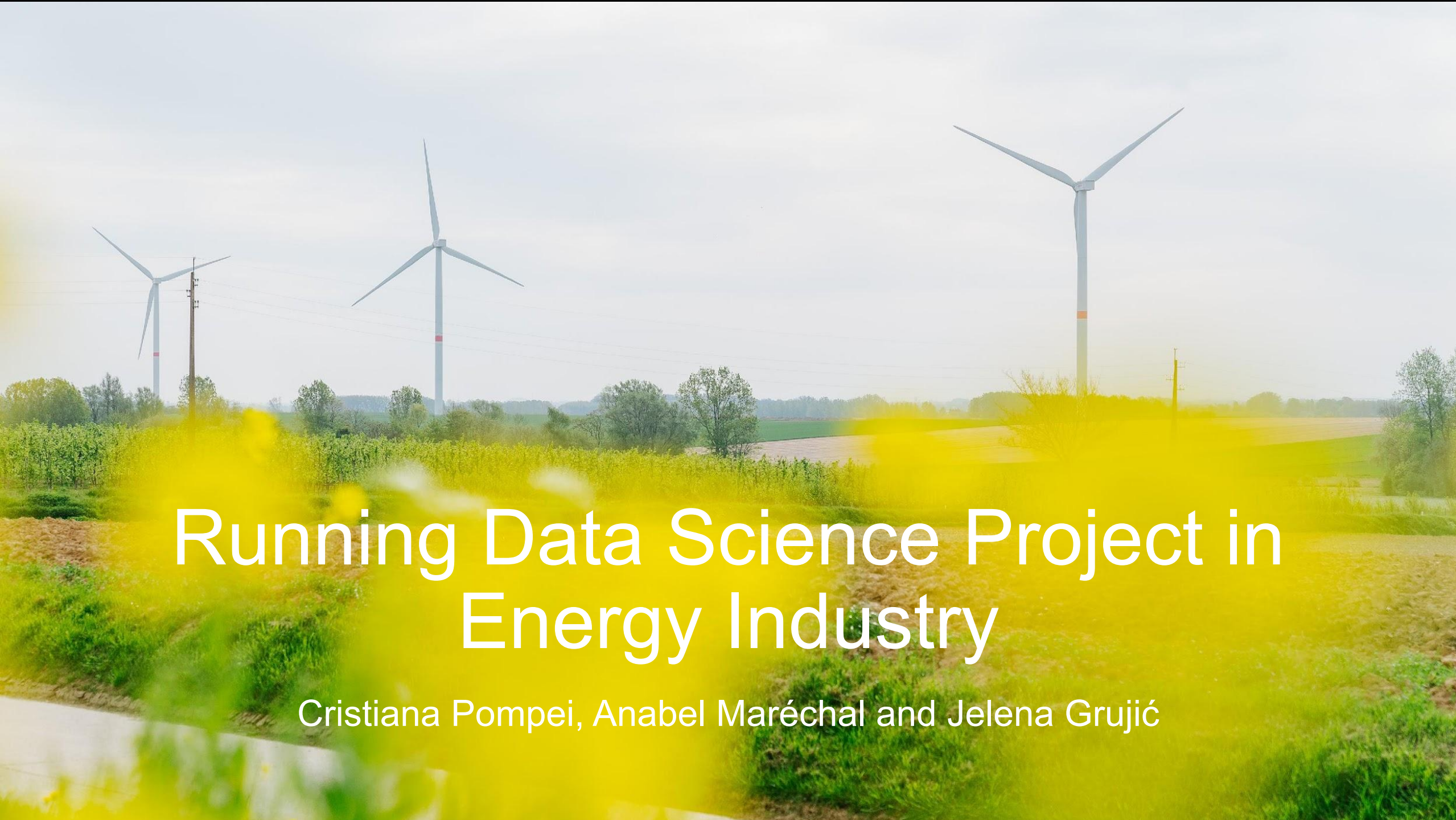
WOMEN IN DATA SCIENCE MAASTRICHT

#WiDSMaastricht2023



Running Data Science Project in Energy Industry

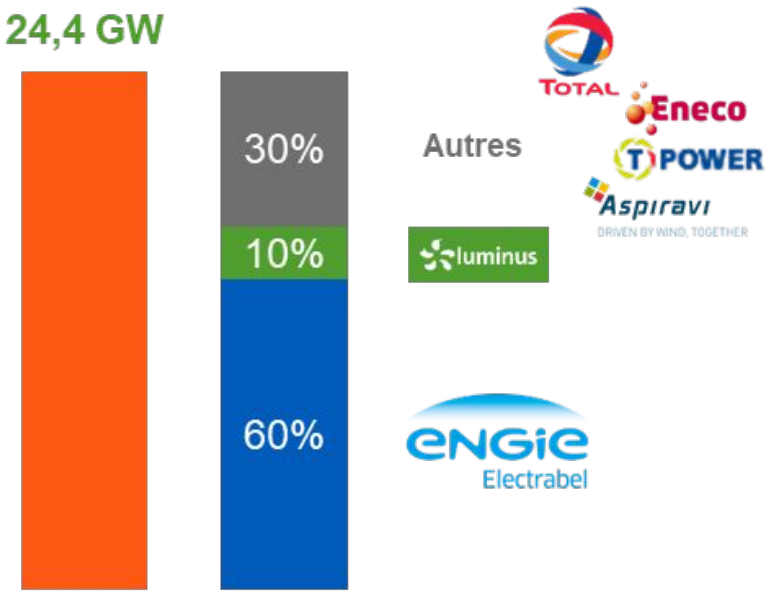
Cristiana Pompei, Anabel Maréchal and Jelena Grujić

Electric power supply chain
A market characterised by both liberalised and regulated activities

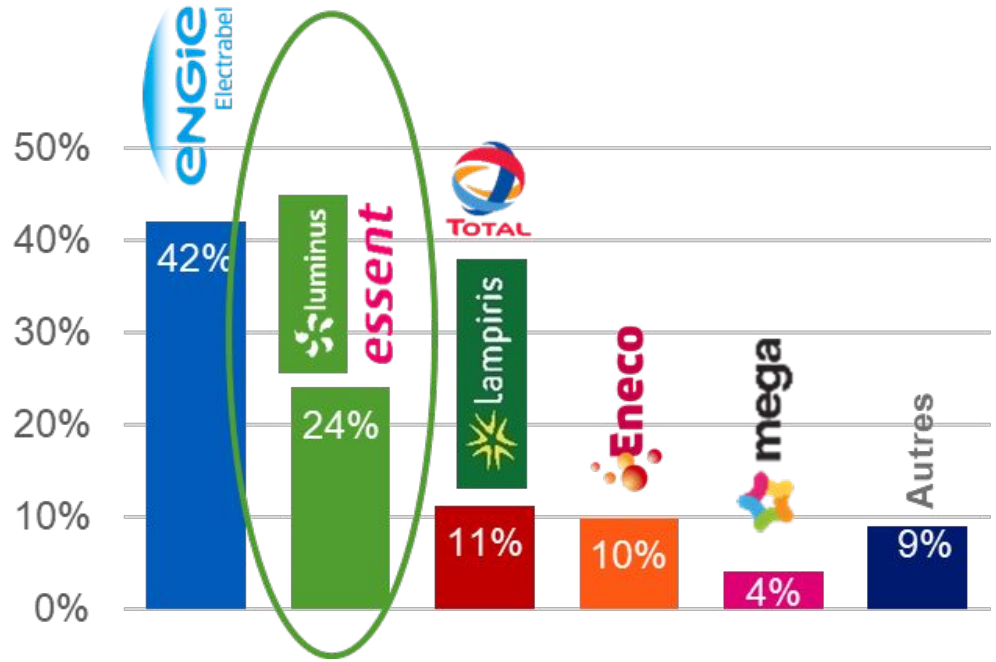
- Liberalisation of the production and supply activities
- Regulated monopoly for transport and distribution

Luminus, an integrated player on the Belgian electricity and gas market

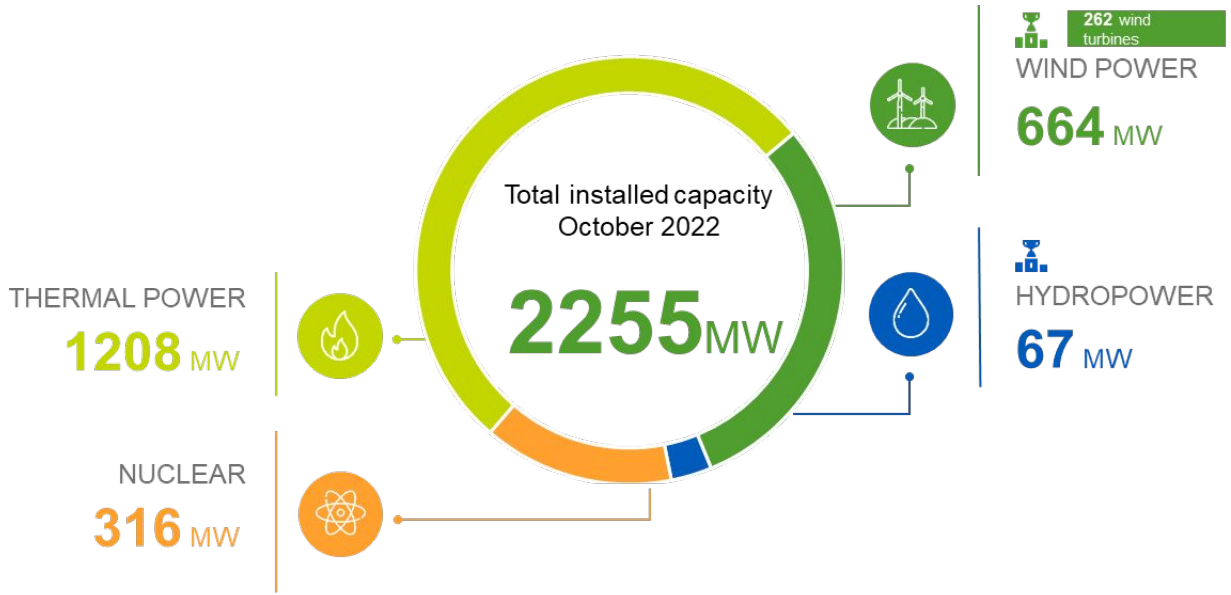
2nd largest electricity producer



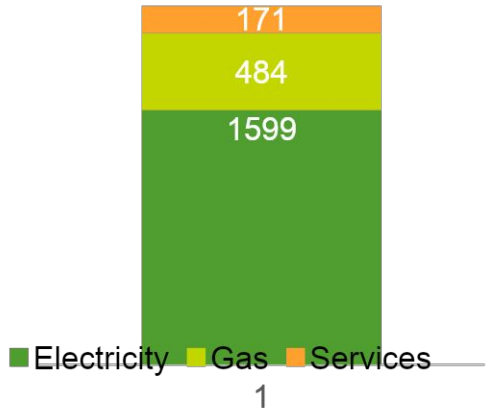
2nd largest supplier of electricity and natural gas



A diversified portfolio of production and 2.25 million contracts



B2C customer portfolio at end of 2021 (k# POD)

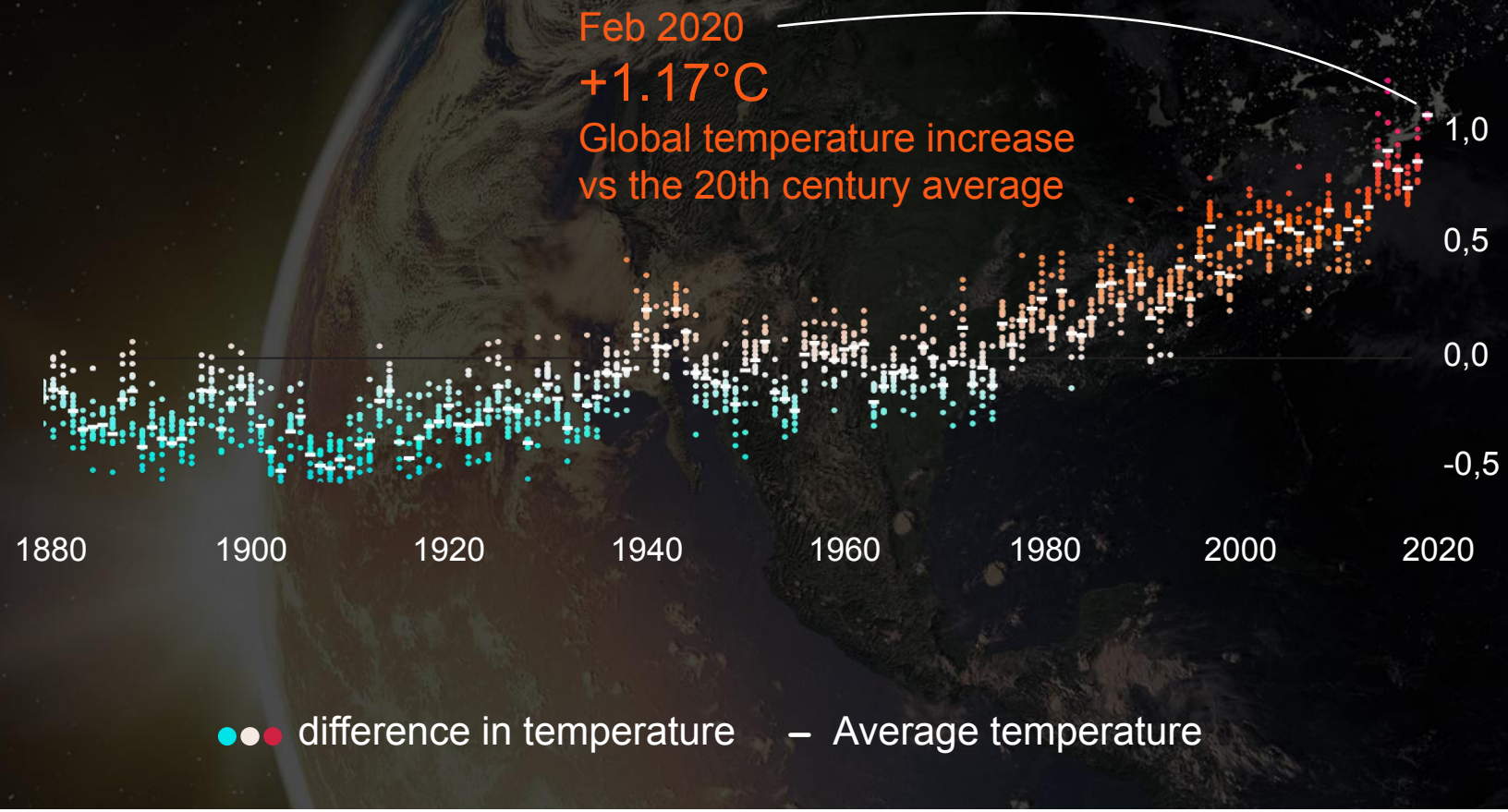


Global warming is a major challenge
Towards carbon neutrality in 2050

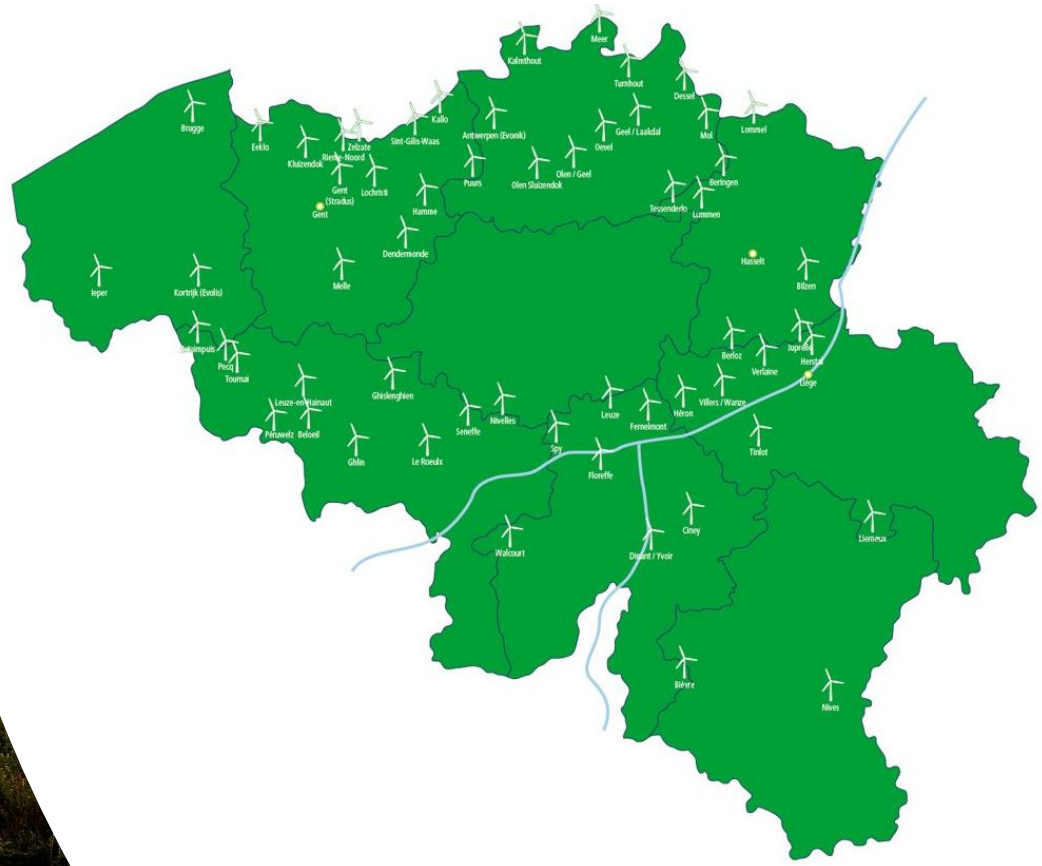
Bloomberg Green

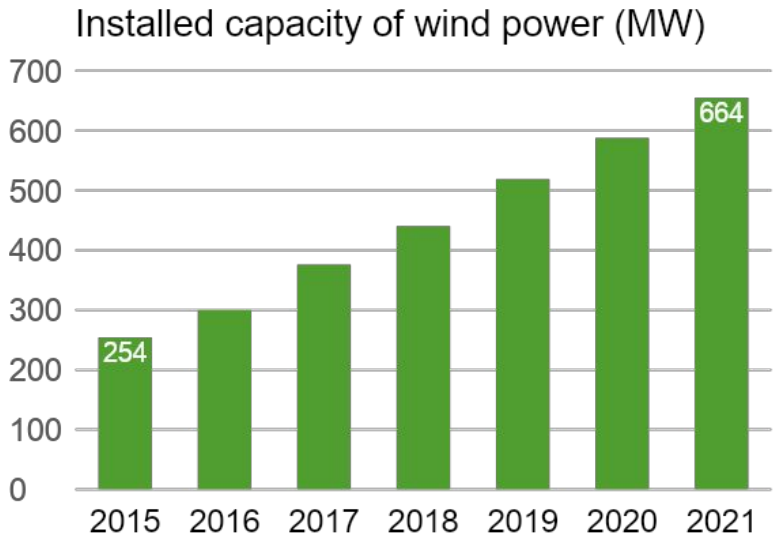
Global Temperature Change

Last year was the second warmest year on record. The five warmest years have all occurred since 2015.

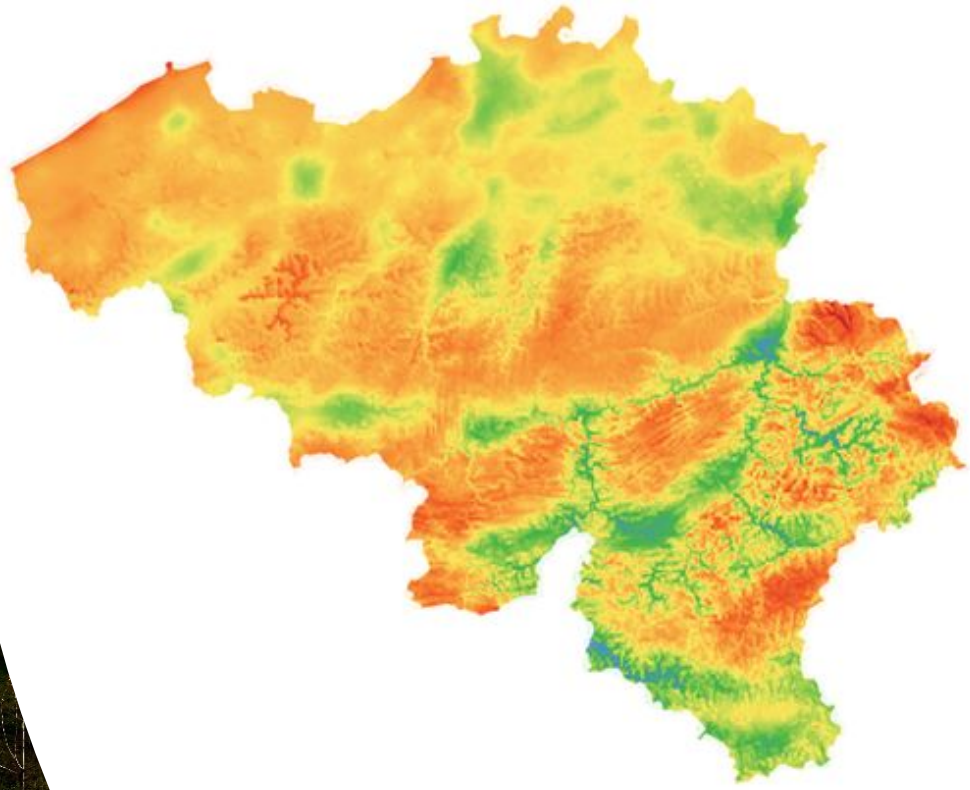


Developing new onshore wind energy





In-house Wind Atlas



- Development of wind resources is expensive and wind resources uncertain
- It allows to evaluate the wind resource for new development projects of wind parks (avoided costs for wind studies *50-70k€/ year*
- Combines measured wind data and meteorological data to evaluate the wind in the past 20 years

Wind Power Forecast

Forecast the power production of each of our wind parks at a granularity of 15 minutes for an update every hour.

Mix of few ML models based on multiple weather forecasts to reduce the risk and increase our accuracy

865 MW Wind

6 Weather Forecasts

600 ML models

XGBoost

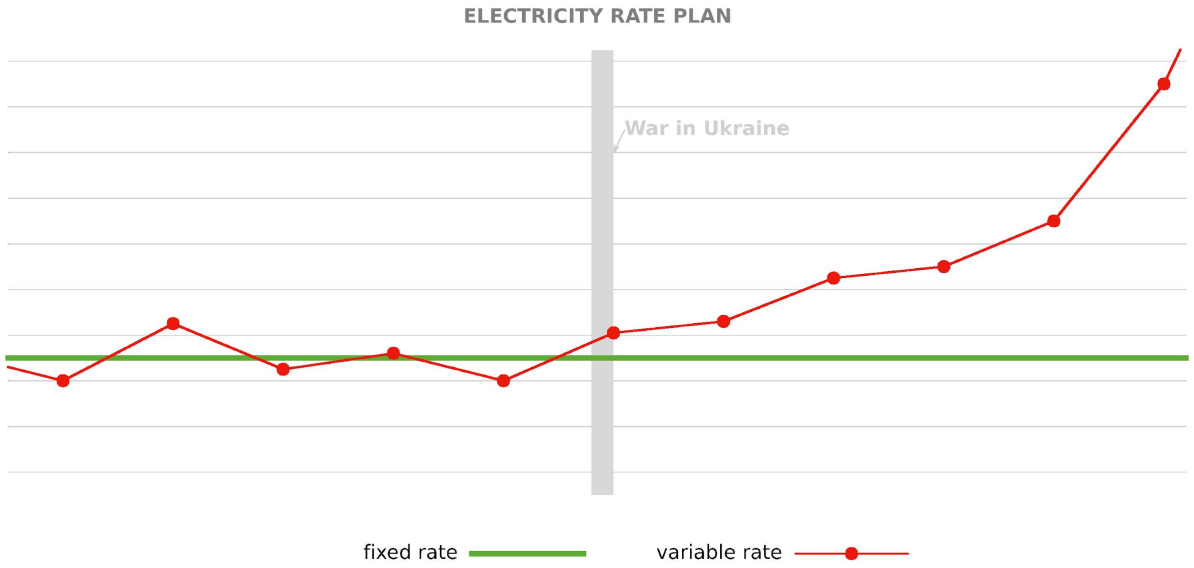
scikit learn

PyTorch

LightGBM

OPTUNA

Long Term Forecast of customer consumption



- A new ML model to forecast Energy demand for the next 3 years
- It allows to evaluate different scenarios to match supply with demand and to offer **fixed price contract** to final customers

LTF Accuracy KPI	FIX POD's
	Avg Abs Deviation
Dec/21 actuals	3,42%
Target 2022	3,25%
Target 2023	2,00%
Target 2024	1,50%

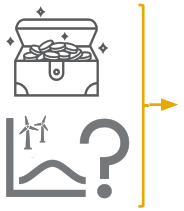
In-house Wind Atlas

WiDS 2023

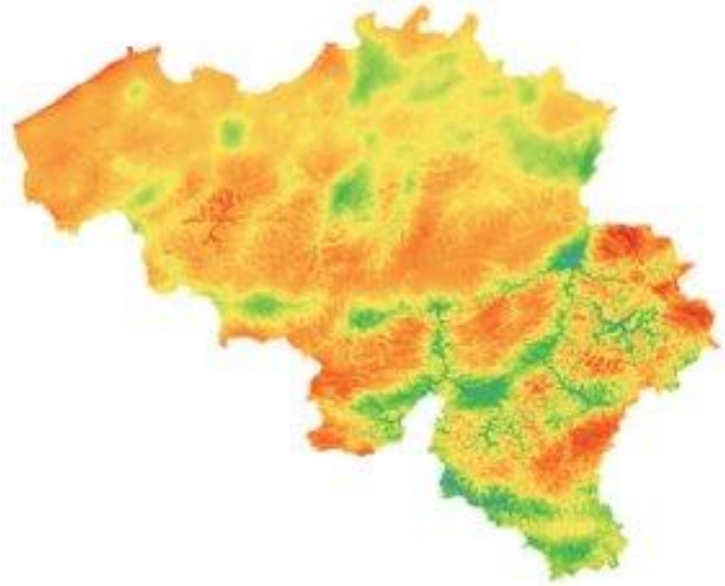
Introduction

Current situation:

- Development of wind turbines is expensive
- Wind resources uncertain



Risky investment decision



Wind Resource Assessment by external parties

BUT the analysis is *expensive* and is a complete *black box*

Goal:
In-house tool to perform Wind Resource Assessments and be able to:

- Avoid costs
- Challenge studies
- Reduce the risk

Introduction

How to proceed ?

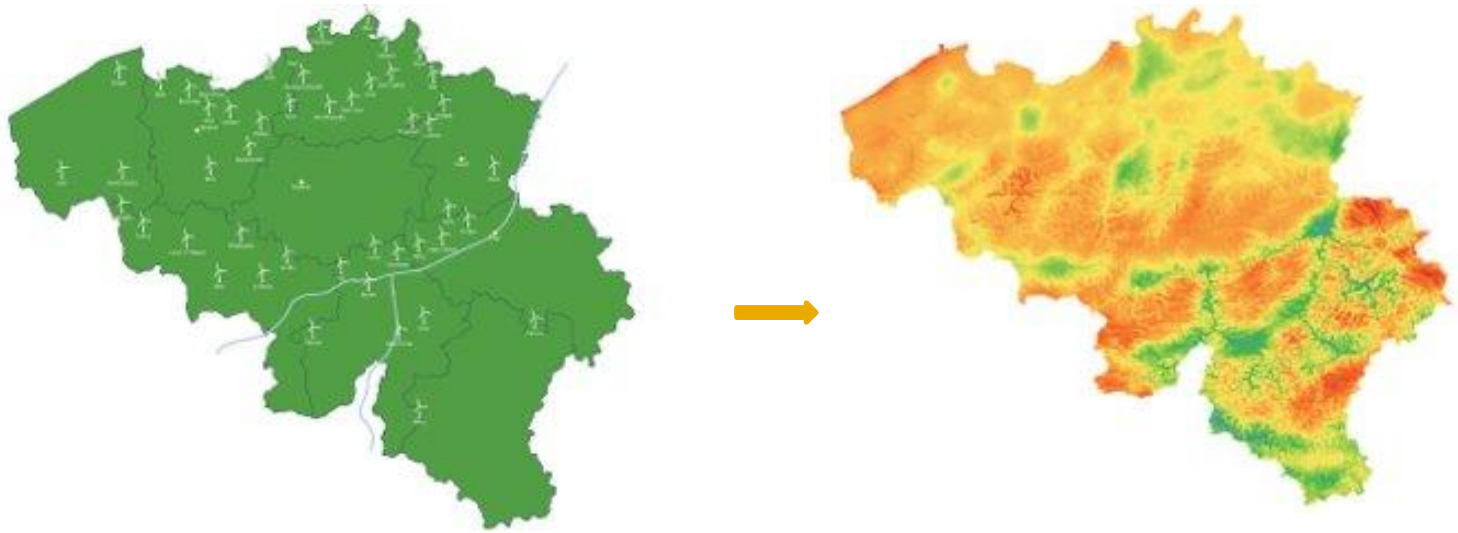
- Tools such as WindPro :

Wind data @ X
+ local roughness
+ topographical info
+ …

→ Wind data @ Y

Luminus has wind data of:

- 250+ wind turbines
- Metering mast campaigns
- LIDAR campaigns



Methodology

Data from

Wind Turbines

+ Well distributed over the territory

+ Measurements at hub height

- 2 years data

Data from

ERA 5

+ 20 years data

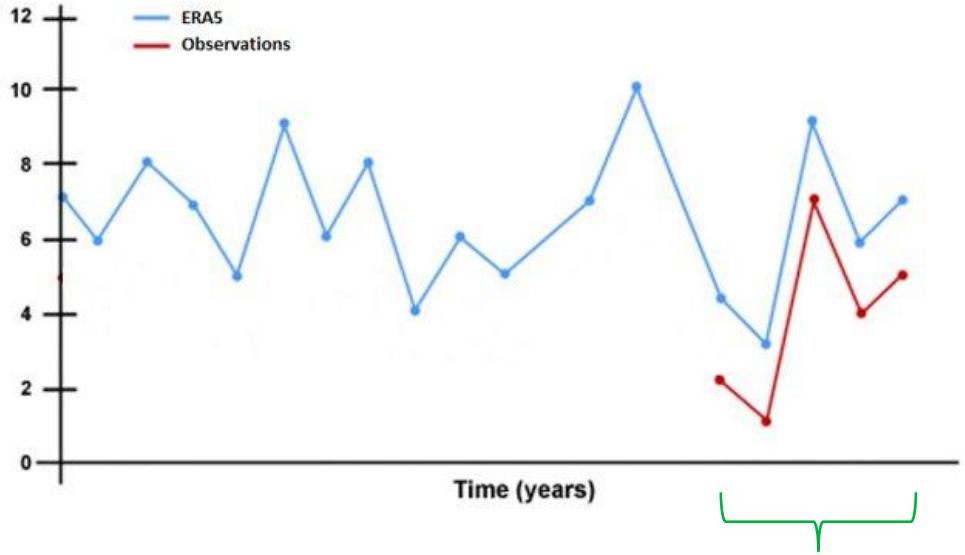
- Fewer measuring points

Calibrate the ERA5 data thanks to the data from the WTGs via a **machine learning** algorithm in order to recreate time series of the wind speed over 20 years

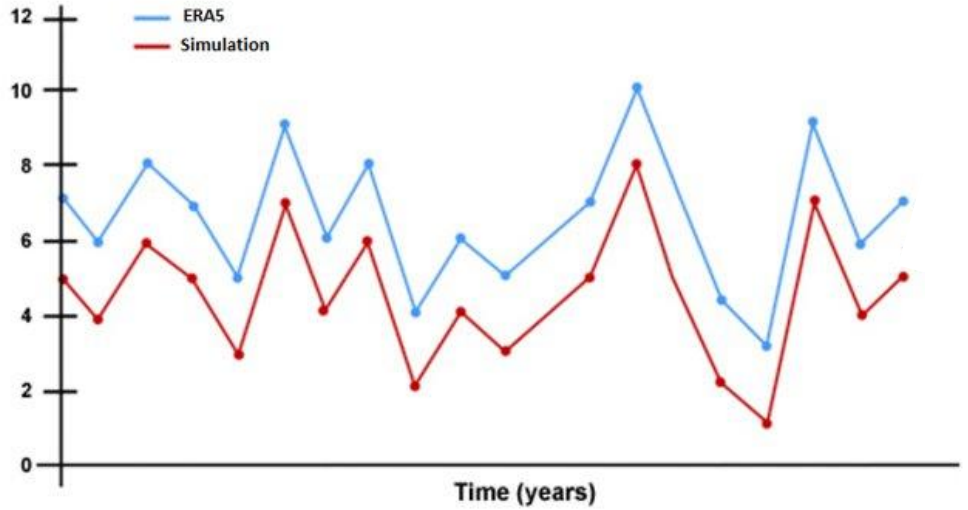
=

Ideal wind speed data for our wind resource assessment !

Methodology



Training &
testing dataset



XGBoost model:
features= Era5 variables
target = Wind Speed from WTG

Cleaning and preprocessing

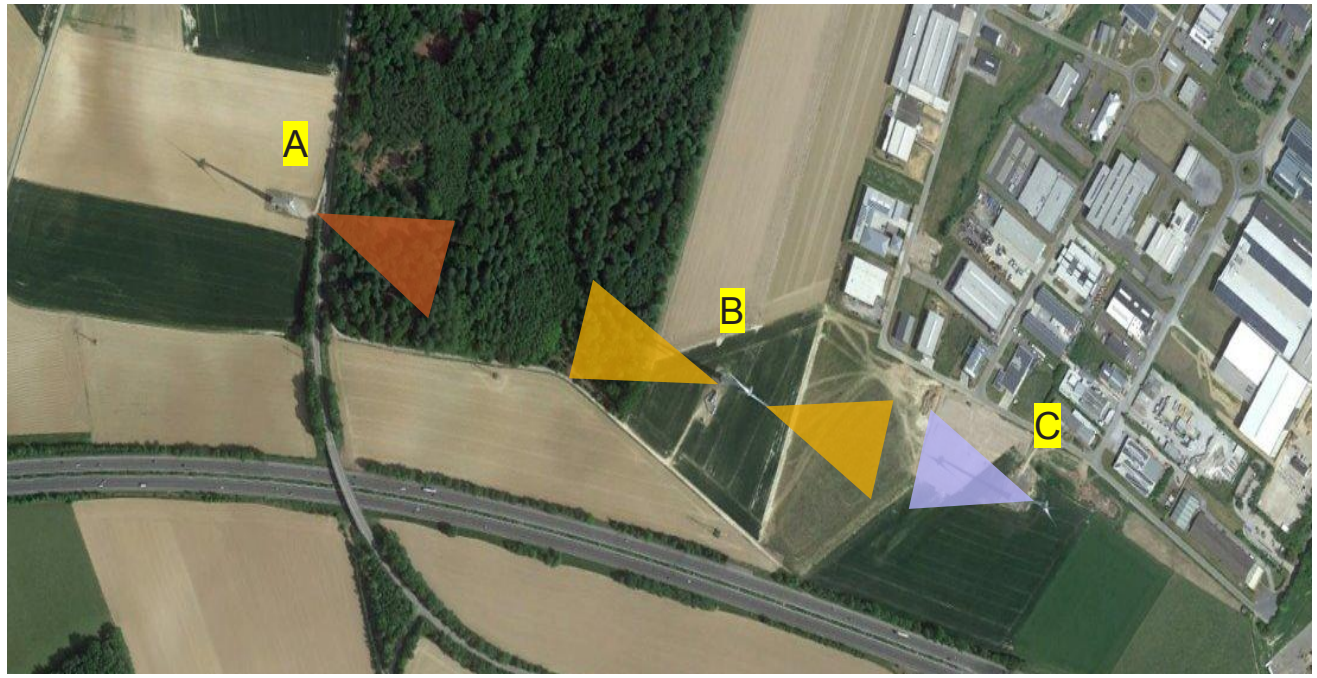
- But first step: CLEAN the data !

1. Delete frozen data, NaN data, …

Timestamp	Wind Speed	Wind Direction	Active Power	Temperature
2022-02-14 00:01:00	5.21	138.7	236	13.82
2022-02-14 00:02:00	6.4	143.1	564	13.71
~~2022-02-14 00:03:00~~	~~6.4~~	~~143.1~~	~~564~~	~~13.71~~
~~2022-02-14 00:04:00~~	~~6.4~~	~~143.1~~	~~564~~	~~13.71~~
~~2022-02-14 00:05:00~~	~~NaN~~	~~137.5~~	~~347~~	~~14.2~~
2022-02-14 00:06:00	6.2	135.4	540	14.5

Cleaning and preprocessing

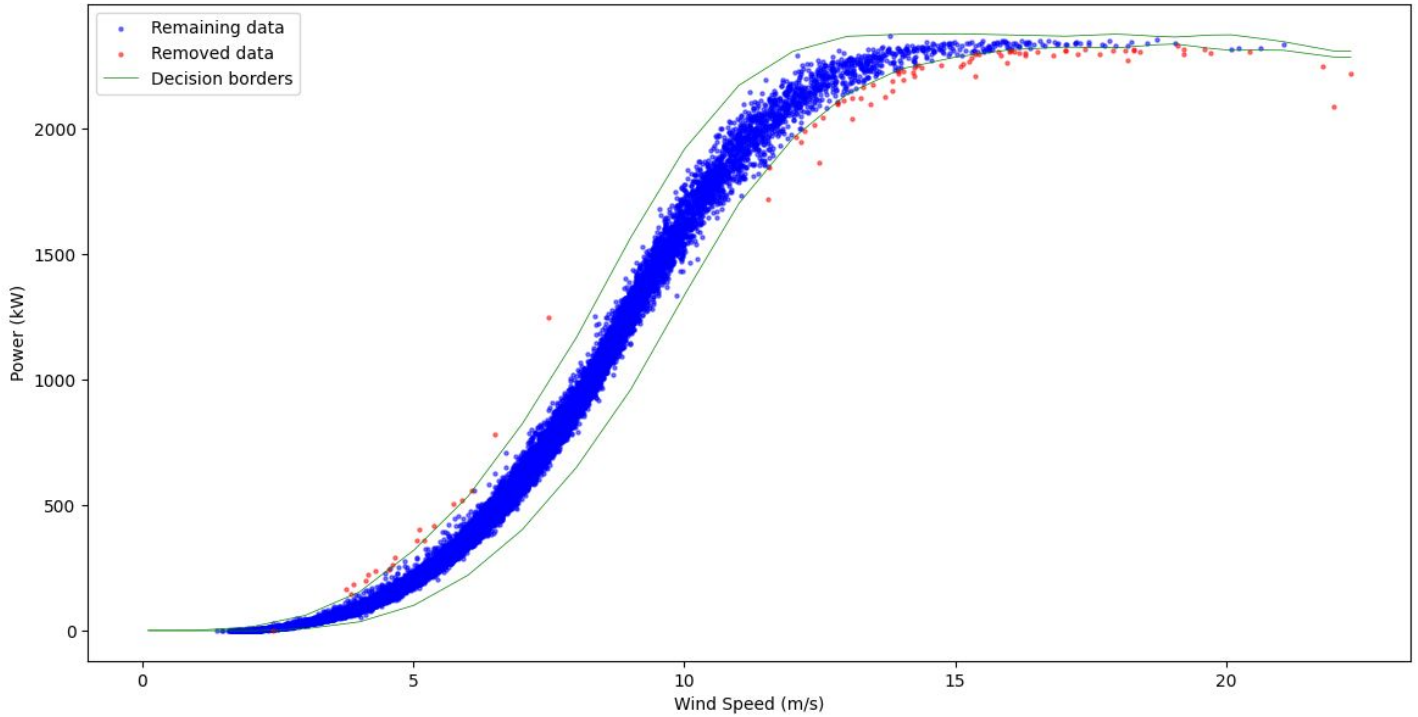
2. Suppress wake effect (= effect of the neighboring wind turbines on the wind)



For relevant result we need to extract the “Free Wind”
= wind without the impact of the wind turbines measuring it

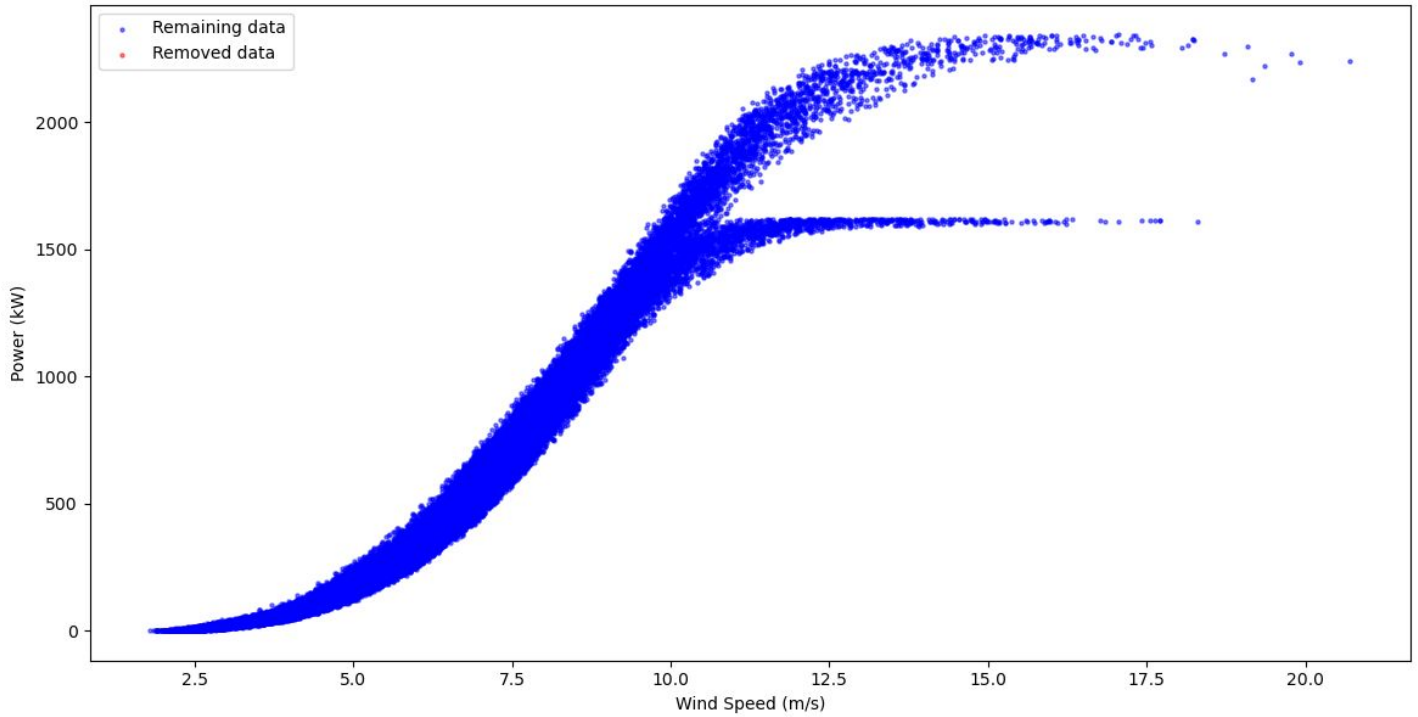
Cleaning and preprocessing

3. Remove uncoherent points based on the power curve of the WTG



Cleaning and preprocessing

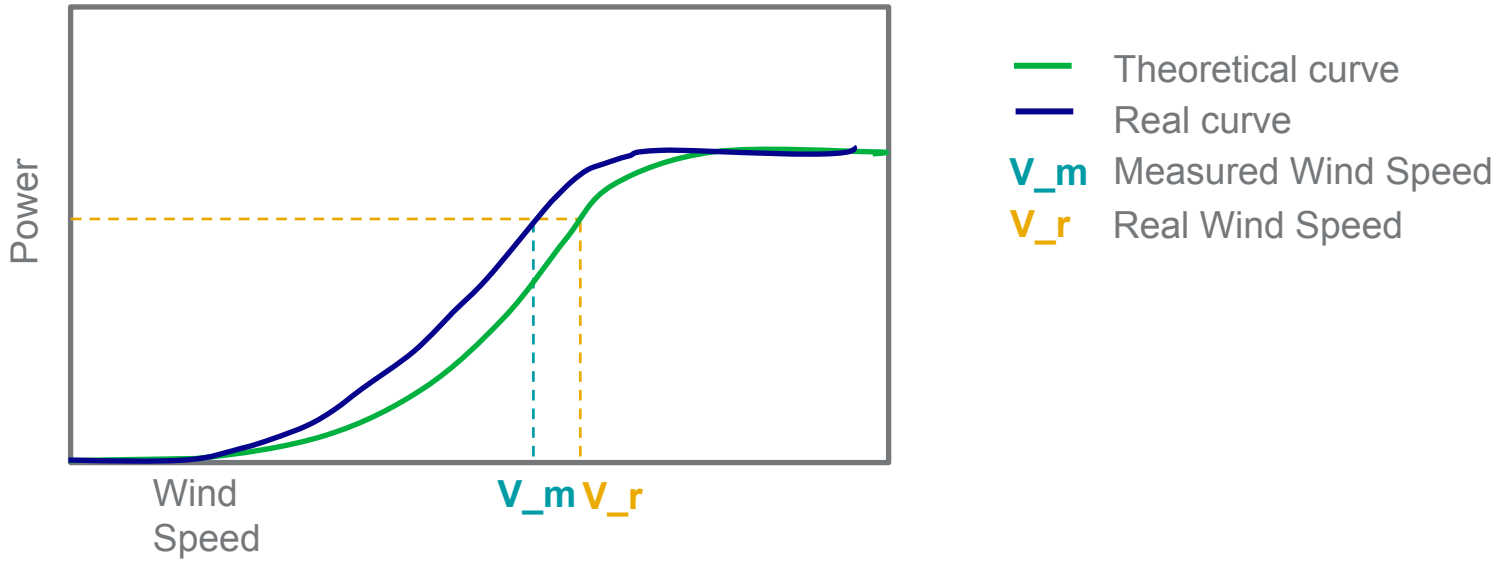
3. Remove uncoherent points based on the power curve of the WTG



Cleaning and preprocessing

4. Wind turbine measurements are not always reliable:

- Bad measurements of the wind (defective anemometers,…)
  - Try to avoid using the wind speed directly
  - Use “Reconstructed wind speed” based on theoretical power curve



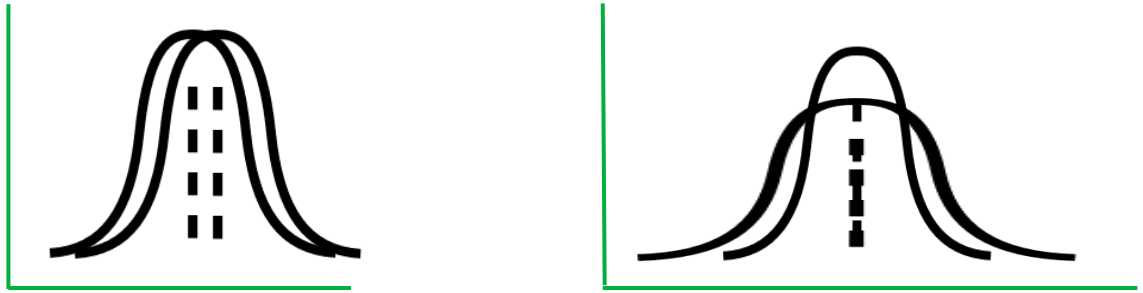
Assessing the performance of the model

Goal : have a good estimation of the wind over 20 years

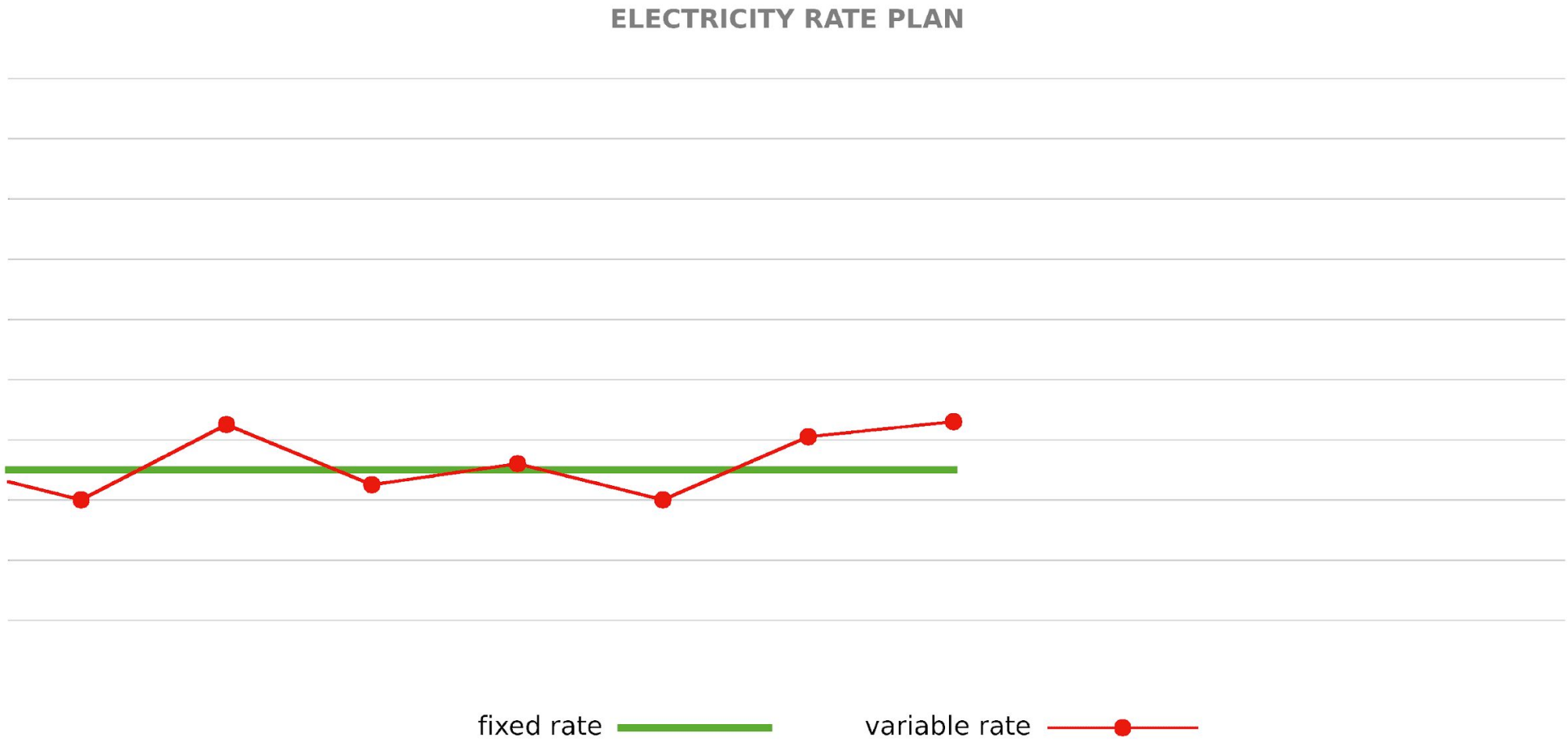
- Model is allowed to not predict perfectly each point

⚠ The model should not be constantly over- or underestimating the wind speed

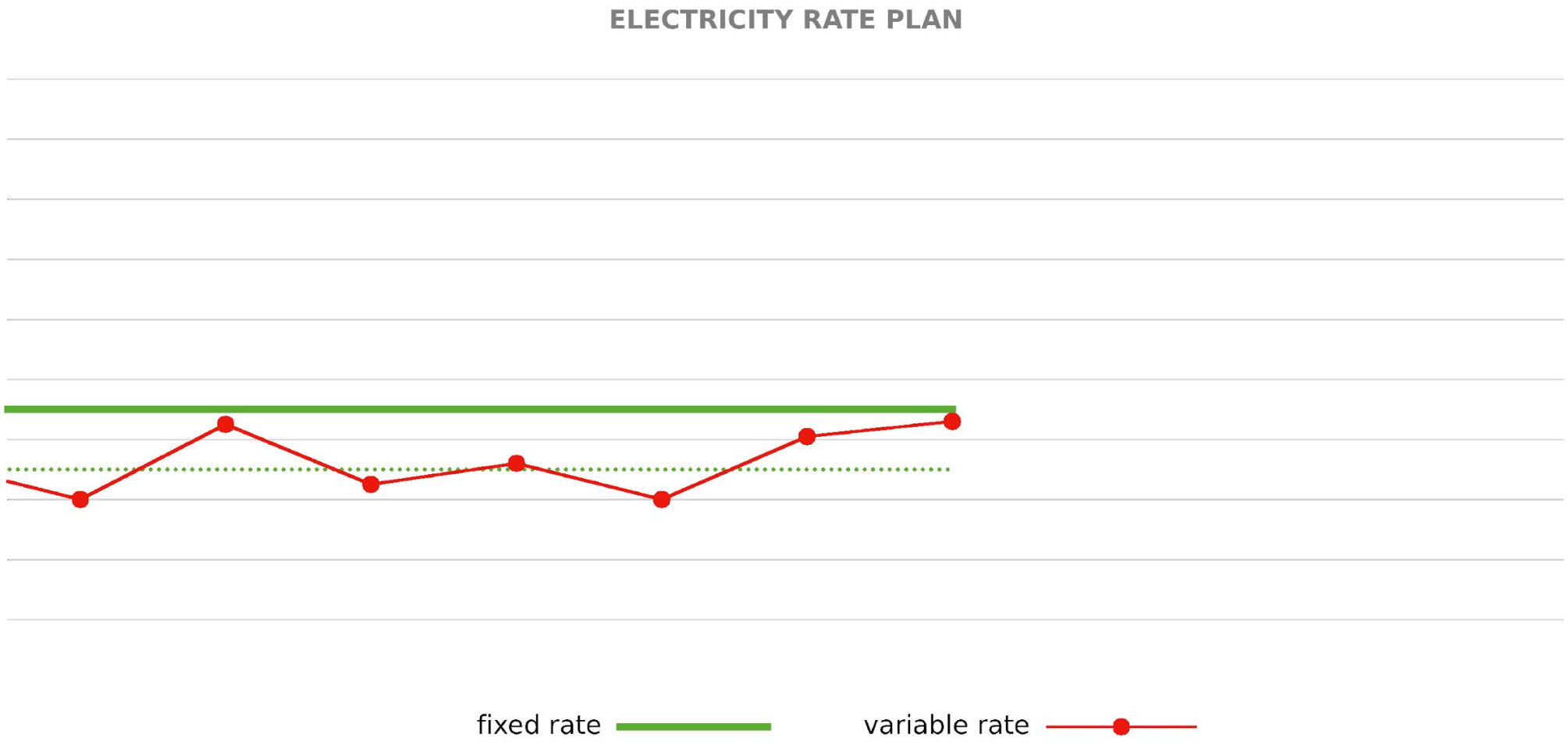
- Special attention to the **distribution** of the *real* vs. *predicted* wind speed



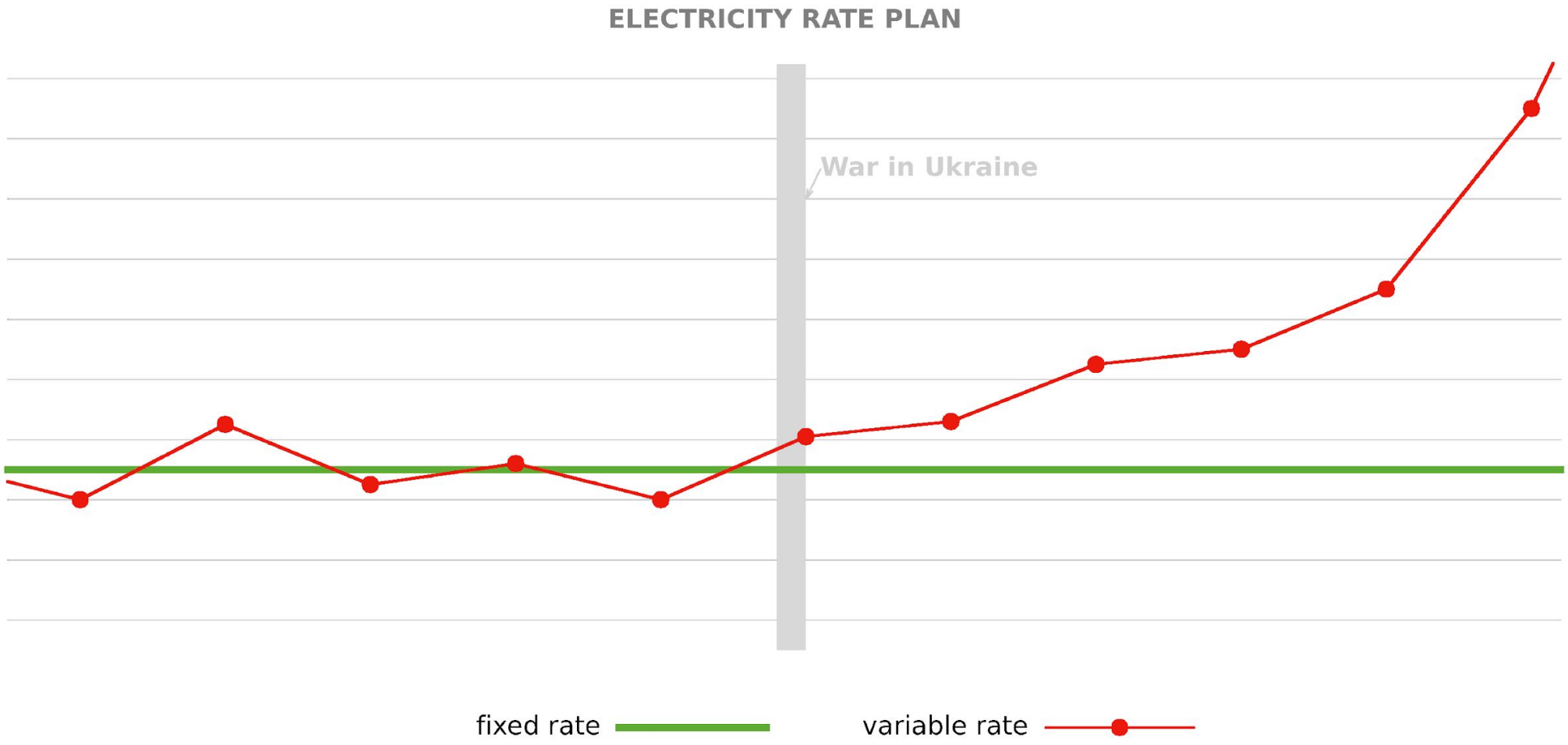
Long Term Forecasting



Long Term Forecasting



Long Term Forecasting



What do we need to change?

Old LTF

- Billing engine (SAP)
- Limited by on premise server
- Black box
- Takes more than 24h
- Manual assumptions
- No ML

Actions

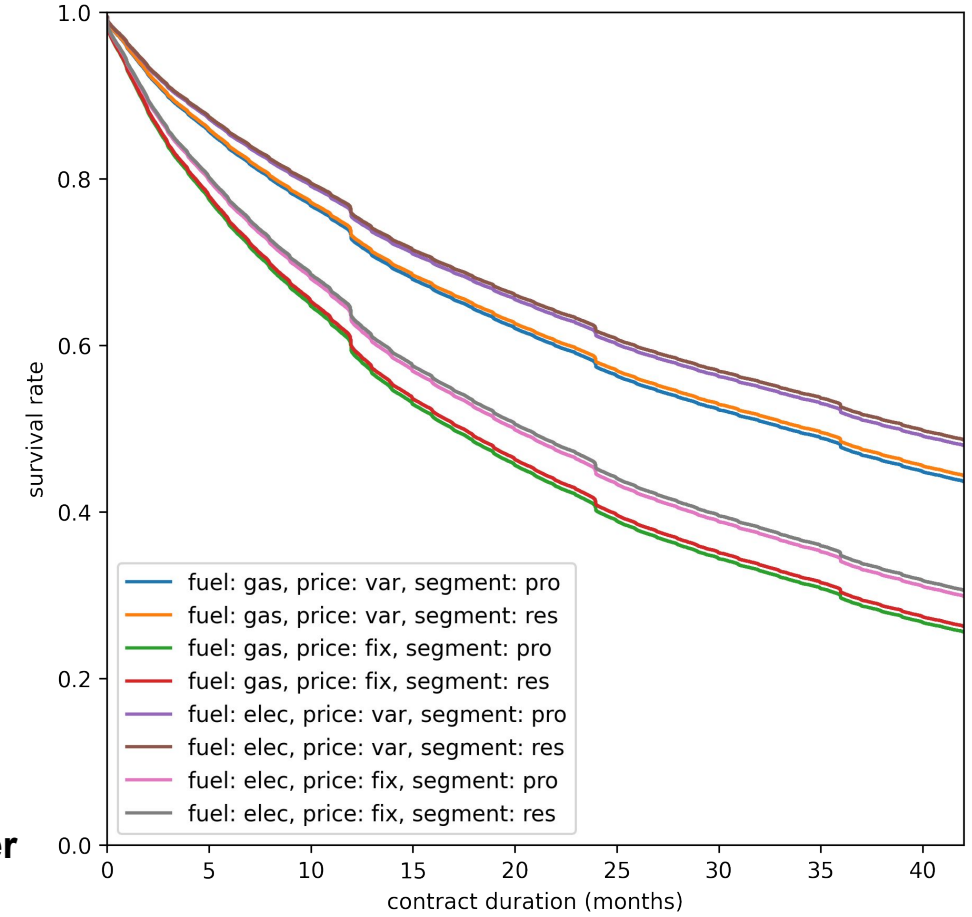
- Improve processes
- Improve predictions

The models – Churn prediction

Survival analysis

Kaplan-Meier curve

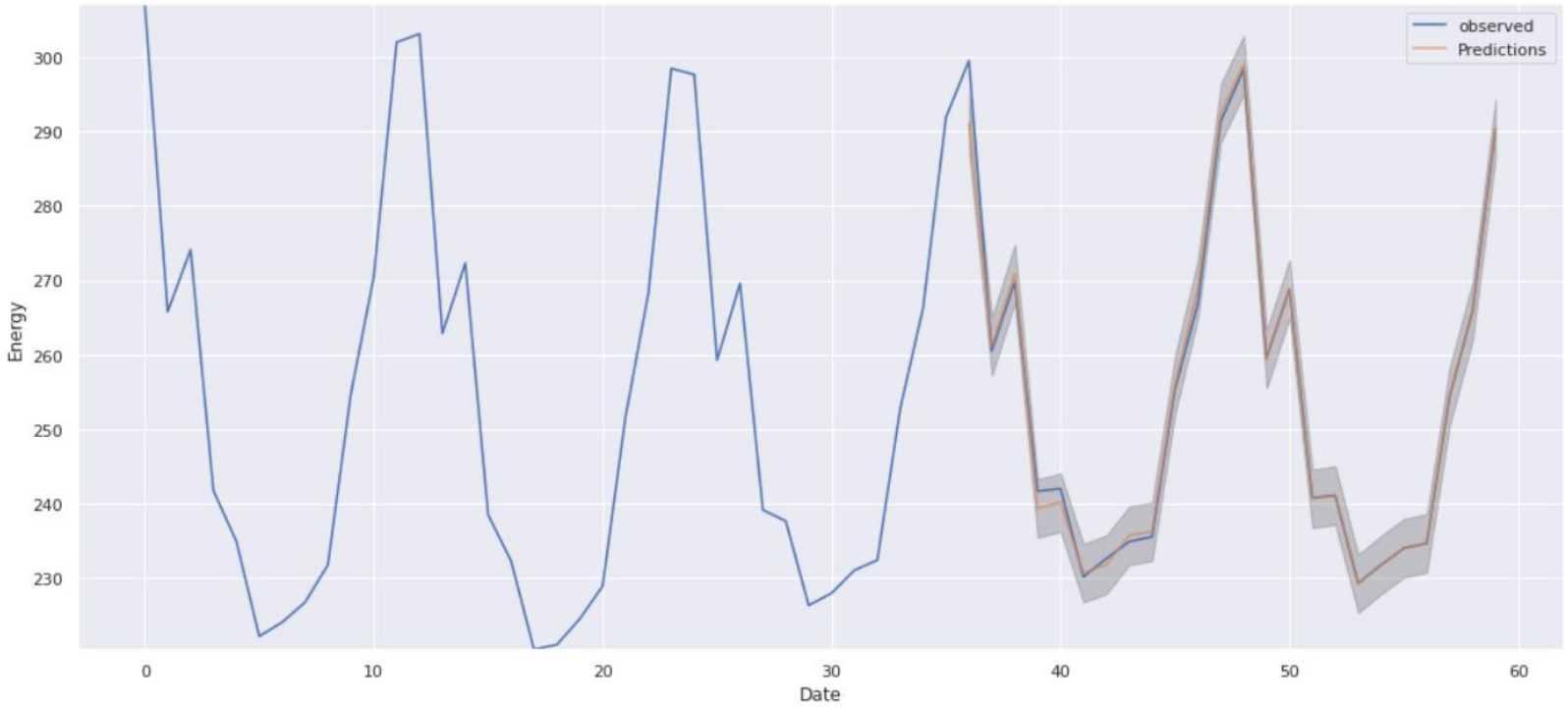
Cox proportional hazards model



by Karen Dedecker

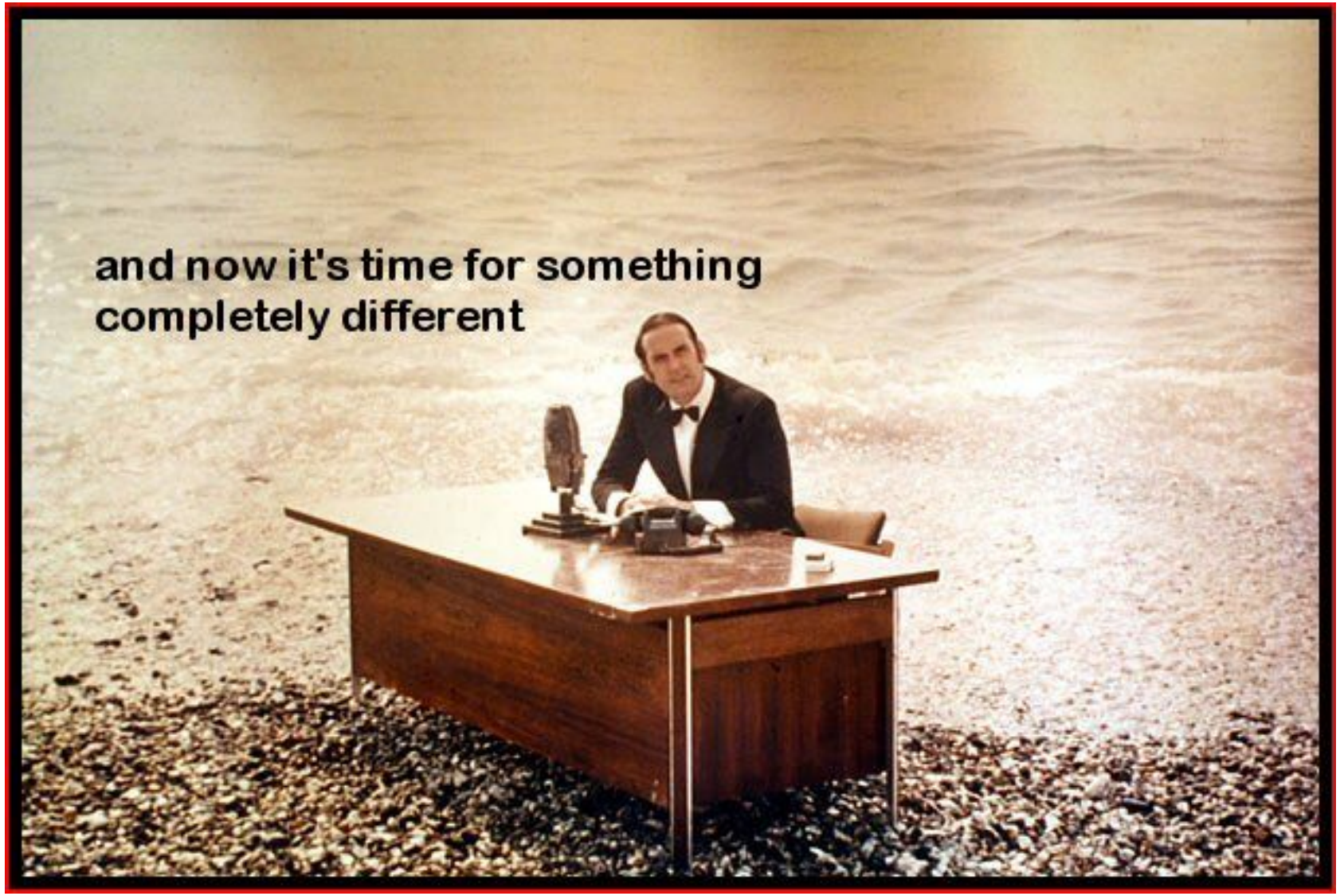
The models – Volume prediction

"**It has been said that history repeats itself. This is perhaps not quite correct; it merely rhymes."** Theodor Reik (not Mark Twain)



- ARIMA model
- Historical tendences
- Yearly fluctuations

by Pierre Brogniet


and now it's time for something
completely different

FIRST IMPROVE PROCESSES!

Analytics And Data Science

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil

From the Magazine (October 2012)

2012

2021

Should You Become a Data Engineer in 2021?

Data Engineering is the new Data Science


Nicholas Leong Mar 1 · 7 min read ★

Doing Data Science at large scale

https://omegapoint.se/devops

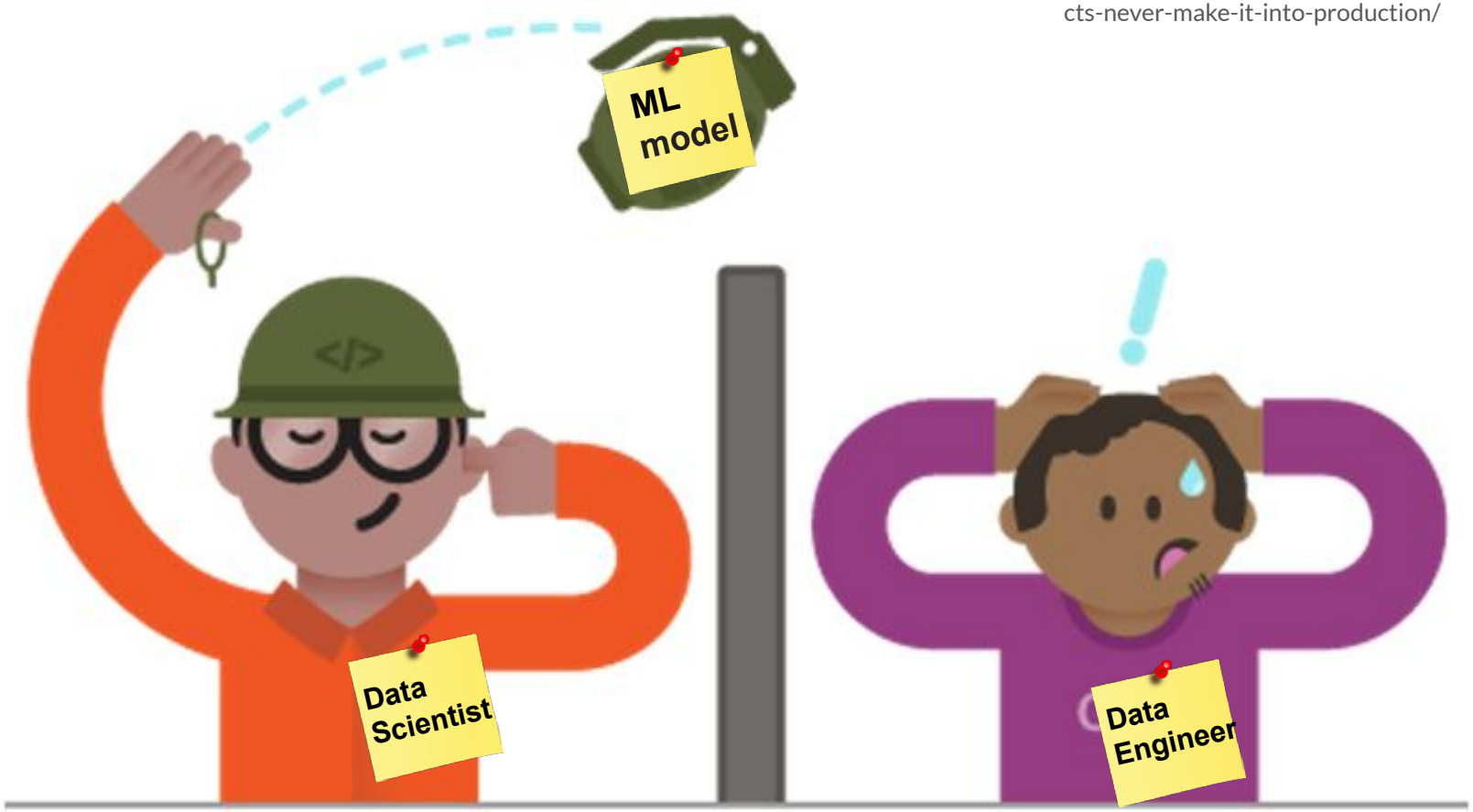
Doing Data Science at large scale

https://omegapoint.se/devops

Doing Data Science at large scale

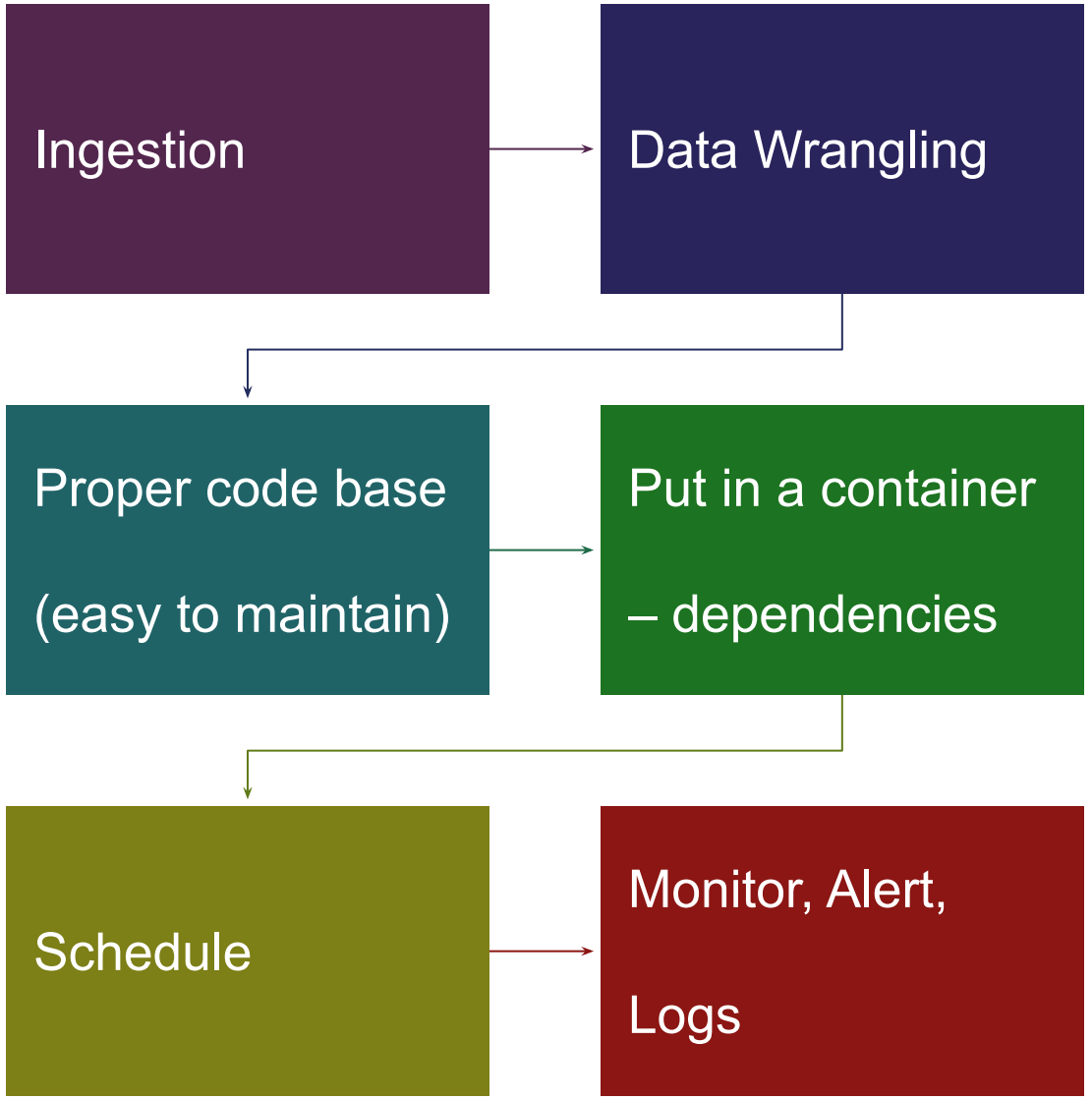
Only 13% of Data Science projects goes into production

https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/



https://omegapoint.se/devops

From the notebook to production model



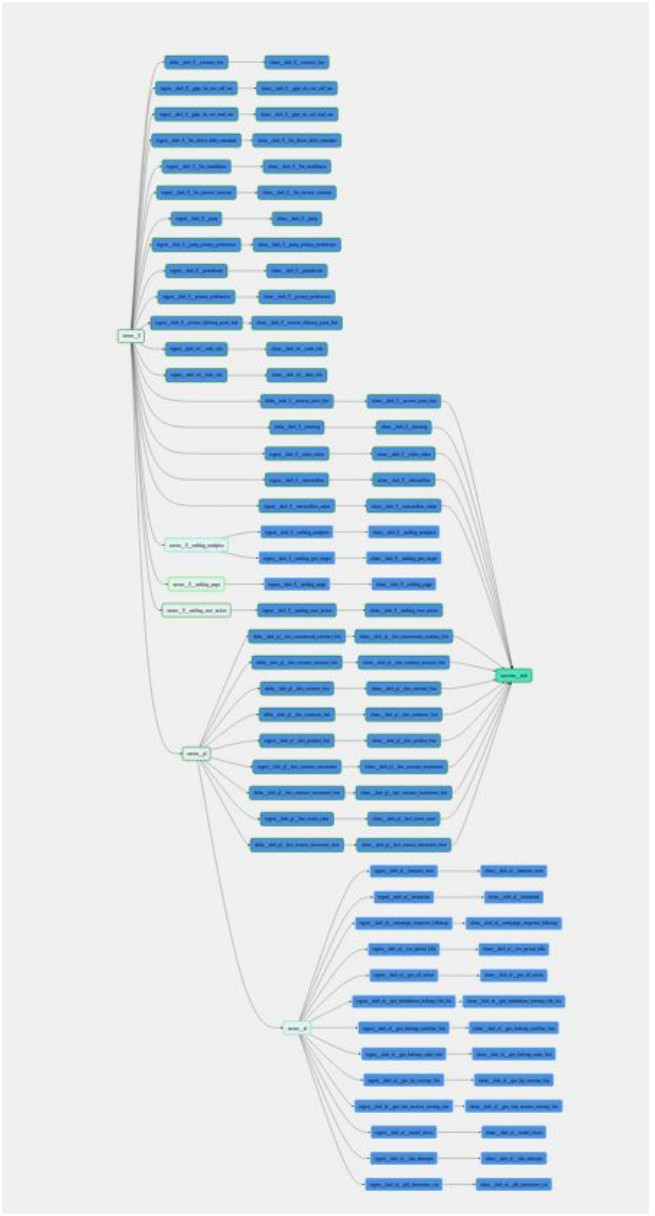
Ingestion

Copy/Paste

Injection

Copy/Paste

- Not so easy with large files
- 154 tasks, code to generate code
- 300 GB every day
- Delta ingestions and full ingestions
- Easy to add new tables



Data Wrangling - Cleaning and preprocessing

```
SELECT
        *
FROM
        "Imaginary_perfectly_clean_dataset_you_think_exists"
```

Copy/Paste

```
SELECT
        *
FROM
        "Imaginary_perfectly_clean_dataset_you_think_exists"
```

Day in a life of veteran data wrangler

How do we make this better?

Data platform

NEO

ML in Cloud

Trinity

BI in Cloud

Machine Learning & AI
Live use cases

Aero Condensor Backpressure
Monitor to avoid trip

Generator Temperature
Monitor to avoid damage

Wind Onshore and Offshore Power Forecast

Weather Alerts:
Saved 20% FTE

Wind Atlas
Avoided costs on wind studies and lower risk for investment decisions

AMR Solar Forecast
accuracy increased by 2,8%

Day Ahead Price Forecast
improved bidding strategy

Day Ahead Imbalance Forecast
improved bidding strategy

Churn
protect customer portfolio

Upsell for services
increased revenu

Lead Generation
cost reduction

Solar panel propensity
new solar panel installation

Long Term Forecasting
fixed price contract

Machine Learning & AI
Live use cases

Aero Condensor Backpressure
Monitor to avoid trip

Generator Temperature
Monitor to avoid damage

Wind Onshore and Offshore Power Forecast

Weather Alerts:
Saved 20% FTE

Wind Atlas
Avoided costs on wind studies and lower risk for investment decisions

AMR Solar Forecast
accuracy increased by 2,8%

Day Ahead Price Forecast
improved bidding strategy

Day Ahead Imbalance Forecast
improved bidding strategy

Churn
protect customer portfolio

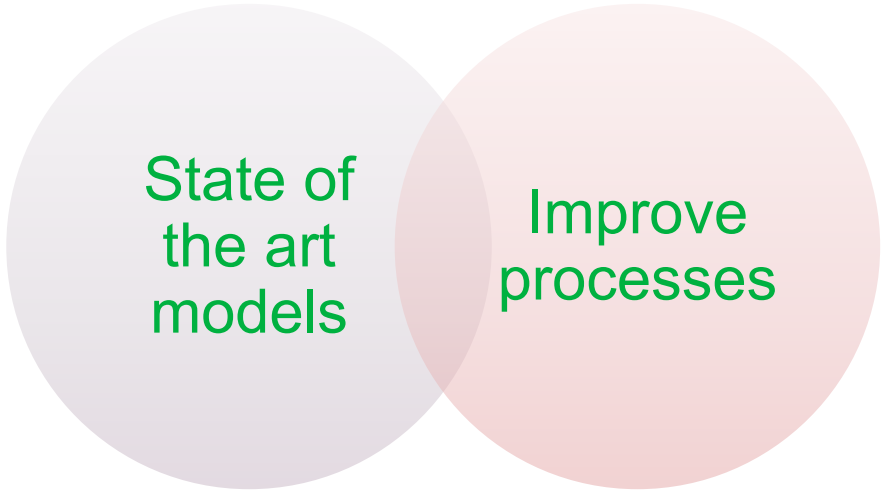
Upsell for services
increased revenue

Lead Generation
cost reduction

Solar panel propensity
new solar panel installation

Long Term Forecasting
fixed price contract

Conclusions



Data bring real value to:

Business

Planet

Customers

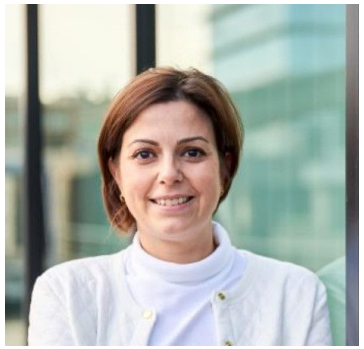
Thank you

Karen Dedecker
Data Scientist

Michalina Igla
Front End developer

Geanina Masgras
Scrum master

Ly Huong Chhor
Data Engineer



Francesca Onofrj
Wind expert

Some of our awesome male colleagues:

Bernard Sacré, Jean-Michel Begon, Philippe Habay, Pierre Brogniet, Bram Stepman, Wouter Bailleul, Kevin Van Hees, Clement Desseyn, Robin Munier

Neo Data Platform Team