With crazy values

RMSE: 0.00018455919375105743 | std: 0.00011954780333337886

Without crazy values

RMSE: 0.00021238300198292626 | std: 0.00014603710696419305

Without night time data

RMSE: 0.0003616055287217457 | std: 0.00024147135040542393

Without nan

RMSE: 0.0005853564876204129 | std: 0.0001893834486798293

FD\_Avg/FG\_Avg with night data

Lenght: 16535 | MIN: 0.06381881887344203 | MAX: 2.2001408837028977 | RMSE: 0.7690764274327034 | std: 0.31841320593485334

RMSE: 0.0013243070353050298 | std: 0.0014534216400252007

FD\_Avg/FG\_Avg without night data

Lenght: 6207 | MIN: 0.06381881887344203 | MAX: 1.0010936936912507 | RMSE: 0.5364909467310378 | std: 0.3317665323866562

RMSE: 0.002101280654350185 | std: 0.000454506184954605

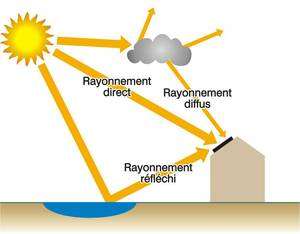
Projet FDEC 2018

Sujet : Défi EGC 2018

# Présentation de la problématique

L’ile de la Réunion est en train d’installer des panneaux solaires afin d’obtenir une autonomie énergétique. Nous avons donc des données sur l’ensoleillement, celle-ci ont été mesure par des capteur SPN1 (Sunshine Pyranometer). Le rayonnement solaire peut être décomposer en 3 flux :

* Le flux global FGlobal
* Le flux diffus FDiffus
* Le flux direct FDirect: FDirect = FGlobal - FDiffus



Les panneaux solaires produisent plus d’énergie lorsque le rayonnement direct est fort. Il pourrait donc être intéressant de prédire FDirect ou l’indice de fraction direct (Kb = FDirect / FGlobal). Cela afin de prédire la production d’énergie solaire et adapter la production d’énergie par d’autres méthode en fonction.

15 stations ont été équipé avec les capteurs SPN1 mais également des capteurs météorologiques nous avons donc 7 mesures par stations sur les 5 dernières années :

* FG\_avg (en W=m2) : le flux global
* FD\_avg (en W=m2) : le flux diffus
* Patm\_avg (en hPa) : la pression atmosphérique
* RH\_avg (en %) : le taux d’humidité dans l’air
* Text\_avg (en °C) : la température extérieure
* WD\_MeanUnitVector (en degré) : la direction du vent
* WS\_avg (en m/s) : la vitesse du vent

Les mesures sont effectuées toutes les minutes, nous avons donc un cas de prédiction de séries temporelles.

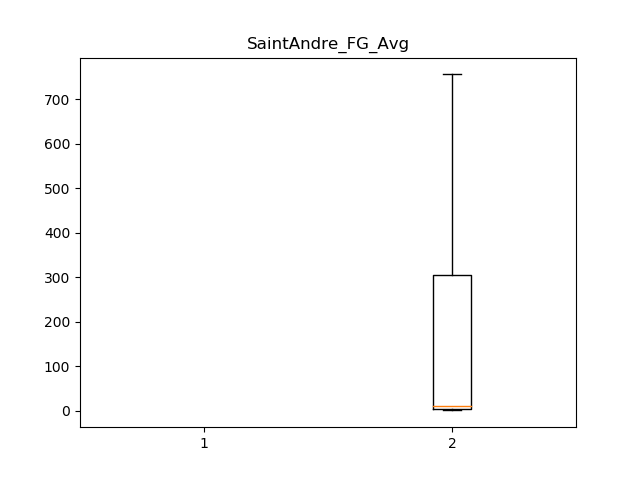
Pour le défi ouvert seulement 2 années de données de 5 stations étaient disponibles. Il faudra dans un premier temps pré traiter les données, enlever les valeurs aberrantes et gérer les valeurs manquantes. Il sera ensuite possible de dériver les attributs existant en nouveaux (ex : Kb).

Une fois ce nettoyage effectué diverse analyse graphiques pourront être effectuées afin de trouver les attributs les plus intéressants. Celles-ci peuvent également permettre de détecter des problèmes de qualité des données nous amenant a rajouter une étape au pré traitement.

La méthode la plus simple pour effectuer des prédictions sur des série temporelle est la méthode ARIMA.

# Pre traitement

A l’aide d’un rapide affichage graphique on observe que la colonne FG\_Avg de la station Saint André a des valeurs manquantes. Les lignes ou il manque une valeur sont supprimées des dataset.



On observe sur un autre graph qu’il y a des valeurs aberrantes dans les dataset Moufia et Saint Andre au mois de Janvier 2014