

# OÙ TROUVER DES DONNÉES CLIMAT ?

## LES DONNÉES MÉTÉO FRANCE

**MOTS CLES : METEO - CLIMAT - PROJECTIONS CLIMATIQUES  
TRAITEMENT DE DONNEES - OPEN DATA**

Depuis plusieurs années, la prise de conscience des enjeux liés aux risques climatiques émerge, et notamment dans le secteur assurantiel. Les données sont indispensables pour une gestion appropriée du risque climatique, **mais où trouver des données fiables en open data ?** Cette série de notes, rédigées par le Data Lab avec l'appui des experts du groupe de travail Climat du cabinet Galea, a pour objectif de présenter des données climatiques d'accès et d'usage libres, d'initier un premier traitement et de mettre en avant quelques utilisations possibles.

### LES DONNÉES MÉTÉO FRANCE

Météo France met à disposition sous licence ouverte les données « SYNOP Essentielles OMM », une extraction des données SYNOP circulant sur le SMT (système mondial de télécommunication) de l'Organisation Météorologique Mondiale (OMM) pour 62 stations météorologiques françaises.

### LES DONNÉES DISPONIBLES

Ces données ouvertes correspondent à des paramètres atmosphériques mesurés ou observés toutes les 3h depuis janvier 1996 pour la plupart des 42 stations météorologiques en France métropolitaine et 20 en Outre-Mer. La volumétrie de ces données est donc assez importante, ce qui nécessite un outil de traitement adapté (réalisable sous Python ou R) puisque le recueil de l'ensemble des données disponibles pour toutes les stations représente près de 5 millions de lignes et une soixantaine de variables.

L'ensemble des constantes et paramètres des relevés couvre notamment :

- // la température actuelle et températures minimale et maximale sur les dernières heures ;
- // les taux d'humidité, température du thermomètre humide et point de rosée ;
- // les précipitations dans les dernières heures ;

- // des indicateurs sur la vitesse et la direction du vent ;
- // des indicateurs sur les dernières rafales ;
- // des indicateurs sur la visibilité ;
- // des indicateurs sur les nuages et la nébulosité ;
- // des indicateurs sur la pression, le niveau et la variation barométrique ;
- // des indicateurs sur l'état du sol, la couche de neige, glace ou autre le cas échéant.

Les consultants de Galea ont étudié la base de données à disposition dans le cadre de travaux sur l'intégration du risque climatique. L'objectif de cette note est de partager un bref retour d'expérience sur l'état de ces données et leur traitement liminaire pour en faciliter la manipulation.

### TRAITEMENT DES DONNÉES

Dans tout projet faisant appel à des données *a priori* non maîtrisées, l'une des étapes essentielles est le contrôle de leur qualité par l'étude des données manquantes.

La figure 1 illustre le taux de données manquantes par variables (en ordonnées) et par stations (en abscisses) sous forme de *heatmap*, elle permet de constater un taux de valeurs manquantes assez élevé au global et d'autant plus important sur les stations d'outre-mer (numéros de stations à 5 chiffres, concentrées à droite du tableau).

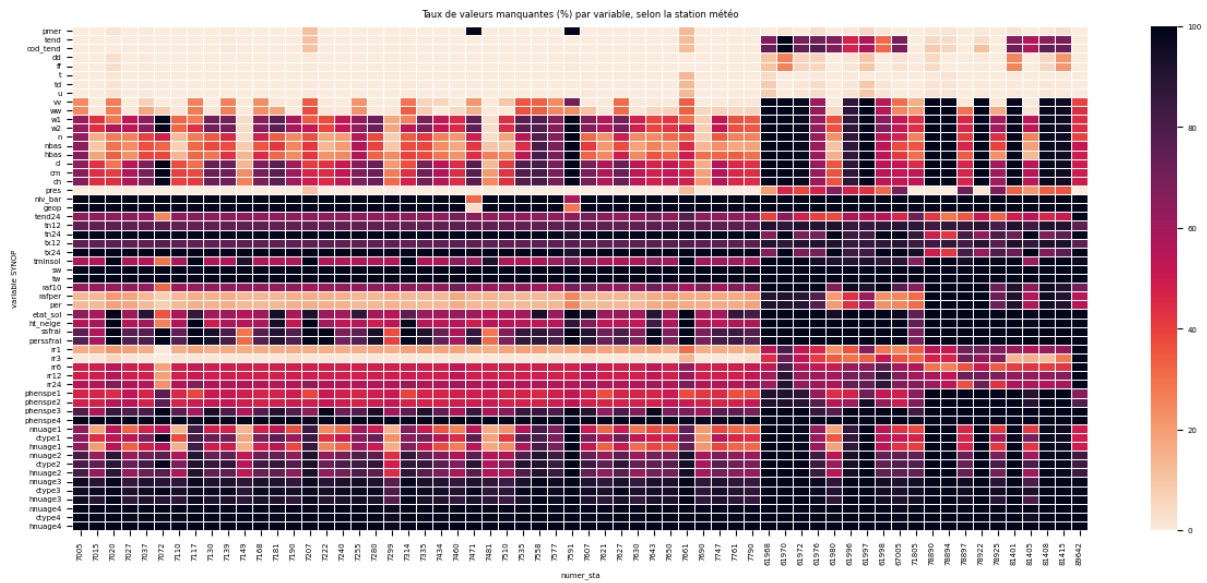


Figure 1 – Taux de valeurs manquantes pour l'ensemble des stations et des variables  
Informations créées à partir de données de Météo-France

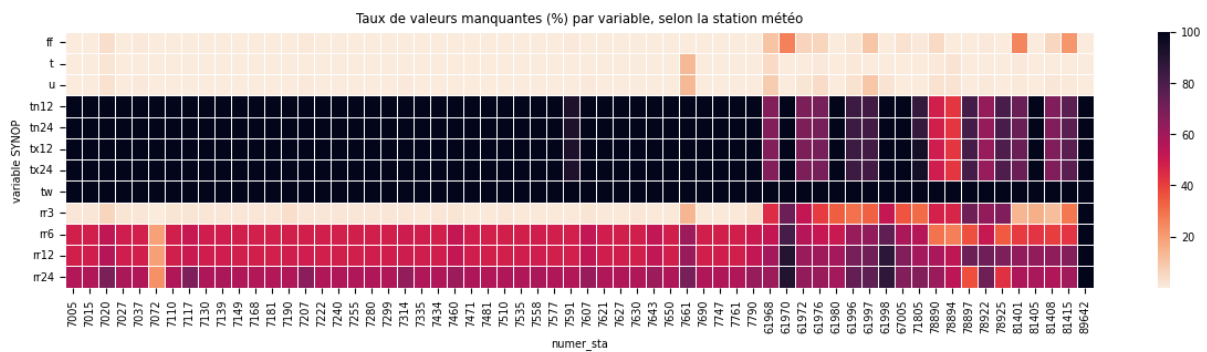


Figure 2 – Taux de valeurs manquantes sur une sélection de variables pour l'ensemble des stations  
Informations créées à partir de données de Météo-France

Pour plus de lisibilité, il est représenté dans la figure 2 les taux de valeurs manquantes sur une sélection de variables d'intérêt sur des risques pouvant être assurés dans le cadre d'une assurance paramétrique par exemple : la vitesse du vent (ff), l'humidité (u), la température (t) ainsi que les températures minimales (tn) et maximales (tx) sur les dernières 12 et 24h, la température du thermomètre mouillé (tw) et les précipitations (rr) sur les 3 à 24 dernières heures. Ce tableau permet d'écarter quelques variables non utilisables et de se concentrer sur les mesures de température, humidité, vitesse du vent et précipitations sur 3h sur la France métropolitaine.

Concernant les valeurs manquantes restantes, il convient de les remplacer à l'aide de méthodes usuelles comme l'utilisation de la mesure du jour précédent, ou des traitements plus spécifiques liés à la connaissance des phénomènes météorologiques que représentent ces variables.

L'étude des données manquantes permet également de détecter les stations dont la période de relevés diffère des autres. Ainsi, les traitements mettent en évidence que la station située près de Reims n'a ouvert qu'en mai 2012, et que la station de Toulon n'enregistre plus de températures (entre autres paramètres) depuis janvier 2020. Ces observations justifient d'exclure ces stations dans la suite de cette note.

## JOINTURE PAR ZONE GÉOGRAPHIQUE

La plupart des études climatiques actuarielles nécessitent le couplage de ces données météorologiques avec des données assurantielles (sinistralité, population assurée, ...) propres à un portefeuille et/ou *open source* afin d'étudier l'influence de la météo ou de mesurer l'exposition à différents risques climatiques. Ces jointures peuvent être réalisées à différentes mailles (région, département, commune, adresse), exigeant différents traitements de la donnée.

Pour une utilisation par région, il est possible de d'agréger les données Météo France et d'utiliser des indicateurs de distribution des données météo (température moyenne/médiane par exemple) de l'ensemble des stations d'une région afin de d'estimer la météo relative à une région donnée.



Figure 3 - Localisation des stations essentielles OMM et des centres géographiques départementaux  
Informations créées à partir de données de Météo-France et de l'IGN

Dans le cas d'une utilisation à une maille plus fine, une étude de proximité est requise.

A titre d'exemple, une méthode pour lier chaque département de France métropolitaine à une station météorologique est de retenir la station la plus proche de son centre de gravité<sup>1</sup>. La figure 3 reprend la localisation des 42 stations météorologiques en France métropolitaine (en bleu les stations dont les données sont utilisables, en rouge les 2 dont les périodes de relevés sont réduites) et les centres géographiques des départements calculés par l'IGN (Institut national de l'information géographique et forestière) en corail.

Une telle méthode entraîne le zonage représenté dans la figure 4, regroupant les départements associés à la même station météo.

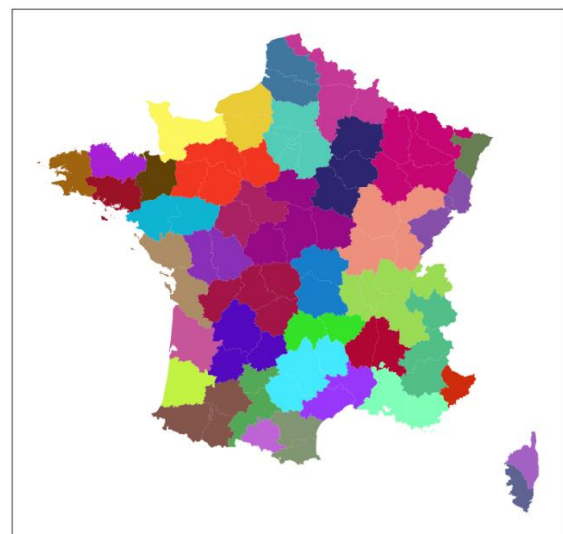


Figure 4 - Zonage induit par un regroupement par station

Ce type de méthode peut également être utilisée pour établir une correspondance entre commune et station météo afin d'obtenir une appréciation de la météo à une maille communale.

<sup>1</sup><https://www.ign.fr/reperes/centre-geographique-des-departements-metropolitains>

## UTILISATION POUR L'ÉTUDE DU RÉCHAUFFEMENT CLIMATIQUE

Une fois les données maîtrisées et mises en qualité, il est intéressant d'étudier la distribution des variables sélectionnées dans l'optique d'une approche risque, la compréhension des phénomènes extrêmes étant essentielle dans un contexte assurantiel.

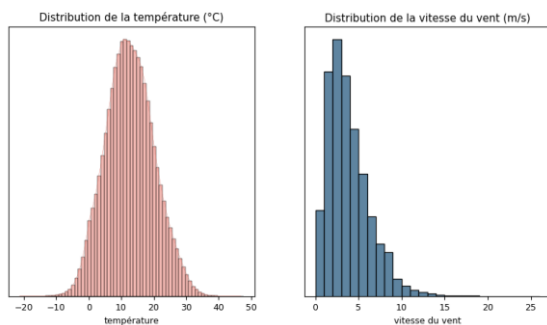


Figure 5 - Distributions des température et vitesse du vent - Informations créées à partir de données de Météo-France

La figure 6 permet de visualiser les écarts de températures moyennes entre les mois de juillet-août 2001-2002 et les mois de juillet-août 2021-2022.

Ainsi, il apparaît que la température moyenne des 2 mois d'été les plus chauds a augmenté de près d'un degré Celsius à minima sur l'ensemble de la France métropolitaine en 20 ans, ce réchauffement atteint près de 3 degrés dans le sud-ouest.

Il peut également être intéressant de constater l'écart entre les phénomènes extrêmes sur différentes périodes.

La base de données fournie par Météo France en accès libre est donc exploitable mais nécessite un traitement et une appropriation des données préalable. Cela passe nécessairement par la mise en qualité des données, la réalisation de statistiques descriptives et des rendus graphiques appropriés à l'aide d'outils de

programmation et de *data visualisation* adéquats.

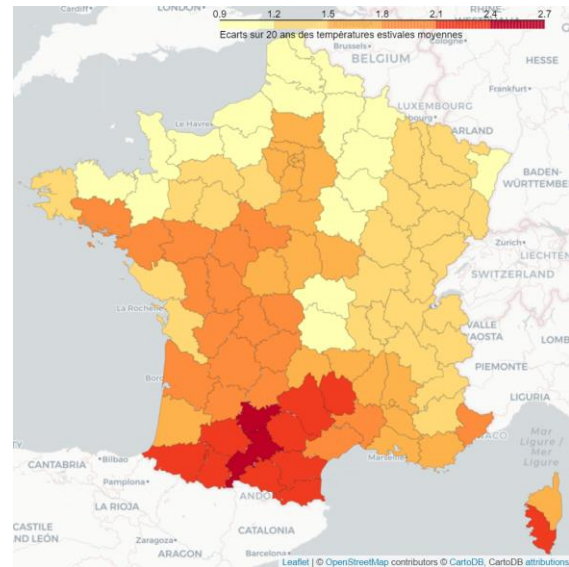


Figure 6 - Ecart des températures estivales (juillet-août) moyennes entre 2001-2002 et 2021-2022 - Informations créées à partir de données de Météo-France et de l'IGN

Les consultants de GALEA sont à votre disposition pour plus de précisions sur le contenu de cette note et pour vous accompagner dans l'intégration des nouveaux risques relatifs aux changements climatiques, de l'utilisation des données à la mise en place de modèles adaptés, en conformité avec les dernières recommandations réglementaires et sectorielles sur le risque climatique, telles que celles émises par l'EIOPA et l'AAI.

<https://www.galea-associes.eu/>

//galea