

# Projet MIASHS

Marc Bourotte & Romain Chailan

6 octobre 2017

## Présentation du projet

Pour ce projet, nous allons nous intéresser à différentes variables climatiques mensuelles pour les stations de métropole et d'outre-mer appartenant au Réseau Climatologique Régional de Base (RBCN) de l'Organisation Météorologique Mondiale (OMM). Ces données Météo-France sont disponibles sur la page suivante : [https://donneespubliques.meteofrance.fr/?fond=produit&id\\_produit=115&id\\_rubrique=38](https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=115&id_rubrique=38)

Je vous ai envoyé l'extraction du mois d'août 2017. L'extraction est composée de 3 fichiers que je vais détailler.

- fichier **climat.201708.csv** : il contient les valeurs des différentes variables climatiques en plusieurs stations ;

##	NUM_POSTE	DAT	PSTATM	PMERM	TMMOY	TMSIGMA	TXMOY	TNMOY
## 1	7005	2017-08-01 00:00:00	1008.7	1017.2	17.4	2.0	22.1	13.6
## 2	7015	2017-08-01 00:00:00	1011.3	1016.9	18.5	2.2	23.7	13.9
## 3	7020	2017-08-01 00:00:00	1016.8	1017.8	17.1	0.8	19.2	15.0
## 4	7027	2017-08-01 00:00:00	1010.0	1017.9	17.5	2.0	22.9	13.1
## 5	7037	2017-08-01 00:00:00	999.6	1017.5	17.6	2.6	23.5	12.8
## 6	7072	2017-08-01 00:00:00	1006.2	1017.4	18.3	2.4	24.6	12.4

- fichier **postesSynop.csv** : il contient des informations sur les stations ;

##	NUM_POSTE	Nom	Latitude	Longitude	Altitude
## 1	7005	ABBEVILLE	50.13600	1.834000	69
## 2	7015	LILLE-LESQUIN	50.57000	3.097500	47
## 3	7020	PTE DE LA HAGUE	49.72517	-1.939833	6
## 4	7027	CAEN-CARPIQUET	49.18000	-0.456167	67
## 5	7037	ROUEN-BOOS	49.38300	1.181667	151
## 6	7072	REIMS-PRUNAY	49.20967	4.155333	95

- fichier **doc\_mess\_climat\_92.pdf** : il décrit les différentes variables du fichier **postesSynop.csv**.

La variable **FXAB** du jeu de données **climat.201708.csv** sera à mettre de côté. Nous l'utiliserons ultérieurement dans une optique de prédiction.

```
## [1] 15.6 18.8 18.9 16.6 16.4 19.9
```

1. Premièrement, vous devrez réaliser une analyse exploratoire du jeu de données à l'aide d'une analyse en composantes principales. Les données nécessiteront très certainement un retraitement. L'interprétation de l'ACP doit être concise (20aines de lignes voire plus si nécessaire) et illustrée par des sorties graphiques pertinentes
2. La deuxième partie du projet a pour objectif de prédire les valeurs de la variable **FXAB** en plusieurs sites. Vous devrez donc considérer, dans un premier temps, que vous ignorez la valeur de **FXAB** en ces sites. Voici la liste de ces sites :

```
## [1] 7110 7149 7222 7481 7535 7591 7630 7747 7790
```

2. (suite) Tous les autres sites (*lignes*) pourront être utilisés en apprentissage. La modélisation de la prédiction se fera en utilisant différents modèles linéaires comme dans le TP3 : régression linéaire classique, régression sur composantes principales, et *partial least square regression*. Enfin, pensez à utiliser les vraies valeurs de **FXAB** pour calculer les scores de prédiction comme le RMSE (cf TP3) et

comparer les performances de vos différents modèles. **Les groupes obtenant un RMSE meilleur (inférieur) au mien auront un bonus de points.**

### Consignes

- Le rapport peut être réalisé en groupe de 3 personnes maximum. Je vous demanderai de me communiquer la composition de votre groupe le plus rapidement possible.
- Vous devrez me rendre un rapport (5-10 pages environ) ainsi qu'un fichier source R permettant d'obtenir les graphiques et les résultats obtenus. La solution R Markdown est fortement conseillée mais non obligatoire.
- Je vous rappelle que ce qui m'intéresse est de suivre votre démarche. Il est donc nécessaire que vous argumentiez votre préparation de données et votre modélisation, et que vous discutiez de votre ACP et de vos résultats graphiques et numériques.
- Un plan *passe-partout* serait : 1/ Préparation des données 2/ Analyse descriptive par ACP 3/ Modélisation de la prédiction de FXAB par modèles linéaires (classique, pcr et pls) 4/ Discussion
- La qualité de rédaction et de présentation seront fortement prises en compte.
- Le rapport et le code R devront m'être envoyés par mail à [marc.bourotte@twinsol.com](mailto:marc.bourotte@twinsol.com) avant le 19 novembre 2017 (23h59).