

## Introduction

- Nous avons choisi de travailler avec les données de la société **Engie**, à partir du travail de notre projet fil rouge. Ces données représentent les **séries temporelles de la production d'électricité solaire** de la ferme de Blond en Haute-Vienne (Figure 1).
- La ferme solaire est composée d'environ **25 000 panneaux photovoltaïques**, dont la production est regroupée en **8 zones** indépendantes. Les données principales correspondent à :
  - La **production électrique** de chaque zone toutes les minutes de 4h à 20h, du 31 mai 2017 au 4 novembre 2018, comme illustrée lors d'une journée typique sur la Figure 2.
  - L'**irradiance**, qui mesure l'énergie du rayonnement solaire.



Fig. 1: La ferme solaire de Blond

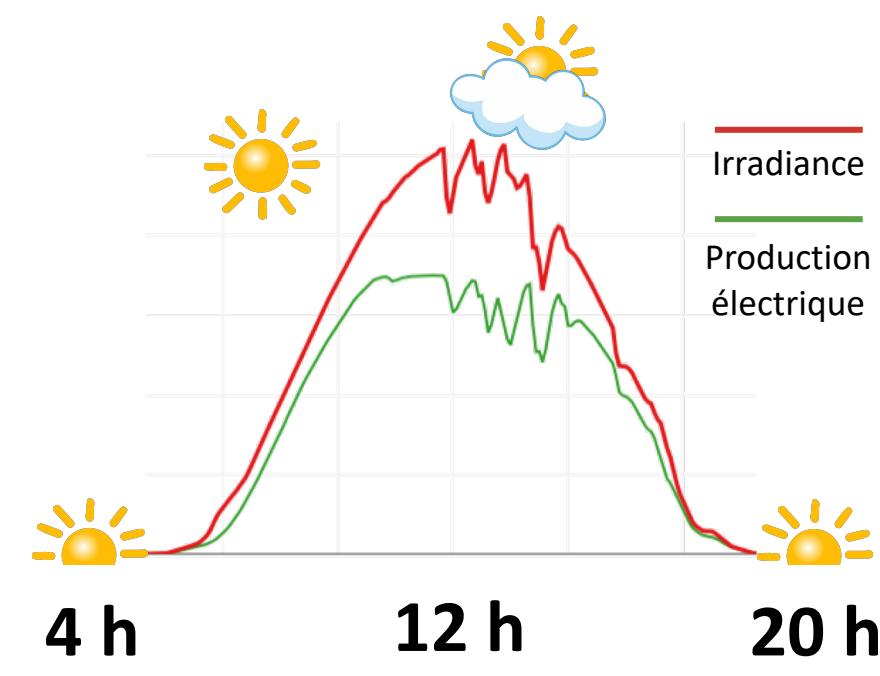


Fig. 2: Une journée typique d'irradiance et de production électrique solaire

## Objectifs

- Créer un outil d'analyse pour les **data scientists**.
- **3 niveaux de visualisation** pour :
  - Faciliter l'analyse de l'historique
  - Évaluer et comparer les sous-performances
  - Évaluer la détection d'anomalies

➤ **Technologies :**

**Auteurs :** Alba Ordoñez, Charles Théron, Ioan Catana, Karine Pétrus, Stéphane Mulard

## 01 Explorer l'historique de la production

### Analyse de l'historique en hiver

- Présence de 5 jours consécutifs d'anomalies (6 au 10 février) pour toutes les zones.
- Corrélation de ces anomalies à la présence de neige sur les panneaux après vérification de la météo.
- Production d'énergie qui augmente graduellement en avançant vers le mois d'avril où les journées sont de plus en plus ensoleillées.



**1** Sélection d'une zone de la ferme et mise à jour de l'historique de production et de la description sur la zone choisie.

**2** Sélection d'une période de temps de 2 mois et mise à jour de l'historique de production sur la période choisie.

**3** Bouger le curseur de gauche à droite permet de naviguer à travers les jours de la période sélectionnée.

**4** Sélection d'un jour d'historique et mise à jour de la description des caractéristiques de la journée quant à la puissance mesurée et à la présence / type d'anomalie.

## 02 Comparer les performances des 8 zones

### Sous-performances récurrentes

- Ces graphiques ont permis de détecter une période de sous-performance significative pour la zone 2\_2 la semaine du 30 août 2018.
- Comme on peut le voir sur le détail de la production du 2 septembre, cette zone affiche une sous-performance tout au long de la journée par rapport aux autres zones, qui ne s'explique pas par un phénomène météorologique.



**1** Sélection d'une ou plusieurs zones de la ferme et mise à jour de la production globale hebdomadaire, de la production par jour sur une semaine et de la production totale par zone sur la période sélectionnée.

**2** Sélection d'une période de temps et mise à jour des graphiques de la production hebdomadaire, quotidienne et de la production totale par zone sur la période.

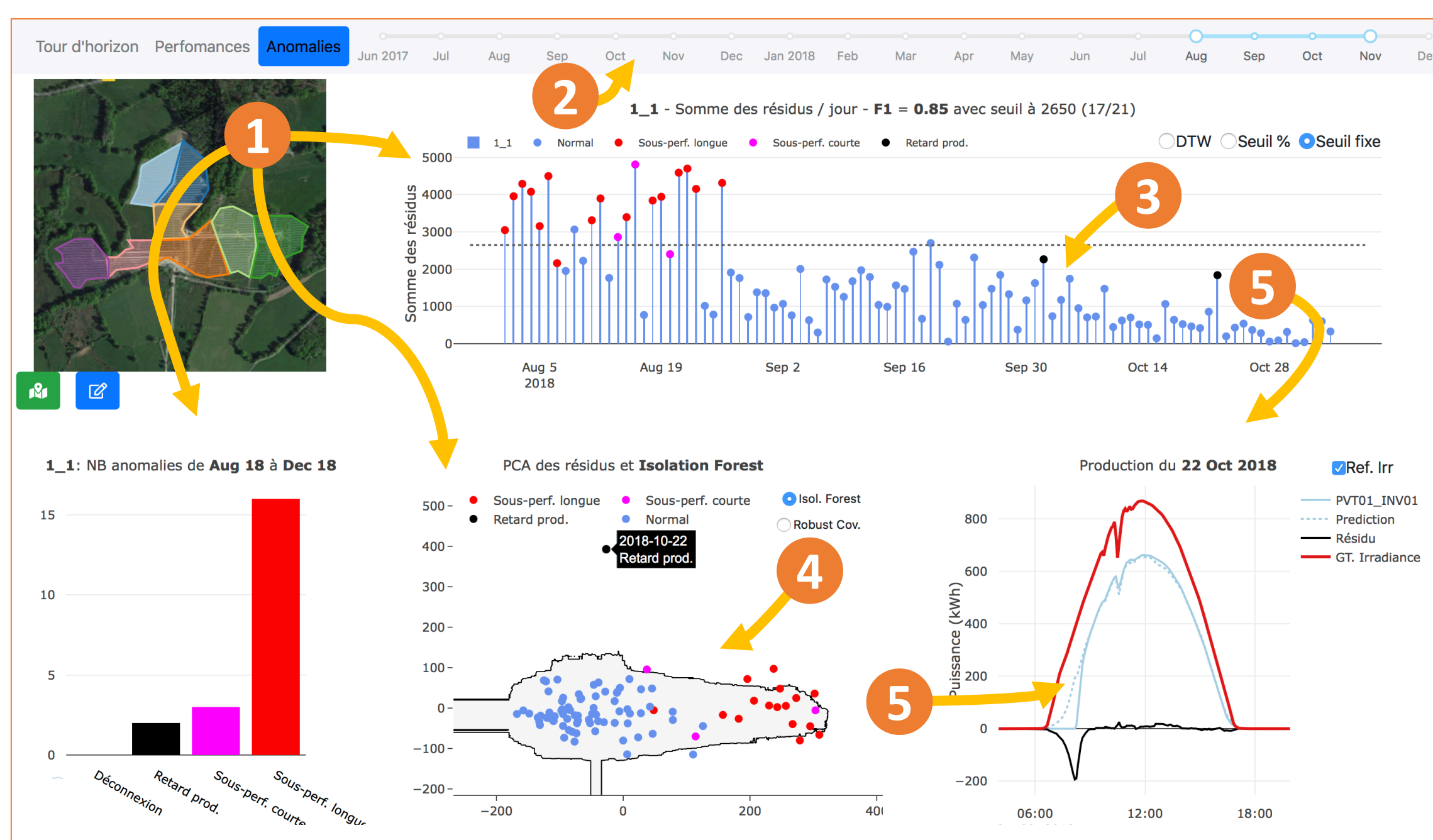
**3** Sélection d'un « point semaine » et mise à jour de la production hebdomadaire sur la semaine et du détail de la production du premier jour de la semaine.

**4** Sélection d'un « point jour » mise à jour du détail de la production pour cette journée.

## 03 Explorer les jours en sous-performance

### Limites des méthodes par seuil

- Sur cet exemple, le seuil fixe absolu donne un F1-Score de 0.85, ce qui est bon malgré 4 faux négatifs, avec 2 retards de production les 2 et 22 octobre dont le détail est présenté en bas à droite.
- L'algorithme d'Isolation Forest montre des limites pour la séparation des points sur cet exemple, mais le retard de production du 2 octobre est bien identifié.



**1** Sélection d'une zone et mise à jour de la somme des résidus, de la distribution d'anomalies et du nuage de points obtenu par PCA des résidus sur la période choisie.

**2** Sélection d'une période de temps et mise à jour de la somme des résidus de la distribution d'anomalies et du nuage de points obtenu par PCA des résidus.

**3** Sélection de la méthode de seuillage pour détecter les anomalies. Les jours situés au-dessus du seuil sont les anomalies détectées, les jours avec des pastilles colorées représentent les vraies anomalies.

**4** Sélection de l'algorithme de clustering et mise à jour du contour de détection des anomalies.

**5** Sélection d'un « point jour » mise à jour du détail de la production de ce jour.