

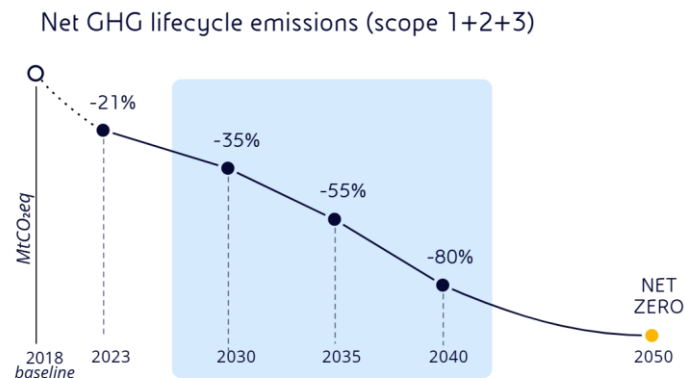
L' Intelligenza Artificiale per accelerare lo sviluppo di tecnologie R&D innovative per la transizione Net-Zero

Francesco Cannarile
Roberto Grana
Giacomo Gorni

9 Ottobre 2025

This document is the property of Eni, which holds all rights. The material may not be reproduced, distributed, altered or used for purposes other than consultation by webinar participants.

Digital for R&D: Motivation & Context



- We have embarked on an **industrial transformation** involving all business lines to decarbonize the entire company. To this end, we invest in researching, developing, and implementing **transition technologies**.
- According to the **technological neutrality principle**, there is no single solution to achieve the energy transition, we need a **technological mix** that can be adapted to different applications and needs. This is why we are developing a wide range of technologies that **support** the **decarbonisation** of each sector of the economy and our daily lives.

Eni applications in R&D to achieve Net Zero Emission



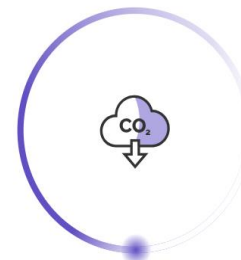
Energy from
renewables
Electric and
thermal energy



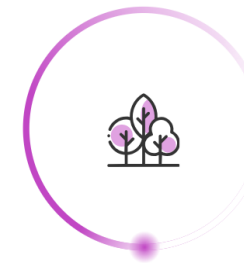
Fusion
The electricity and
thermal energy of
tomorrow



Circular
Economy
Biofuels, sustainable
chemistry and critical
materials

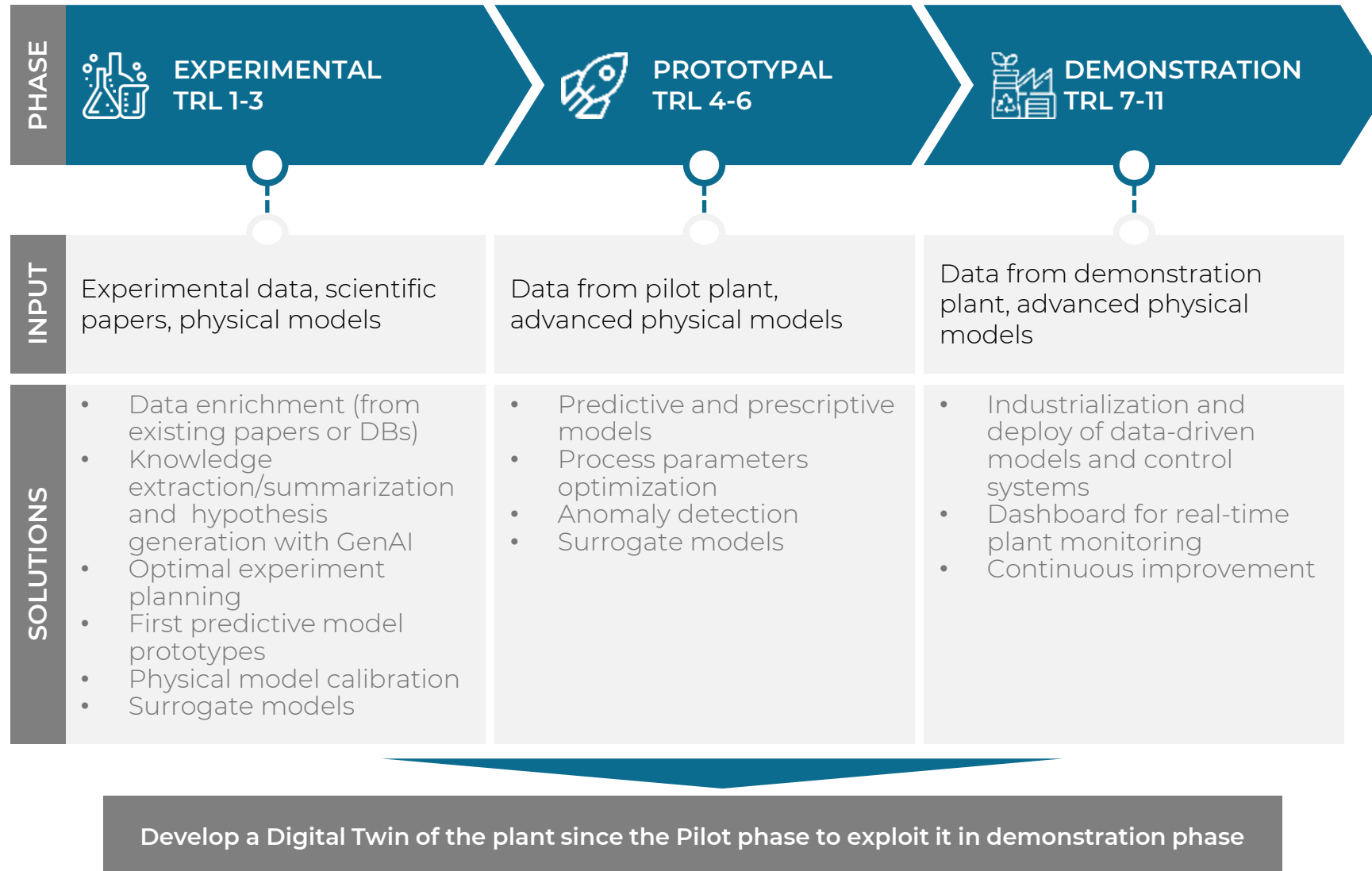


CCUS
Decarbonization of
processes



Environment
Ecosystem
restoration

How AI can boost experimentation



Example – AI for material discovery

● Artificial Intelligence
● Advanced Modeling & Simulation



EXPERIMENTAL PHASE

PHASES



STUDY

HYPOTHESIZE

SCREENING

TESTING

DESCRIPTION

Defining the problem and desired material properties based on the intended application. **Perform literature review** and create a first database of material

Formulating hypothesis and **proposing** a set of candidate materials with target properties for further evaluation

Evaluate material candidates by analyzing their **predicted properties** and **stability**, prioritizing those with the highest potential for success in the intended application

Determining **synthesis pathways**, synthesizing materials, **measuring** properties at atomic and macroscopic scales, and **refining processes** based on experimental results

CAPABILITIES

- **Reviewing** prior research, patents, and material databases to gather relevant insights.
- **Extracting** and organizing data from existing experimental or computational investigations
- **Simulating** by **first principle methods** data set when these are not available or not reliable.

- **Proposing** a promising set of candidates for further evaluation via iterative exploration
- **Formulating hypotheses** using **theoretical or computational model** chemical physical-based

- ● Running **computational chemical physical-based models** and/or **Artificial Intelligence** based ones to assess candidates' properties.
- **Assessing** key factors s by focusing on those with the highest likelihood of success.

- ● Extract **synthesis recipes** from literature (AI) and propose **potential routes** using advanced computational modeling techniques.
- **Analyze** experimental data
- Using computational model to **validate** experimental results

The **discovery process** involves four iterative phases: **study** (analyzing known data and materials), **hypothesize** (generating candidate structures), **screening** (evaluating candidates), and **testing** (promising materials through experiments).

Example - AI for accelerating carbon capture technologies



PROTOTYPAL PHASE

Objective

Support research and development in **designing, validating and industrializing a new technology for CO2 capture** through developing **artificial intelligence tools**



FIRST-PRINCIPLES MODELING

Provide a **mathematical description** of the system behavior for a better understanding of the system and simulate scenarios of interest

- Experimental data
- Data of literature
- Subject Matter Experts knowledge

- **First-Principles model** based on differential equations that allows simulating the trajectory of an experiment
- Meta-heuristic optimization algorithms (**differential evolution**) to calibrate physical model parameters



OPTIMIZATION AND CONTROL

Prescribing the **optimal set-points** for the experimental setup in order to performance in **real-time**

- Simulation from the calibrated first-principles model

- **Reinforcement learning** model to prescribe the **best set points** for the experimental setup to optimize system performance in real-time using **simulations** from the first-principles model



PREDICTIVE MODELING

Providing real-time and automated **assessment** and **prediction** of quantities of interest

- Experimental data
- Calibrated first-principles model

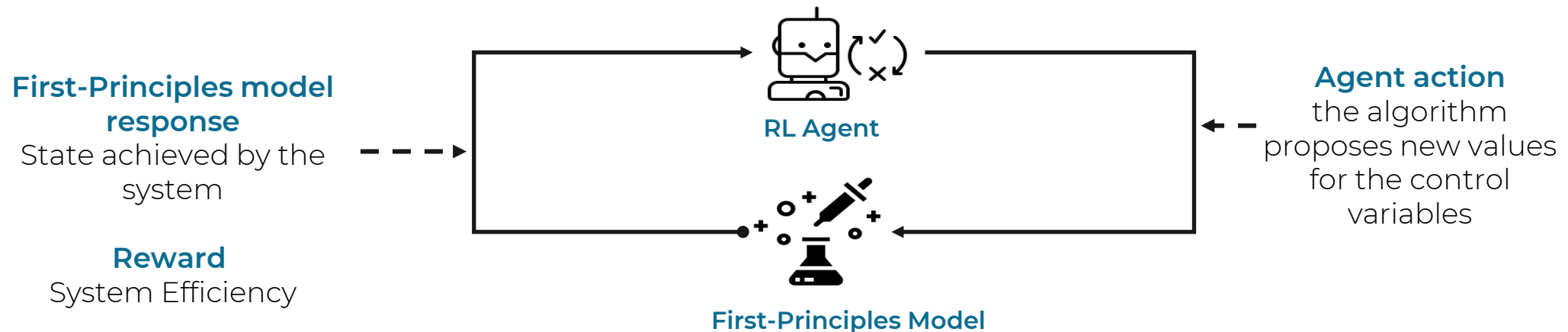
- Machine learning (deep **neural networks**) model to predict quantity of interest to be monitored

Reinforcement Learning for optimal system control



PROTOTYPAL PHASE

- **Modeling approach:** the system is represented by the **first-principles model**. The agent interacts instant by instant with the first-principles model by proposing new values of the variables to be optimized (actions) and causing the model to evolve (new state) and return a new value of **system efficiency*** (reward).



- **RL Algorithm:** Deep Deterministic Policy Gradient (DDPG)

Results

System performance obtained with RL is **doubled** with respect to the not optimized setting



Focus on Design of Experiment & Optimal Experimental Design

Why Design of Experiments (DoE)?

CHALLENGE IN R&D

- Developing new technologies or processes involves **many variables**
- Resources are limited: experiments cost **time, money, and materials**.
- **Trial-and-error** or “**one factor at a time**” testing is:
 - Too slow
 - Cannot reveal **interactions** between factors
 - Often gives **incomplete or misleading results**

THE NEED

R&D teams require:

- **Learning** more with fewer experiments.
- **Building** a quantitative understanding of how factors affect outcomes.
- **Predicting** and optimize results before scaling up.

THE ANSWER: DoE

- DoE provides a **statistical structured** approach to experimentation
- Each test is designed to contribute **maximum information** to the overall picture
- **Fewer runs, deeper insights, faster and more reliable** decision making

What is DoE?



A SMARTER WAY TO EXPERIMENT

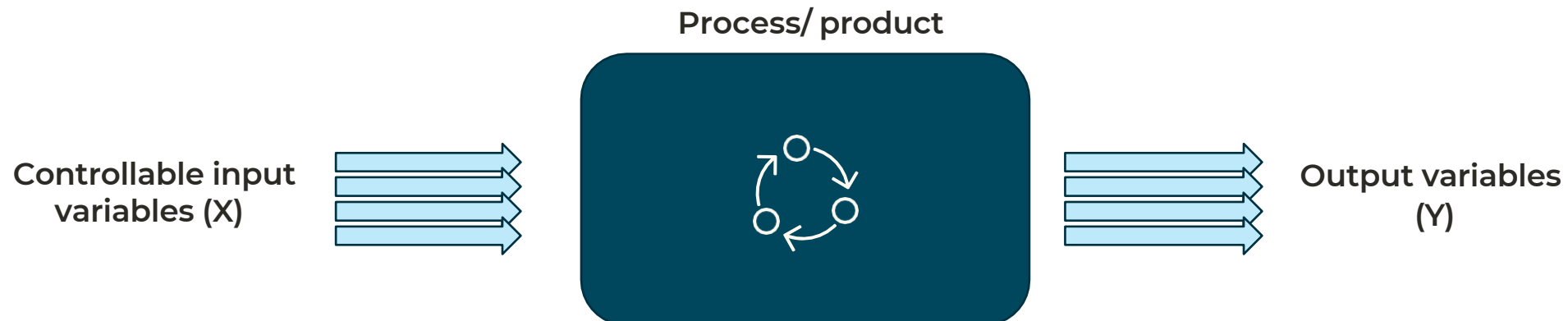
- DoE is a **structured method** to plan and run experiments
- Instead of changing one variable at a time, it tests **multiple factors together** in a systematic way.
- Each experiment is chosen so that, combined with the others, it builds a **complete picture** of how the system behaves.

WHAT IT DOES

- Identifies the **factors that really matter** ("vital few")
- Reveals how factors **interact** (e.g. temperature may matter more at high pressure)
- Builds a **predictive model** linking inputs → outputs
- Helps find the **best settings** for performance, quality, or cost.

WHY IT MATTERS

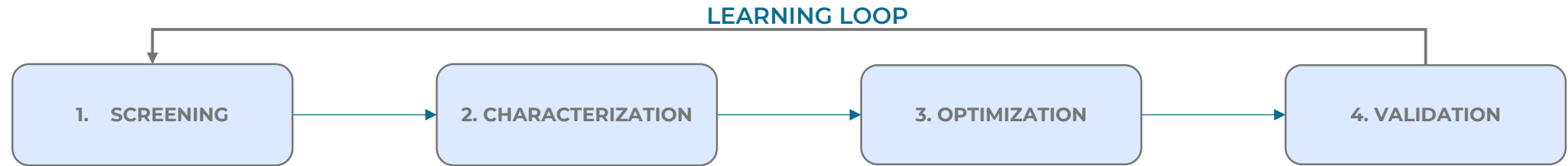
- Saves time and resources by avoiding redundant trials.
- Provides **scientific confidence** instead of guesswork.
- Creates knowledge that can be reused and scaled up.



The DoE Journey



An iterative process where each step builds on the results of the previous one



Efficiently **identify** the most influential factors (the "**vital few**") from a wide range of possibilities. This reduces complexity and focuses subsequent efforts on a smaller, more manageable set of variables.

Method: Fractional factorial design (linear screening model)

Quantify how significant **factors behave collectively** by modeling their main effects and, crucially, their interactions. The outcome is a **first-pass mathematical model** of the process that explains how variables work together.

Method: Full factorial design (linear model with interaction terms)

Locate the optimal settings for process variables by **modeling response curvature** (i.e., quadratic effects). This involves running new, targeted experiments to precisely **estimate the factor levels** that maximize or minimize the output.

Method: Response Surface Methodology (RSM) with quadratic models (CCD, Box-Behnken).

Formally **validate** the final model. Use statistical analysis to test the significance of all model terms, then conduct **confirmation runs** to empirically verify the model's predictive accuracy and ensure the results are reproducible

Method: Model diagnostics, confirmation runs

This workflow is an example of an iterative learning loop: in practice, steps can also be revisited if needed, and the cycle continues until the model is accurate enough and the optimum is confirmed with confidence.

From DoE to Optimal Experimental Design

Evolving from a fixed plan to a dynamic learning strategy



Classical DoE *The statistic blueprint*

Optimal Experimental Design *The adaptive strategy*

PLANNING

Fixed and pre-defined (e.g., Factorial).

Model-based and **built sequentially**

FLEXIBILITY

All runs are **decided upfront** with no adaptation (rigid)

Works as a loop in an **adaptive way** (*run → update model → choose next*).

APPLICABILITY

Limited to simple, linear, or polynomial models.

Handles complex, **non-linear systems** and **practical constraints**.

EFFICIENCY

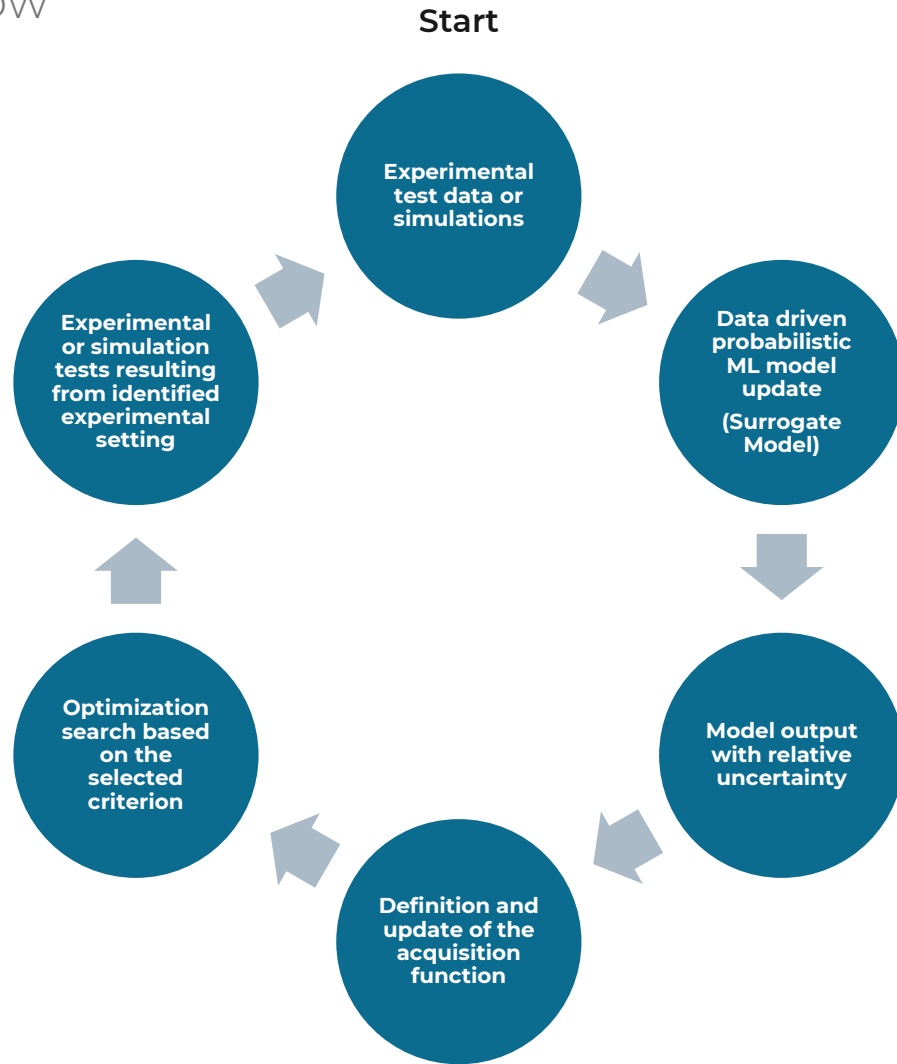
Some tests may bring **little** new information.

Leads to **faster** decisions with **fewer** experiments.

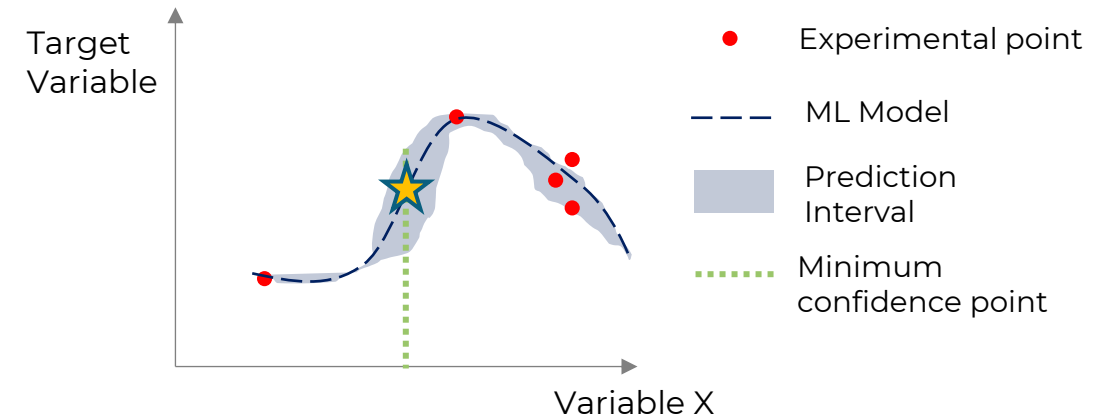
OED transforms experimentation from following a *fixed map* to using a *smart GPS* that constantly finds the fastest route to the objective

Optimal Experimental Design

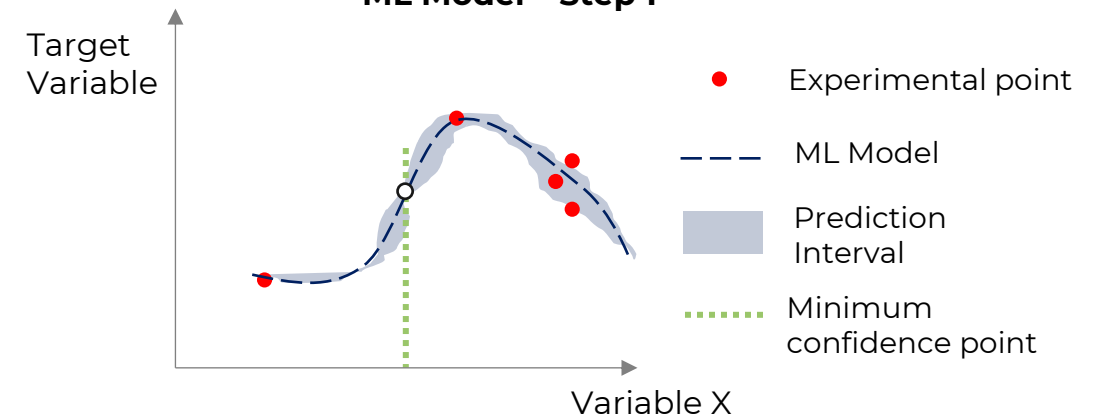
Workflow



ML Model – Step 0



ML Model – Step 1



Experimental tests

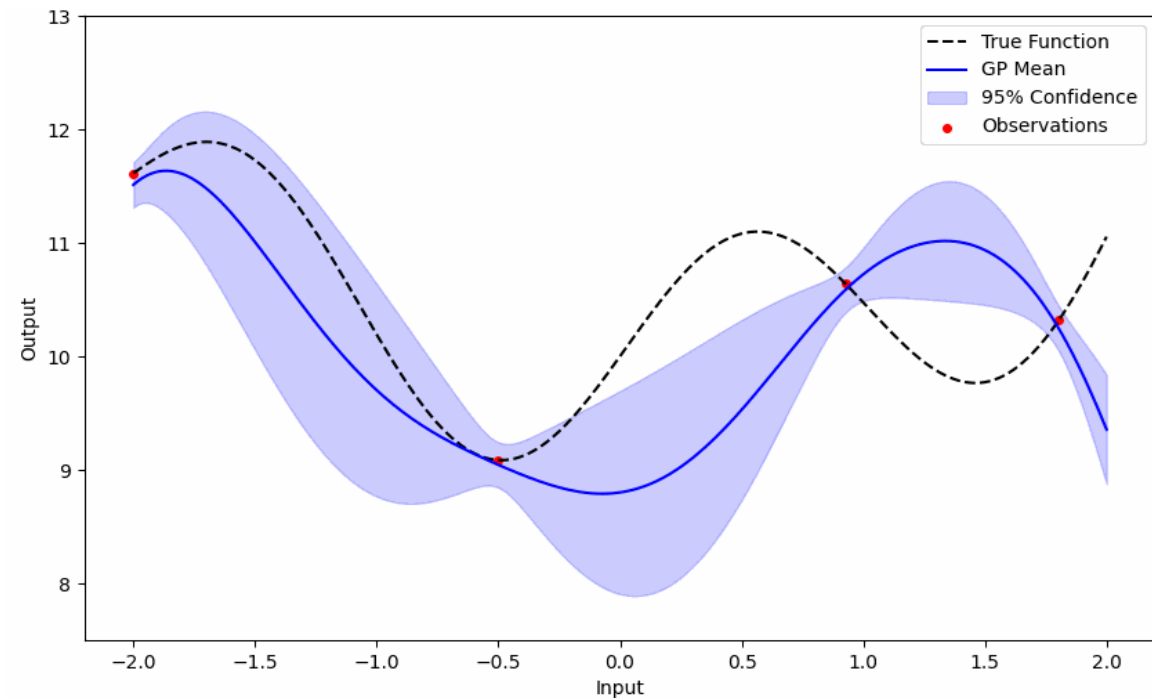
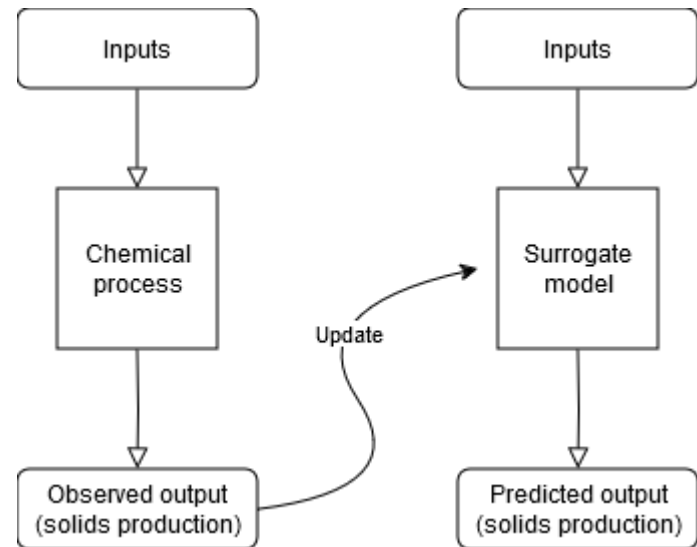
Tests suggested by the OED methodology

Surrogate Model - Data driven probabilistic ML model



A probabilistic ML model is used to approximate the unknown process that relates the inputs and the outputs. The model should be able to provide

1. **A prediction of the output**, which is more accurate the more data we have to train the model on
2. **A prediction interval around the prediction** to represent the uncertainty of the model about the prediction



Gaussian Processes (GPs) are typically used for the development of the surrogate model

Making decisions: the acquisition function



An acquisition function defines how to select the inputs for the next experiment. Different acquisition functions allow for different experiment design strategies.

Exploitation

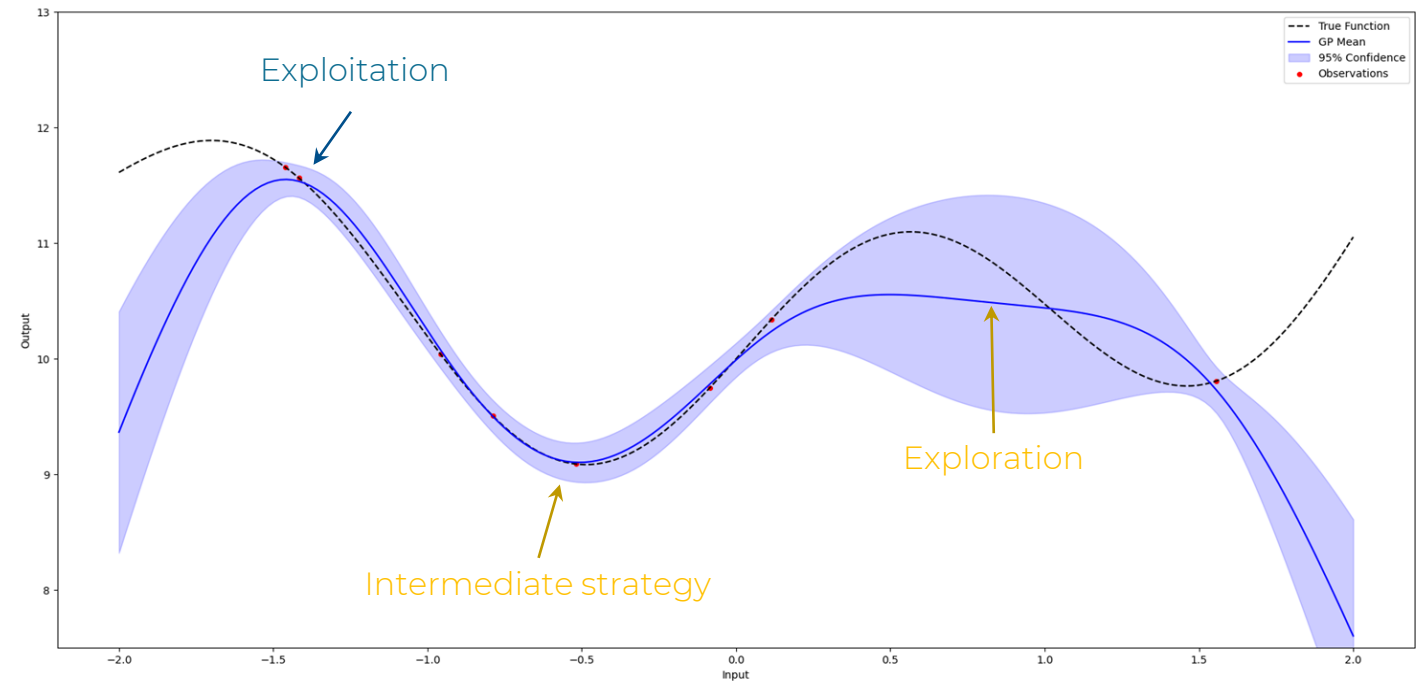
- "Go where we think the best results are"
- Like going down the lowest visible valley

Exploration

- "Go where we are most uncertain"
- Like mapping unknown territory

Intermediate strategy

- "Go where we are confident of the best results, but allow some exploration"
- Balances exploration and exploitation

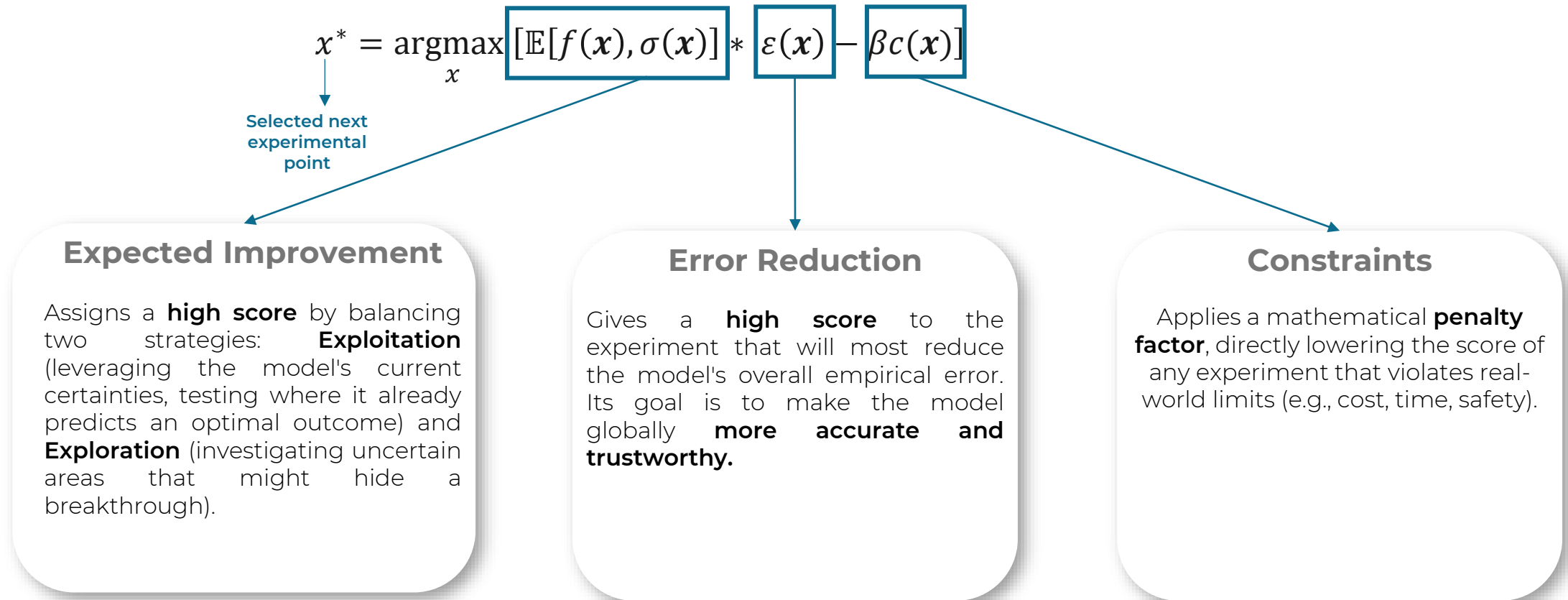


Each acquisition function has its advantages and disadvantages, and its use depends on the objective.

Choosing the next experiment: the acquisition function



An example of acquisition function



This acquisition function combines these competing scores to find the experiment with the highest overall rating

The optimization algorithm

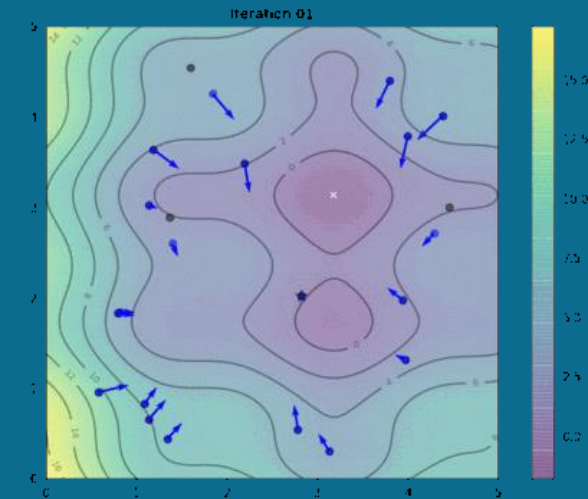
The optimization algorithm is the **numerical solver** that finds the most promising next experiment, given the surrogate model and the acquisition function.

HOW IT WORKS

1. The acquisition function (derived from the surrogate model) defines the objective function **to be optimized**.
2. The algorithm **explores** the input space to find its maximum.
3. The result is the **next experimental point** to test.

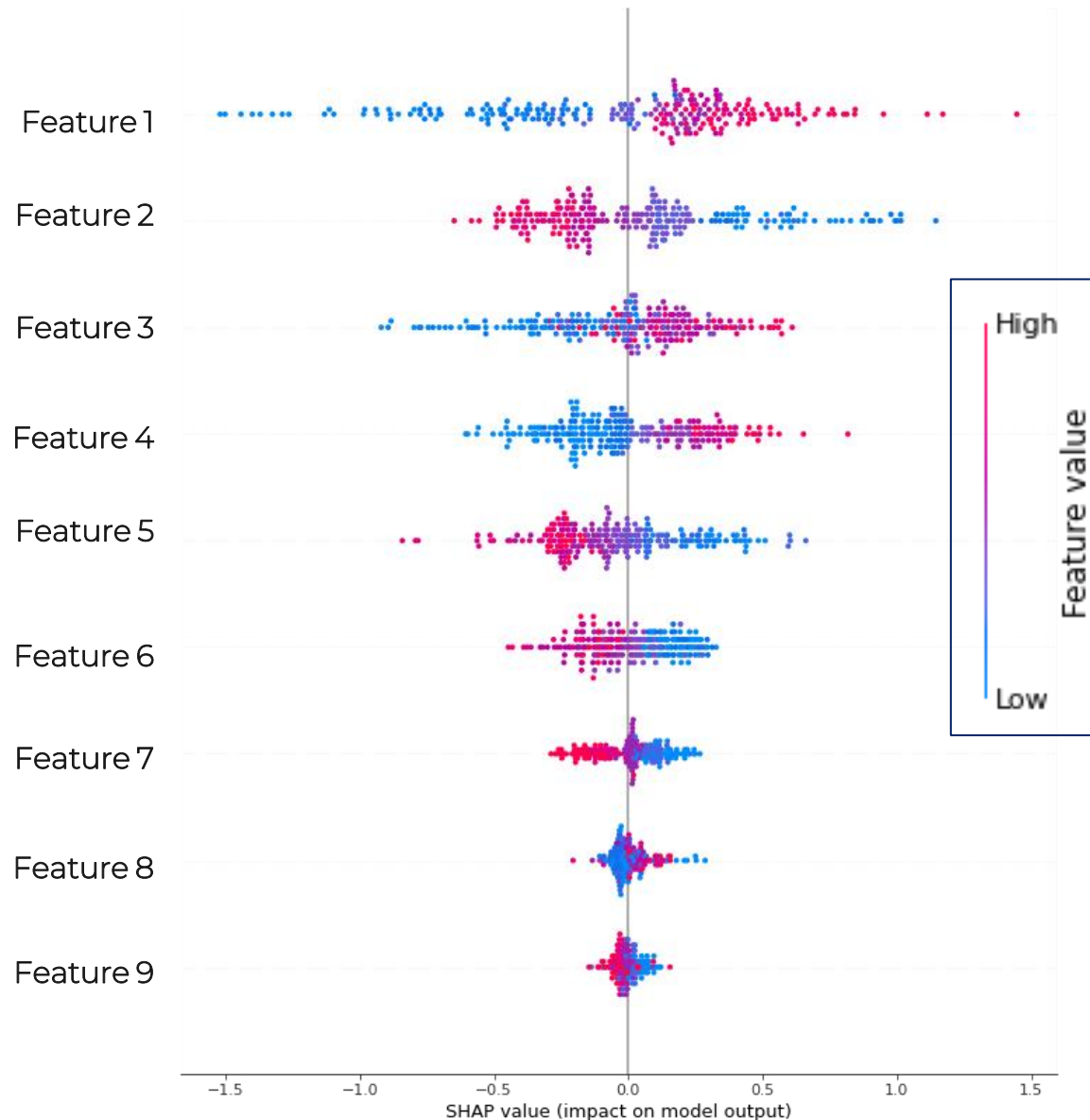
TYPICAL SOLUTION METHODS

- **Gradient-based search:** fast if the function is smooth, but may get stuck in local optima.
- **Global search heuristics:** genetic algorithms, simulated annealing, particle swarm — slower but robust.
- **Bayesian optimization routines:** combine Gaussian Processes with acquisition functions; popular for expensive experiments.



Example of iterations using meta-heuristic optimization
(Particle swarm optimization)

OED - Feature contribution



The surrogate mode can be used to analyze the **most impactful** features of the system and what is their possible effect on the prediction.

Shapley values are typically used to infer the feature contribution.

- *Feature 1*
- *Feature 3*
- *Feature 4*
- *Feature 8*
- *Feature 7*
- *Feature 6*
- *Feature 9*
- *Feature 2*

Positive impact
on target

Negative
impact on
target



Optimal Experimental Design in Bio-feedstocks pretreatment processes

BioRefining Value Chain

Enhancing value across processes



BIOFEEDSTOCK

A wide variety of raw materials such as:

- vegetable oils
- tallow
- waste or used cooking oil (UCO)
- wastes or residues such as nonfood-grade vegetable oils, animal fats, sludge palm oil mill effluent (POME)

BIOMASS TREATMENT

Pretreatment unit is necessary to remove impurities such as phosphorous, metals, polyethylene, nitrogen and chlorine-containing components that are naturally present in some raw materials

ECOFINING

Ecofining™ is proprietary technology to convert cooking oil, tallow and non-edible vegetable oils to produce biofuels

BIOFUELS

Hydrotreated vegetable oil (HVO) is a newly developed renewable diesel that uses renewable feedstocks via the hydrotreatment process

Development of an Unconventional Pretreatment Process



Definition of UnConventional Bio-Feedstocks

The unconventional raw bio-feedstocks usually contain high amounts of pollutants such as:

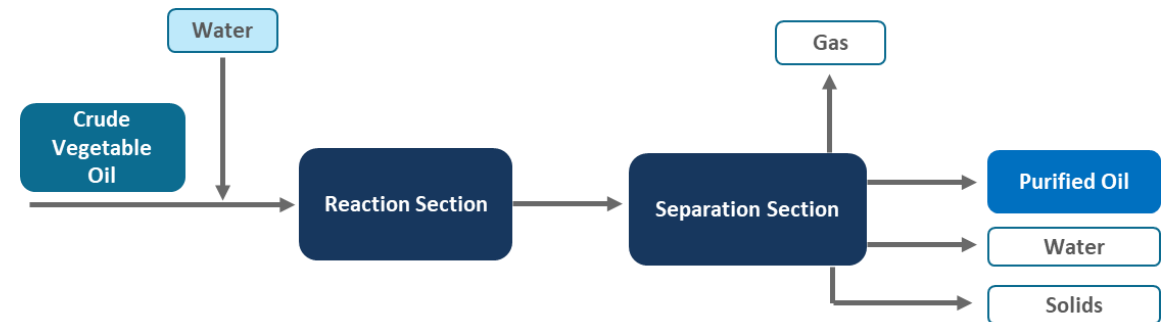
- metals,
- phosphorous,
- salts, etc.



Definition of UnConventional Pre-treatment Process

The unconventional pre-treatment process is based on the right selection of:

- Operating conditions (contact time, temperature, and pressure)
- Ratio between reagents and phases (oil, water, other chemicals agent, etc..)

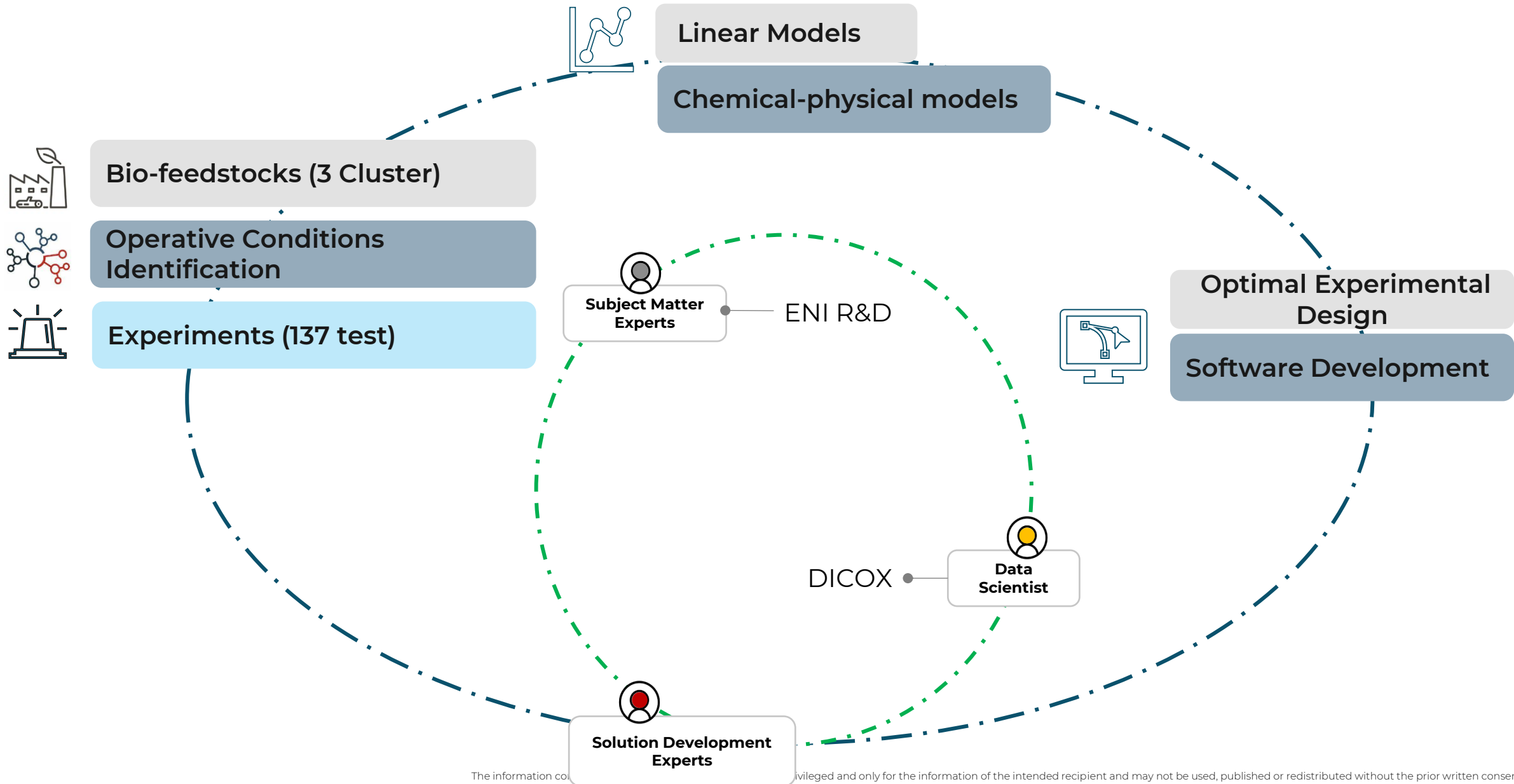


To overcome the conventional pretreatment processes battery limits and to widen the bio-feedstock portfolio, Eni R&D department has developed an innovative pre-treatment technology.

The UNCONVENTIONAL pre-treatment has to be :

- Effective in the contaminants removal;
- Simple with low chemical agents e few reaction steps
- Flexible capable to purified a wide range of biofeeds

Unconventional Pretreatment Process: Project Team and Workflow



Unconventional Pretreatment Process: Digital Boost



Data Aggregation

Create an experimental dataset and apply advanced analytics techniques to achieve deeper insights and knowledge.



Predictive Model and Optimization

Starting from Eni R&D models, develop predictive data-driven linear models to estimate reaction efficiency based on process conditions and bio-feedstock properties



Optimal Experimental Design

Apply Optimal Experimental Design techniques to identify new optimal experimental tests based on different criteria



Tool Development

Develop a shared tool to allow all R&D users to independently integrate DoE within daily work

Unconventional Pretreatment Process: Data Driven Models

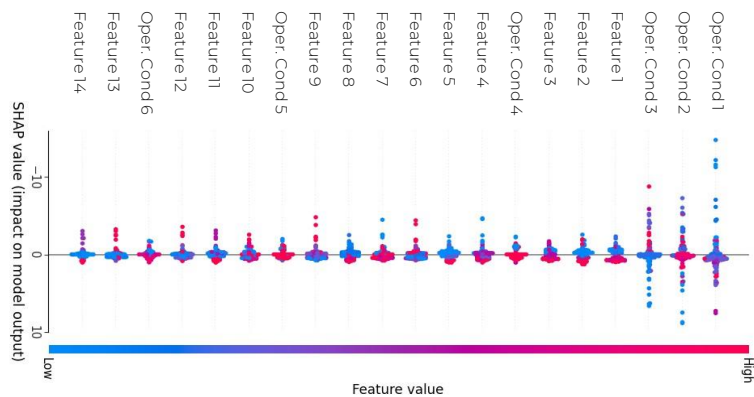


Data Aggregation

The model shall be fed with proper input data.

In details:

- 6 Parameters of process operative conditions
- 17 Parameters of bio-feedstocks (e.g. chemical and physical properties such as acidity, density, carbon %)



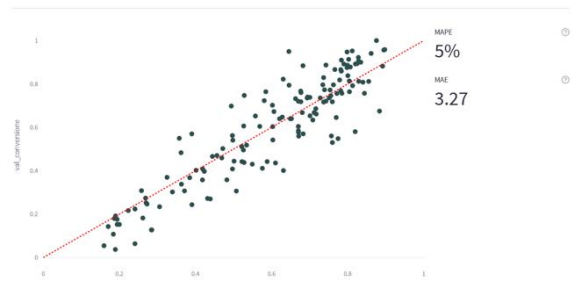
At this stage is deemed necessary:

- analysis of input data through ML techniques
- to share *main outcomes* with the *subject experts*

Predictive Model and Optimization

Different strategies have been implemented:

- Different bio-feedstock clustering
- Different predictive models



Campaign	Model	MAE	MSE
Set 1	linear	0.63	0.72
Set 2	forest	0.38	0.29
Set 3	mlp	3.17	28.57
Set 4	gradient	8.46	152.70
Set 5	forest	2.11	8.72
Set 6	gpr	0.80	0.87
Set 7	forest	2.93	12.66
Cluster 1	gradient	4.21	59.30
Cluster 2	forest	2.54	10.04
Cluster 3	mlp	0.86	2.18
Multi set	gpr	3.27	42.85

Model training stage:

- Multi-set cluster selected
- Gaussian Process Regressor selected

Unconventional Pretreatment Process: Optimal Experimental Design (OED)



Optimal Experimental Design goal is to define a new set of experimental test which allow to:

- Filtering inputs relevant to output (**screening**)
- Build a model to correlates inputs and output (**modelling**)

In the first stage of analysis, **the goals of the experimental investigation** shall be defined in order to define the constraints.
The constraints may comprise many aspects, even not strictly related to the technical activity such as time and cost efforts.

Selection Criteria

The goal of this specific case study (which is not general) is to **minimize the pretreatment efficiency**.

In order to have different new experimental points, it was decided to combine points with **lower pretreatment efficiency** and **higher model uncertainty**.

For the *Gaussian Process Regressor (GPR)*, model uncertainty is the variance between model predictions and experimental results

Experimental Design

Parameters	Base Case	EXP 1	EXP 2	EXP 3	EXP 4
Oper Cond 1	0%	-1%	7%	-3%	13%
Oper Cond 2	0%	-41%	-41%	-41%	6%
Oper Cond 3	0%	64%	0%	94%	200%
Oper Cond 4	0%	-6%	25%	25%	25%
Pretreatment Efficiency	0%	14%	-30%	-1%	-11%
Standard Deviation	0%	-72%	-68%	-37%	-45%

- **The new experiments are defined through the OED procedure based on:**
 - ✓ **Criterion 1: efficiency minimization**
 - ✓ **Criterion 2: maximum model uncertainty**

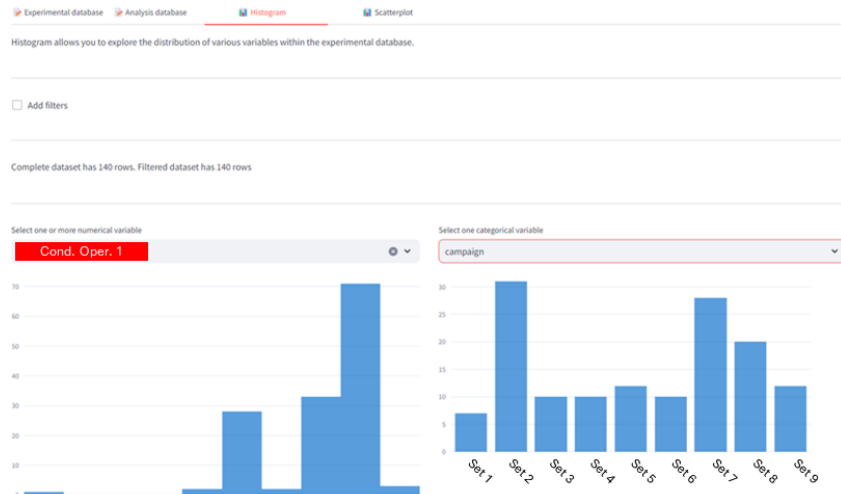
Unconventional Pretreatment Process: Web APP



The final stage is the development of a tool (Web APP) with the features:

- Collection, management and aggregation of experimental data (database);
- Prediction by the **data driven model** of new experimental data;
- Application of **Optimal Experimental Design** to efficiently boost the entire development process by means of upgrading the predictive model.

Database analyses and experiments



Data driven model



OED

Parametri da OED

alpha	beta	n_doe_points
1.00 - +	0.20 - +	5 - +

Alpha: peso nel DoE all'exploration (visitare regioni inesplorate)

Beta: peso nel DoE all'exploitation (visitare regioni ad alta/bassa conversione)

Obiettivo: alta conversione/bassa conversione



Development and Validation of a Performance Analysis Platform for Utility-scale Photovoltaics

Giacomo Gorni

Introduction



Objective:

- Reduce capacity downtime
- Speed up underperformance detection
- Speed up O&M troubleshooting

Problem:

- Multiple masking factors overlapping
- Handle 100ks of data points per plant each day
- Utility-scale PV are huge

Solution:

- Identify where the losses are located
- Categorize and rank issues
- Fully automatized process
- User-friendly application

Methodology Adopted

From a “basic” performance monitoring at plant level

$$PR \propto \frac{\text{Production}}{\text{Insolation} \cdot \text{Capacity (DC)}}$$



To a set of Key Performance Indicators dedicated to main devices (inverters, trackers, combiner boxes)

- Automatically calculated
- Cleared of all the masking factors
- Comprehensive of suggested corrective actions

Scale problem



Milano

Utility-scale plant

- 250 MW
- 60 inverters
- 1000 CBs
- 500,000 modules
- 2,500,000 m²
- 3000 signals/plant
- 1.7 million datapoints/day



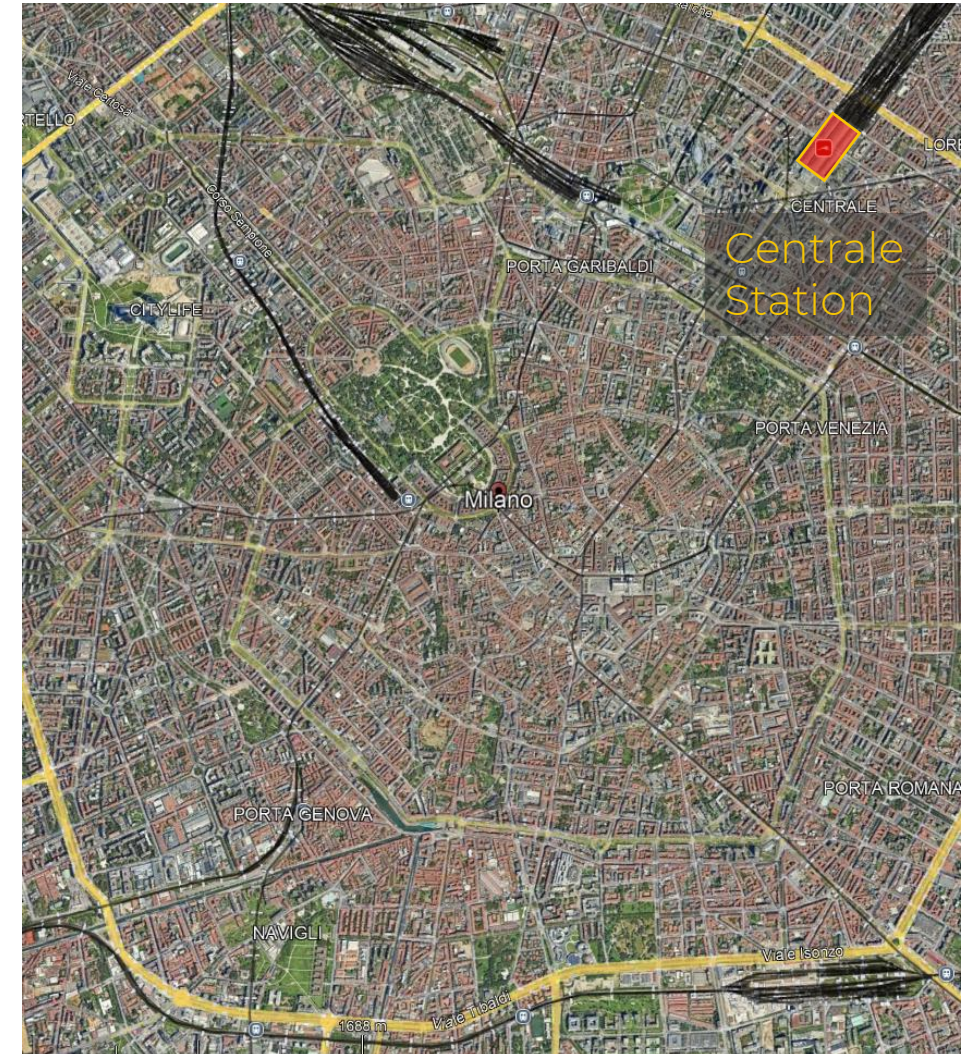
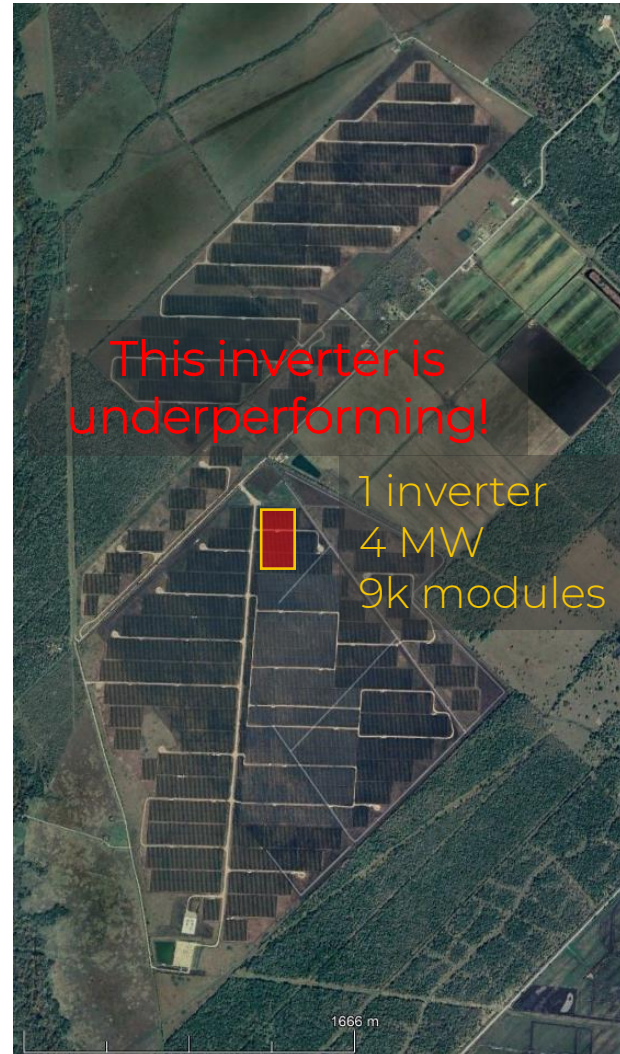
Traditional monitoring tools



Milano

Inverter

- 4 MW
- 16 CBs
- 8,000 modules
- 40,000 m²

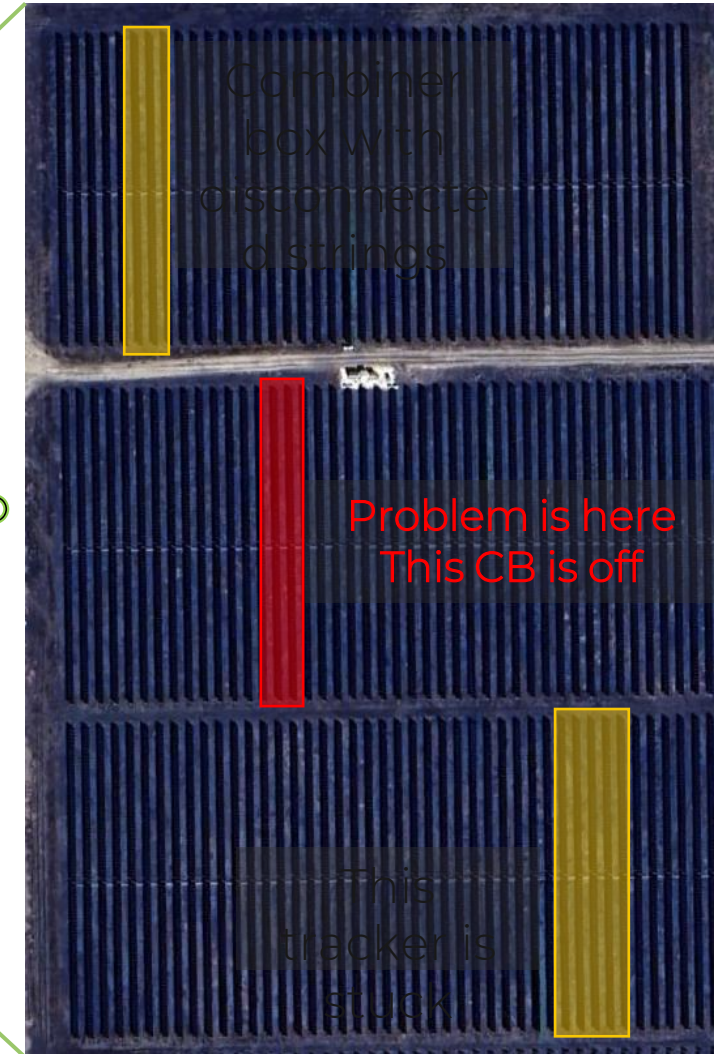
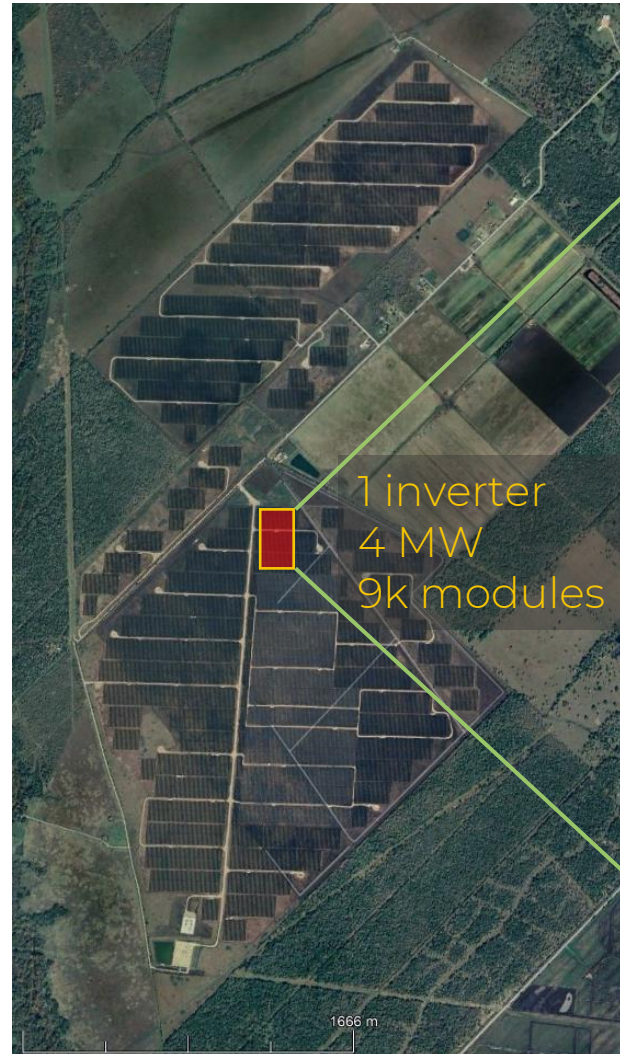


Digital tools as a solution to the scale problem



Combiner box

- 250 kW
- ~20 strings
- 500 modules
- 2,500 m²



From the MW scale...

...to the 100s kW scale

Data quality and event labeling

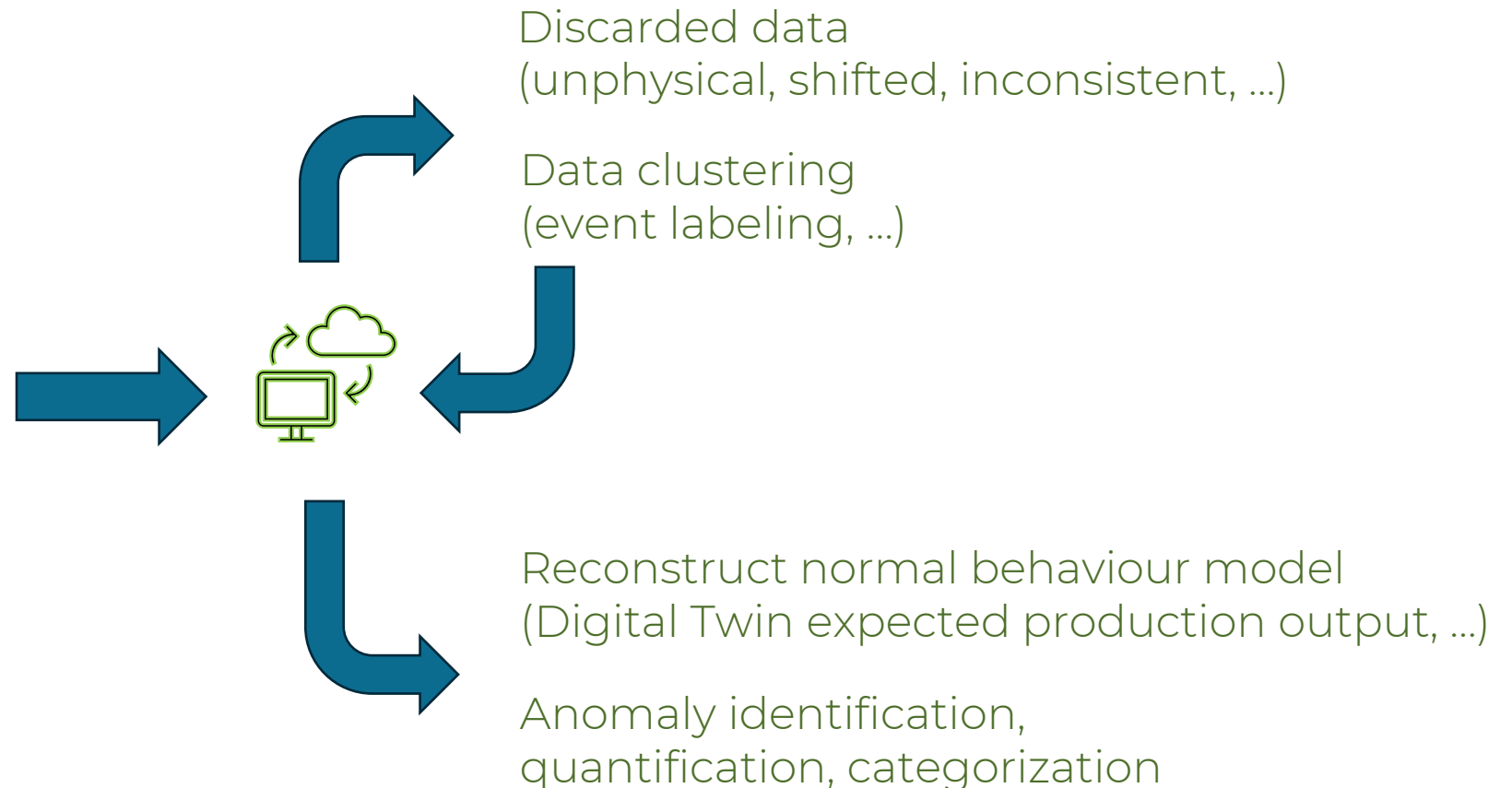


Data analytics tools (AI, ML, physical models...) give access to a solid data quality preprocessing and event labeling.

Plant metadata
(GPS, nominal power, layout...)

Actual Weather data
(Temp, Irr, Wind, RH...)

Actual production data
(power, current, voltage...)

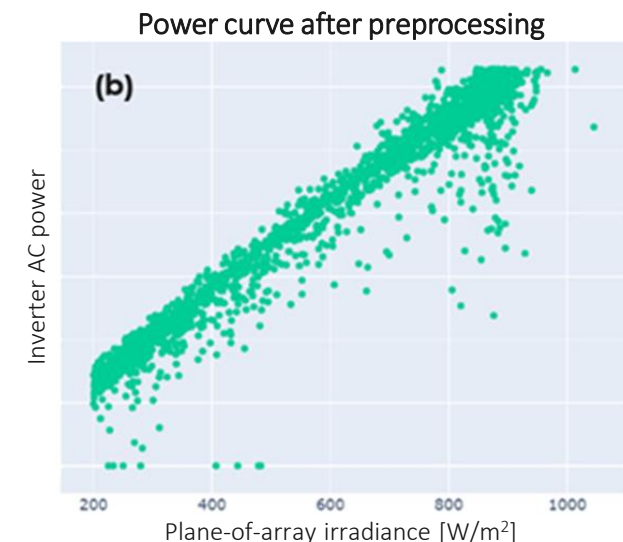
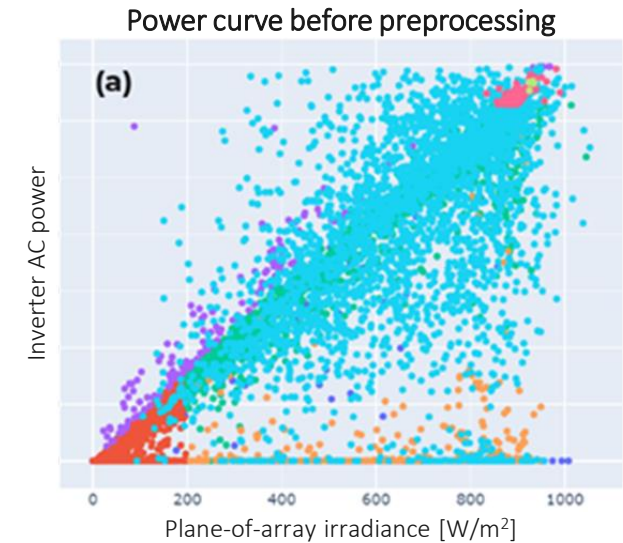


Data quality and event labeling



Data analytics tools (AI, ML, physical models...) give access to a solid data quality preprocessing and event labeling.

- Irradiance sensors **anomalies**.
- MPP deviations (technical or economical **curtailment**).
- Tracker's wind **stow** mode.
- **Temperature** derating.
- **Rear** Irradiance correction for bifacial systems.
- Tilted irradiance correction by **angular** response.
- Interrow **self-shading** effects.
- Inverter **clipping** detection and correction.



Key Performance Indicators (KPIs)

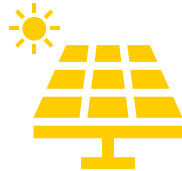


“Advanced” PR

Target:

- Inverter

Comprehensive metric to have a bird-eye view of the macro trends occurring in the plant.



DC power health

Target:

- Combiner box

Detectable issues:

- CB off / unavailable
- Non-communication
- Disconnected strings
- Modules degradation
- Misalignment with as-built documentation



Tracker health

Target:

- Motor

Detectable issues:

- Non-communication
- Stalled trackers
- Suboptimal angle
- Wind stow position



DC-AC health

Target:

- Inverter

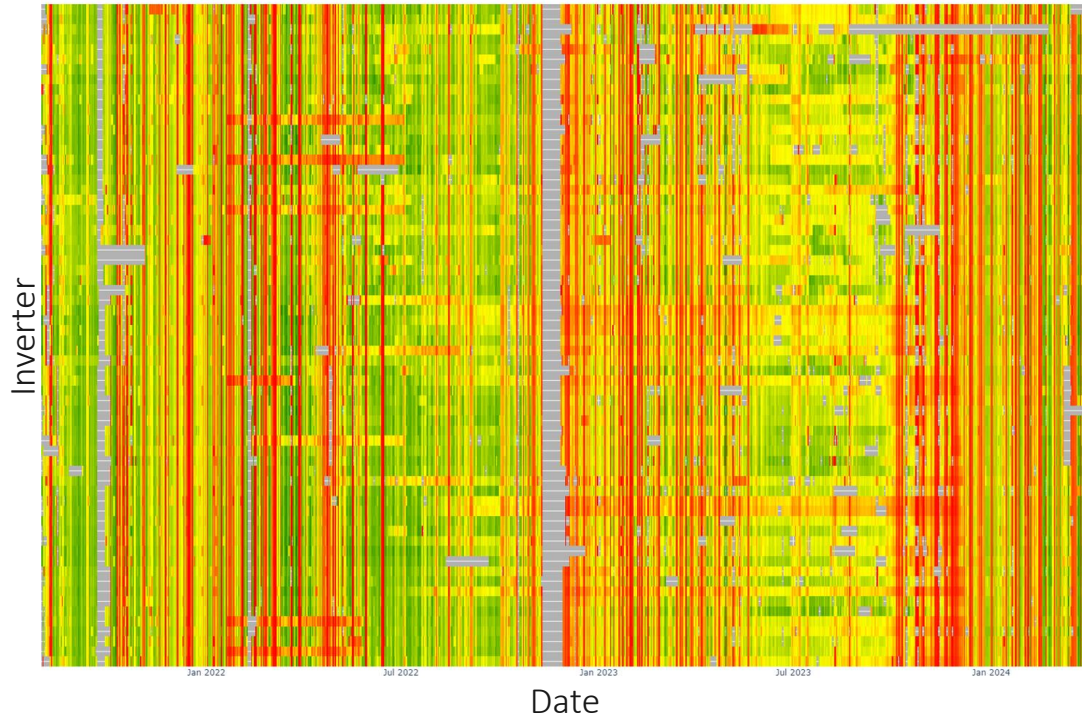
Detectable issues:

- Inverter off / unavailable
- Non-communication
- Low efficiency
- Derating

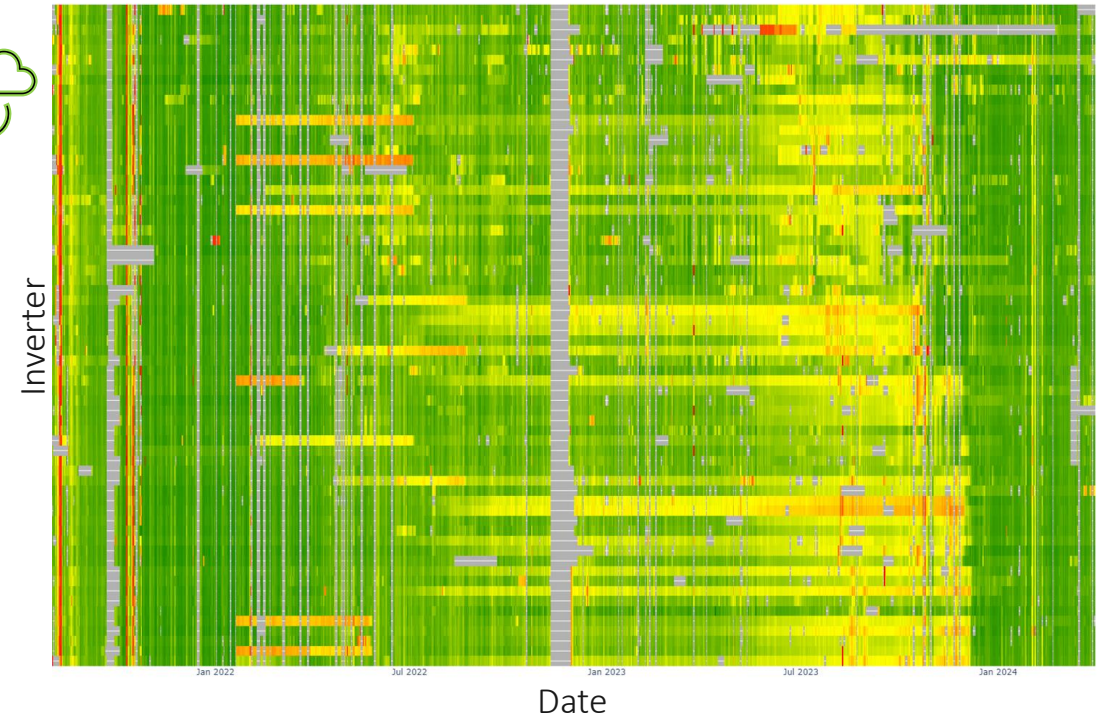
KPIs developed – Performance Ratio (PR)



“Basic” Performance Ratio – Inverter level



“Advanced” Performance Ratio – Inverter level



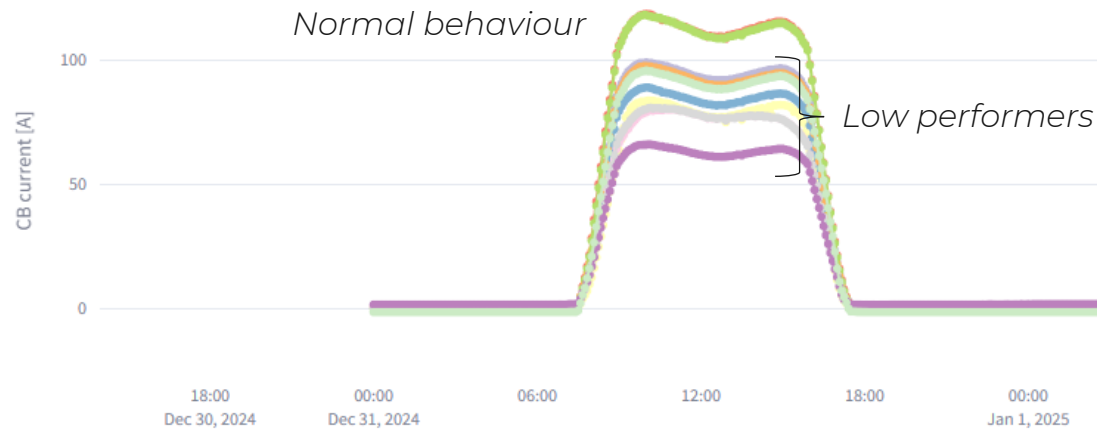
- Any possible underlying issue is masked by a heavy curtailment from the grid operator that artificially reduced the daily standard PR

- Results are clearer and more reliable
- Specific pattern have completely different meaning
- Helpful in understanding the more affected areas

From detailed “top-down” Data Analytics...

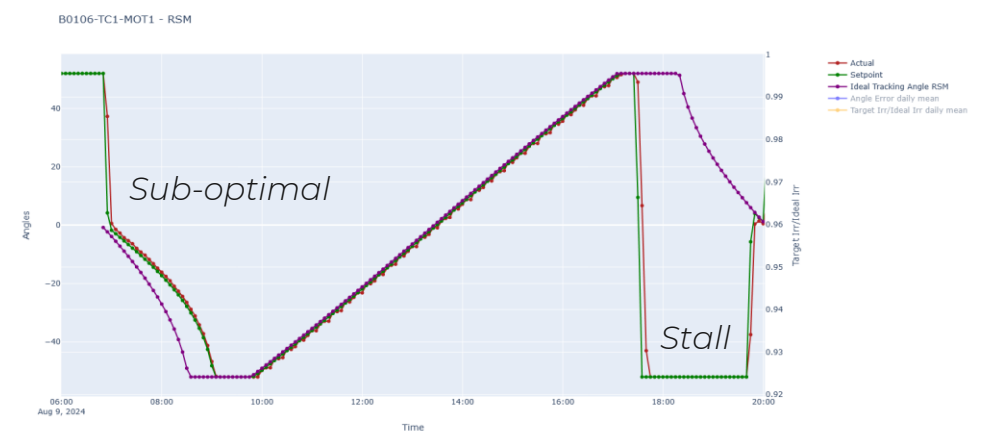


DC power – Combiner Box level



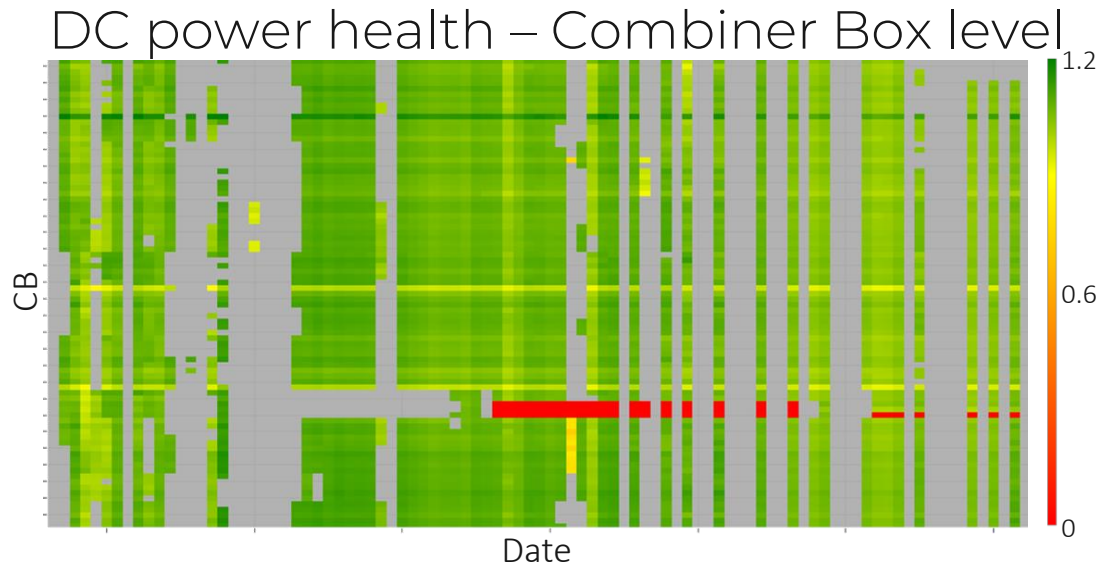
- DC capacity at each Combiner Box (CB) is regressively fitted and normalized to actual weather data.
- A daily peer-to-peer CB comparison identifies underperformers.

Tracker angles – Motor level



- Captured irradiance at each tracker is compared to the optimal under ideal orientation.
- A daily analysis identifies setpoint issues (i.e., software) and actual angle issues (i.e., mechanical).

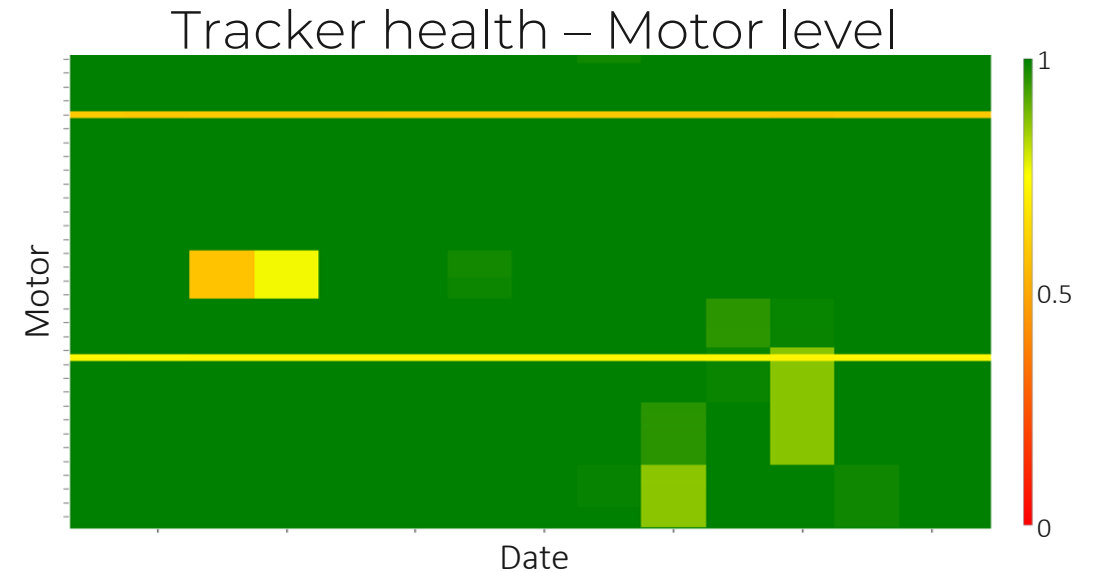
... to “bottom-up” new compact KPIs



Information at the 100s kW scale

Clear anomaly identification:

- CB off (red)
- Disconnected strings (yellow)
- More strings than as-built (dark green)



Information at the 100s kW scale

Clear anomaly identification:

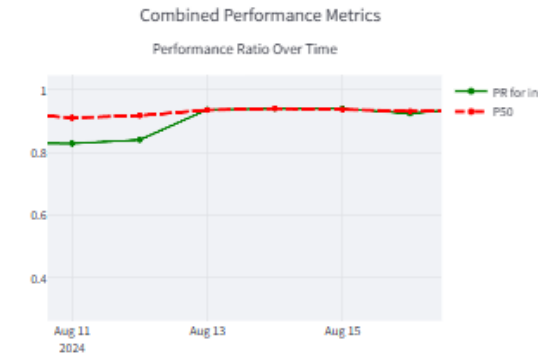
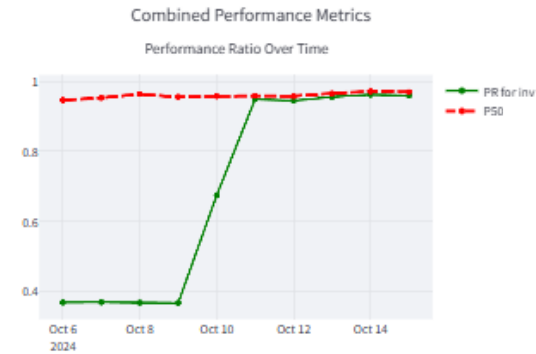
- Stalled trackers (continuous yellow/orange)
- Ongoing maintenance activities (yellow/orange spots)

KPIs cross analysis

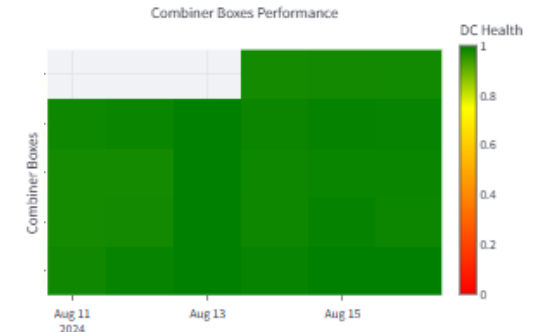
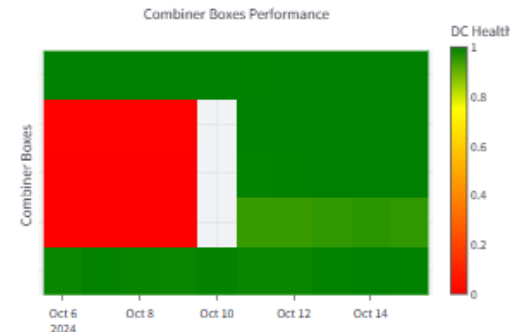
Comparing KPIs at inverter level:

- Immediate imputation of the loss category and identification of the anomalous equipment
- Verification over time of the effectiveness of the corrective actions

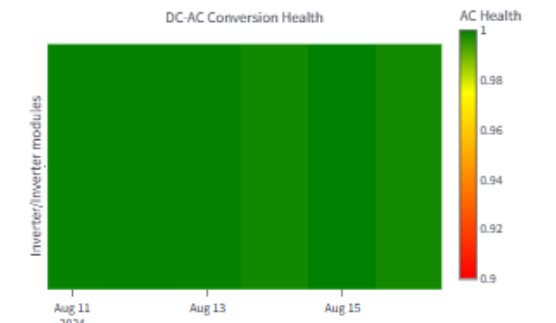
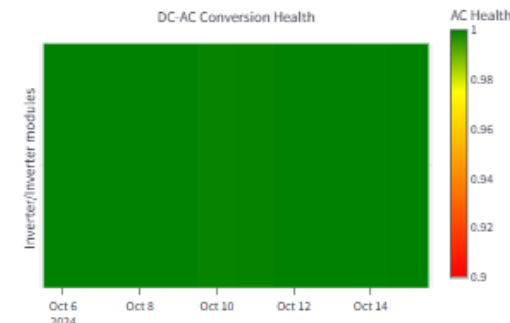
PR



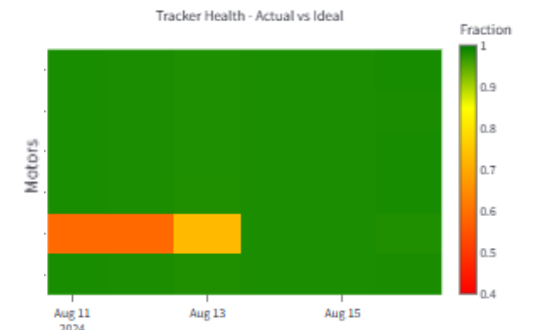
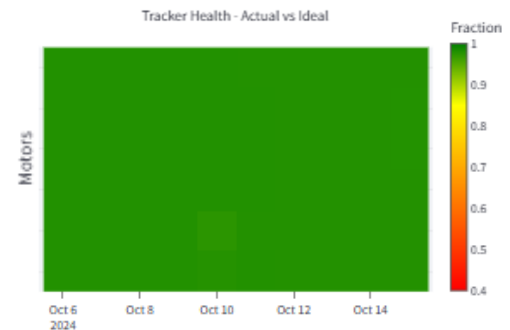
DC health



DC-AC health



Tracker health



Automated Weekly Reporting



Actionable Items

Table

Total actionable capacity: 2.84 MW

	Total capacity loss [MW]	AC health	DC losses	Tracker health
B0402	0.64		CBs with no DC health: B0402-DC4-CB3 - B0402-DC4-CB4 - B0402-DC4-CB5;	B0402-TC1-MOT5 -
B0209	0.38		CBs off: B0209-DC3-CB5 - B0209-DC5-CB3;	
B0311	0.42		CBs with no DC health: B0311-DC4-CB1 - B0311-DC4-CB2 - B0311-DC4-CB3 - B0311-DC4-CB4;	B0311-TC1-MOT3
B0314	0.23		CBs off: B0314-DC1-CB5;	
B0207	0.23		CBs with no DC health: B0207-DC1-CB1 - B0207-DC1-CB2 - B0207-DC1-CB3 - B0207-DC1-CB4;	B0207-TC1-MOT2
B0404	0.09		CBs with possible disconnected strings: B0404-DC5-CB1;	
B0112	0.04		CBs with possible disconnected strings: B0112-DC4-CB2;	
B0104	0.05		CBs with possible disconnected strings: B0104-DC4-CB4;	
B0206	0.04		CBs with possible disconnected strings: B0206-DC2-CB4;	
B0403	0.06		CBs with possible disconnected strings: B0403-DC4-CB1 - B0403-DC4-CB2;	

Inverters off or non communicating for the entire week

No inverters off or non communicating for the entire week.

Combiner boxes with DC health < 10% or Nan

- B0209-DC3-CB5
- B0209-DC5-CB3
- B0314-DC1-CB5

Suggested action: These CBs are most likely disconnected, non-communicating, or with severe issues.

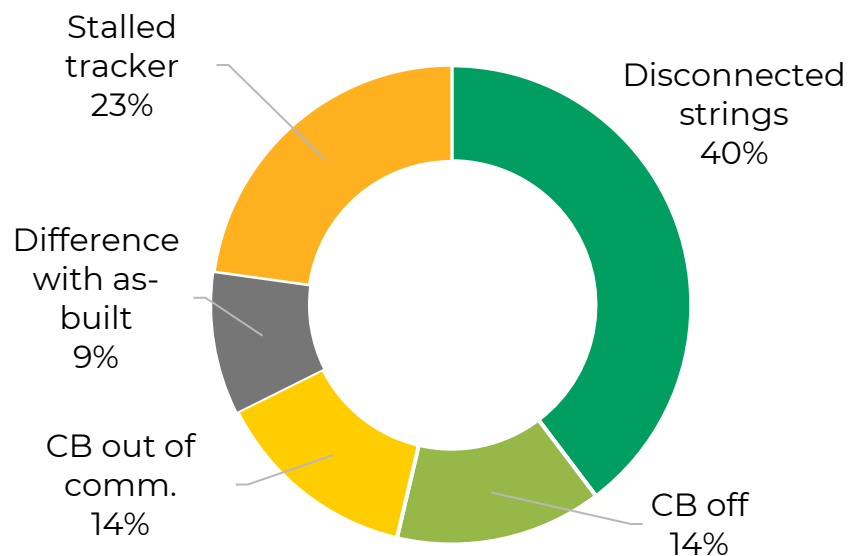
- Every week report only the **outstanding** anomalies.
- **Quantify** the anomalies in terms of potential capacity lost to prioritize the activities.
- Clearly list the **devices** on which the O&M need to focus their inspections.
- Suggest possible **root causes**.
- Interactive and downloadable.
- Faster transition from the analysis of the anomalies to the action on site.

Results

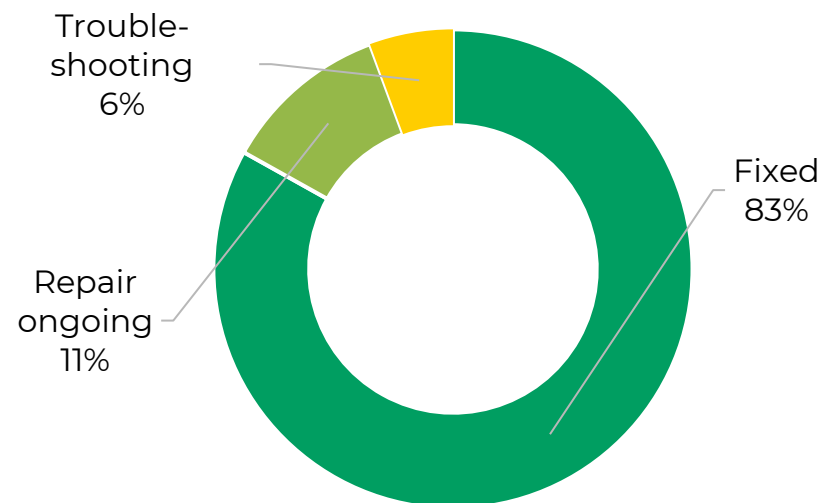


- Platform deployed on 4 assets
- DC Capacity monitored: ~ 650 MW
- Technologies: mono and bifacial; c-Si and CdTe; back-tracking and true-tracking
- Anomalies identified (Nov 2024 - Feb 2025): ~140
- No false positives so far

Anomaly type



Resolution status



Conclusion



- Developed and **validated on ~650 MW** a digital platform designed to enhance the performance monitoring of large-scale PV plants.
- **Preprocessing and data filtering** via physics-driven and data-driven algorithms is key to obtain clear and reliable technical KPIs.
- KPIs designed to **analyze specific devices and phenomena** enabling faster troubleshooting and issue resolution.
- **No false positives** after four months since deployment.



Le nostre iniziative di Open Innovation

Valorizziamo idee, progetti e tecnologie per trovare le migliori soluzioni a livello globale. L'approccio Eni all'Open Innovation prevede attività diversificate con 4 entità:



For more details, please visit

<https://www.eni.com/it-IT/azioni/innovazione-tecnologica-collaborazioni/open-innovation.html>