OÙ TROUVER DES DONNÉES CLIMAT ?

LES DONNÉES MÉTÉO FRANCE 2024 (RADOME)

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

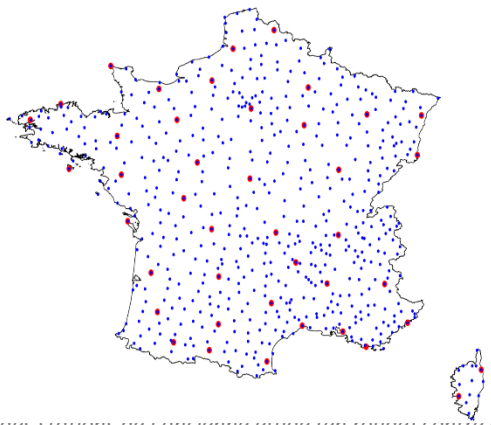
Depuis plusieurs années, la prise de conscience des enjeux climatiques s'intensifie, particulièrement dans le secteur de l'assurance. Pour gérer efficacement les risques que le changement climatique engendre, des données fiables sont essentielles et une question se pose : Où peut-on trouver des données fiables en open data ? Ce quatrième article de cette série de notes (**liens dans l'encadré de fin**) rédigées par le DataLab avec l'appui des experts du groupe de travail Climat du cabinet Galea, a pour objectif de présenter l'évolution de l'accès aux données climatiques en usage libre, et de mettre en lumière les changements quantitatifs et qualitatifs de ces données. L'objectif est aussi de rendre compte d'un possible traitement des nouvelles données Météo France (plus de données différentes et quantité de stations disponibles augmentées) et de mettre en avant quelques utilisations possibles de ces nouvelles données en comparant les résultats précédemment obtenus avec les données SYNOP[1].

LES DONNÉES MÉTÉO FRANCE

Météo France met à disposition sous licence ouverte, une quantité importante de données météo, disponibles par département et par station. Ces données sont accessibles depuis le début d'année 2024 sur la plateforme meteo.data.gouv.fr. Les données sont mises à jour quotidiennement. Les périodes des différents jeux de données sont détaillées ci-dessous :

Base quotidiennes	1852 - 2024
Base mensuelles	1852 - 2024
Base horaires	1850 - 2024
Base décadaires agro	1852 - 2024
Base 6 minutes	2000 - 2024
Base décadaire	1852 - 2024

La différence la plus importante entre les données avant le 1er janvier 2024 et après est la quantité d'informations disponibles. L'ordre de grandeur est le suivant on passe de 42 stations exploitables en France métropolitaine à plus de 800 stations exploitables sur l'ensemble de la période d'étude de cet article à savoir 1950-2022. Ci-après la carte comparative des stations disponible avant et après le premier janvier 2024.



Carte des stations de l'ancienne étude (en rouge) contre celles d'aujourd'hui (en bleu)

LES DONNÉES DISPONIBLES

Ces données ouvertes correspondent à des paramètres atmosphériques mesurés ou observés périodiquement. Cela représente 1801 stations météorologiques en France métropolitaine et 304 en Outre-Mer. La volumétrie de ces données est donc très importante, ce qui nécessite un outil de traitement adapté (réalisable sous Python ou R), en effet, l'ensemble des données disponibles pour les données quotidiennes (les données retenues sont sur 73 ans entre 1950 et 2022) pour toutes les

[1] https://www.galea-associes.eu/2023/02/ou-trouver-des-donnees-climat-les-donnees-meteo-france/

stations, représente par exemple plus de 500 millions de lignes et une soixantaine de variables. Les données sont issues de deux réseaux différents : le réseau RADOME qui inclut la majorité des stations SYNOP (stations de l'Organisation Météorologique Mondiale déjà disponible auparavant) et plus de 500 stations supplémentaires fonctionnant sur le même principe : données météorologiques horaires ou infrahoraires (toutes les 6 min) prisent sous abris, de manière automatique. Le réseau étendu (environ 1500 stations) ne relève que les données de pluie et/ou de température sur le même rythme que le réseau RADOME. Certaines stations du réseau étendu sont des stations amateures.

L'ensemble des constantes et paramètres des relevés par le réseau RADOME couvre notamment :

- la température minimale, moyenne et maximale sur la plage de relevé;
- les taux d'humidité, de 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 maitrisées, l'une des étapes essentielles est le contrôle de leur qualité par l'étude des données manquantes.

Les *figure 1 et 2* illustrent la proportion de données manquantes par variables (en ordonnées) et par stations (en abscisses) pour 60 stations prises aléatoirement sur le territoire métropolitain respectivement sur le réseau ETENDU et RADOME. Elles permettent de constater un taux de valeurs manquantes assez élevé sur les stations du réseau ETENDU (*figure 1*). A l'inverse, les stations du réseau RADOME, dépendant directement du réseau Météo France, semblent être bien plus fiables et avec très peu de valeurs manquantes

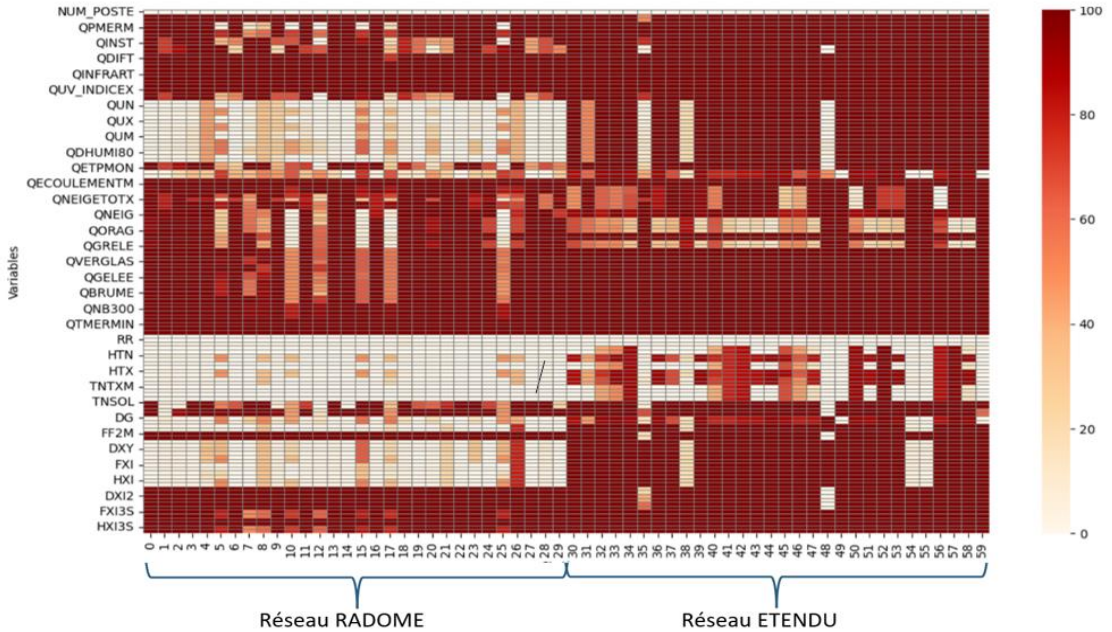


Figure 1 – Taux de valeurs manquantes sur une sélection de variables pour 60 stations 30 du réseau RADOME et 30 du réseau étendu
Informations créées à partir de données de Météo-France

Note réalisée par le Data Lab Galea en collaboration avec le Groupe de Travail climat

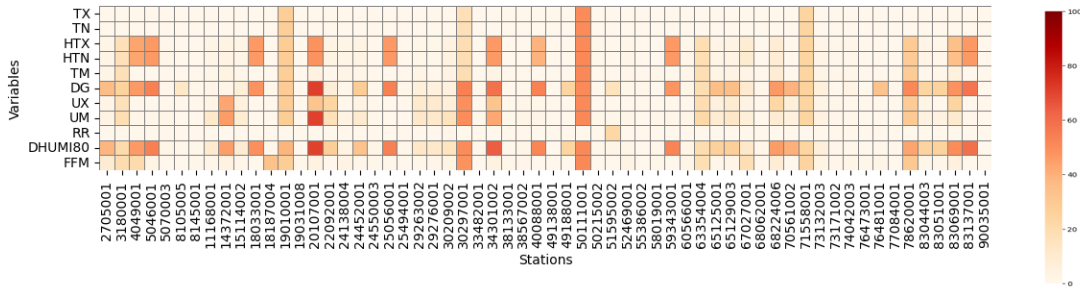


Figure 2 – Taux de valeurs manquantes sur une sélection de variables pour 60 stations du réseau RADOME

(*figure 2*). La suite de cet article se concentrera sur l'étude des données RADOME. 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 (FFM), l'humidité moyenne (UM), l'humidité maximal (UX), la durée ou l'humidité est supérieure à 80% (DHUMI80), la température moyenne (TM) ainsi que les températures minimales (TN) et maximales (TX) et les heures passées à TN (HTN) et TX (HTX), la durée de gel (DG) et les précipitations (RR). Ce tableau permet d'écarter les variables non utilisables et de se concentrer sur les mesures de température, humidité, vitesse du vent et précipitations sur 24 dernières heures 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 et du jour suivant (moyenne glissante), la moyenne du jour J dans le même département, ou des traitements plus spécifiques liés à la connaissance des phénomènes météorologiques que représentent ces variables.

CRÉATION D'UNE CARTE SPATIALEMENT CONTINUE

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.

Au vu de la quantité de stations et de données disponibles, il est possible de d'interpoler 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 zone prédéfinie afin de d'estimer la météo relative à un point donné. Le choix de la maille pour l'interpolation a été la maille de 10km de la DRIAS.

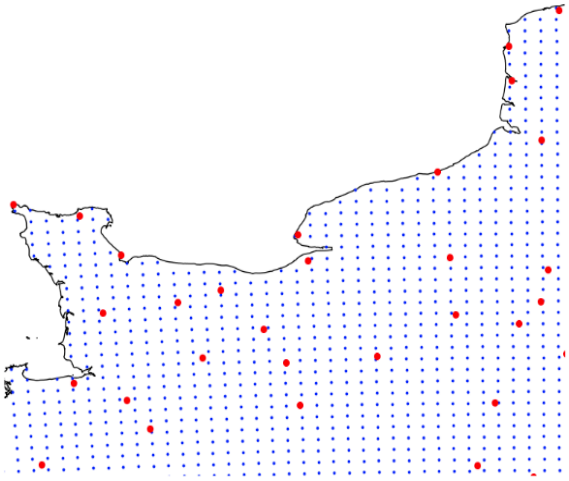


Figure 3 – Maille de 10 km utilisée pour l'interpolation des donnée (station météo en rouge) (Source cadrillage : DRIAS)

Pour l'interpolation de données météorologiques, la triangulation de Delaunay est une méthode couramment utilisée pour estimer les valeurs entre des points de mesure dispersés géographiquement. Elle consiste à diviser la région d'intérêt en une série de triangles de telle sorte que le cercle circonscrit à chaque triangle n'inclue aucun autre sommet. Les sommets de chaque triangle correspondent à une station météorologique. En interpolant les données météorologiques sur les sommets de ces triangles, on peut ensuite estimer celles-ci à n'importe quel point à l'intérieur de ces triangles. Cette méthode permet ainsi de générer

des cartes continues et lisses de données météorologiques sur l'ensemble du territoire.

Comme représenté sur la figure 4 il existe une certaine quantité de points à l'extérieur de la triangulation

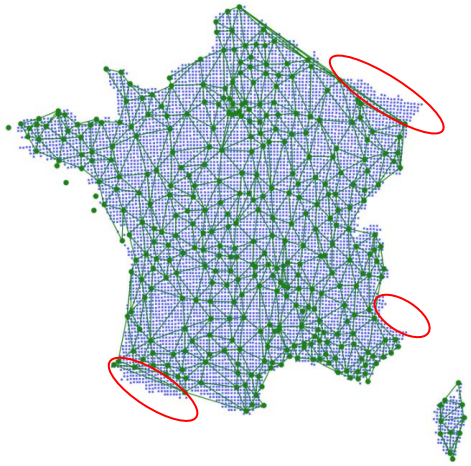


Figure 4 - Triangulation Delaunay (en vert) et principales zones non couvertes par la triangulation (en rouge)

Ces points ont été identifiés et traités différemment pour permettre une estimation globale optimale. La méthode retenue pour ces points est la méthode des 3 plus proches voisins (ou 3NN) pour garder la même quantité d'informations que lors de la triangulation avec 3 stations. La méthode kNN consiste à estimer les valeurs d'un point en se basant sur les mesures des k points les plus proches. En calculant une moyenne pondérée des valeurs de ces voisins en fonction de leur distance, on obtient une estimation précise des données météorologiques pour ce point. Cela permet de générer des cartes continues de données météorologiques avec une résolution spatiale élevée, même dans des zones où les mesures directes sont limitées.

Une fois que tous les points ont 3 stations rattachées, les distances du point aux trois stations sont recalculées pour permettre la même mise en forme qu'avec la triangulation. Ensuite pour l'ensemble des points (3NN et triangulation), nous avons pondéré par la somme des distances pour avoir une distance relative qui servira de pondérations pour les mesures météorologiques.

UTILISATION POUR L'ÉTUDE DU RÉCHAUFFEMENT CLIMATIQUE

Une fois les données maitrisées et mises en qualité, il est intéressant d'étudier la distribution des variables sélectionnées dans l'optique d'une appréhension du risque. La *figure 5* permet de visualiser les écarts de températures moyennes entre les mois de juillet-aout 2001-2002 et les mois de juillet-aout 2021-2022.

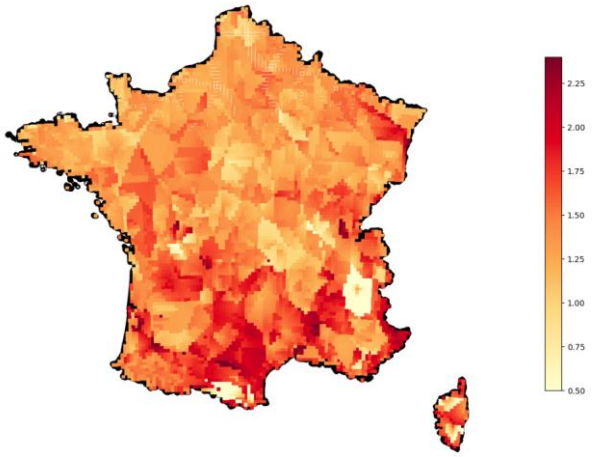


Figure 5 - Ecart des températures estivales (juillet-aout) moyennes entre 2001-2002 et 2021-2022 - Informations créées à partir de données de Météo-France et de la DRIAS

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

La *figure 6* met en lumière la différence du nombre de jour de pluie entre les années 2001-2002 et 2021-2022. »

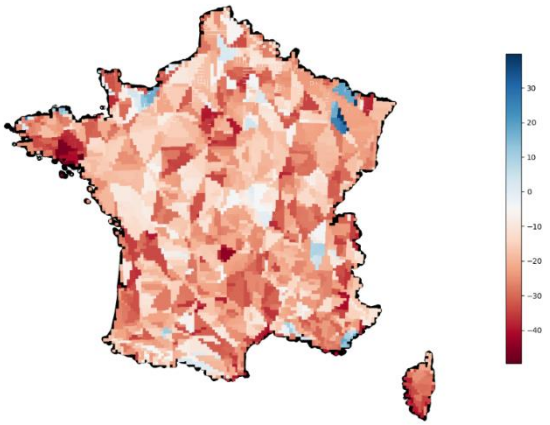


Figure 6 - Ecart des du nombre de jours de pluie annuels entre 2001-2002 et 2021-2022 - Informations créées à partir de données de Météo-France et de la DRIAS

On remarque aussi une baisse nationale du nombre de pluie allant jusqu'à 50 jours de pluie en moins par an sur 20 ans.

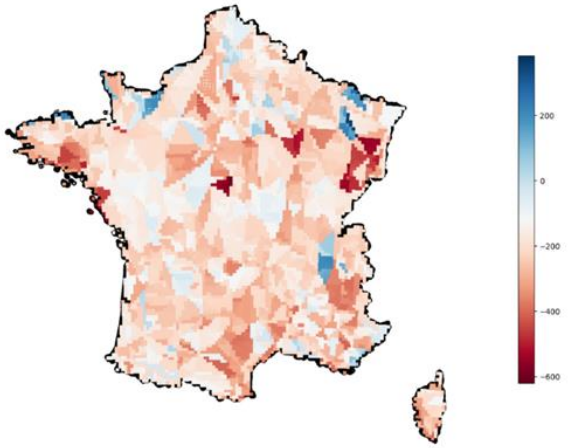


Figure 7 - Ecart du volume de pluie annuel en mm entre 2001-2002 et 2021-2022 - Informations créées à partir de données de Météo-France et de la DRIAS

Cette baisse du nombre de jours de pluie n'est pas compensée par une hausse en volume sur l'année comme le montre la figure 7.

Des études plus poussées sur de durées plus longues peuvent aussi être menées dans le cas d'étude climatique. Les études menées ici ont été effectuées dans l'idée d'expliquer des variations de la sinistralité en assurance.

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 visualisations* adéquates.

Liens vers les articles précédents :

1. Où trouver des données climat ? Les données Météo France
2. Où trouver des données climat ? Les données climatiques DRIAS
3. Quelques utilisations des données météo : Impact de la météo, et plus particulièrement des pics de chaleur humide, sur la Surmortalité

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/