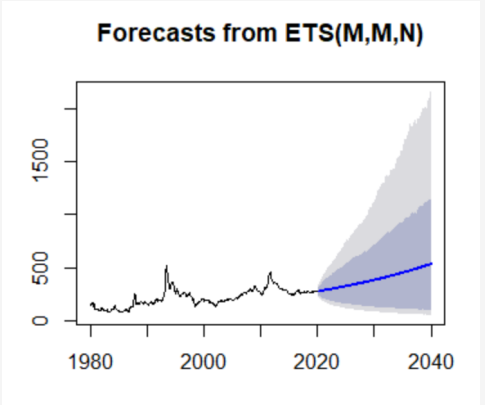
Time series predictions are possible for historical monthly commodity price data

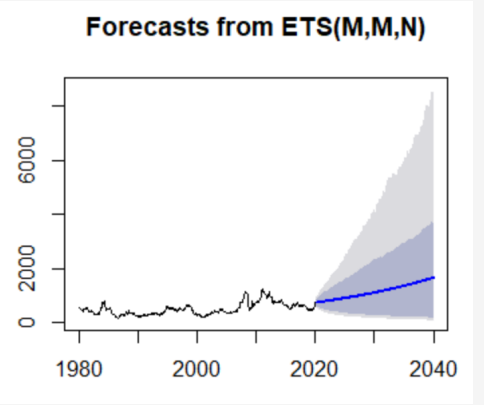
- In order to make predictions about the future forest area, we need to obtain the future values for all relevant variables
- For commodity prices: Soft Log, Hard Log, and Palm oil, we have decided to run the time series regression
- Time series makes sense in this context since commodity prices follow certain cycles and trends

	Soft Log	Hard Log	Pail Oil
Models considered	• AAN, AAA, MMN, MMM	• AAN, AAA, MMN, MMM	• AAN, AAA, MMN, MMM
Model with least mean error	• AAN	• MMN	• MMN

Plots of selected model







C Prediction consisted of 3 steps

Step One

- **Split data** into training (1995 – 2012) and testing (2013 – 2015) sets
- **Train** model on training set
- **Validate** model on testing set
- **Calculate** the expected absolute and percentage loss for each country based on the prediction above

Step Two

- **Refine** model by introducing interaction terms, feature engineering, and non-linear terms
- **Pick the model** with the highest MAPE and use it to generate predictions

Step Three

- **Use** the chosen model to predict forest loss in years 2016-2024
- Calculate the expected absolute and percentage loss for each country based on the predictions above

C

A regression model can assist in making forest loss predictions by country

Initial regression model	**Y-o-Y change in Forest Area ~** GDP per capita + GDP growth + population density + population growth + Gini coefficient + Agriculture share of GDP + Soft Log + Hard Log + Palm Oil + Price Shock dummies **R^2 = 15%**

In order to capture more variation in forest area changes country fixed effects were introduced

Improved regression model	**Y-o-Y change in Forest Area ~** GDP per capita + GDP growth + population density + Gini coefficient + Agriculture share of GDP + Forested country dummy + Country Fixed effects **R^2 = 74%**

C

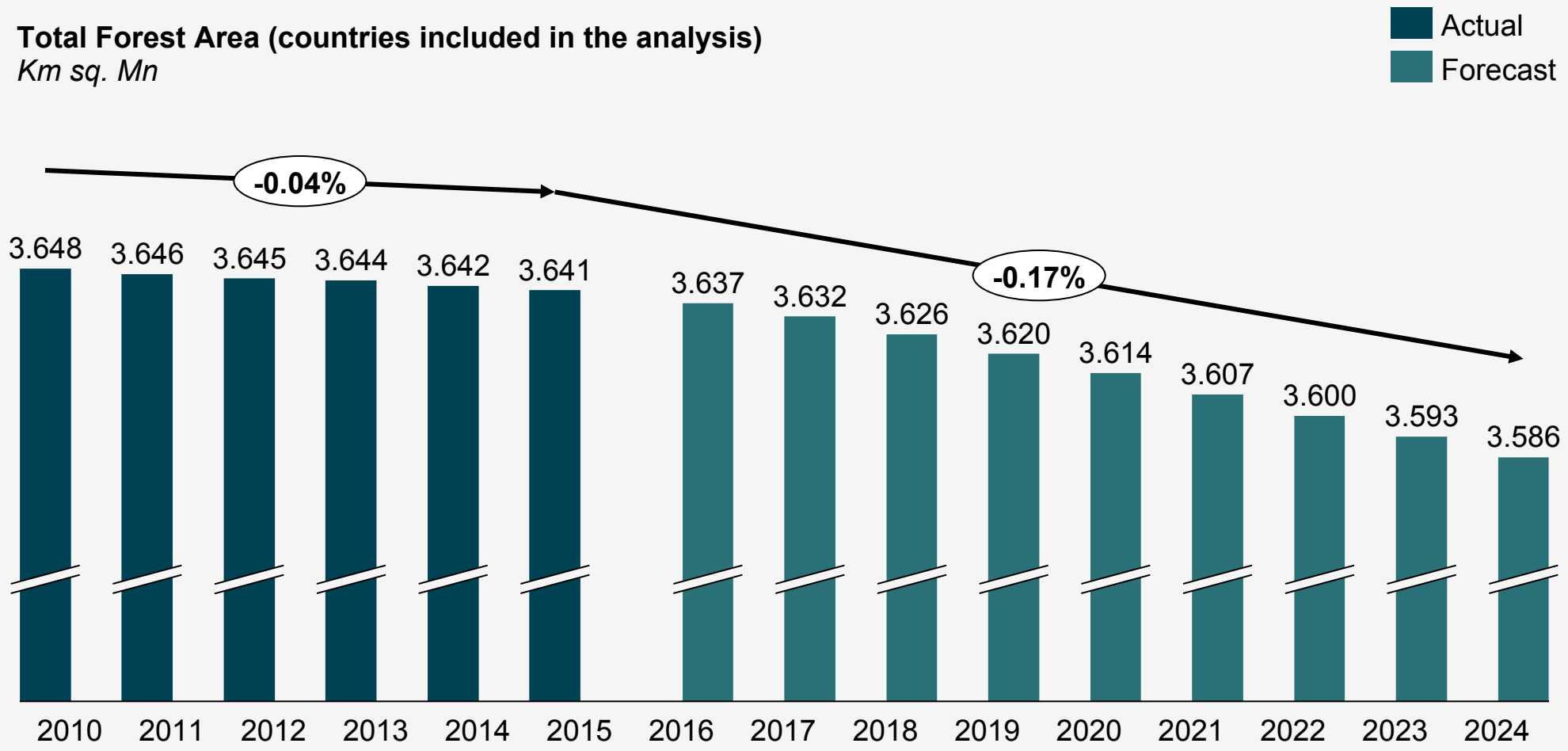
The most significant indicators of expected forestry loss are related to the economic profile of the country

Variable	Agriculture % of GDP	GDP Growth %	GDP Level
Significance (P-value)	$1.39*10^{-5}$	0.0096	0.0353
Directionality	Negative (-)	Positive (+)	Negative (-)
Rationale	Countries with high agriculture % of GDP are most likely to de-forest for expansion of this crucial industry	High GDP growth countries are transitioning from natural resource use to other industries and services	Controlling for Ag % of GDP, the higher income countries are predicted to experience higher annual forest loss

D

Forest loss will not be stopped and is predicted to decline by 0.17% p.a. between ’16-’24

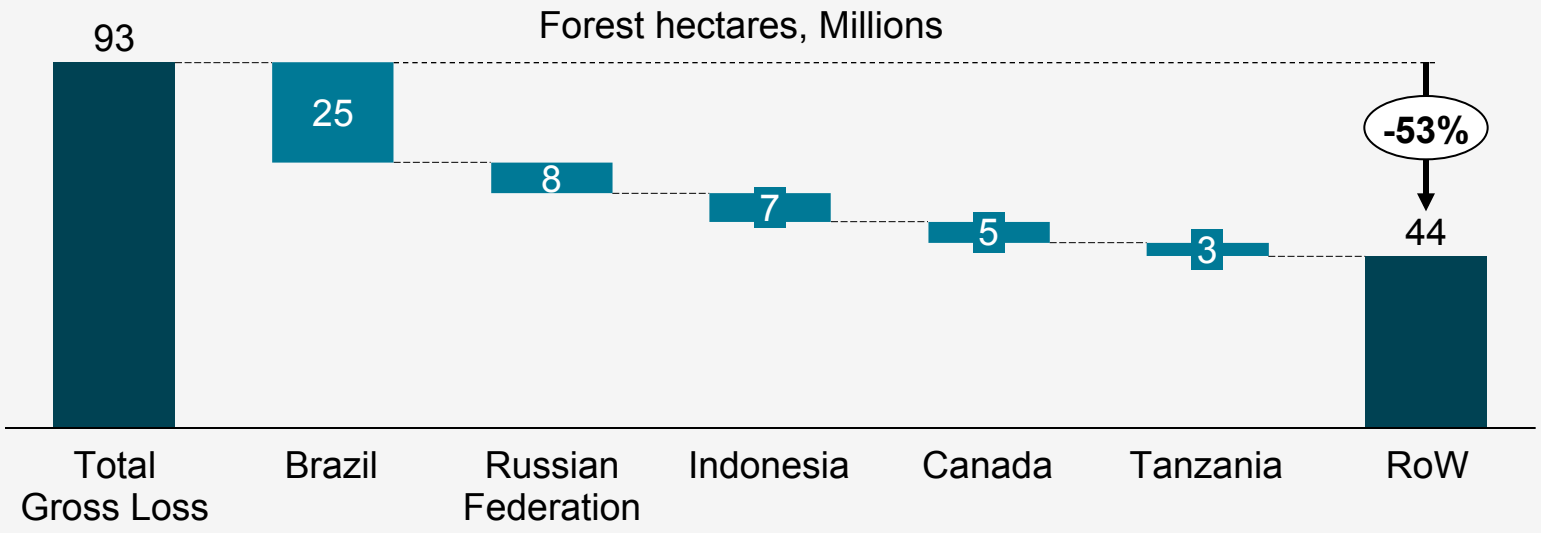
Total Forest Area (countries included in the analysis)
Km sq. Mn



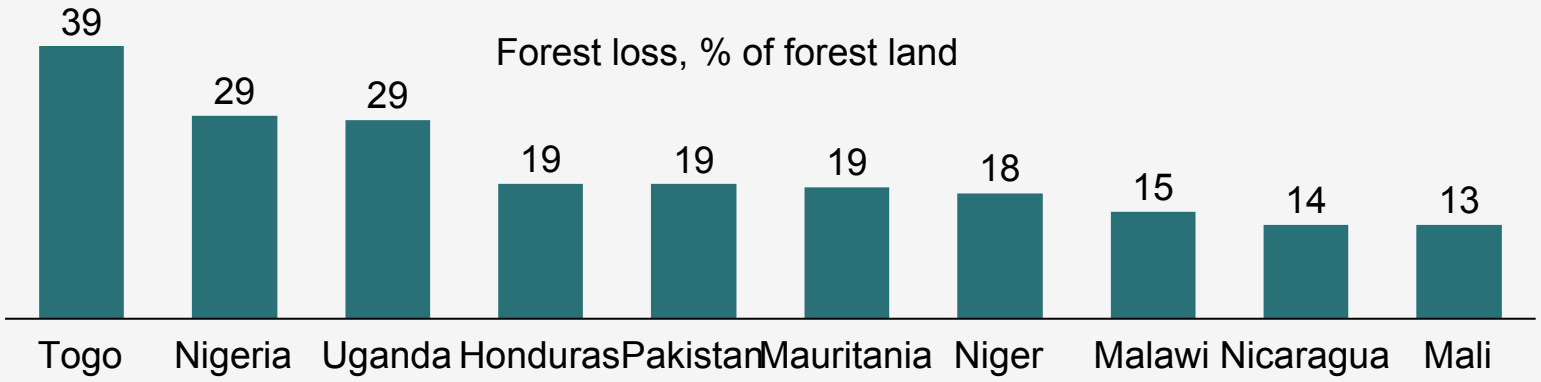
D

More than 50% of de-forestation is expected to occur in just 5 countries

Greatest absolute expected loss



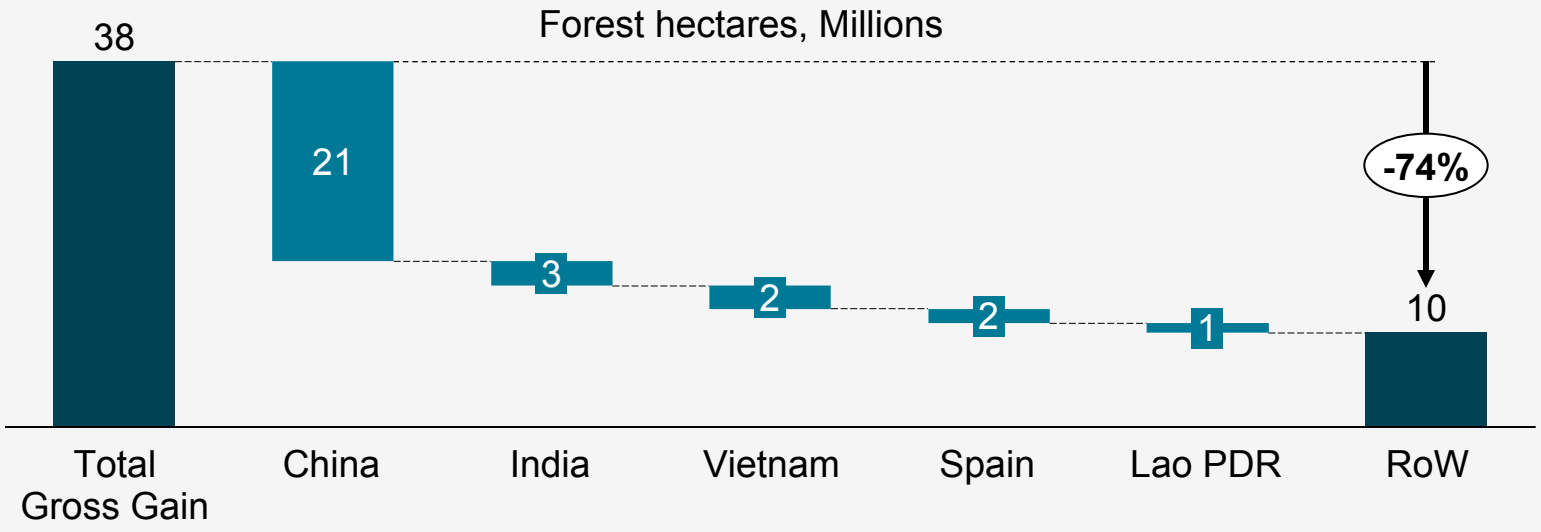
Greatest expected % loss



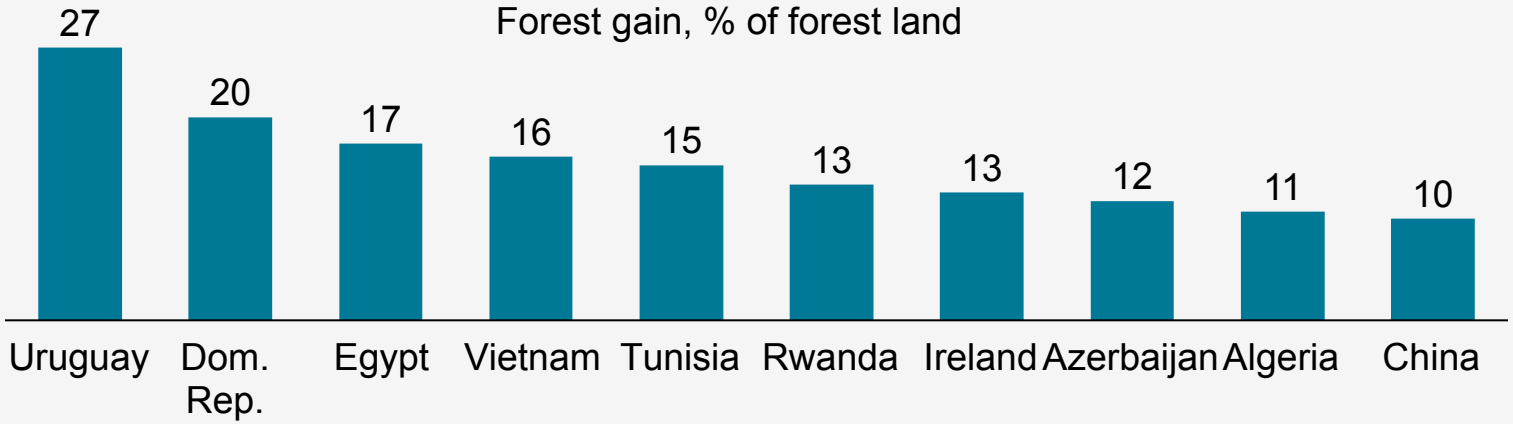
- 74 countries expected to lose forest area between ’16 and ‘24
- 5 countries will account for more than 50% of all forest loss between 2016-24
- In terms of greatest % loss, African countries take 7 of 10 countries with greatest expected forest loss

China leads among the countries which will manage to increase number of forests

Greatest expected absolute gain



Greatest expected % gain



- 51 countries expected to increase their total forest area
- Only 5 countries will account for more than 75% of all forest loss between 2016-24
- Greatest gain among mid-income countries (exc. Ireland) which experience transition from agriculture to service industry

Recommended next steps for deforestation analysis

Suggestions for de-forestation alleviation

- Target **top loss countries** for prospective de-forestation avoidance countries
- Take **lessons from forest land gains** for application at “neutral” loss countries
- **Communicate the impact of agriculture focus and GDP growth** to governments and corporations, so that these relevant factors can be addressed
- Analyze **positive economic models of “sustainable forestry”** as an alternative asset class for investors and asset managers

Additional data for collection and analysis

- Further **segment agriculture activities** and separate from fishing activities
- Understand the impact of **different major crops** such a rice, maize, and wheat
- Variables on the **“type of forest”** such as temperate vs. tropical to understand if they have different outcomes
- **Average temperature**
- **Expected emission** captured by type of forest

Helpful modeling steps for further evaluation

- Further understand **possible correlation** between variables (Ag % GDP vs. Absolute GDP levels)
- **Segmentation of different countries** to identify those with similar characteristics that could implement more targeted programs
- **Additional feature engineering** related to country level of development