

Projet sur les énergies solaires et éoliennes

Ministère de la Transition écologique et solidaire



Olivier Brunet - Présentation du 09/01/2020

- | | | |
|------------------|-----|---------------------------|
| - Hamza XXX | XXX | Data Analyst |
| - Sabrina XXX | XXX | Data Analyst |
| - Flora XXX | XXX | Data Analyst |
| - Olivier BRUNET | XXX | Data Engineer / Scientist |



- ESN / cabinet de conseil et d'ingénierie
- Fondé en 2002, certifié ISO 27001
- ~500 collaborateurs (forte croissance)
- C.A 2019 ~50 M€
- 4 cœurs de métier : l'Infra, la MOE - la AMOA/MOA - la B.I
- Secteurs d'activités variés : la santé, l'agriculture, le retail, la finance & l'industrie.
- Un acteur Datacentric
- Smart Lab dédié à l'innovation

1. **Contexte & enjeux**
 - *Contexte, enjeux liées aux EnR*
 - *Besoins & uses cases finaux*
 - *Etapes du projet*
2. **Organisation mise en place**
 - *Travail collaboratif / outils*
 - *Mode agile – framework scrum*
3. **Réponse technique – Big Data**
 - *Méthodologie*
 - *Sources de données*
 - *Architecture retenue*
 - *Exploratory Data Analysis*
 - *Machine Learning & prédictions*
4. **Résultats et conclusions**
 - *Perspectives / préconisations*

1. Contexte et enjeux



1. Introduction : contexte & enjeux

- La transformation de l'énergie : **secteur stratégique**
- La **transition écologique** : enjeu majeur, comment passer :
 - d'une énergie fossile aux EnR,
 - d'une production centralisée à un système décentralisé
- Les EnR :
 - proviennent de phénomènes naturels,
 - se renouvellent assez rapidement pour être considérées comme inépuisables à l'échelle du temps humain.
- **Projet : *prévisions du rendement des installations photovoltaïques (PV) & solaires existantes***

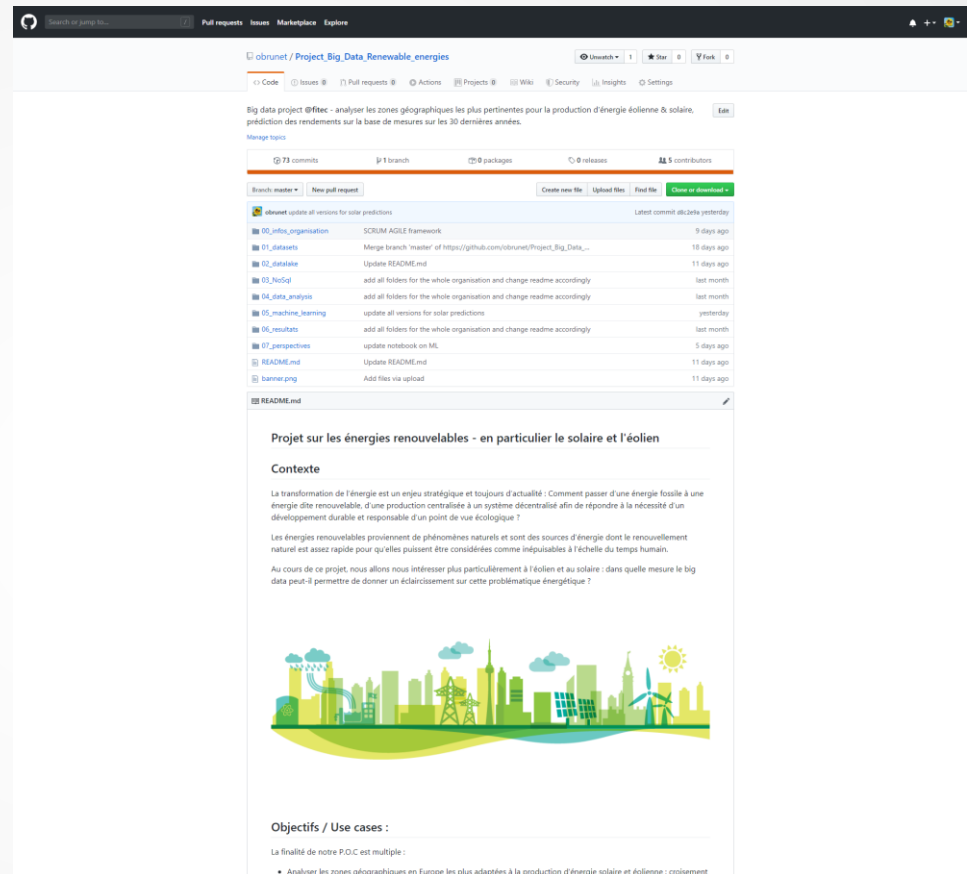
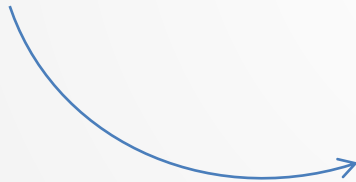
1. Introduction : besoins & uses cases finaux

- Use cases « business » :
 - **Anticiper les rendements** pour investir intelligemment
 - Développer la part des EnR → Union Européenne
 - Ajuster plus finement le reste de la production → EDF
 - Équilibrer l'offre & la demande pour faciliter la gestion du réseau de distribution → RTE
 - **Diminuer le cout de l'énergie** → particuliers & PME / PMI

2. Organisation mise en place

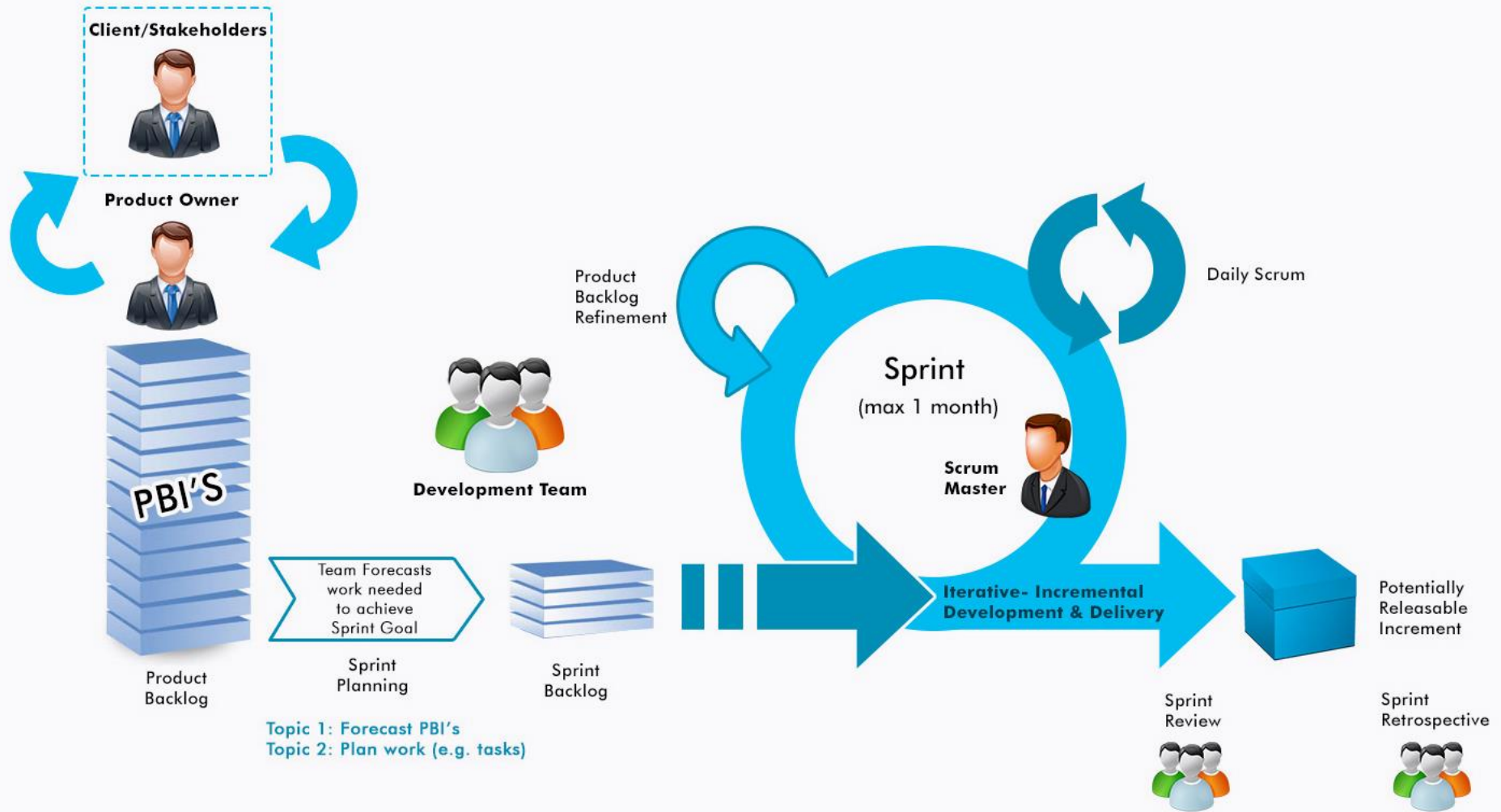


2. Organisation : travail collaboratif



github.com/obrune/Project_Big_Data_Renewable_energies

2. Organisation : mode agile

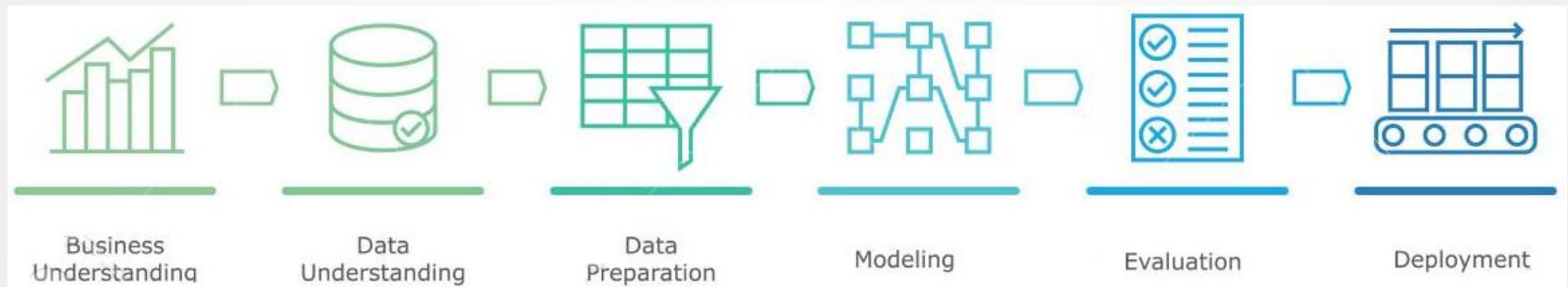


3. Réponse technique / Big Data



3. Réponse technique : méthodologie

1. **Cerner le problème** et rechercher une vision d'ensemble.
2. **Récupérer** et **analyser** les données → mieux les comprendre.
 - *Clustering (apprentissage non supervisé)*
3. **Préparer** les données pour mieux exposer leurs structures sous jacentes aux modèles de ML et les **ingérer**.
4. Essayer **plusieurs modèles différents** → short list.
 - *Prévision de valeurs de rendement dans le temps*
 - *Régression - apprentissage supervisé*
 - « *Metric* » : RMSE
5. **Régler finement** les modèles → solution performante.
6. **Présenter** la solution.
7. **Déployer**, surveiller et **maintenir**



3. Réponse technique : sources des données

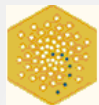


SETIS

Strategic Energy Technologies Information System



THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP



European Climate Assessment & Dataset

ECA&D

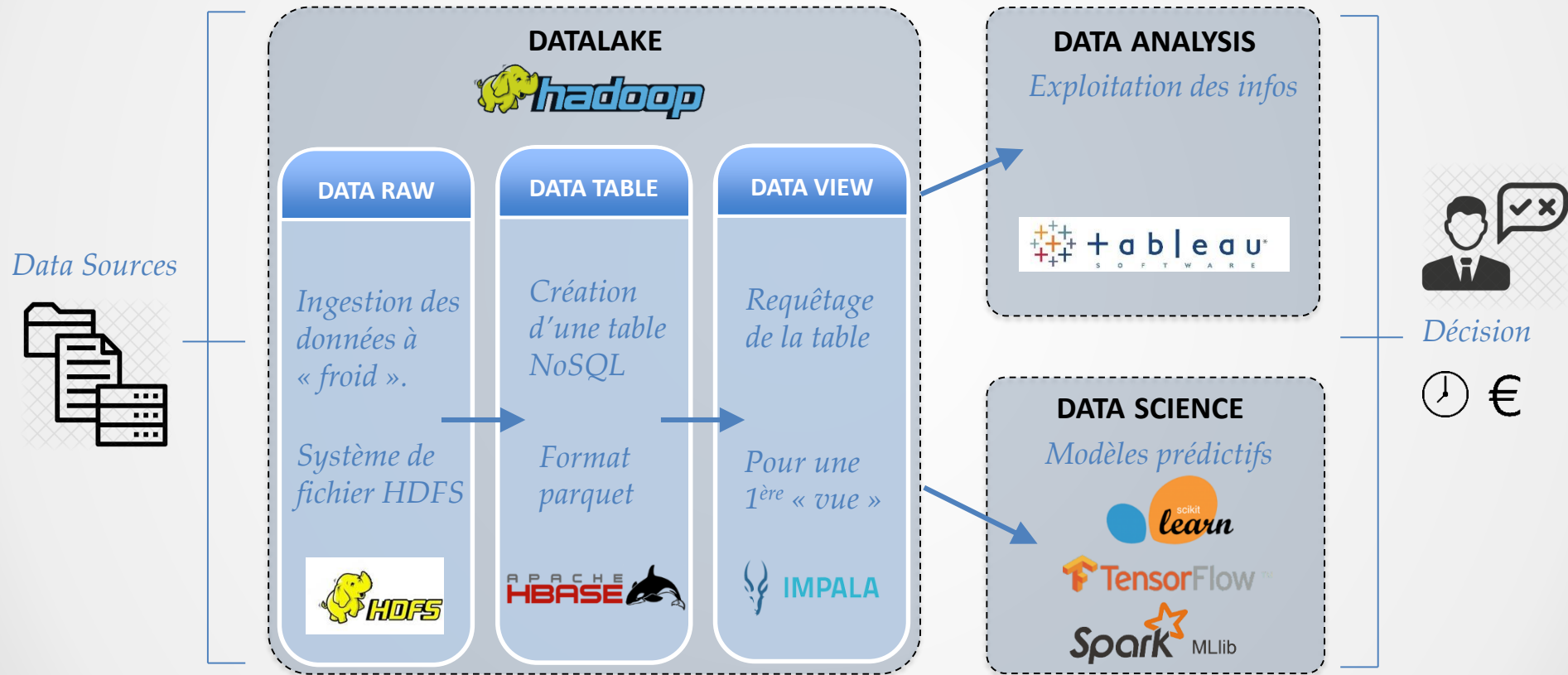


Our World In Data

Sources	Contenu	Type	Format	Licence
Kaggle & SETIS	Rendement des installations PV & solaires - historique sur 30 ans, heure par heure https://setis.ec.europa.eu/	Structuré	csv	CC
European Climate Assessment	Mesures météo extrêmes ainsi que des infos journalières https://www.ecad.eu/	Structuré	csv	spécifique
The World Bank	Indicateurs énergétiques pour chaque pays, sur 50 ans https://data.worldbank.org/	Structuré	csv	CC
Open Power Data System	Plateforme pour la modélisation de système de production d'énergie https://open-power-system-data.org/	Structuré	csv	CC
Web scraping de site météo	La météo en temps réel & historique https://www.infoclimat.fr/	Structuré	csv	?
API REST open data réseaux énergies	Données du gouvernement : production, infrastructure, consommation, marchés, météorologie... https://opendata.reseaux-energies.fr/	Structuré	json	CC
Our World in Data	Données collectées de différentes sources et très complètes : production, consommation, par pays et récentes. https://ourworldindata.org/renewable-energy	Structuré	csv	CC

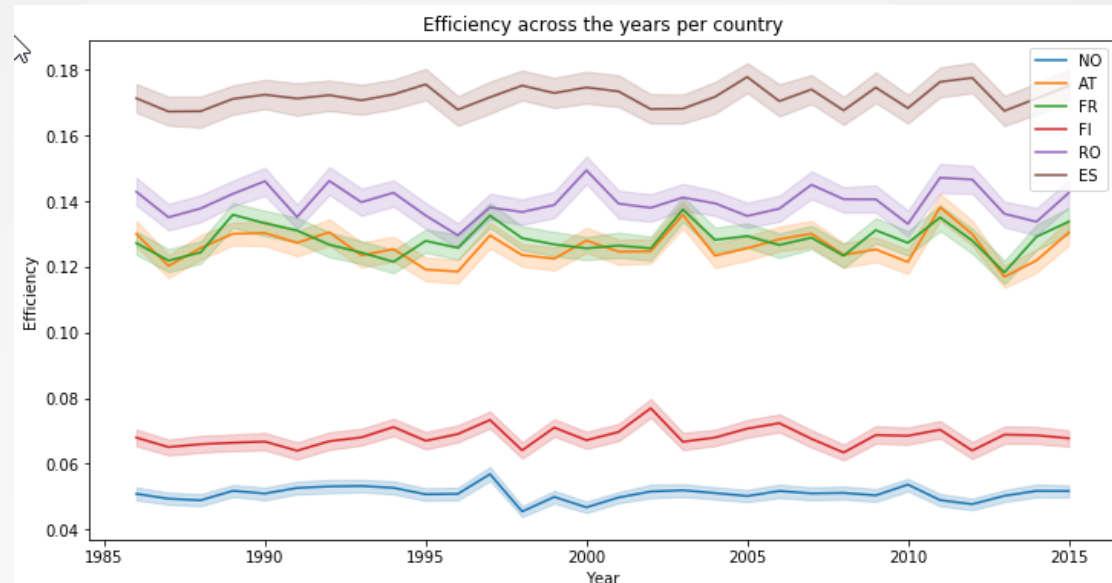
3. Réponse technique : architecture big data retenue

Batch processing avec la distribution CDH **cloudera**

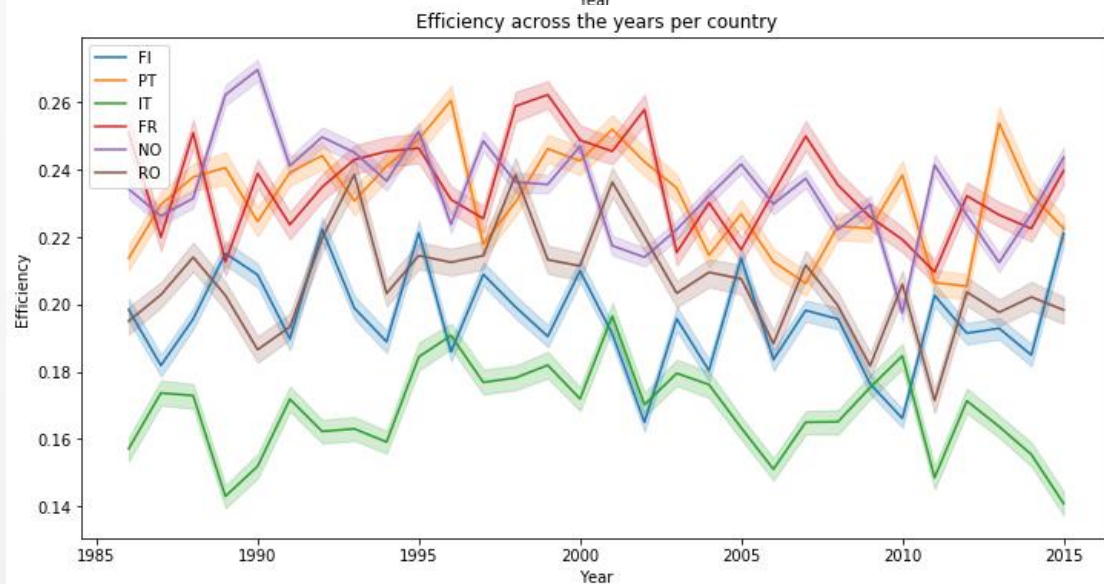


3. Réponse technique : E.D.A / comparaison des rendements moyens annuels

SOLAIRE
Relativement stable

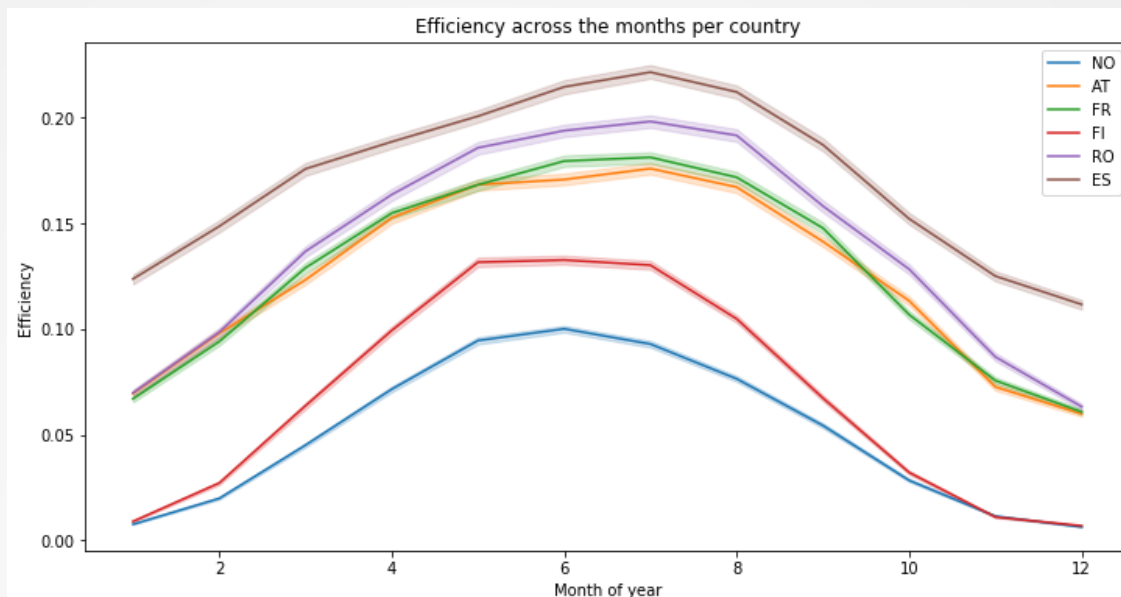


EOLIEN
Rendements variables

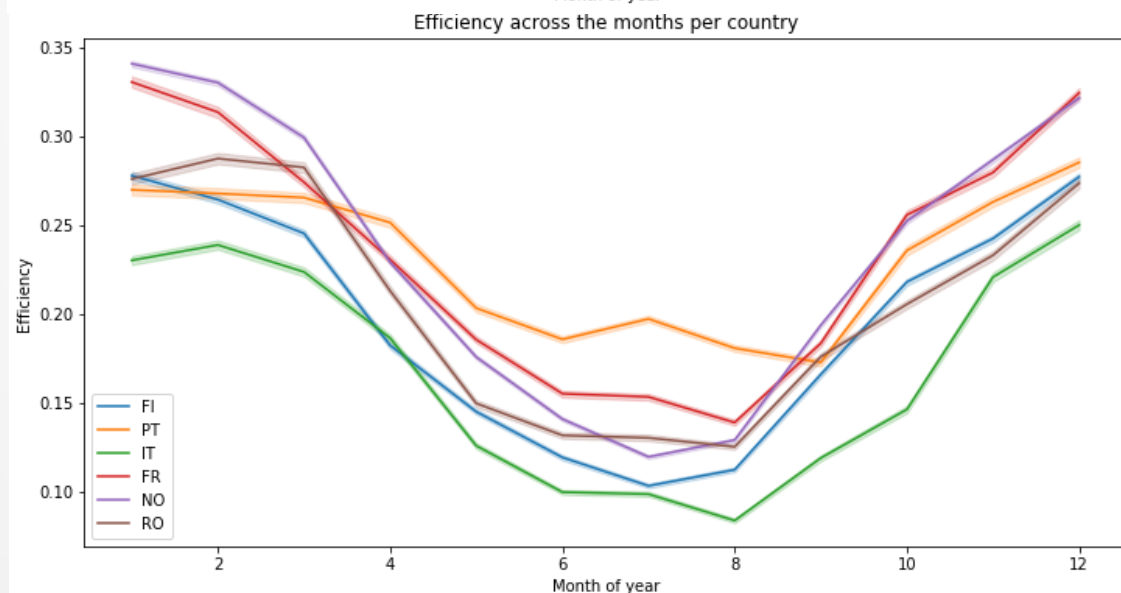


3. Réponse technique : E.D.A / comparaison des rendements moyens mensuels

SOLAIRE
Meilleurs rendements
en été

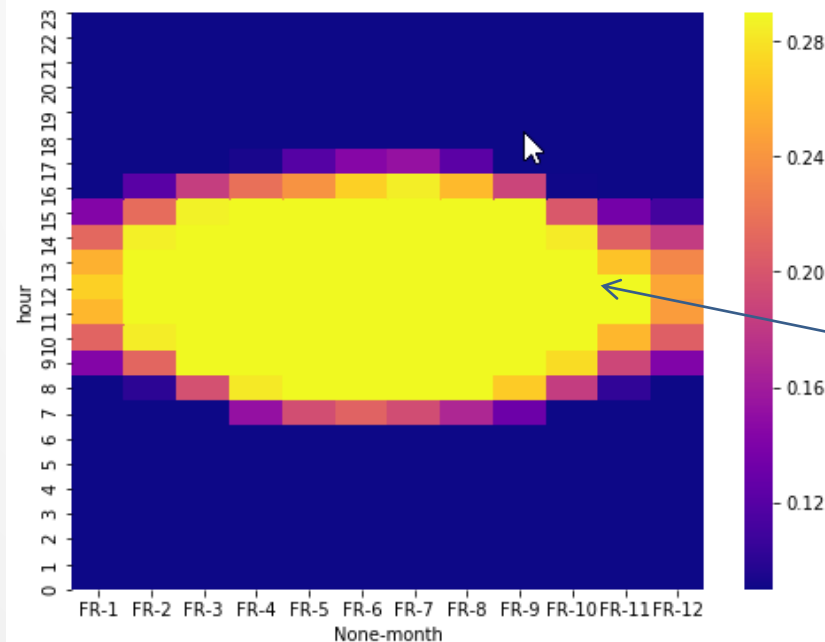
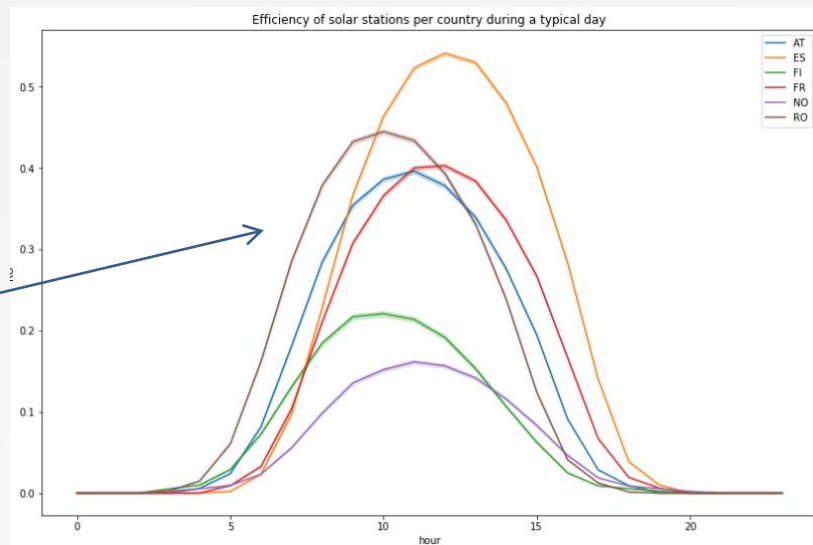


EOLIEN
Meilleurs rendements
en hiver



3. Réponse technique : focus sur les installations photovoltaïques (P.V)

Pics de rendement en fonction des heures de la journée selon les pays



Heures de la journée présentant le plus de rayonnement en fonction des mois de l'année.

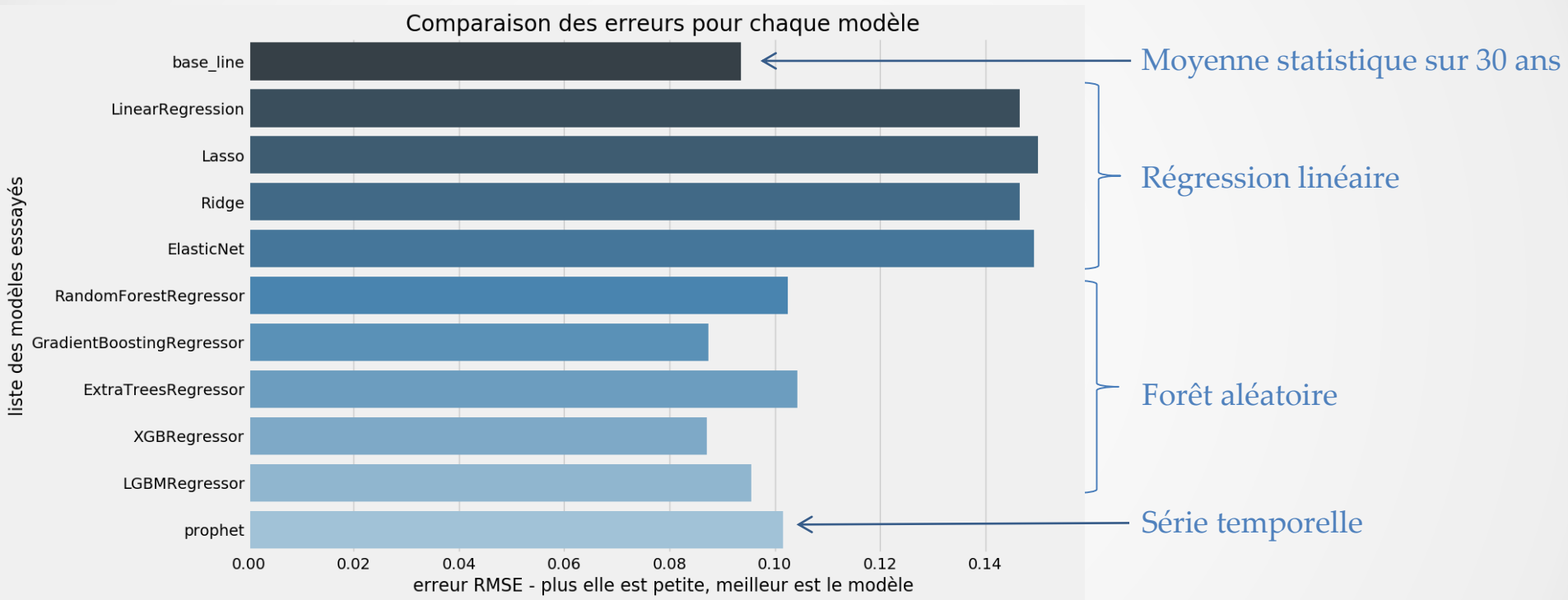
4. Résultats et conclusions



4. Résultats & conclusions : pertinence du meilleur modèle de prédiction

Erreurs des différents modèles :

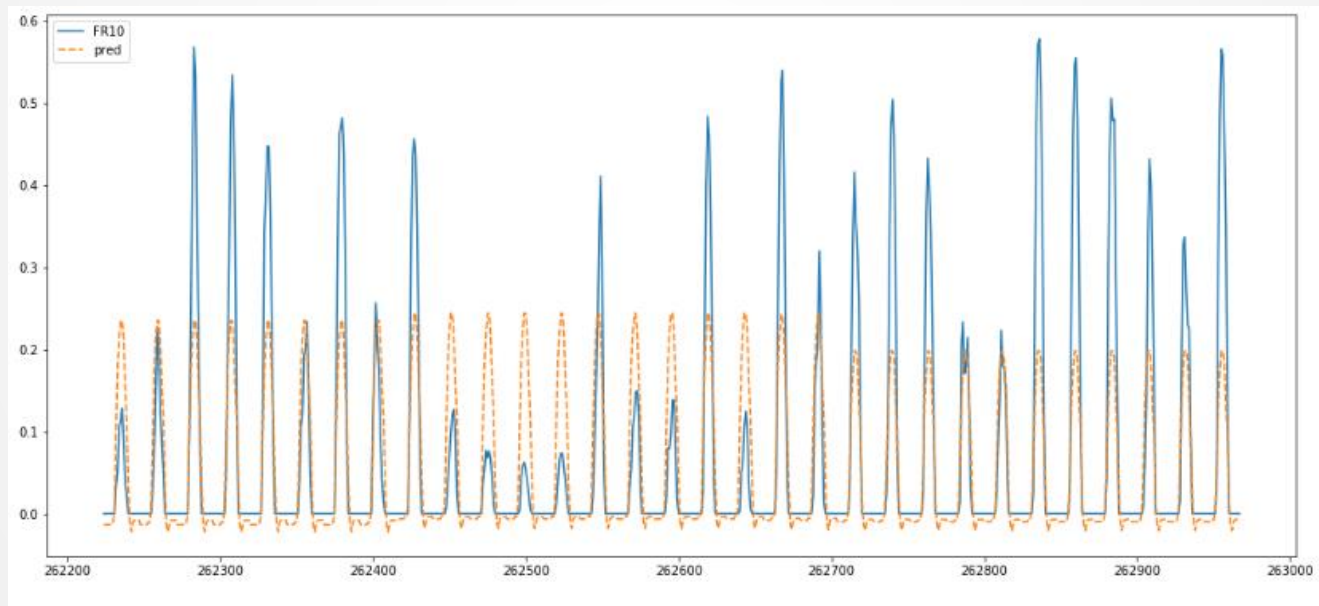
4 Familles de modèles :



4. Résultats : prédictions vs. valeurs réelles de rendements P.V

Visualisation des rendements prédits & réels

- Abscisse = temps (1 mois) :
 - 30 pics (rendements > 0) → 30 jours
 - 30 plateaux (rendements = 0) → 30 nuits
- Avantage : bonne fréquence quotidienne
- Inconvénient : du mal à appréhender la bonne valeur de rendement (hauteur du pic) → valeurs prédites « lissées »



Améliorer les modèles :

1. **Enrichir** la donnée avec les infos sur la météo.
2. **Intégrer l'historique** des rendements pour la prédiction à une heure donnée
3. Essayer un modèle de **deep learning** (type RNN).
4. Croiser les rendements **éoliens et photovoltaïques**

Perspectives :

1. Voir plus grand : **à l'échelle nationale** → scalabilité de l'architecture.
2. Faire un modèle **prédictif de la consommation**.
3. Réaliser un **modèle global intégrant consommation & production**

Merci pour votre attention :)



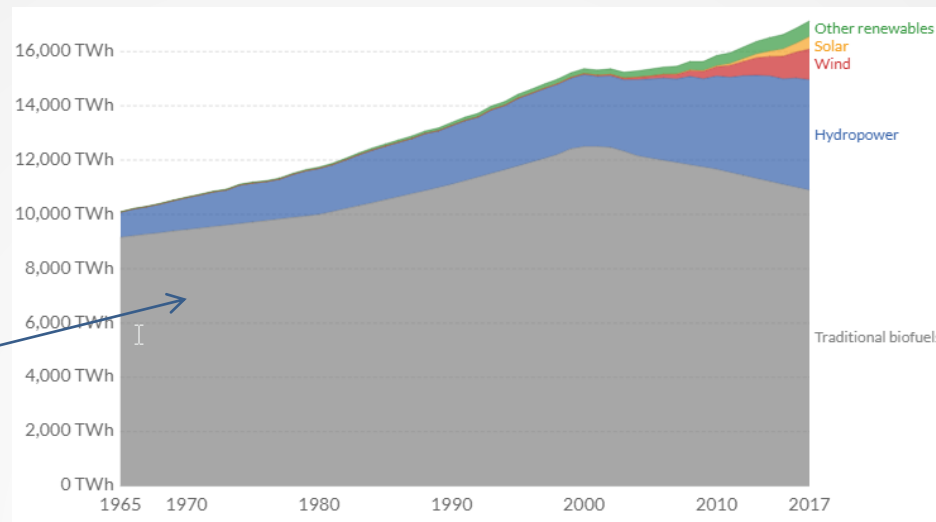
Olivier Brunet – 09/01/2020

Annexes

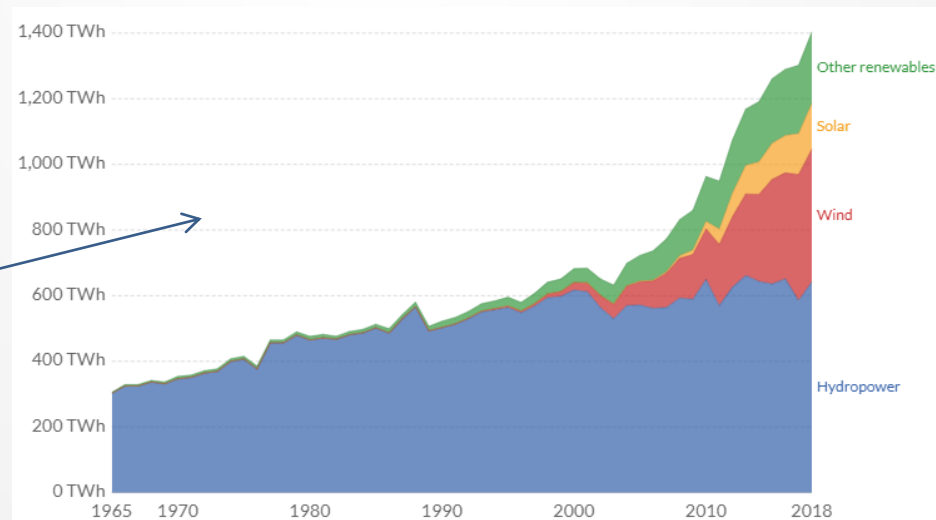


Annexe : contexte – évolution de la consommation d'EnR / monde vs U.E

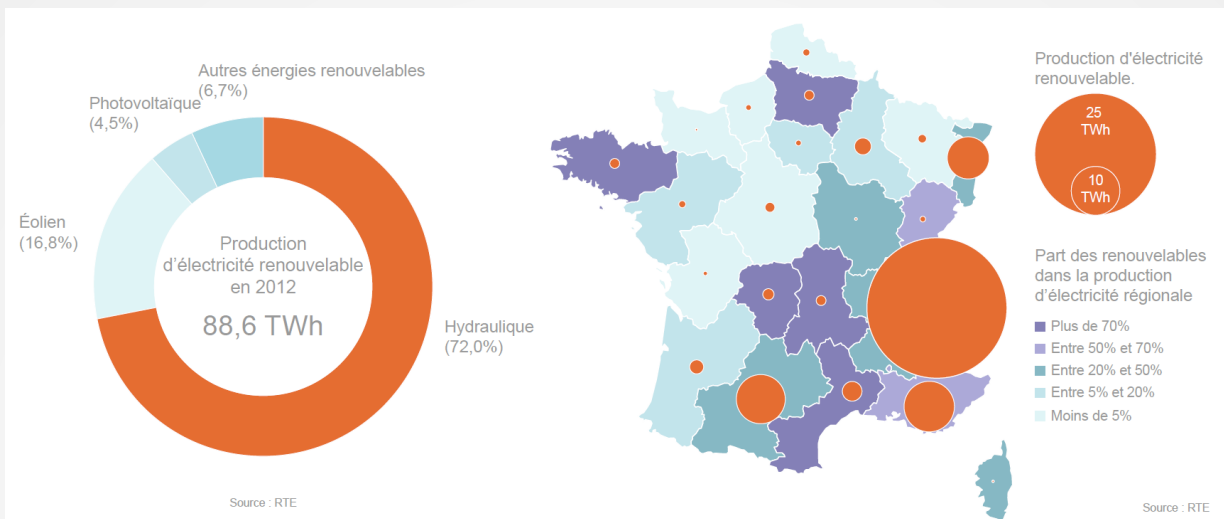
MONDE:
les biocarburants sont
prépondérants mais
déclinent.



EUROPE :
pas de biocarburant -
le solaire et l'éolien en
pleine expansion.



Annexe : contexte – focus sur la France (données RTE 2012)

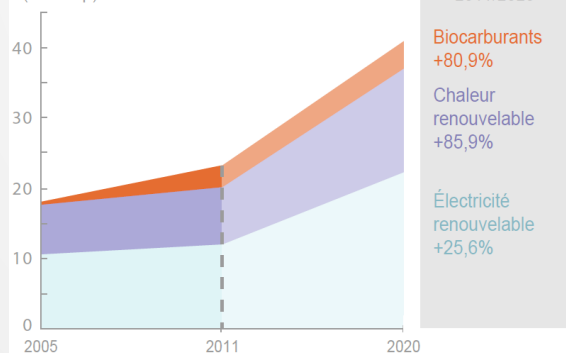


23%

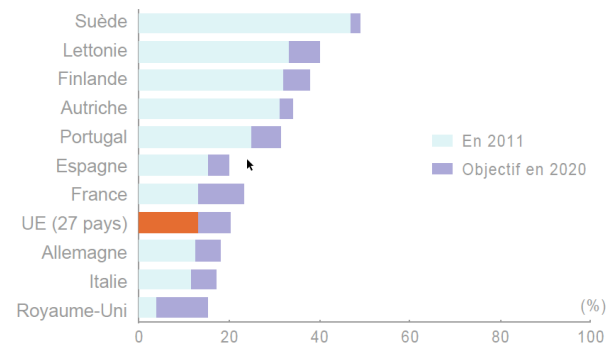
Objectif de consommation finale d'énergie satisfaite par les EnR en France en 2020 (paquet énergie climat)

Consommation finale brute d'énergie renouvelable par filière

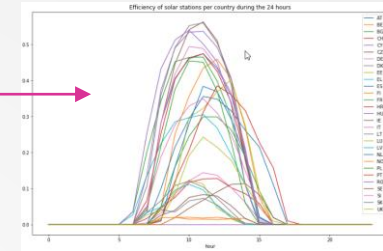
(en Mtep)



Part des énergies renouvelables dans la consommation finale brute d'énergie



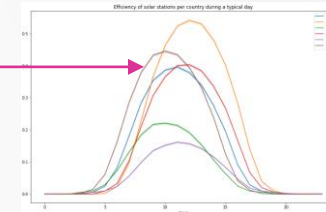
1ère analyse – 30 pays.



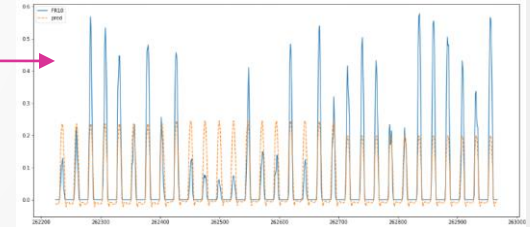
Clustering (ML non supervisé).

cluster nb : 0 CY IE NO SE
cluster nb : 1 AT CH CZ HR HU IT SI SK
cluster nb : 2 BE DE DK FR LU NL UK
cluster nb : 3 EE FI LT LV PL
cluster nb : 4 BG EL RO
cluster nb : 5 ES PT

2ème analyse – 1 pays par cluster

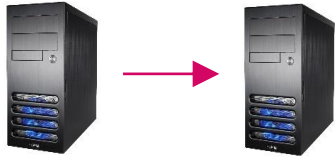


Proof of concept (P.O.C)



Prototype (Big Data)

2 nœuds de référence



Specs par nœud :

- Stockage : 4 TB
- RAM : 32 GB
- CPU : 4 cores 4 GHz

La compression donne
une sécurité

Dimensionnement :

Stockage :

- Data : 5TB x 3 (réplication) = 15 TB
- DD par nœuds 4 TB dont 30% non HDFS, 25% data provisoire → 1,8 TB dispo
- $15 / 1,8 = 8,6$ soit 9 nœuds

RAM :

- Process memory : 4GB
 - Task tracker : 4 GB
 - OS : 4GB
 - 4 cores x 4 GB / core
- Total 28 arrondi à 32 GB

9 nœuds esclaves

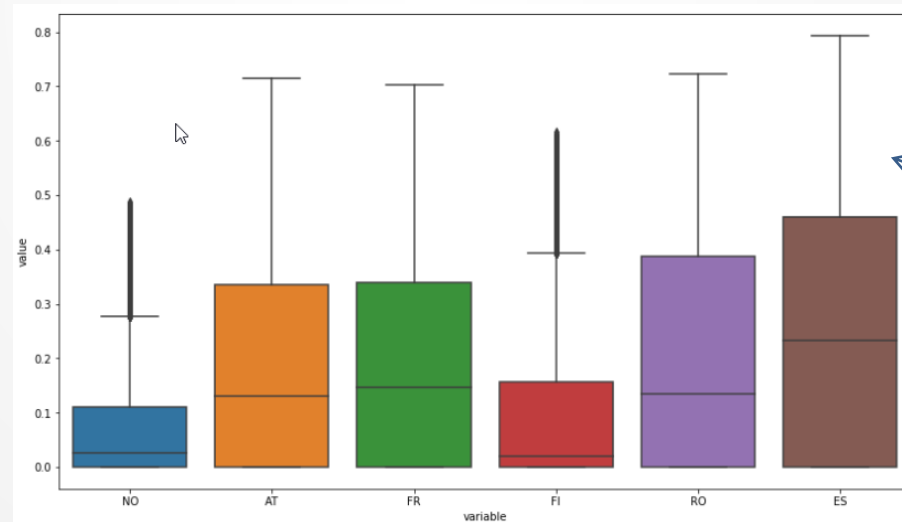
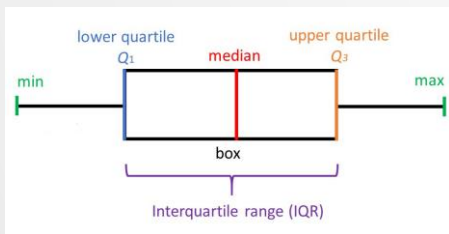
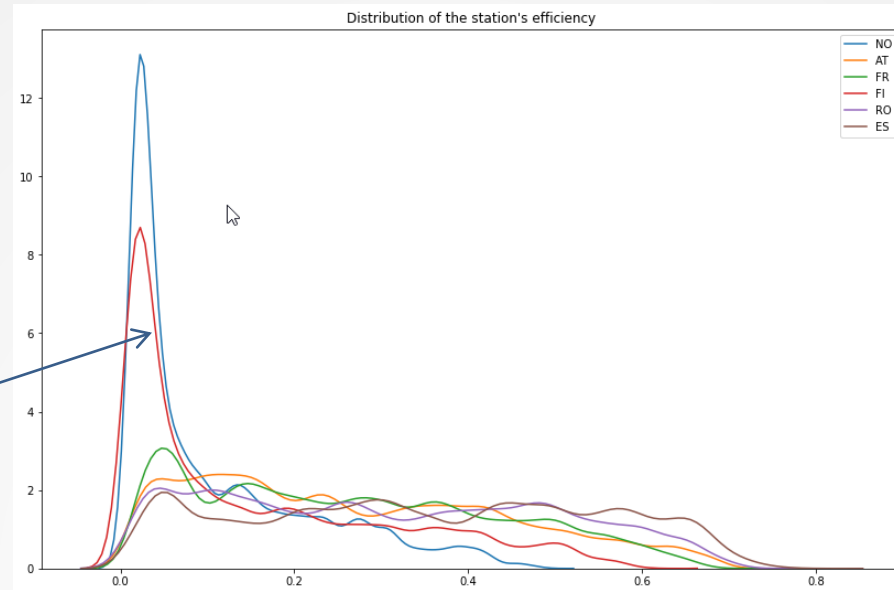


Specs par nœud :

- Stockage : 4 TB
- RAM : 32 GB
- CPU : 4 cores 4 GHz

Annexe = distribution statistiques des valeurs de rendement photovoltaïque

Distribution des
valeurs

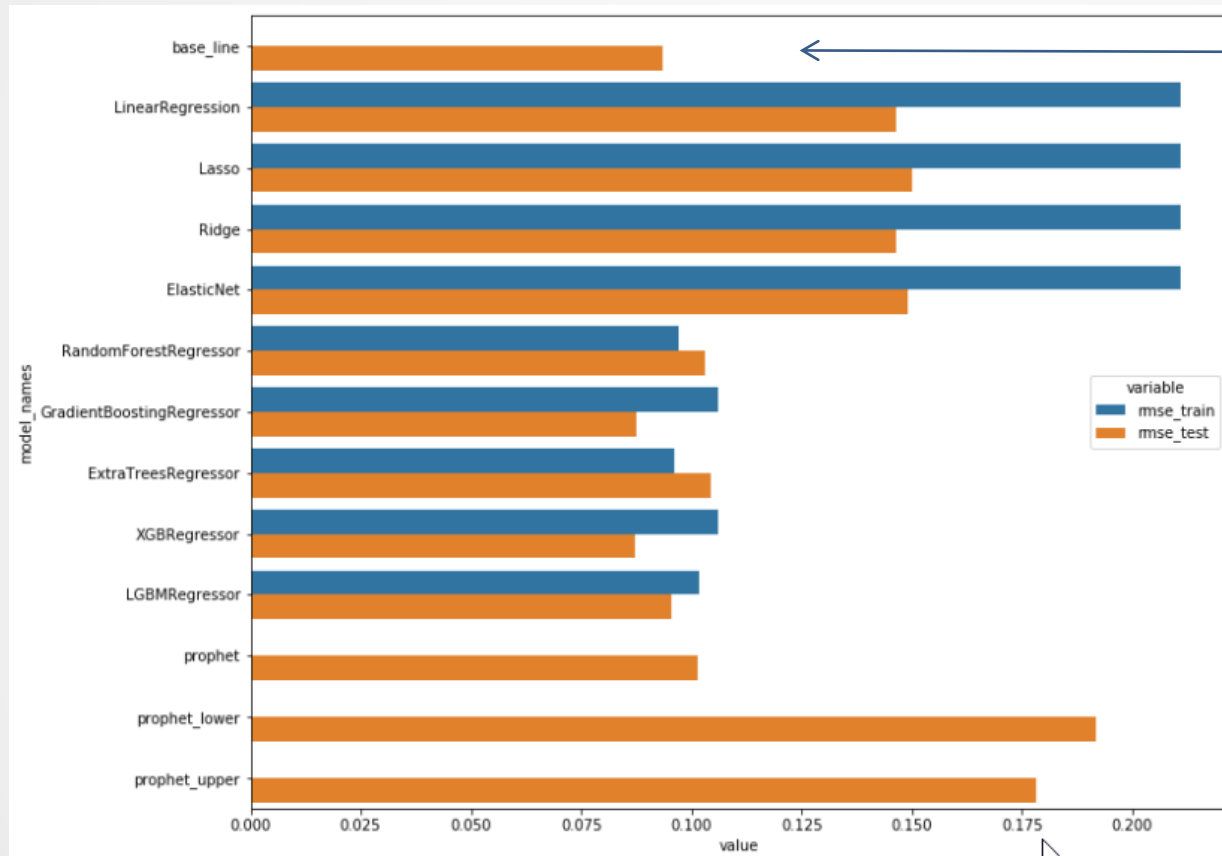


Boîtes à
moustaches

4. Résultats & conclusions : pertinence du meilleur modèle de prédiction

Erreurs des différents modèles :

4 Familles de modèles :



Moyenne statistique sur 30 ans

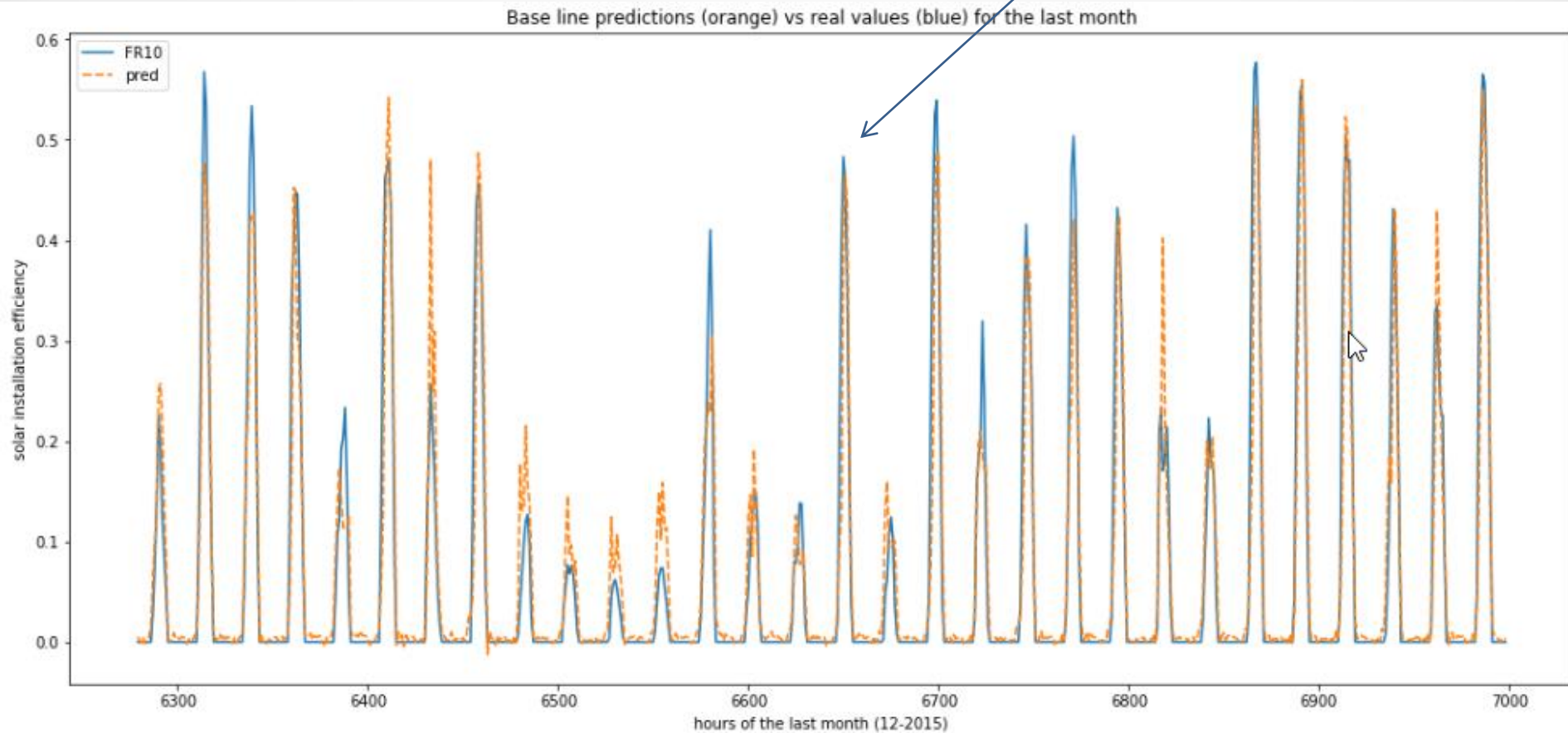
Régression linéaire

Forêt aléatoire

Série temporelle

4. Résultats & conclusions : Utilisation d'un R.N.N (deep learning)

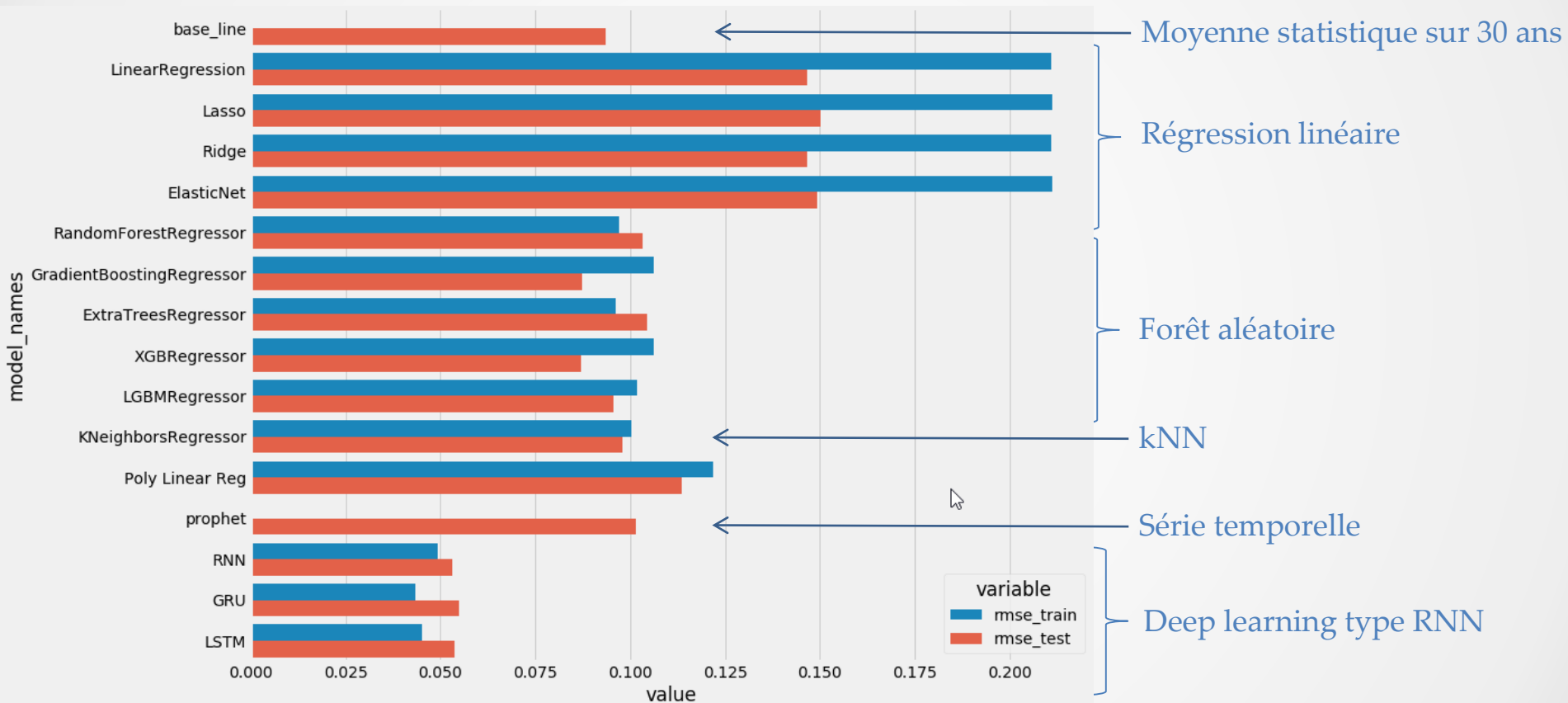
Meilleure anticipation des pics journaliers de rendement



4. Résultats & conclusions : pertinence du meilleur modèle de prédiction

Erreurs des différents modèles :

Familles de modèles :



4. Résultats & conclusions : pertinence du meilleur modèle de prédiction

Erreurs des différents modèles :

4 Familles de modèles :

