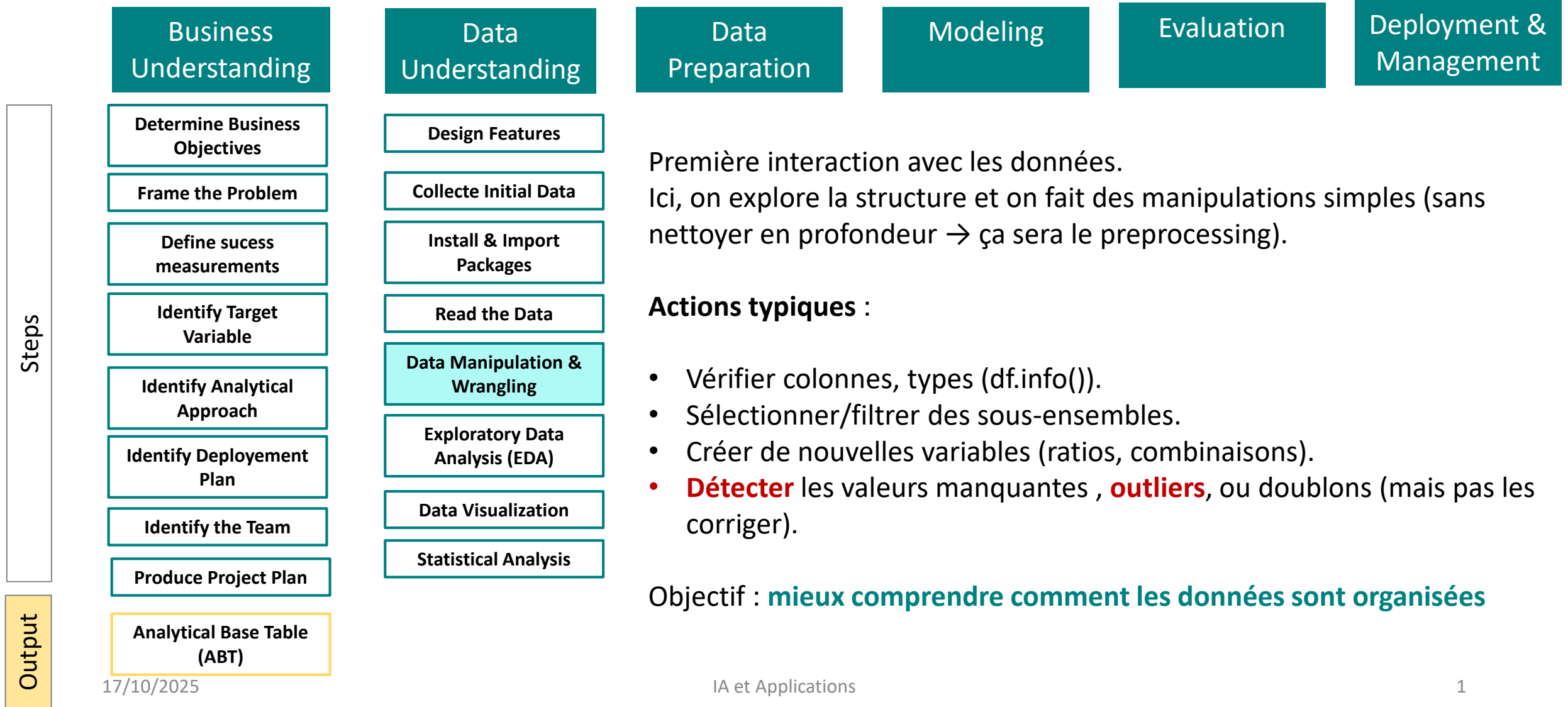


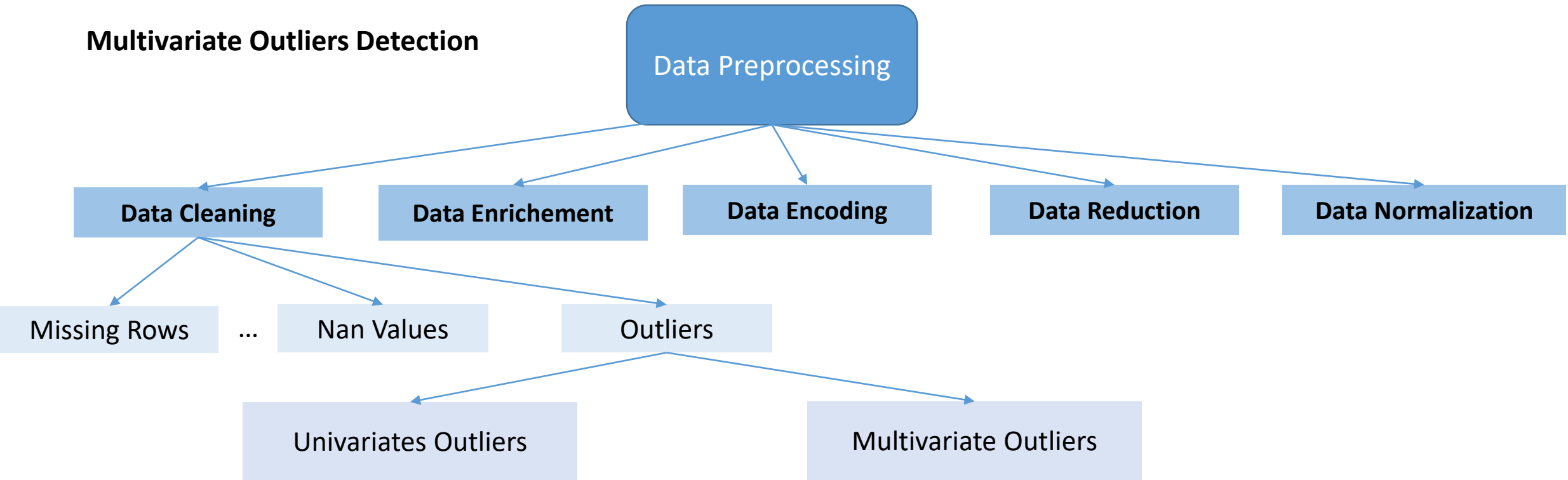
Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)



Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)



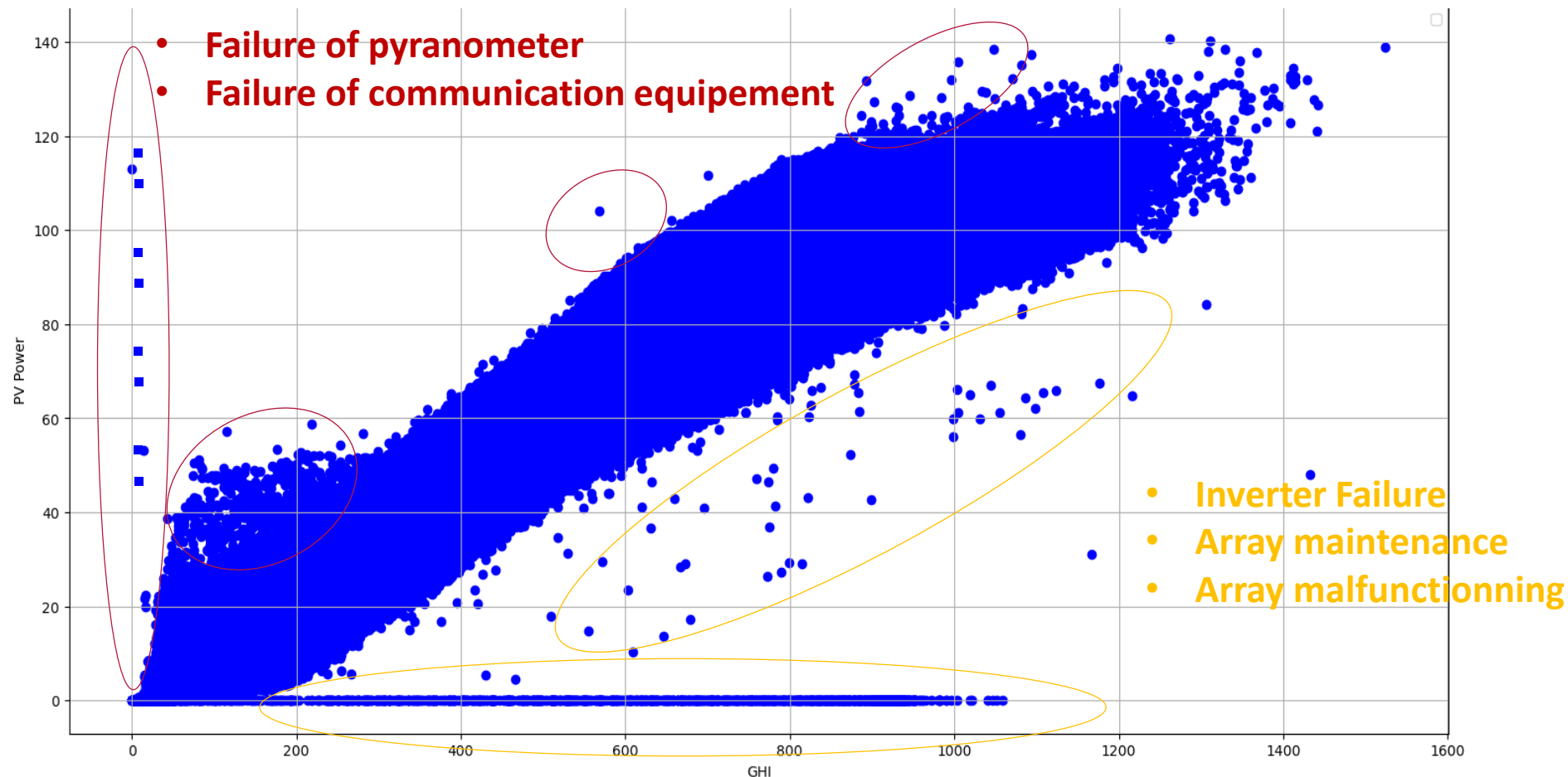
Temperature	-1	-60	10	30	55	25
-------------	----	-----	----	----	----	----

Solar radiation	0	450	1000	54	1000	15
PV Power	100	4.9	100.1	5.6	0	0

Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection



Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

SOA

- DBSCAN

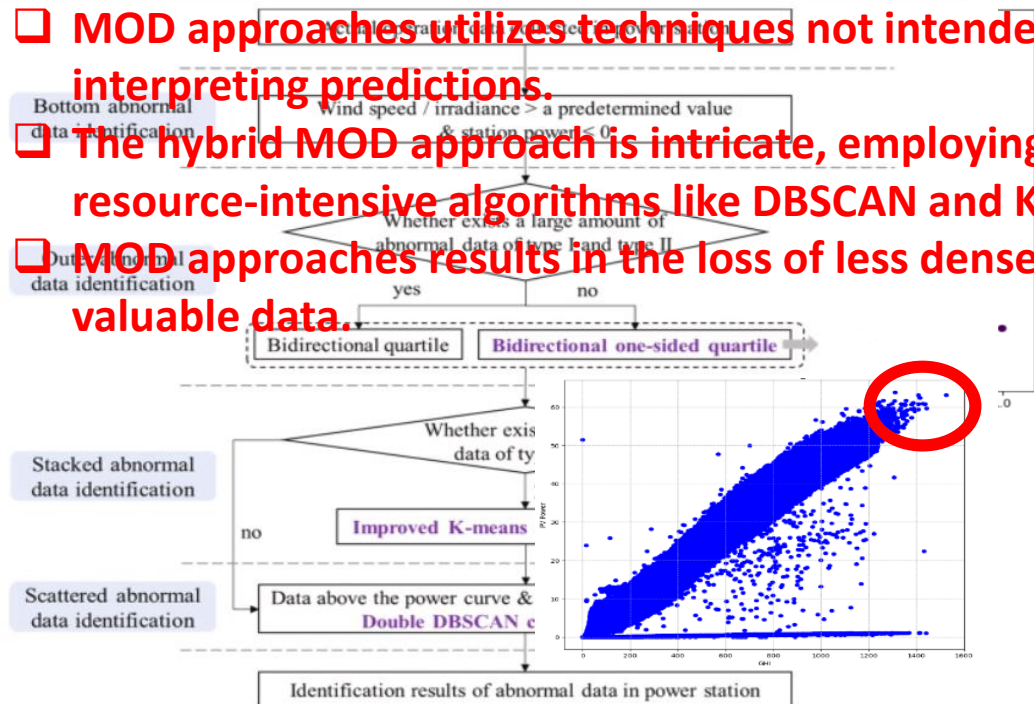
- Hybrid Approach

This approach relies on **density** to distinguish between **genuine data** points clustered closely together and **outliers** that are **sparsely distributed**. It employs a combination of multiple methods to identify a wide range of outliers.

❑ MOD approaches utilizes techniques not intended for interpreting predictions.

❑ The hybrid MOD approach is intricate, employing resource-intensive algorithms like DBSCAN and K-means.

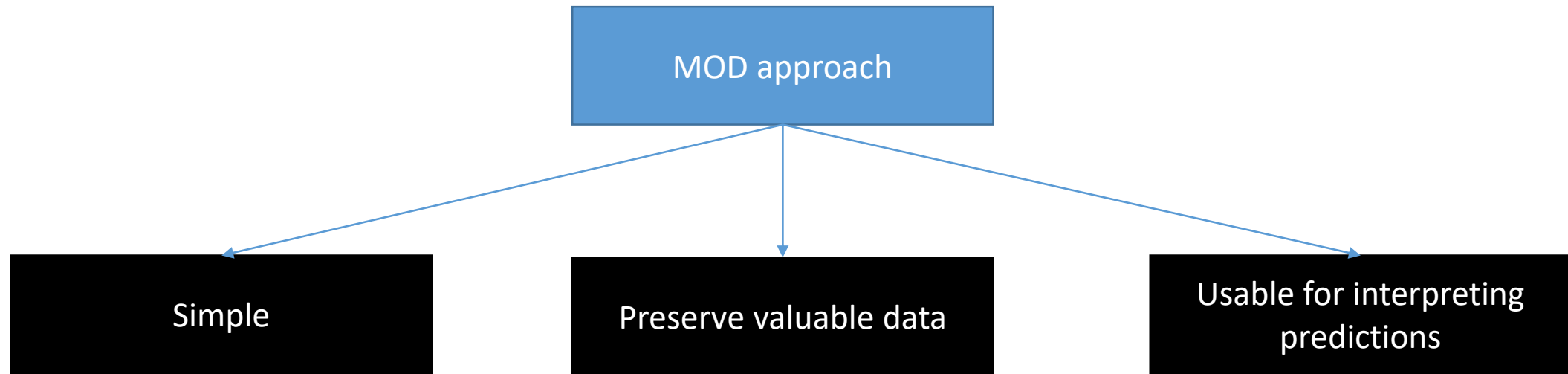
❑ MOD approaches results in the loss of less dense but valuable data.



Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

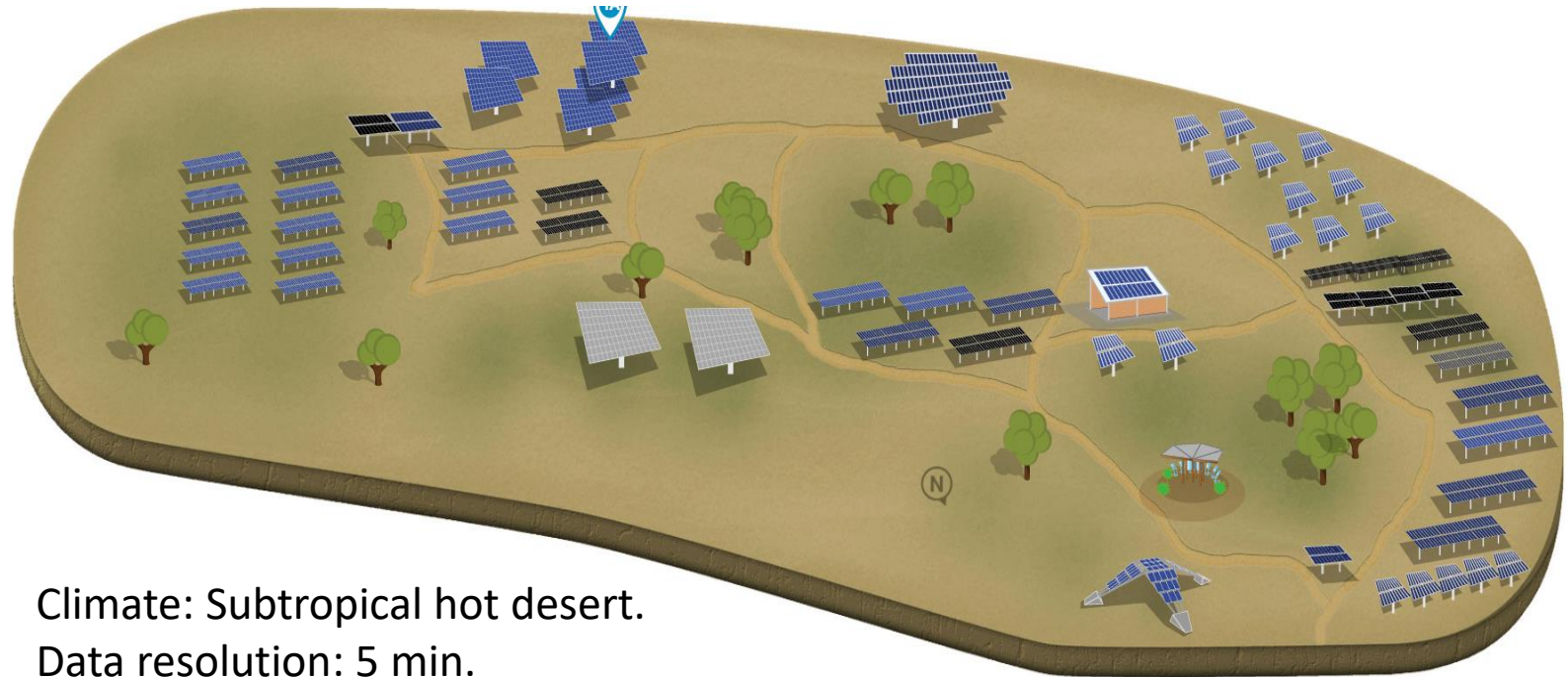


Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

Location: **Alice Springs, Northern Territory, Australia**

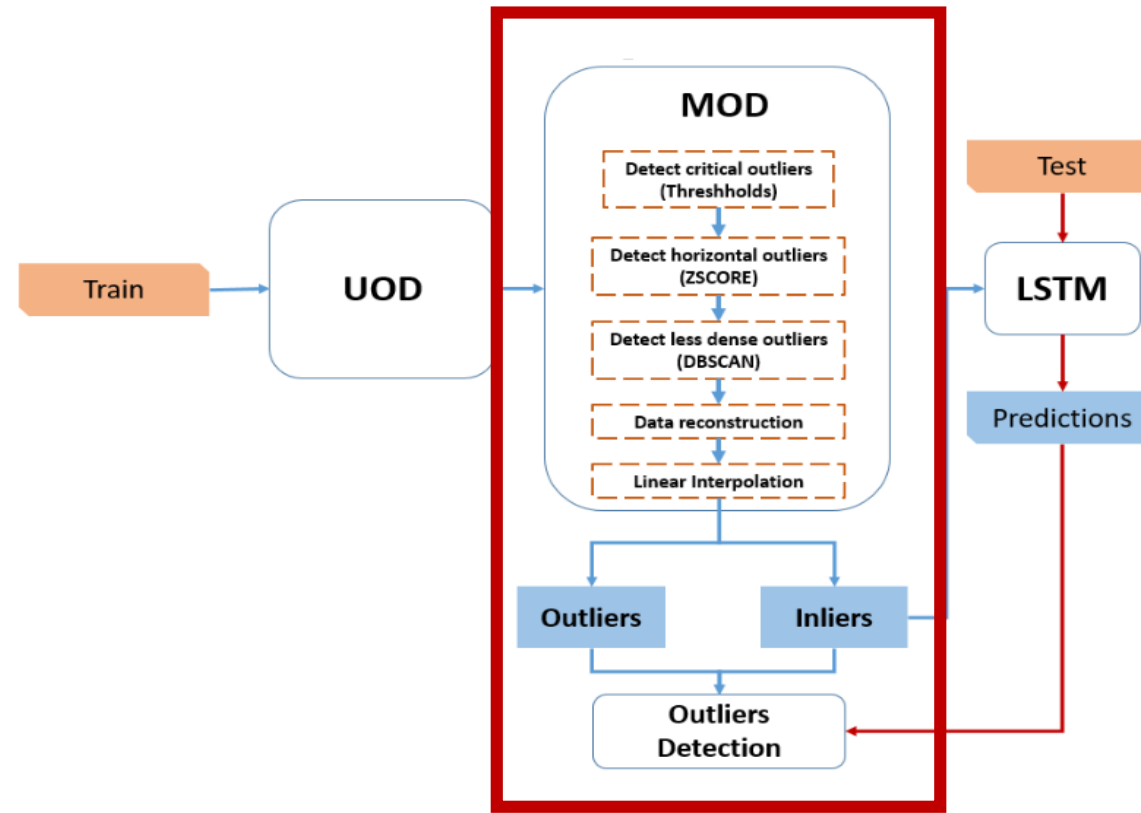


- Climate: Subtropical hot desert.
- Data resolution: 5 min.
- Samples: 55 M.
- Array technologies: 7 technologies.
- Number of arrays: 39.
- Data period: 04-2008 to Now

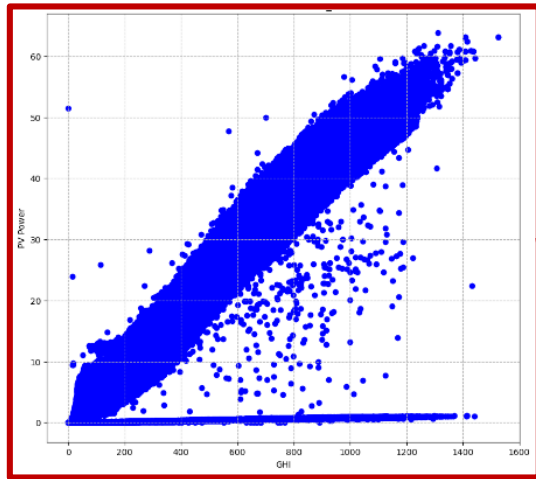
Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

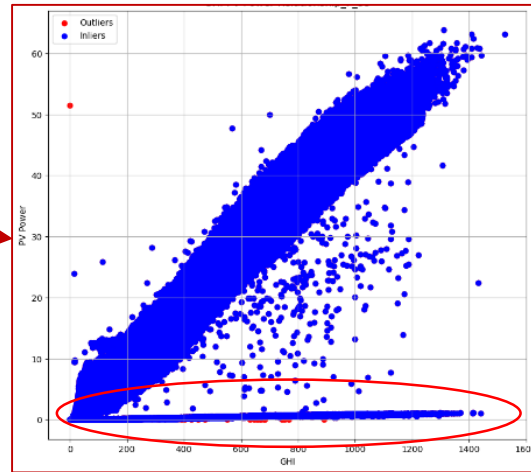
Multivariate Outliers Detection



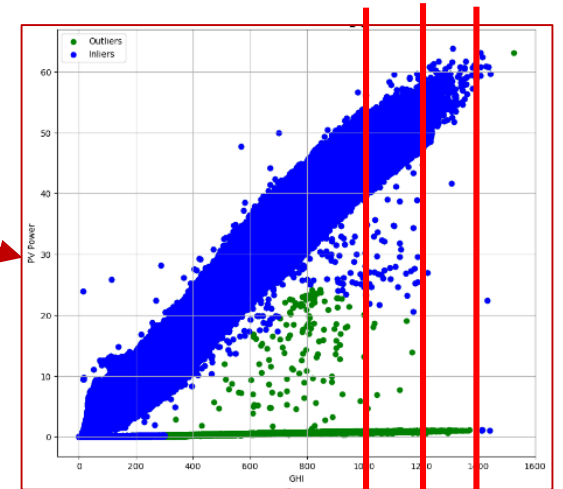
Multivariate Outliers Detection



Detect critical outliers (Thresholds)



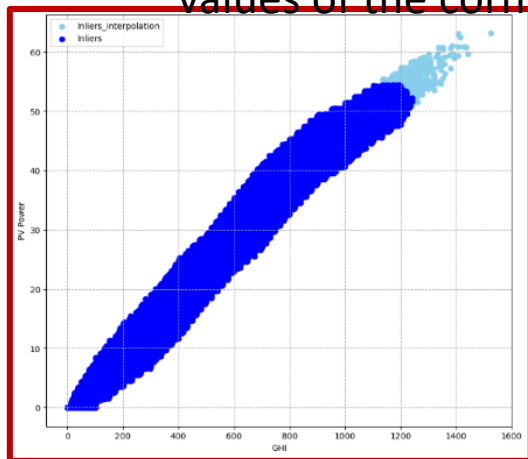
Detect horizontal outliers (ZSCORE)



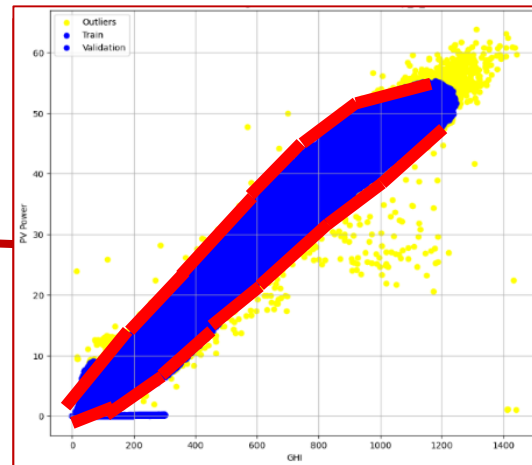
Detect rows where:

1. Sliding the graph into windows of 100 W/m^2 of GHI. Beyond 1000 W/m^2 of GHI, the upper and lower bounds of the graph maintained consistent slopes.
2. Interpolating the two minimum and two maximum values of the corners of each interval

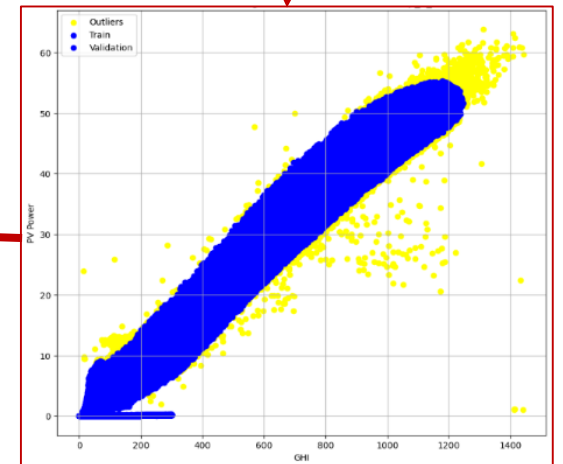
1. Segment the data into windows of 100 GHI units
2. Calculate Z-scores for each point
3. Consider points with absolute Z-scores lower than $Q3 + 1.5 * QR$ of absolute Z-scores as outliers



Linear Interpolation

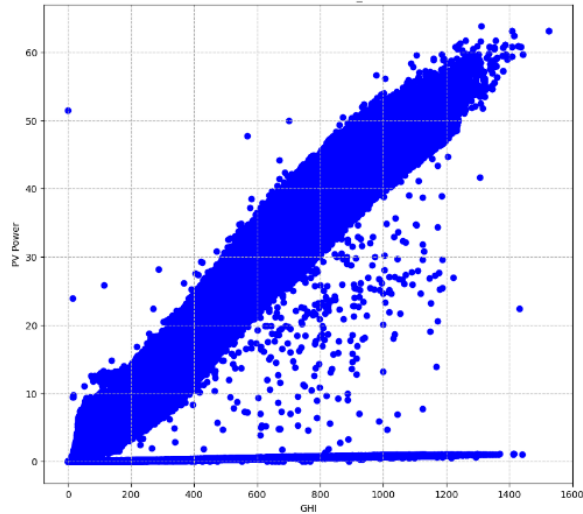


Data reconstruction

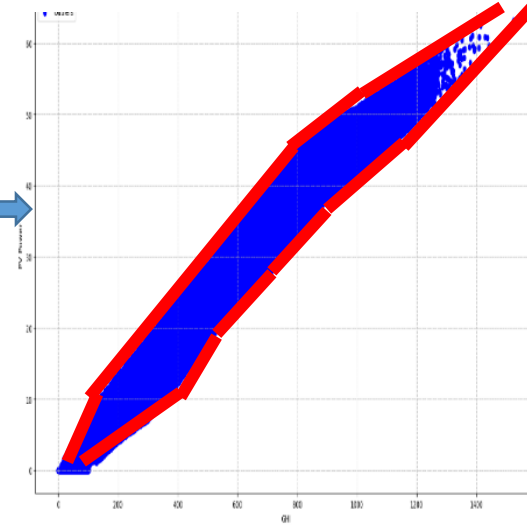


Multivariate Outliers Detection

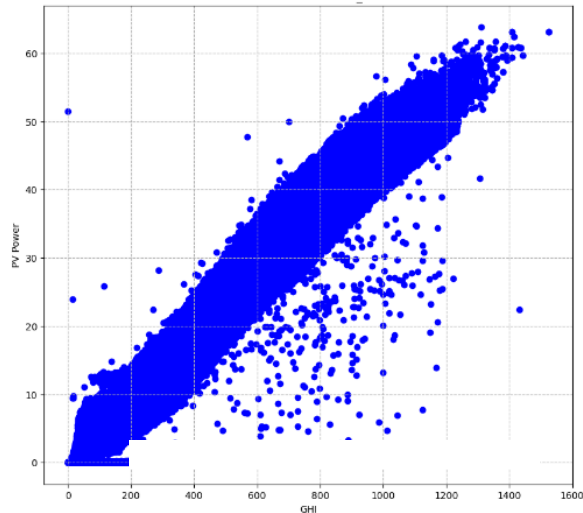
Train Data



MOD



Test Data



The proposer MOD is:

- ☐ Simple.
- ☐ It preserves the majority of valuable data.
- ☐ It is used to interpret predictions.

Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

Ce que vous allez produire?

Un **code fonctionnel** de la méthode MOD

Le Gain:

- Le MOD est une composante du pipeline de prédiction PV.
- Le meilleur code sera intégré comme bibliothèque Python (PyPi).
- L'étudiant retenus sera co-auteurs d'un article scientifique issu du travail.

Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

Critères d'évaluation du meilleur code

1. Performance technique

Critère	Description
Temps d'inférence	Rapidité d'exécution du code pour de nouvelles données
Temps d'entraînement	Optimisation du traitement global
Efficacité	Pourcentage d'inliers et d'outliers détectés correctement (sur chaque array et globalement)

Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

Critères d'évaluation du meilleur code

2. Qualité scientifique

Critère	Description
Généralisation	Testé et performant sur d'autres bases de données
Production graphique	Génération automatique des graphes de détection (GHI vs PV power)
Indicateurs produits	Index des outliers, résumé statistique clair

Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

🏆 Critères d'évaluation du meilleur code

3. Qualité logicielle

Critère	Description
Documentation	Code bien commenté, clair et reproductible, nom significatif de la bibliothèque
Structure	Modularité (fonctions, paramètres configurables)
Bibliothèques	Utilisation judicieuse (pas de dépendances inutiles)

Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

Critères d'évaluation du meilleur code

4. Innovation & Explicabilité

Critère	Description
Amélioration du MOD	Ajout d'une étape ou d'un paramètre pertinent
Détection d'outliers nouveaux	Capacité à identifier de nouveaux points non valides dans des jeux non vus
Interprétabilité	Visualisation claire et explication du comportement du modèle

Méthodologie de développement de projets d'IA

Cross-Industry Standard Process for Data Mining (CRISP-DM)

Multivariate Outliers Detection

🏆 Critères d'évaluation du meilleur code

5. Temps

