A REPRODUCIBLE ENSEMBLE MACHINE LEARNING APPROACH TO FORECAST DENGUE OUTBREAKS

ESA Φ-lab: Rochelle Schneider, **Alessandro Sebastianelli**, Raquel Carmo, James Wheeler

UNICEF Giga: Do-Hyung Kim

UNICEF Latin America and Caribbean Regional Office: Hanoch Barlevi, Zoraya El Raiss Cordero

Italian Space Agency / Sapienza University of Rome: Dario Spiller

Warsaw University of Technology: Artur Nowakowski

Wellcome Trust: Felipe de Jesus Colon Gonzalez

University of Sannio: Silvia Ullo

Fluminense Federal University (Brazil): Ludmilla Viana Jacobson

Barcelona Supercomputing Centre: Rachel Lowe

OUTLINE

Data Collection

DATA COLLECTION

Climatic

ERA5-Land

Altitude

Shuttle Radar Topography Mission

Dengue cases

Sistema de Informação de Agravos de Notificação (SINAN)

Forest Loss

Landsat 7 and 8 satellites

Population

Instituto Brasileiro de Geografia e Estatistica (IBGE)

NDVI

NASA Moderate Resolution Imaging Spectroradiometer (MODIS)

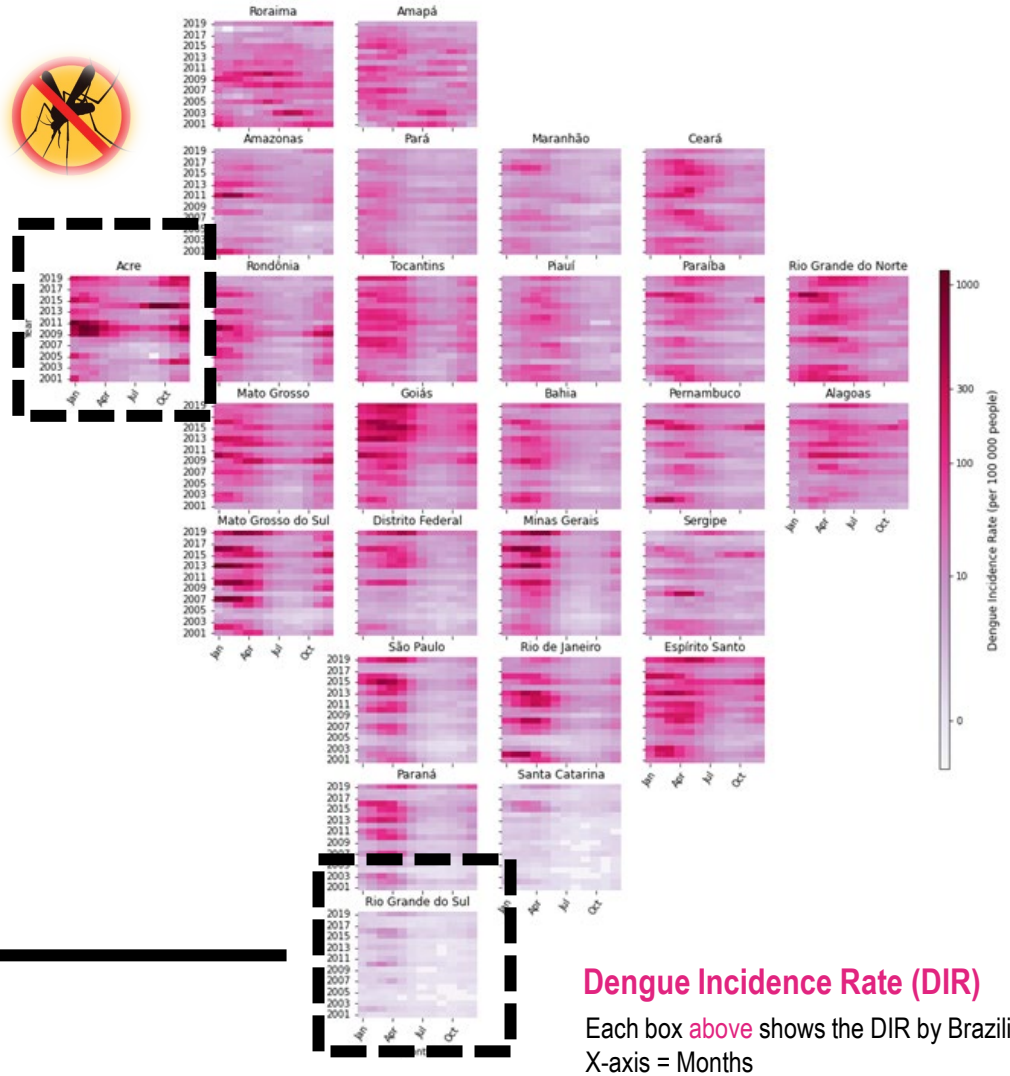
Settings

Period of analysis:
19 years
(2001/01 – 2019/12)

Time step:
1 month

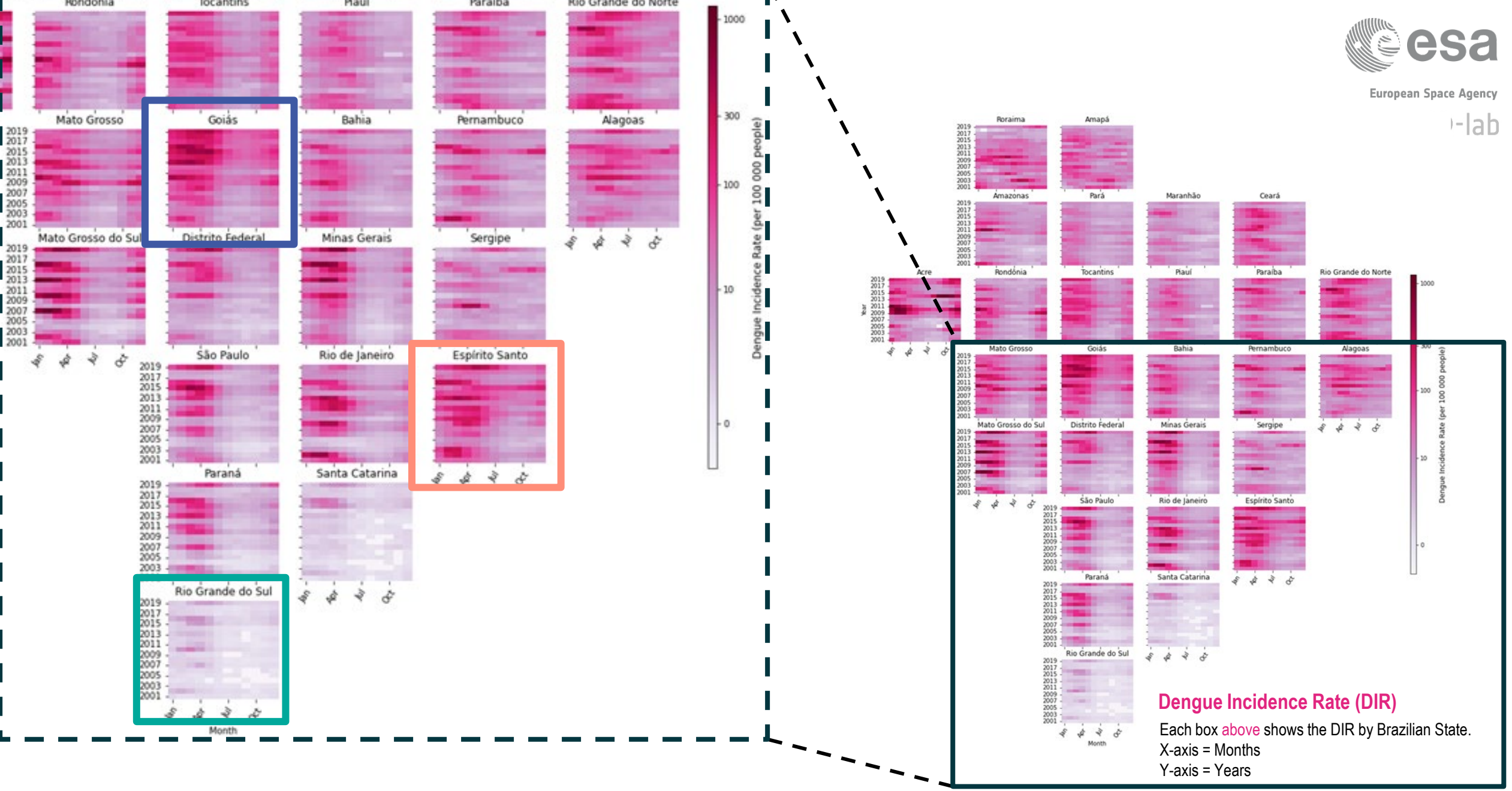
Spatial Resolution:
Federal Unit

PRODUCT EXAMPLE



Dengue Incidence Rate (DIR)

Each box above shows the DIR by Brazilian State.
X-axis = Months
Y-axis = Years

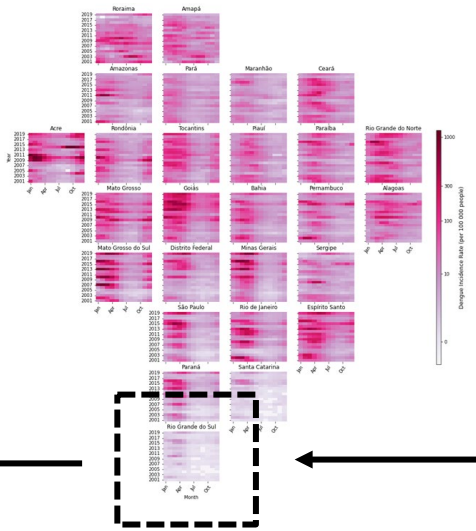

Rondônia
Tocantins
Piauí
Paraíba
Rio Grande do Norte
Mato Grosso
Goiás
Bahia
Pernambuco
Alagoas
Mato Grosso do Sul
Distrito Federal
Minas Gerais
Sergipe
São Paulo
Rio de Janeiro
Espírito Santo
Paraná
Santa Catarina
Rio Grande do Sul
Roraima
Amapá
Amazonas
Pará
Maranhão
Ceará
Acre
Dengue Incidence Rate (per 100 000 people)
Year
Month
Jan
Apr
Jul
Oct
Dengue Incidence Rate (DIR)
Each box above shows the DIR by Brazilian State.
X-axis = Months
Y-axis = Years

PRODUCTS IN OUR DATASET

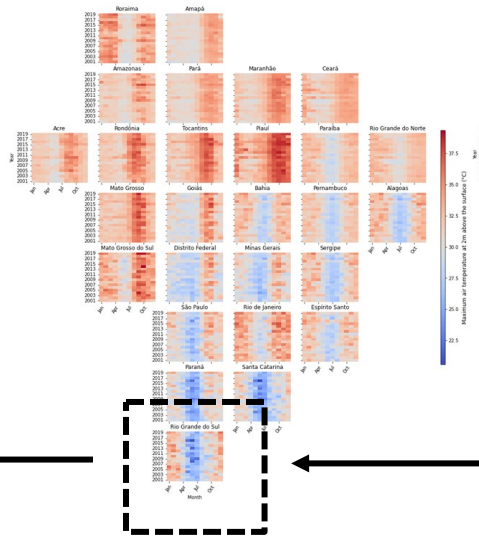
European Space Agency

27 Brazilian States

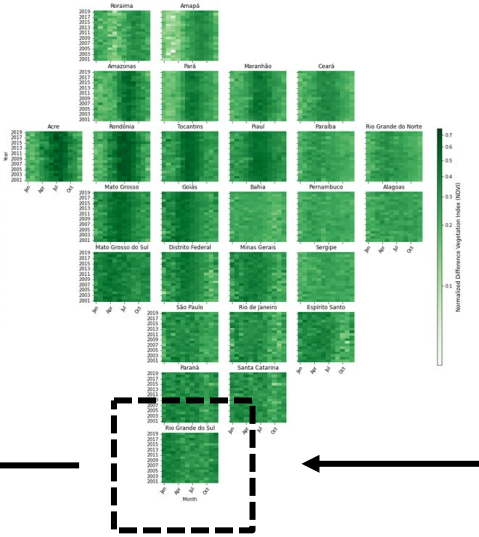
Dengue Incidence Rate (DIR)



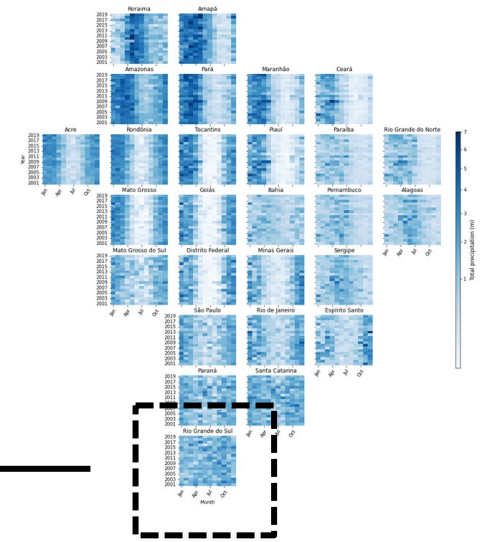
2M MAX AIR TEMPERATURE



NDVI

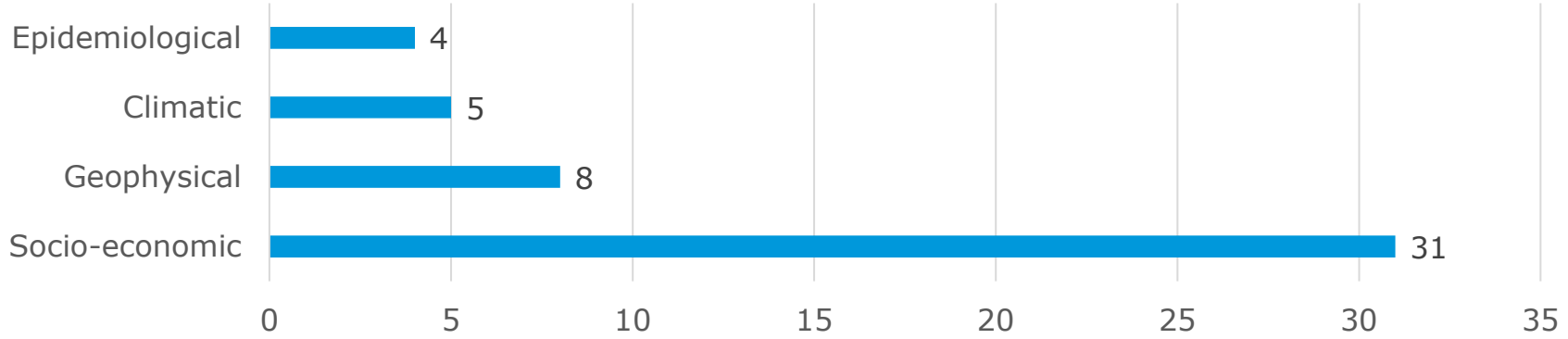


PRECIPITATION

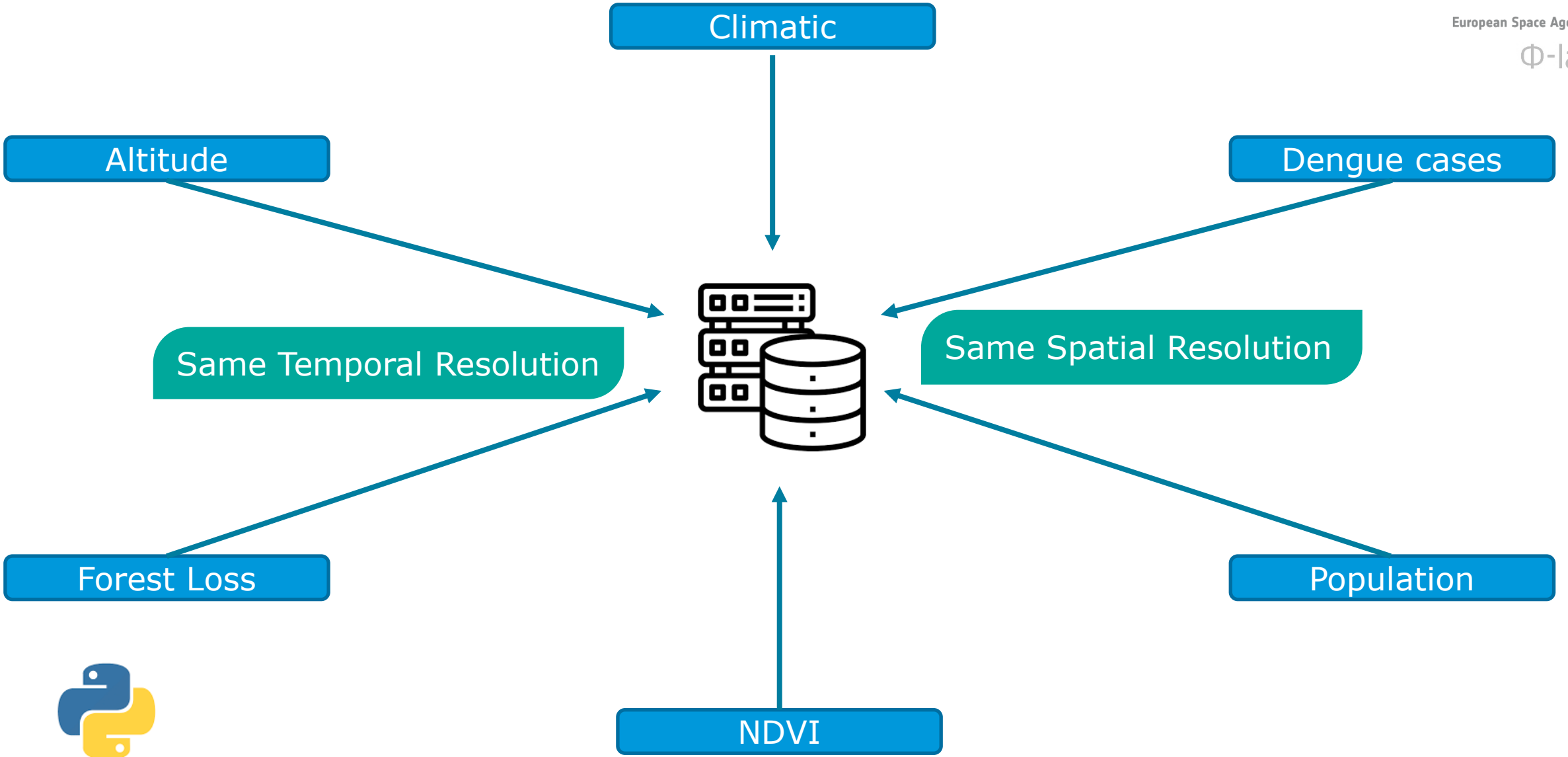


Each box shows the EO products by Brazilian State.
X-axis = Months
Y-axis = Years

Number of products for each cluster

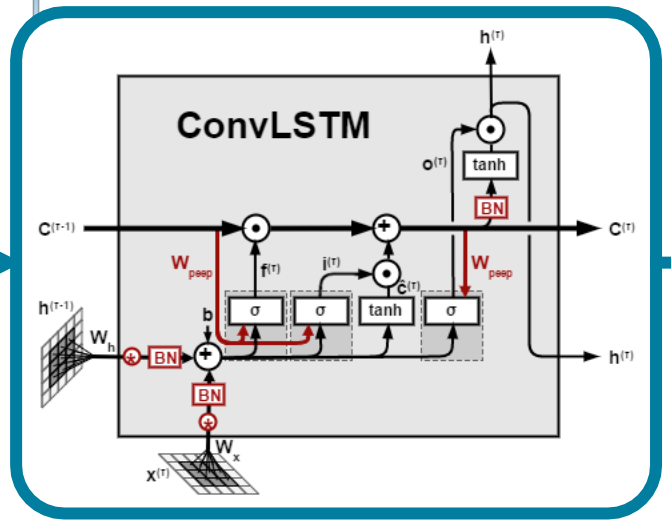


DATA SYNCHRONIZATION



OUTLINE

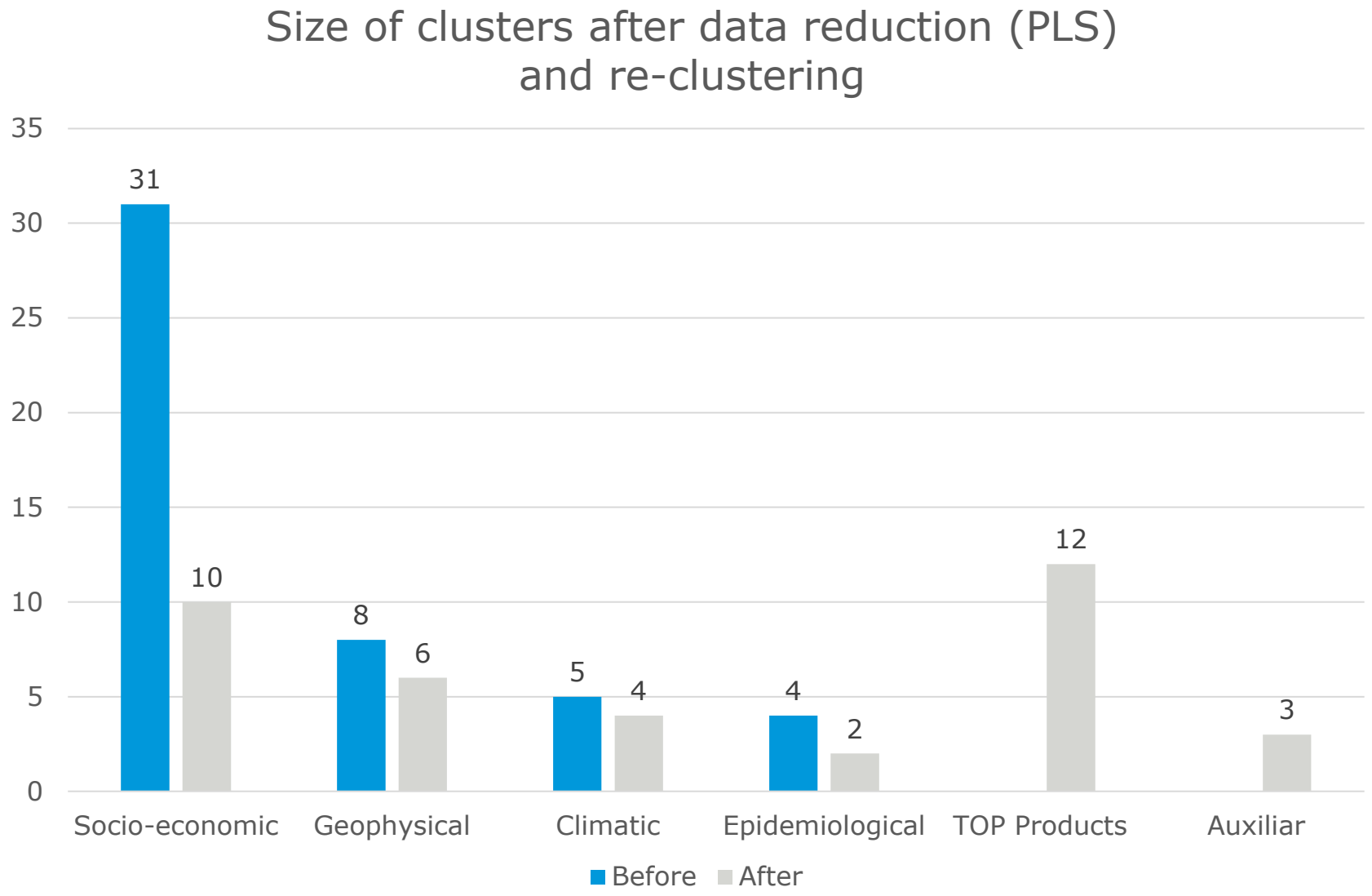
SPATIAL CORRELATION



Rough estimation of Dengue

The Convolutional LSTM looks at the spatio-temporal correlation among the data and returns a first guess/rough estimation of Dengue cases.

DATA REDUCTION AND RE-CLUSTERING

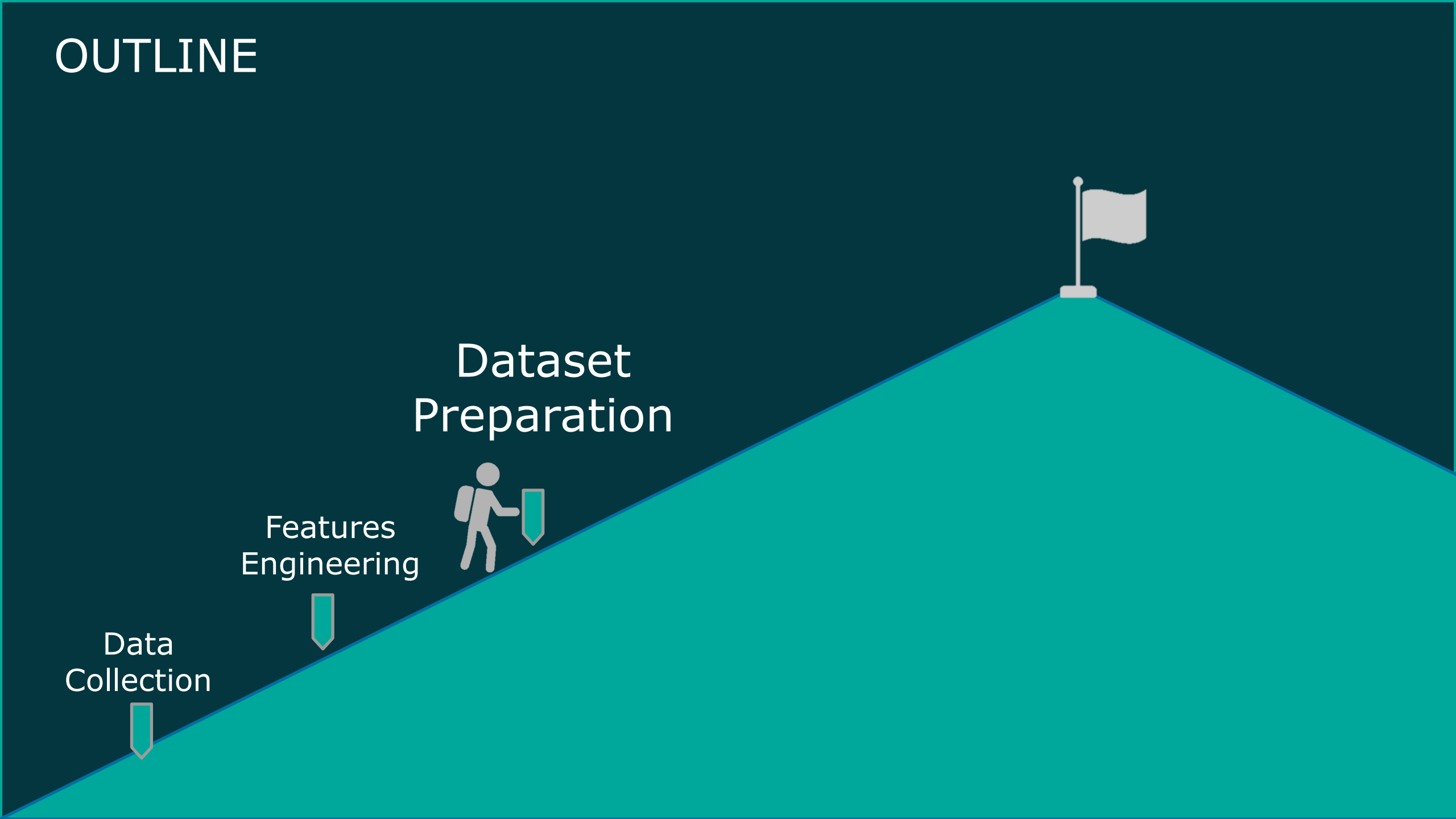


TOP Products

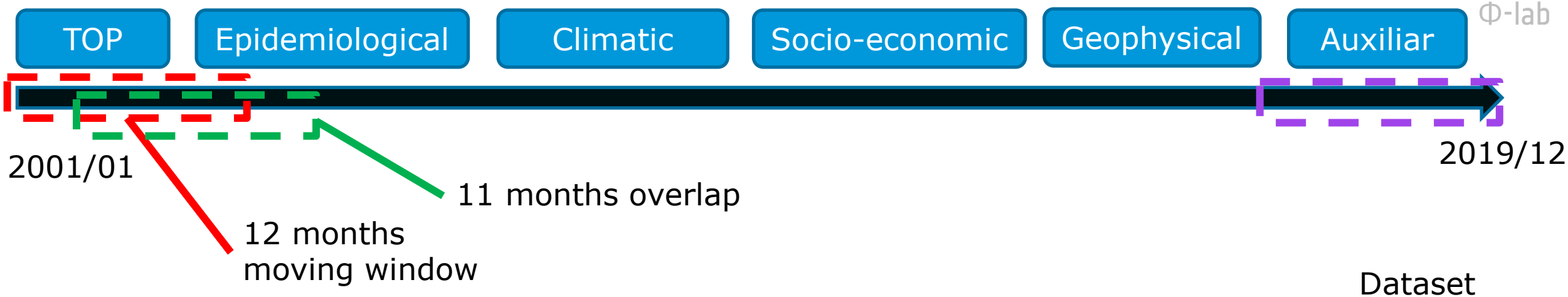
These are products that have been excluded from the PLS operation.

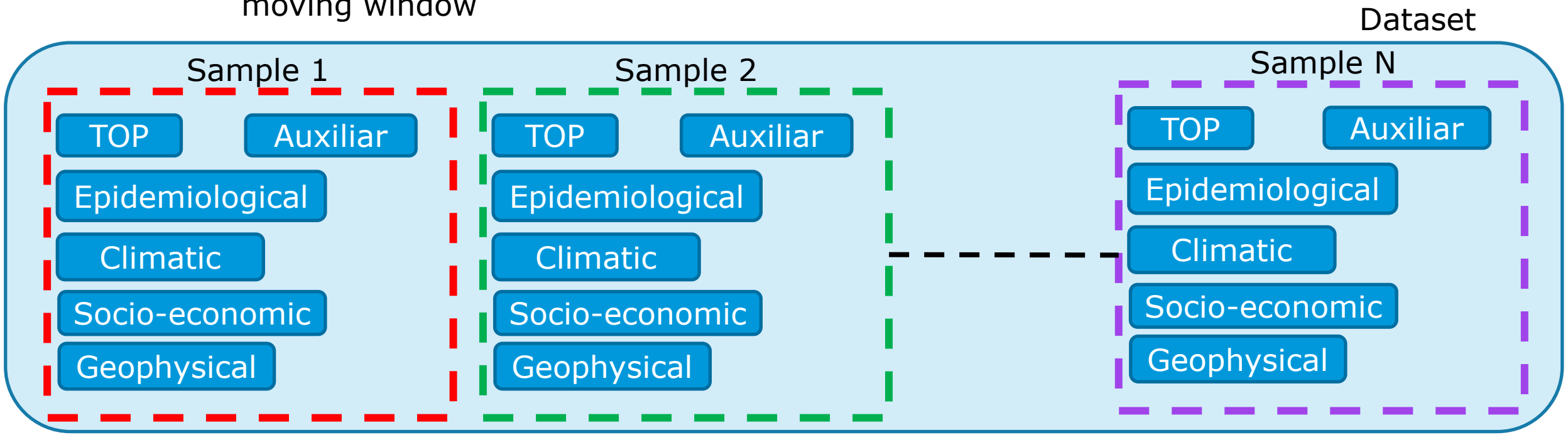
They represents the most important features in our dataset, that better correlates with the Dengue Incidence Rate.

OUTLINE

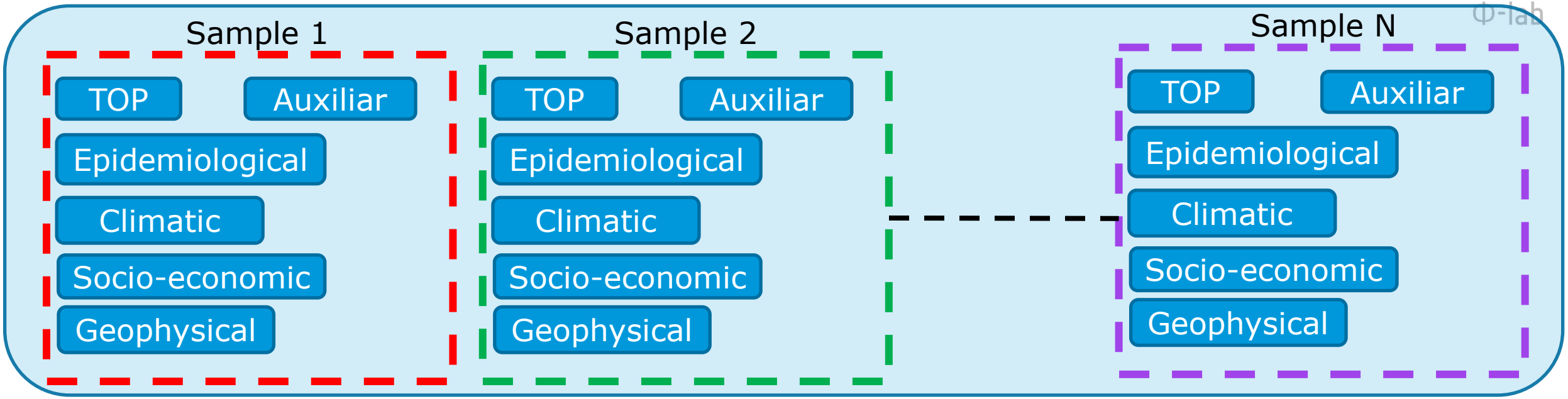

Dataset
Preparation
Features
Engineering
Data
Collection

DATASET AUGMENTATION



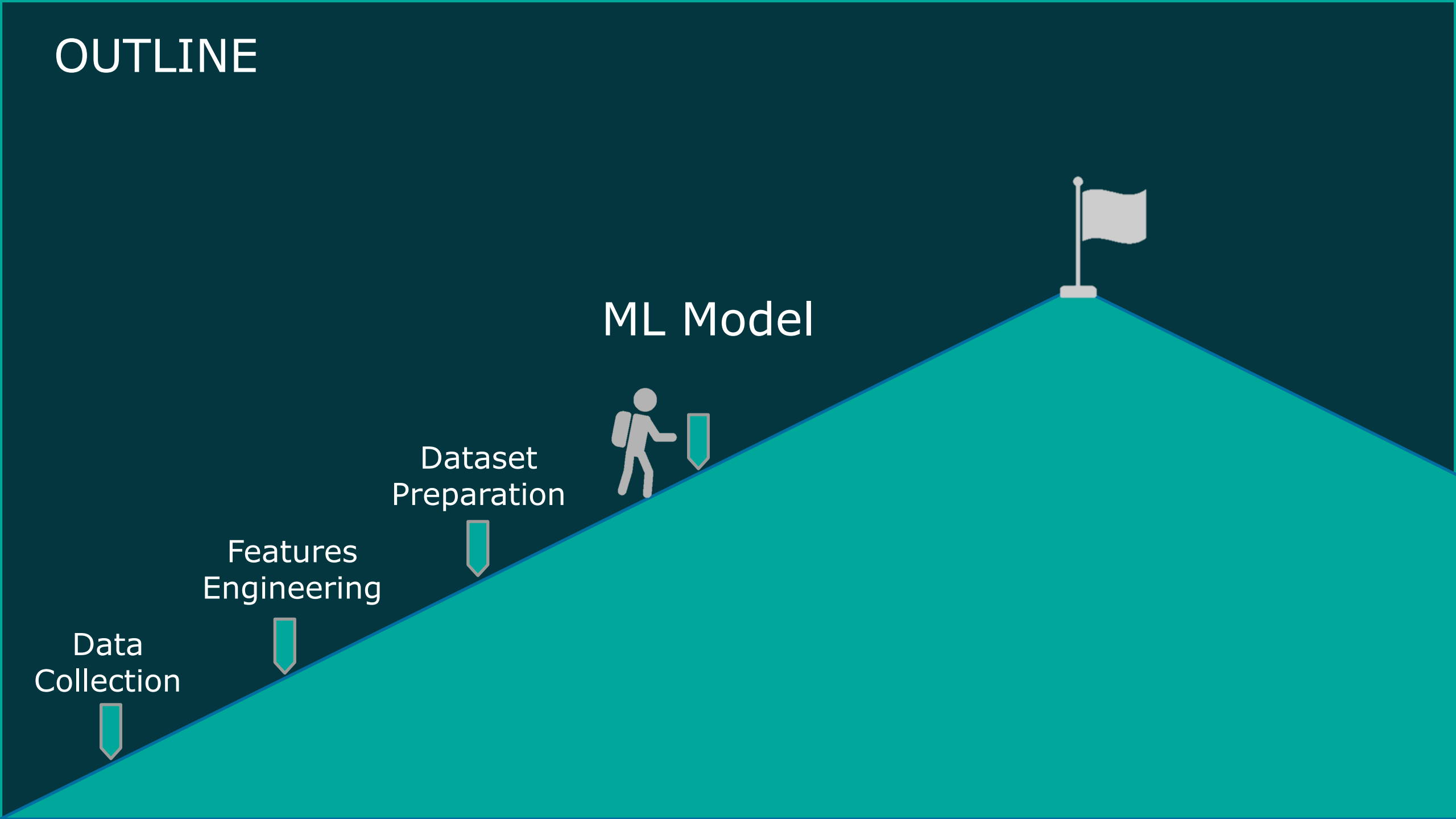


DATASET SPLIT



Training	Validation	Testing
2001/01 – 2016/12	2017/01-2019/12	Peru

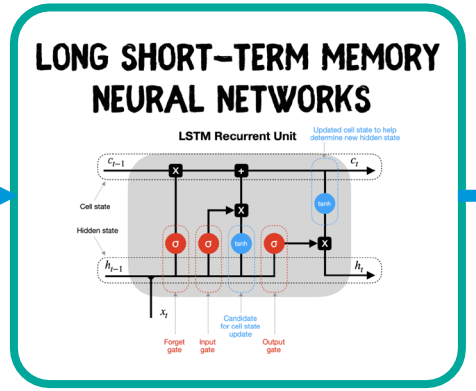
OUTLINE


ML Model
Dataset Preparation
Features Engineering
Data Collection

MACHINE LEARNING MODEL

European Space Agency
Φ-lab

Dataset



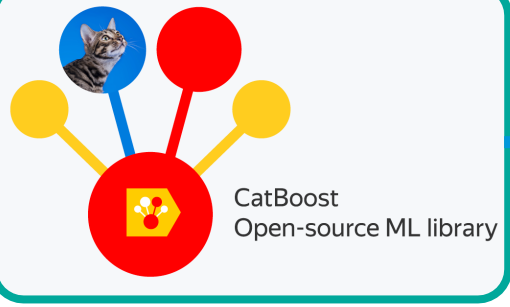
Dengue Forecast

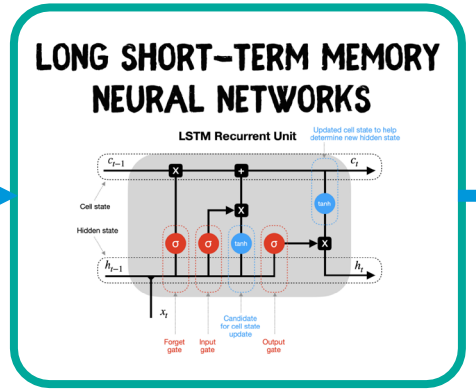
- 0-19 Population
- Total Population

PROs	CONs
Good temporal data handling	General trend, no outliers

MACHINE LEARNING MODEL

Dataset

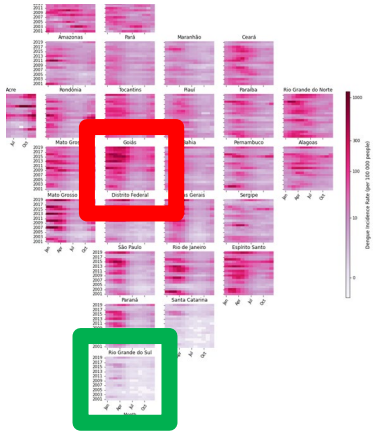




Ensemble Strategy
Weighted Mean

Dengue Forecast

- 0-19 Population
- Total Population

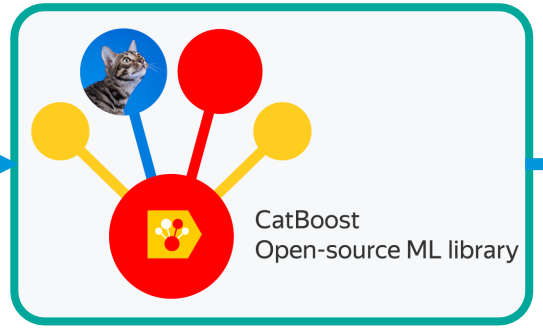


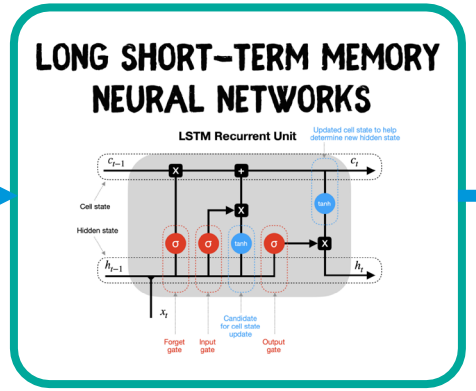
PROs	CONs
Good with almost zero-cases series	Not good with outliers (peak)

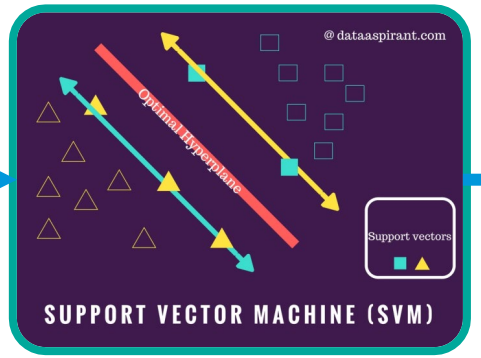
MACHINE LEARNING MODEL

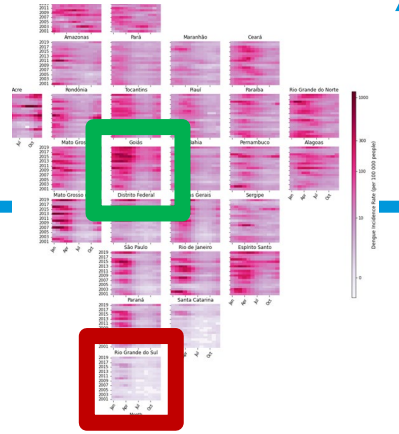
European Space Agency
Φ-lab

Dataset









Ensemble Strategy
Weighted Mean

Dengue Forecast

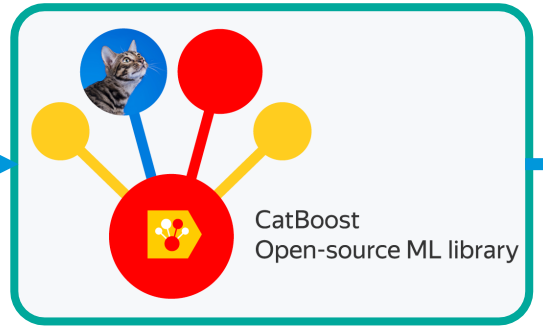
- 0-19 Population
- Total Population

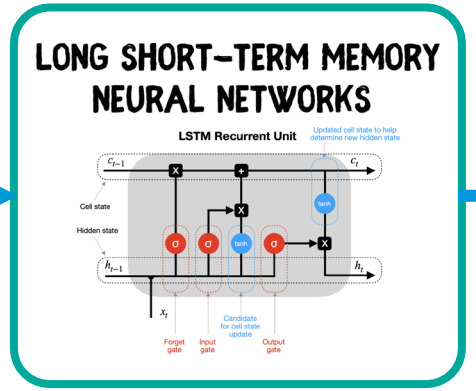
PROs	CONs
Good with outliers (peak)	Sometimes loses general trends

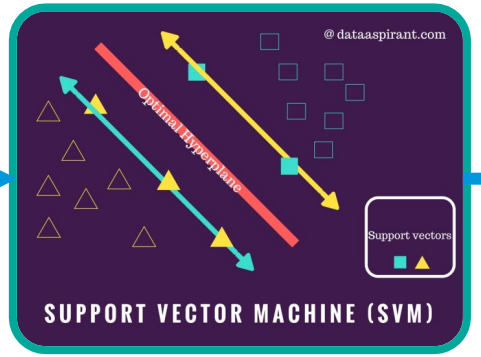
MACHINE LEARNING MODEL

European Space Agency
Φ-lab

Dataset







Dengue Forecast

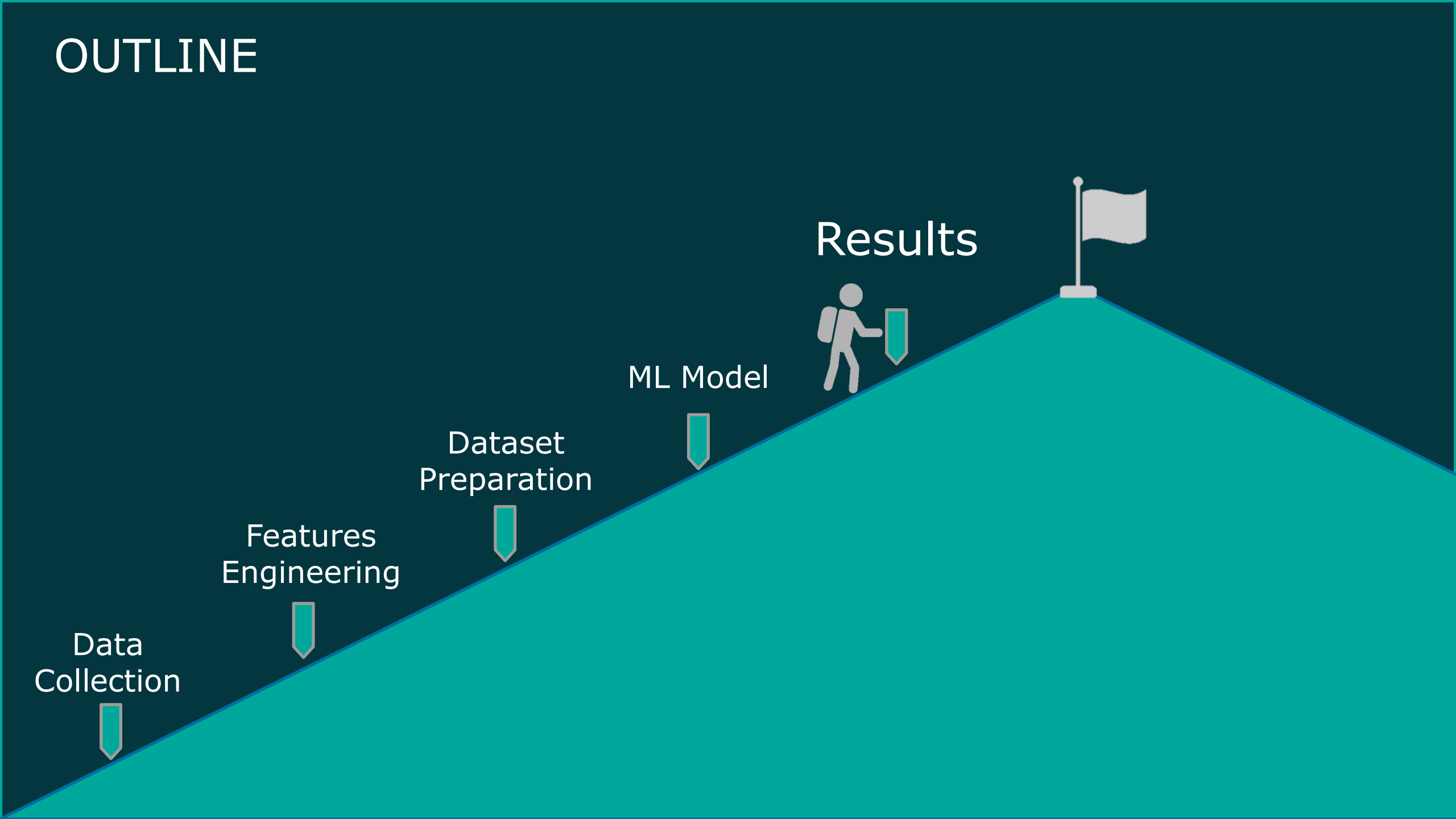
- 0-19 Population
- Total Population

Confidence Interval/Model Variability

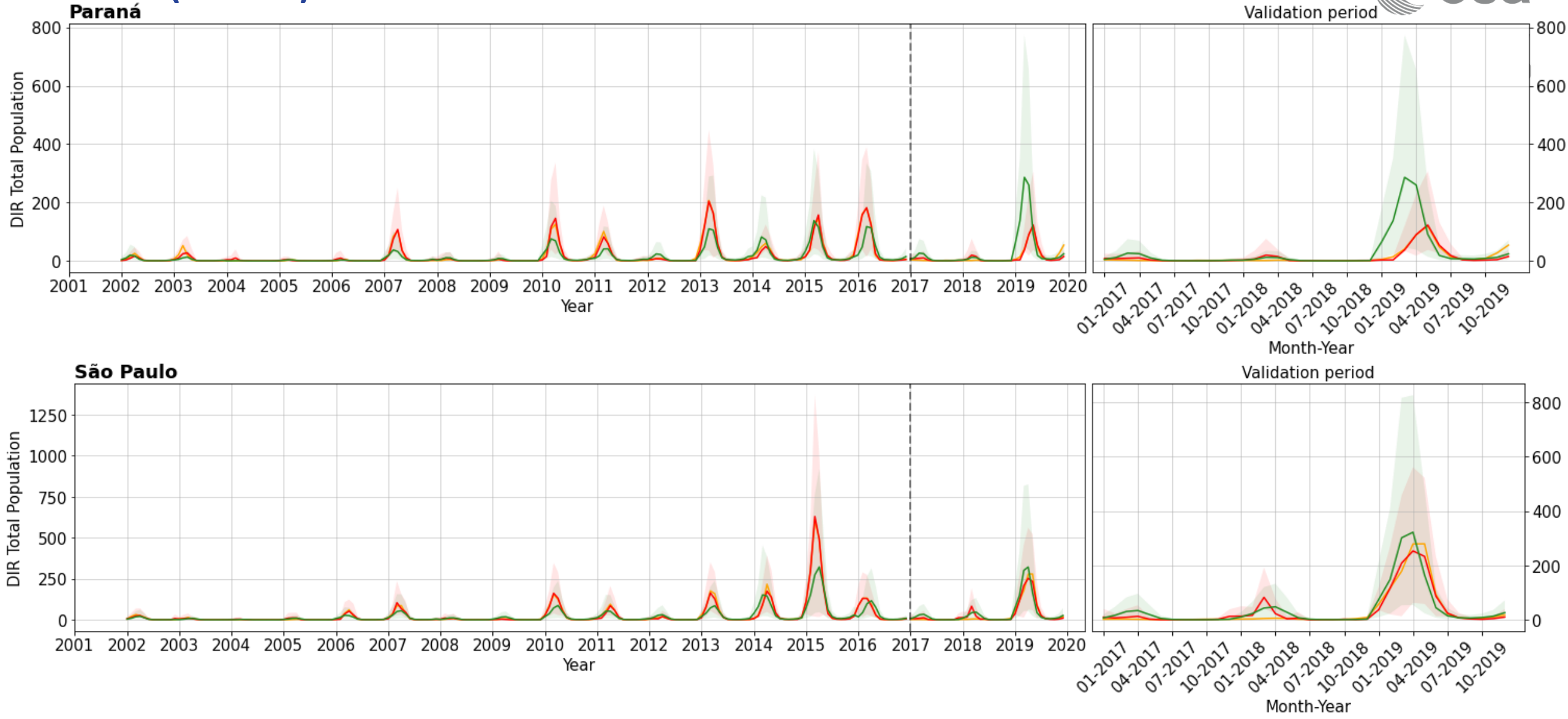
PROs

More robust than weighted mean and gives a sort of confidence interval

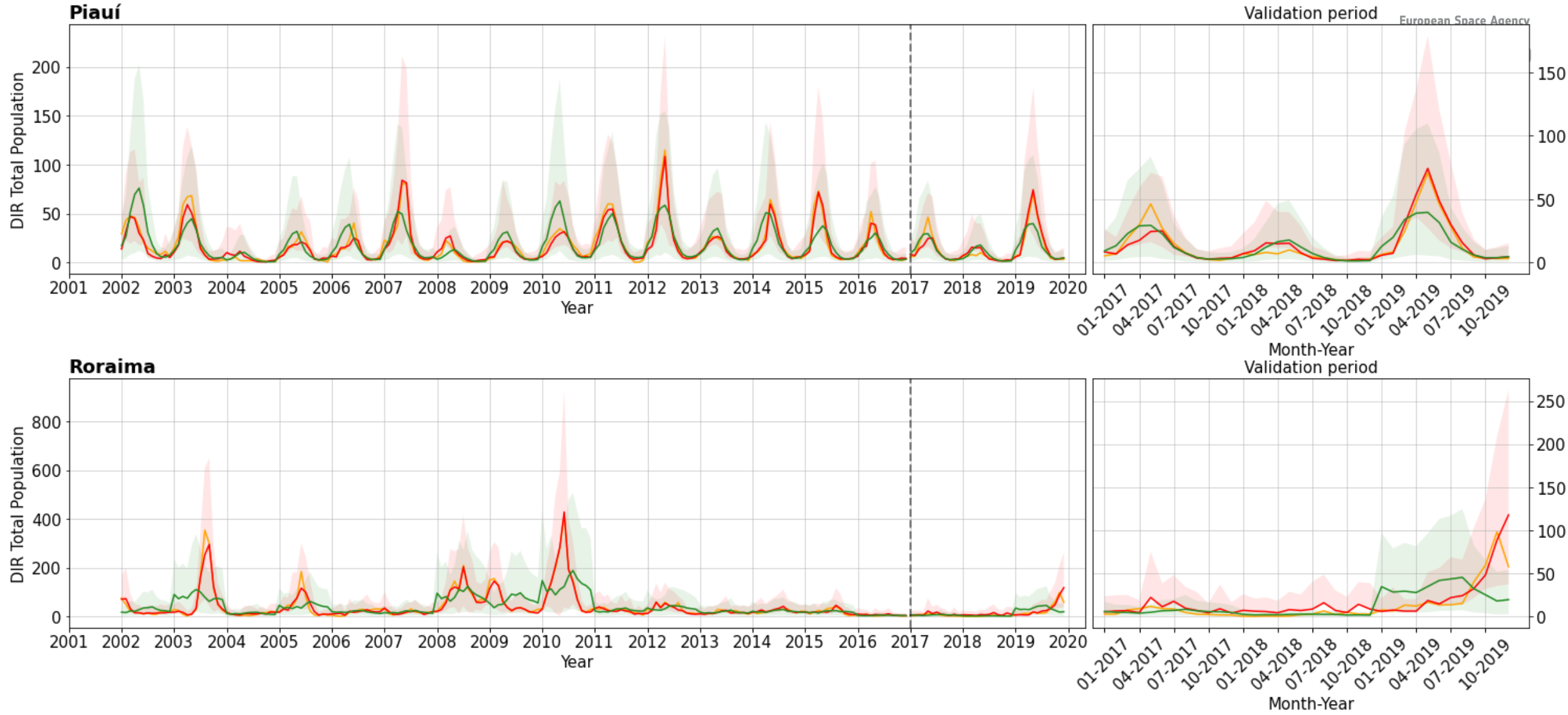
OUTLINE


Results
ML Model
Dataset Preparation
Features Engineering
Data Collection

RESULTS (BRAZIL)



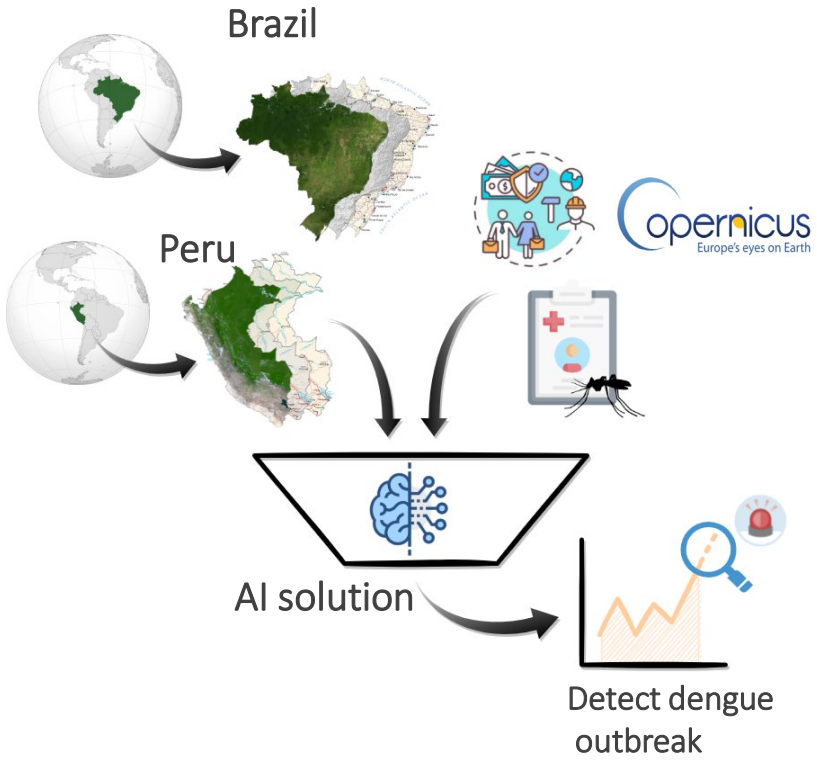
RESULTS (BRAZIL)



Conclusions

Summary

1 Developed an AI model to detect dengue outbreaks in Brazil and transferred to Peru



2 Develop a 'do-it-yourself' guidebook: a fully-reproducible framework

Replicable AI solution

GitHub

*"This project is a **perfect example** of **collaboration** between a humanitarian organisation and a research entity to support the UN SDGs."*

Dohyung Kim
Data Scientist at the UNICEF Office of Global Innovation.

unicef INNOVATION

3 **Multi-Award Winning Project**

1 – UNESCO – IRCAI

United Nations Educational, Scientific and Cultural Organization · International Research Centre of Artificial Intelligence under the auspices of UNESCO

GLOBAL TOP 100 AI solutions for SDGs

2 - Best of UNICEF Research

showcase the most rigorous, innovative and impactful research produced by UNICEF offices worldwide

3 – Wellcome Trust support

wellcome

Wellcome trust is an independent charitable foundation