



Mémoire présenté devant le jury de l'EURIA en vue de l'obtention du
Diplôme d'Actuaire EURIA
et de l'admission à l'Institut des Actuaire

le 08 septembre 2023

Par : Baptiste ALLAIRE

Titre : Impact prospectif de la pollution de l'air en assurances de personnes

Confidentialité : Non

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

**Membre présent du jury de l'Institut
des Actuaire :**

David DUBOIS, Pierre CORREGE et

Florence PICARD

Signature :

Entreprise :

Optimind

Signature :

Membres présents du jury de l'EURIA :

Pierre AILLIOT

Directeur de mémoire en entreprise :

Guillaume BESSON et Victor DEQUATRE

Signature :

Invité :

Signature :

**Autorisation de publication et de mise en ligne sur un site de diffusion
de documents actuariels**

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise :

Signature du candidat :

Citation

Première loi de la géographie : « Tout interagit avec tout, mais deux objets proches ont plus de chances de le faire que deux objets éloignés. »

Waldo Tobler cartographe et géographe américain né en 1930 à Portland et mort le 20 février 2018, est considéré comme un pionnier de la cartographie assistée par ordinateur et a permis le développement de plusieurs modèles de cartogramme.

Résumé

Il est aujourd'hui légitime de penser que le changement climatique viendra sensiblement impacter les assureurs et leurs prochains exercices via l'apparition de nouveaux risques (risques physiques ou de transition) ou en renforçant la sévérité de ceux déjà existants.

C'est dans ce contexte singulier que le régulateur prudentiel français a publié son premier exercice pilote climatique en 2020, reconduit en 2023, ayant pour objectif d'encourager les acteurs bancaires et assurantiels à opter pour une gestion prospective à court, moyen et long terme de ces deux familles de risques.

Dans cet exercice, l'assurance de personnes a fait l'objet de la quantification des impacts du changement climatique au travers de trois études indépendantes : une première au sujet des conséquences des maladies à transmission vectorielle, une deuxième relative à l'impact des épisodes caniculaires sur la mortalité et une troisième analysant les effets de la pollution de l'air sur la santé.

En se basant en partie sur ce dernier écrit, le présent mémoire a pour ambition d'étudier l'incidence prospective du fardeau que peut représenter la pollution de l'air sur la santé humaine. Le péril quantifié est la mortalité induite par une exposition à une concentration de différents polluants estimée par interpolation spatiale à la maille pixel d'1km² sur l'ensemble de la France métropolitaine (à l'exclusion de la Corse).

De plus, une étude prospective projetant la concentration des polluants à horizon 2050 selon différents scénarios climatiques, réglementaires et énergétiques est réalisée afin d'anticiper l'évolution des risques prospectifs liés à la pollution atmosphérique.

A la suite d'une évaluation de la matérialité des risques que représente chaque polluant, les résultats obtenus consistent en l'élaboration de zoniers et à la communication de tableaux de facteurs additifs majorant les hypothèses classiques de mortalité d'un assureur, selon différentes géographies, divers scénarios prospectifs et tranches d'âge.

Enfin, une comparaison des résultats obtenus via la méthode proposée dans ce mémoire et ceux publiés dans le cadre de l'exercice pilote climatique de l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) de 2020 est réalisée.

Mots-clés : Pollution de l'air, assurance de personnes, interpolation spatiale, machine learning, évaluation de la matérialité, ORSA, carroyage.

Abstract

Today, it's fair to assume that climate change will have a significant impact on insurers and their next financial years, through the emergence of new risks (physical or transition risks) or by increasing the severity of existing ones.

It is against this singular backdrop that the French prudential regulator has published its first climate pilot exercise for 2020, which is to be repeated in 2023, with the aim of encouraging banking and insurance players to opt for forward-looking short-, medium- and long-term management of these two families of new risks.

As part of this exercise, the impacts of climate change on life and health insurance were quantified through three independent studies: one on the consequences of vector-borne diseases, a second on the impact of heat waves on mortality and a third analysing the effects of air pollution on health.

Based in part on the latter, the aim of this report is to study the prospective impact of air pollution on human health. The peril quantified is the mortality induced by exposure to a concentration of different pollutants estimated by spatial interpolation on a pixel grid of 1km² over the whole of mainland France (excluding Corsica).

In addition, a prospective study projecting the concentration of pollutants to 2050 according to different climatic, regulatory and energy scenarios has been carried out in order to anticipate changes in the prospective risks associated with air pollution.

Following an assessment of the materiality of the risks represented by each pollutant, the results obtained consist of the development of zoniers and the communication of tables of additive factors increasing an insurer's traditional mortality assumptions, according to different geographies, prospective scenarios and age groups.

Finally, the results obtained using the method proposed in this paper are compared with those published as part of the ACPR's 2020 climate pilot exercise.

Keywords: Air pollution, life and health insurance, spatial interpolation, machine learning, materiality assessment, ORSA, raster.

Note de Synthèse

Mise en contexte

Le contexte actuel est marqué par la convergence d'enjeux cruciaux où le secteur de la régulation, représenté par l'Autorité de Contrôle Prudentiel et de Résolution (ACPR), engage des initiatives novatrices pour évaluer la résilience du système financier et assurantiel face aux risques climatiques. Dans ce cadre singulier, les exercices pilote climatique menés par l'ACPR revêtent une importance stratégique en qualité de socle méthodologique permettant d'appréhender les vulnérabilités potentielles du secteur financier face aux perturbations climatiques. Cependant, il est essentiel de noter que le changement climatique transcende les seuls aspects financiers, ayant également des répercussions substantielles sur la qualité de l'environnement de vie des assurés et, par extension, sur l'assurance de personnes.

L'un des risques majeur identifié en assurance de personnes concerne la pollution de l'air et ses répercussions sur la santé, les arrêts de travail et la mortalité. Le régulateur prudentiel a ainsi été destinataire d'une étude indépendante, conduite par le courtier AON, sur le sujet. Cette étude vise à évaluer les conséquences potentielles de la pollution atmosphérique sur ces différentes garanties dans un contexte prospectif. En effet, le lien complexe entre le changement climatique et la pollution atmosphérique peut affecter la solvabilité des acteurs en induisant des évolutions indésirables dans les taux de sinistralité. Ces interactions sont traduites sous forme de facteurs additionnels, destinés à renforcer les hypothèses conventionnelles des assureurs pour chaque garantie évoquée précédemment.

Ces facteurs additionnels synthétisés sous forme de tables sont déclinés en fonction de divers paramètres tels que l'horizon de projection, les groupes d'âge et les scénarios climatiques considérés (notamment RCP8.5 pour l'exercice 2020 et respectivement RCP4.5 pour celui de 2023). Les résultats sont communiqués selon deux niveaux de granularité : une granularité globale pour la France entière et une granularité fine pour neuf villes urbaines.

Granularité	2023 - RCP4.5			2020 - RCP8.5		
	2024-2030	2031-2040	2041-2050	2021-2030	2031-2040	2041-2050
France	0,02 %	0,02 %	0,02 %	0,02 %	0,02 %	0,03 %
Île de France	0,02 %	0,03 %	0,03 %	0,02 %	0,03 %	0,04 %
...
Toulouse	0,01 %	0,01 %	0,02 %	0,01 %	0,02 %	0,02 %

TABLE 1 – Facteurs additionnels de la mortalité communiqués par l'ACPR dans ses exercices pilote climatique de 2020 et 2023.

Ces résultats aident les assureurs à anticiper les impacts financiers des problèmes de santé associés à la pollution atmosphérique future et renforcent leur compréhension. *In fine*, cette étude illustre l'importance d'intégrer les facteurs environnementaux dans les processus d'évaluation des risques et de conception de produits d'assurance, en réponse aux enjeux de santé et de durabilité actuels.

Le présent mémoire vise à compléter l'étude réalisée par AON en identifiant des pistes d'améliorations dans l'analyse des impacts de la pollution de l'air sur la mortalité, seul risque retenu au sein de cette étude. En se concentrant sur des aspects tels que la modélisation spatiale à une granularité plus fine, la méthodologie s'efforce de calculer des chocs plus représentatifs et de mettre en lumière d'autres dimensions cruciales en introduisant la notion d'évaluation de la matérialité. Ainsi, ce travail ambitionne d'approfondir la compréhension des effets de la pollution atmosphérique en apportant des perspectives nouvelles et une approche méthodologique avancée par le biais de méthodes dites d' « impulsion ».

Méthode d'impulsion première : La base carroyée de l'INSEE

La première méthode d'impulsion adoptée dans cette étude repose sur l'utilisation de la base carroyée de l'Institut National de la Statistique et des Études Économiques (INSEE), qui fournit une infrastructure de données spatiales exhaustive et rigoureuse.

Cette base carroyée, constituée d'une grille régulière de carrés de surface unitaire d'1 km² (soit 371 821 carreaux dans le périmètre retenu), offre une représentation homogène de l'espace géographique du territoire français métropolitain. En utilisant les données imputées à chaque carreau, notamment le nombre de personnes par tranche d'âge, il est alors possible de coupler démographie et concentration de pollution atmosphérique avec une résolution fine. Cette approche granulaire offre la capacité d'identifier les zones à forts risques présentant des populations fortement exposées à des niveaux de pollution élevés. La suite de l'étude repose donc sur l'utilisation de la projection de ce plan spatial fourni par l'INSEE.

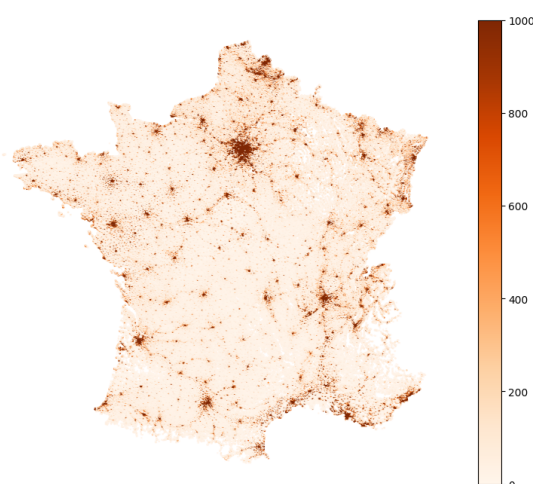


FIGURE 1 – Représentation de la densité de population par pixel d'1 km² selon la base « Revenus, pauvreté et niveau de vie en 2017 - Données carroyées », INSEE

Données *in situ* de pollution de l'air : Base GEOD'AIR

Afin de construire le module d'exposition de la démographie française aux différents polluants, la présente étude s'est appuyée sur les données *in situ* de la base *Open-data* appelée GEOD'AIR. Cette base compile les mesures de concentration de polluants atmosphériques tels que les particules fines ($PM_{2.5}$, PM_{10}), le dioxyde d'azote (NO_2) et l'ozone troposphérique (O_3). L'intégralité des données annuelles, couvrant la période de 2013 à 2022, a été collectée pour former un ensemble exhaustif de données.

Pour garantir la cohérence et la qualité des données, une phase de retraitement a été mise en œuvre. Ces ajustements visent à éliminer les stations ne disposant pas d'une profondeur temporelle suffisante, à homogénéiser les mesures et à minimiser les éventuels biais. De plus, les concentrations d'ozone ont été converties en valeurs de l'indicateur réglementaire *SOMO35*, un indice désormais largement adopté pour quantifier le fardeau de la pollution associé à ce composant. Enfin, les concentrations historiques ainsi corrigées sont agrégées par site de mesure, ce qui a conduit à la création d'un échantillon d'observations moyennées. Cette approche offre une vision précise des niveaux de pollution historiques en différents points géographiques spécifiques du territoire français.

Méthode d'impulsion deuxième : Interpolation spatiale

La deuxième méthode d'impulsion adoptée repose sur la conception d'un modèle d'interpolation spatiale permettant une quantification précise des niveaux de pollution en tout carreau du plan spatial retenu. On réalise alors une étape préliminaire à l'interpolation spatiale en associant les points géographiques des mesures de concentration de l'échantillon d'observations moyennées et le *raster* (cadrillage) généré par l'INSEE. Les carreaux contenant un site d'observation sont alors utilisés comme points d'échantillonnage pour l'interpolation de la variable régionalisée, permettant ainsi d'estimer les concentrations de polluants dans des zones sans données. À la suite d'une analyse des différents modèles d'interpolation spatiale, les modèles déterministes ont été écartés au profit des modèles géostatistiques. C'est donc un modèle type *kriegerage* qui a été implémenté. Afin de successivement sélectionner, implémenter et valider les modèles d'interpolation spatiale sur les données historiques pour chaque polluant, une procédure automatisée basée sur les principes de l'apprentissage automatique est utilisée et peut être synthétisée de la sorte :

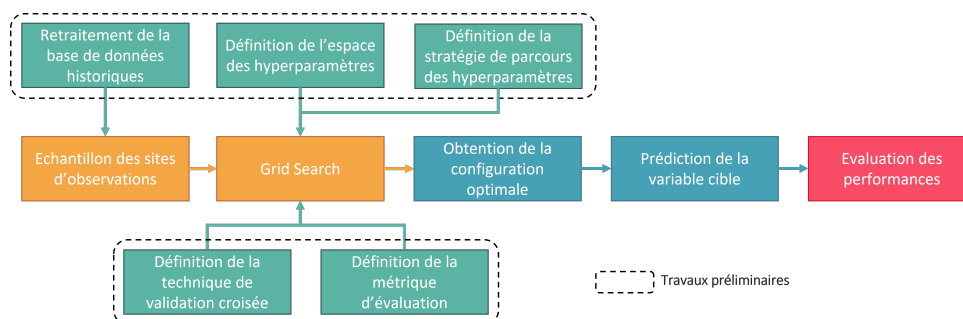


FIGURE 2 – Synthèse schématisée de la méthodologie adoptée pour la sélection, l'implémentation et la validation du modèle d'interpolation spatiale. Chaque étape ayant été réalisée pour chacun des polluants étudiés ($PM_{2.5}$, PM_{10} , NO_2 , *SOMO35*).

Automatisation de l’hyperparamétrage

Pour l’élection du modèle optimal d’interpolation spatiale, une méthode de type *Grid Search* a été employée, impliquant l’exploration exhaustive d’un espace prédéfini d’hyperparamètres. Cette approche consiste à tester différentes combinaisons de paramètres d’un modèle, en cherchant à identifier celle qui génère les résultats les plus performants selon une métrique d’évaluation définie au préalable. Dans le cadre de cette étude, cette dernière est la racine de l’erreur quadratique moyenne (RMSE). L’analyse variographique du modèle de *krigeage* est alors déléguée à la procédure d’hyperparamétrage. Cette dernière permet également au modèle de capter la détection de motifs géographiques et la compréhension des facteurs sous-jacents qui influencent la distribution spatiale des polluants atmosphériques historisés. Afin d’évaluer la performance du modèle, une stratégie de validation croisée de type *Leave-One-Out* a été déployée. Cette méthode consiste à exclure séquentiellement chaque point d’échantillonnage pour ensuite valider le modèle entraîné sur les données restantes. Bien que cette approche puisse être exigeante en termes de calculs, elle permet d’obtenir une évaluation robuste de la capacité prédictive du modèle et offre la possibilité de détecter d’éventuelles erreurs de prédiction.

Validation du modèle

Les métriques de performance telles que l’erreur absolue moyenne (MAE), le biais, la variance et la corrélation entre les valeurs prédites et observées sont ensuite calculées pour évaluer la précision et la cohérence du modèle. En couplant méthode de validation croisée et méthodes d’évaluation du modèle, il est possible de quantifier la fiabilité des estimations de la variable cible pour les carreaux dont on ne connaît pas la valeur. La table suivante présente les résultats obtenus pour les modèles optimaux de chaque polluant étudié :

Polluant	Val. Moyenne	r ²	Corrélation	MAE	Biais	Variance	RMSE
<i>NO</i> ₂	16,22	0,43	0,68	3,84	0,07	25,91	5,09
<i>PM</i> _{2.5}	10,63	0,14	0,37	1,46	0,03	3,81	1,95
<i>PM</i> ₁₀	16,95	0,38	0,62	1,59	0,02	4,87	2,21
<i>SOMO</i> ₃₅	12,28	0,71	0,84	1,56	-0,01	4,69	2,17

TABLE 2 – Résultats obtenus par polluant des métriques pour les modèles optimaux de *krigeage* sur les échantillons historiques.

Le modèle estimant la valeur de *SOMO*₃₅ ressort alors particulièrement en présentant de meilleures valeurs de métriques que pour les autres modèles.

Module prospectif

L’étape suivante de l’étude implique la construction d’un module prospectif visant à anticiper les évolutions futures de la pollution de l’air en tenant compte des variations démographiques et des changements atmosphériques. Dans cette optique, deux volets distincts sont abordés. D’une part, pour capturer l’impact des évolutions démographiques en France, une méthodologie pour faire évoluer la base carroyée de l’INSEE jusqu’à l’horizon 2050 est développée. Cette évolution basée sur le scénario central de l’INSEE projette la démographie de chaque pixel de la base carroyée en effectif et en âge en suivant les tendances démographiques départementales auquel il appartient.

D'autre part, pour la considération de l'évolution prospective de la pollution de l'air, le module de pollution est enrichi en introduisant des variables exogènes. Sur le constat des résultats d'une double analyse causalité-corrélation, et appuyé par la littérature, l'étude pose l'hypothèse que les évolutions de concentrations du cocktail, $PM_{2.5}$, PM_{10} , NO_2 sont indexées sur les données d'émission nationales tandis que le $SOMO35$ (O_3) est lui indexé sur les données de températures moyennes annuelles en France.

Ces variables exogènes sont extraites de différents scénarios climatiques réglementaires tels que RCP, SSP et GEA, qui couvrent une large gamme de perspectives futures à horizon 2050. De plus, une approche de régression exponentielle sur les données historiques des émissions ainsi que des scénarios réglementaires sont ajoutés. La combinaison de ces deux volets prospectifs, à savoir l'évolution démographique et l'évolution de la pollution, permet de créer un module d'exposition complet qui prend en compte les facteurs majeurs influençant la pollution atmosphérique future tout en considérant les dynamiques socio-démographiques anticipées. Ce double volet réalisé, une jointure spatiale est appliquée pour que chaque carreau de la base soit imputé du couple pollution-démographie.

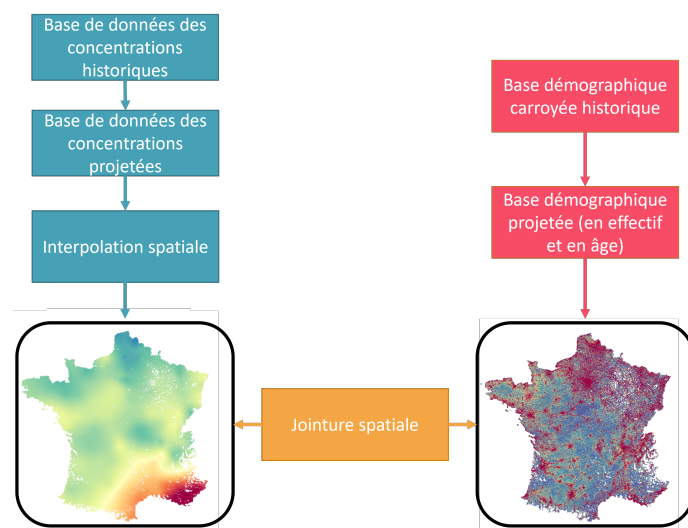


FIGURE 3 – Schéma procédural de l'évolution conjointe du double volet prospectif de l'étude : pollution/démographie. Le module d'exposition est obtenu finalement par jointure spatiale selon la clé d'identification de chaque pixel.

Évaluation de la matérialité

L'évaluation de la matérialité d'un risque consiste à déterminer l'importance et l'impact significatif que peut représenter un risque spécifique et les conséquences que ce dernier peut avoir sur une entité. En s'appuyant sur des critères définis dans le guide d'application de l'intégration de scénarios climatiques dans l'ORSA publié par l'EIOPA, une évaluation de la matérialité des risques est conduite pour chaque polluant. Sur la base de ces critères, il est déduit que les polluants $PM_{2.5}$, PM_{10} et NO_2 présentent des tendances à la baisse et ne généreront pas de mortalité additionnelle par rapport à la situation historique. De plus, la précision des modèles d'interpolation spatiale pour ces mêmes polluants s'est avérée discutable, leur étude quantitative a donc été écartée. En revanche, l'analyse révèle que

l’ozone (O_3) constitue une préoccupation en raison de son potentiel à générer une mortalité accrue dans un contexte de changement climatique. L’analyse quantitative doit donc être menée.

Méthode d’impulsion troisième : Industrialisation de l’outil AirQ+

Le logiciel AirQ+ est un outil pour évaluer l’impact de la pollution atmosphérique sur la santé. Il prend en compte plusieurs paramètres en entrée, tels que les concentrations de polluants atmosphériques, les données démographiques et les facteurs de risques spécifiques à la population étudiée. En utilisant ces informations, le logiciel génère des résultats dont le « nombre de décès attribuables ». Ce dernier correspond au nombre de décès dans une population imputable à l’exposition enregistrée.

La troisième méthode d’impulsion de ce mémoire consiste à « industrialiser » l’outil AirQ+ en le migrant vers un environnement Python. En effet, AirQ+ permet une évaluation de l’impact de la pollution sur chaque pixel de la base carroyée. Cependant, l’utilisation de cet outil s’est avérée inadaptée à cette tâche en raison de contraintes opérationnelles liées au nombre considérable de pixels du *raster*. Pour pallier cette limite, l’outil a été migré vers Python, offrant une meilleure capacité de traitement et une souplesse accrue pour le calcul d’impact à une échelle fine. Cette migration a finalement été validée par un audit des résultats obtenus à partir des deux méthodologies, assurant ainsi la cohérence et la fiabilité des résultats produits par la nouvelle méthode d’analyse.

Mise en application

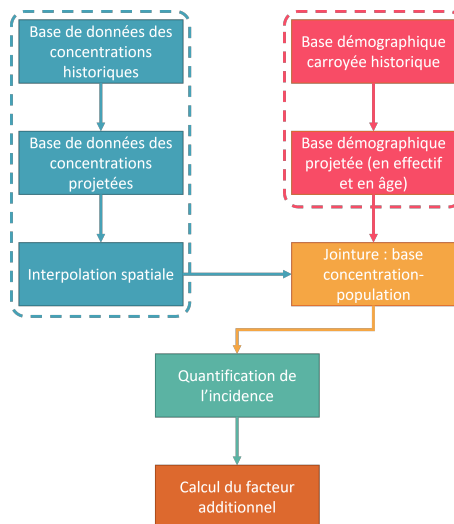


FIGURE 4 – *Compendium* procédural schématisé de la modélisation. L’ensemble des méthodes d’impulsion et des étapes de ce schéma sont implémentées dans un unique outil Python.

La méthodologie illustrée dans la *Figure 4* est mise en pratique, conduisant à la détermination des chocs additionnels. Ces derniers sont évalués à partir de l’estimation actuelle du

nombre de décès attribuables (dans la situation historique ou *baseline*), comparée à l'estimation future de cette même quantité (en fonction d'un horizon et d'un scénario spécifiés). Soit pour l'année de projection k les variables suivantes : F_k , le facteur additif de mortalité, D_0 , le nombre de décès historique par an attribuables à la pollution, en particulier à l'Ozone (valeur *baseline*) et D_k , le nombre de décès attribuables à la pollution à l'Ozone, le facteur additionnel pour l'année k est obtenu en réalisant :

$$F_k = \frac{D_k - D_0}{P_k} \times 100$$

Les résultats de facteurs additionnels moyens à granularité globale obtenus sont résumés dans la table suivante :

Scénario	2030	2040	2050
RCP4.5	0,0024 [0,0009;0,004]	0,0051 [0,0019;0,0087]	0,0062 [0,0022;0,0105]
RCP8.5	0,0021 [0,0008;0,0036]	0,0051 [0,0019;0,0087]	0,0067 [0,0024;0,0113]

TABLE 3 – Chocs additionnels nationaux en % obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Toutes classes d'âges confondues. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en **gras**.

Il convient de noter que l'exclusion des particules fines et du dioxyde d'azote conduit à la sous-estimation des résultats par rapport à l'analyse menée par l'ACPR. Cependant, la méthodologie d'agrégation des incidences de tous les polluants n'est pas communiquée, et les interactions complexes des effets croisés de chaque type de polluant ne présente pas de consensus dans la littérature.

Pour finir, l'avantage majeur de cette méthodologie réside dans sa capacité à fournir une évaluation détaillée des impacts en tout point de la France. En agrégeant les résultats à différentes échelles administratives, des niveaux locaux aux niveaux régionaux, il est possible de mieux appréhender la variabilité spatiale des risques.

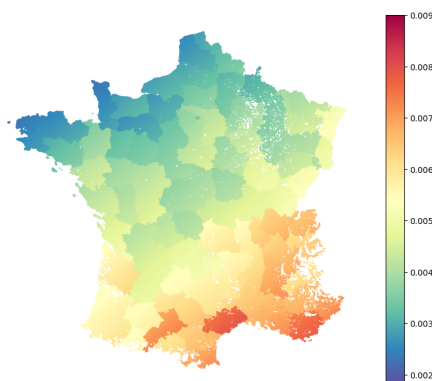


FIGURE 5 – Carte choroplèthe : Zonier - Chocs additionnels (en % de majoration de l'hypothèse initiale de mortalité) par département obtenus pour la pollution à l'Ozone selon le scénario climatique SSP5-8.5 pour l'horizon 2050 - Toutes classes d'âges confondues - Modélisation selon les valeurs médianes de la variable exogène.

La présente illustration révèle que l'impact additionnel sur la mortalité n'est pas seulement influencé par les différences d'âge, mais également par des variations géographiques significatives. Les départements situés dans le sud-est de la France présentent une exposition particulièrement élevée aux valeurs de *SOMO35*, en raison des conditions météorologiques favorables à sa formation. Ces disparités spatiales peuvent représenter une importance capitale pour les compagnies d'assurance et peuvent être incorporées dans l'élaboration d'une table dite « *taylor-made* », visant à procurer une représentation fidèle et complète du support de risque d'un portefeuille en utilisant des données géolocalisées.

Pour conclure, il convient de mettre en évidence que cette étude apporte une contribution significative à l'approche d'une problématique complexe liée à l'impact prospectif de la pollution atmosphérique. Cependant, certaines perspectives d'amélioration demeurent à explorer, notamment une investigation plus étendue de l'espace des hyperparamètres, l'inclusion de l'« *effet cocktail* », qui se caractérise par une potentialisation du facteur de risque en cas d'exposition conjointe à plusieurs polluants simultanément, ainsi que la quantification des incidences autres que la mortalité, telles que les répercussions sur la santé, les arrêts de travail et les effets de la pollution sur les facteurs de comorbidité.

Synthesis note

Contextual setting

The current context is marked by the convergence of crucial issues where the regulatory sector, represented by the *Autorité de Contrôle Prudentiel et de Résolution (ACPR)*, is undertaking innovative initiatives to assess the resilience of the financial and insurance system to climate risks. Within this singular framework, the climate pilot exercises conducted by the ACPR are of strategic importance as a methodological foundation for understanding the potential vulnerabilities of the financial sector to climate disruption. However, it is crucial to note that climate change transcends the financial aspects, as it also has substantial repercussions on policyholders' quality of the living environment and, by extension, on life and health insurance.

One of the major risks identified in life and health insurance concerns air pollution and its repercussions on health, work stoppages and mortality. The prudential regulator has received an independent study on the subject, conducted by the broker AON. The aim of the study is to assess the potential impact of atmospheric pollution on these various types of cover in a forward-looking context. In fact, the complex link between climate change and atmospheric pollution may affect the solvency of stakeholders by inducing undesirable changes in claims rates. These interactions are reflected in the form of additional factors, designed to reinforce the conventional assumptions made by insurers for each of the coverages mentioned above.

These additional factors, summarised in the form of tables, are broken down according to various parameters such as the projection horizon, age groups and climate scenarios considered (in particular RCP8.5 for the 2020 financial year and RCP4.5 for the 2023 financial year). The results are communicated at two levels of granularity : a global granularity for all of France and a fine granularity for nine urban cities.

	2023 - RCP4.5			2020 - RCP8.5		
Granularity	2024-2030	2031-2040	2041-2050	2021-2030	2031-2040	2041-2050
France	0,02 %	0,02 %	0,02 %	0,02 %	0,02 %	0,03 %
Île de France	0,02 %	0,03 %	0,03%	0,02 %	0,03 %	0,04 %
...
Toulouse	0,01 %	0,01 %	0,02 %	0,01 %	0,02 %	0,02 %

TABLE 4 – Additional mortality factors communicated by the ACPR in its 2020 and 2023 climate pilot exercises.

These results help insurers to anticipate the financial impact of health problems associa-

ted with future air pollution and enhance their understanding. This study illustrates the importance of integrating environmental factors into the risk assessment and insurance product design processes in response to current health and sustainability issues.

The purpose of this report is to supplement the study carried out by AON by identifying ways of improving the analysis of the impact of air pollution on mortality, the only risk included in the study. By focusing on aspects such as spatial modelling at a finer granularity, the methodology strives to calculate more representative shocks and to highlight other crucial dimensions by introducing the notion of materiality assessment. The aim of this work is to deepen our understanding of the effects of air pollution by providing new perspectives and an advanced methodological approach using so-called "impulse" methods.

Primary impulse method : The INSEE gridded database

The first impulse method adopted in this study is based on the use of the *Institut National de la Statistique et des Études Économiques* (INSEE) gridded database, which provides an exhaustive and rigorous spatial data infrastructure.

This grid base, consisting of a regular grid of squares with a unit area of 1 km² (i.e. 371 821 squares in the selected perimeter), provides a homogenous representation of the geographical space of mainland France. By using the data attributed to each tile, in particular the number of people per age group, it is then possible to couple demography and air pollution concentration with fine resolution. This granular approach makes it possible to identify high-risk areas with populations that are highly exposed to high levels of pollution. The rest of the study is therefore based on the projection of this spatial plan provided by INSEE.

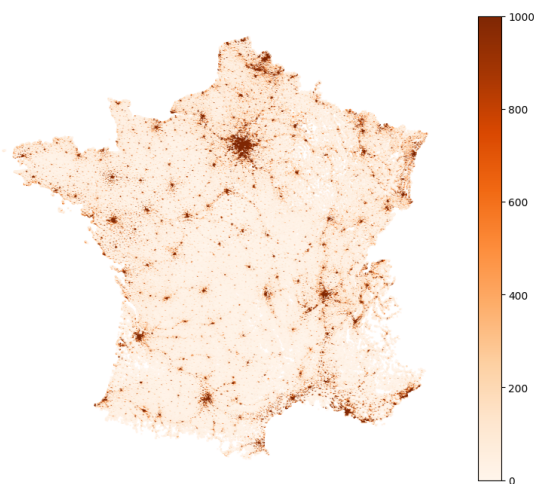


FIGURE 6 – Representation of population density per pixel of 1 km² according to the « *Revenus, pauvreté et niveau de vie en 2017 - Données carroyées* » database, INSEE

In situ air pollution data : GEOD’AIR database

In order to construct the French demographic exposure module for the various pollutants, our study was based on the *in situ* data in an open database called GEOD’AIR. This database compiles concentration measurements of atmospheric pollutants such as fine particles ($PM_{2.5}$, PM_{10}), nitrogen dioxide (NO_2) and tropospheric ozone (O_3). All annual data, covering the period from 2013 to 2022, have been collected to form a comprehensive dataset.

To guarantee the consistency and quality of the data, a reprocessing phase was carried out. These adjustments were aimed to eliminate stations with insufficient temporal depth, homogeneise the measurements and minimise any biases. In addition, ozone concentrations were converted into regulatory indicator *SOMO35* values, an index now widely adopted for quantifying the pollution burden associated with this component. Finally, the historical concentrations corrected in this way were aggregated by measurement site, leading to the creation of a sample of averaged observations. This approach provides an accurate picture of historical pollution levels at specific geographical points in France.

Second impulse method : Spatial interpolation

The second impulse method adopted is based on the design of a spatial interpolation model enabling pollution levels to be accurately quantified in any tile of the spatial plane selected. A preliminary stage in the spatial interpolation is then carried out by associating the geographical points of the concentration measurements from the sample of averaged observations with the grid generated by INSEE. The tiles containing an observation site are then used as sampling points for the interpolation of the regionalised variable, making it possible to estimate pollutant concentrations in areas with no data. Following an analysis of the various spatial interpolation models, deterministic models were discarded in favour of geostatistical models. A model of the kriging type was therefore implemented. In order to successively select, implement and validate the spatial interpolation models on the historical data for each pollutant, an automated procedure based on the principles of machine learning was used, which can be summarised as follows :

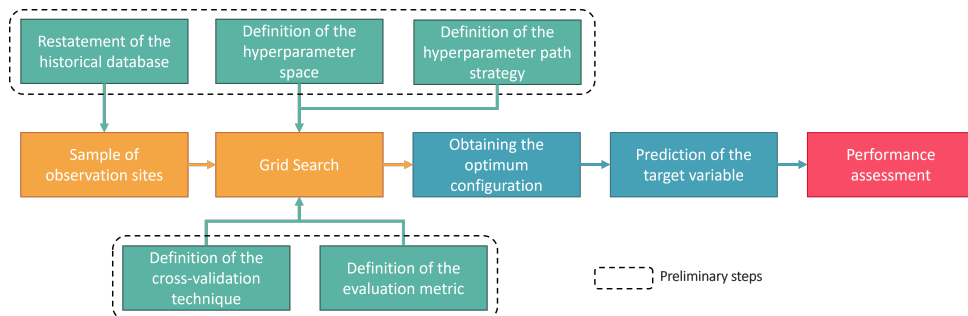


FIGURE 7 – Schematic summary of the methodology adopted for the selection, implementation and validation of the spatial interpolation model. Each step was carried out for each of the pollutants studied ($PM_{2.5}$, PM_{10} , NO_2 , *SOMO35*).

Automation of hyperparameter tuning

To select the optimal spatial interpolation model, a Grid Search type method was used, involving exhaustive exploration of a predefined space of hyperparameters. This approach consists of testing different combinations of model parameters, seeking to identify the one that generates the best results according to a predefined evaluation metric. In the context of this study, this metric is the root mean square error (RMSE). The variographic analysis of the kriging model is then delegated to the hyperparameterisation procedure. The latter also enables the model to capture the detection of geographical patterns and an understanding of the underlying factors influencing the spatial distribution of historically recorded air pollutants. To assess the model's performance, a Leave-One-Out cross-validation strategy was deployed. This method consists of sequentially excluding each sampling point and then validating the trained model on the remaining data. Although this approach can be computationally demanding, it provides a robust assessment of the model's predictive capability and offers the possibility of detecting any prediction errors.

Model validation

Performance metrics such as mean absolute error (MAE), bias, variance and correlation between predicted and observed values are calculated to assess the accuracy and consistency of the model. By combining cross-validation and model evaluation methods, it is possible to quantify the reliability of the estimates of the target variable for tiles whose value is unknown. The following table presents the results obtained for the optimal models for each pollutant studied :

Pollutant	Mean Val.	r^2	Correlation	MAE	Bias	Variance	RMSE
<i>NO</i> ₂	16,22	0,43	0,68	3,84	0,07	25,91	5,09
<i>PM</i> _{2.5}	10,63	0,14	0,37	1,46	0,03	3,81	1,95
<i>PM</i> ₁₀	16,95	0,38	0,62	1,59	0,02	4,87	2,21
<i>SOMO</i> ₃₅	12,28	0,71	0,84	1,56	-0,01	4,69	2,17

TABLE 5 – Results obtained by pollutant metric for the optimal models of kriging on historical samples.

The model estimating the value of *SOMO*₃₅ therefore comes out particularly well, with better values on various metrics.

Forecast module

The next stage of the study involves the construction of a forward-looking module designed to anticipate future changes in air pollution, taking into account demographic variations and atmospheric changes. With this in mind, two distinct aspects are addressed. Firstly, in order to capture the impact of demographic change in France, a methodology is developed for evolving the INSEE grid base up to 2050. This evolution, based on the INSEE's central scenario, projects the demographics of each pixel in the grid base in terms of numbers and age, following the demographic trends of the « *département* » to which it belongs.

Secondly, to take into account the future evolution of air pollution, the pollution module is enriched by introducing exogenous variables. Based on the results of a dual causality-correlation analysis, and supported by the literature, the study assumes that changes in

the concentrations of the cocktail, $PM_{2.5}$, PM_{10} , NO_2 are indexed to national emission data, while $SOMO35$ (O_3) is indexed to average annual temperature data for France.

These exogenous variables are extracted from various regulatory climate scenarios such as RCP, SSP and GEA, which cover a wide range of future outlooks to 2050. In addition, an exponential regression approach on historical emissions data and regulatory scenarios are added. The combination of these two forward-looking components - demographic change and pollution trends - makes it possible to create a comprehensive exposure module that takes into account the major factors influencing future air pollution, while also considering anticipated socio-demographic dynamics. Once this dual component has been achieved, a spatial join is applied so that each tile in the database is imputed with the pollution-demography pairing.

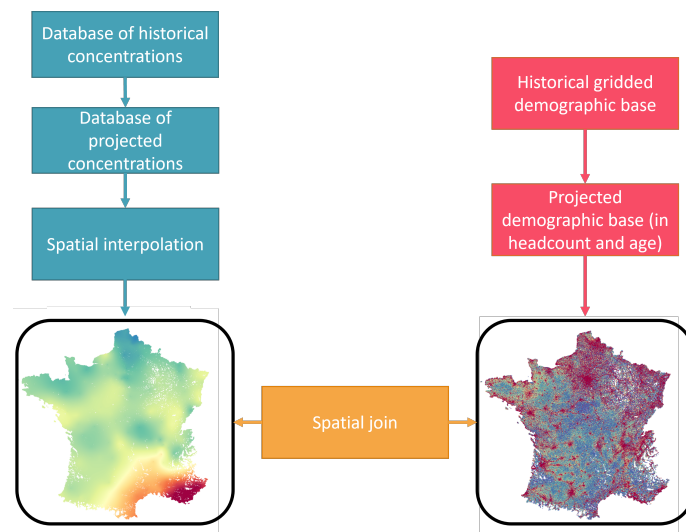


FIGURE 8 – Procedural diagram of the joint evolution of the two prospective aspects of the study : pollution/demography. The exposure module is finally obtained by spatial join according to the identification key of each pixel.

Materiality Assessment

Assessing the materiality of a risk consists of determining the importance and significant impact that a specific risk may represent and the consequences that it may have on an entity. Based on the criteria defined in the application guide for incorporating climate scenarios into ORSA published by EIOPA, an assessment of the materiality of risks is carried out for each pollutant. On the basis of these criteria, it has been deduced that the $PM_{2.5}$, PM_{10} and NO_2 pollutants are on a downward trend and will not generate any additional mortality compared with the historical situation. In addition, the accuracy of the spatial interpolation models for these same pollutants proved questionable, so their quantitative study was discarded. On the other hand, the analysis reveals that ozone (O_3) is a cause for concern because of its potential to generate increased mortality in a context of climate change. A quantitative analysis must therefore be carried out.

Third impulse method : Industrialisation of the *AirQ+* tool

The AIRQ+ software is a tool for assessing the impact of air pollution on health. It takes into account several input parameters, such as air pollutant concentrations, demographic data and risk factors specific to the population under study. Using this information, the software generates results including the "number of attributable deaths". The latter corresponds to the number of deaths in a population attributable to the exposure recorded.

The third method used in this dissertation is to 'industrialise' the AirQ+ tool by migrating it to a Python environment. AirQ+ can be used to assess the impact of pollution on each pixel of the grid base. However, the use of this tool proved unsuitable for this task due to operational constraints linked to the considerable number of pixels in the grid. To overcome this limitation, the tool was migrated to Python, offering greater processing capacity and increased flexibility for fine-scale impact calculations. This migration was finally validated by an audit of the results obtained using the two methodologies, thus ensuring the consistency and reliability of the results produced by the new analysis method.

Implementation

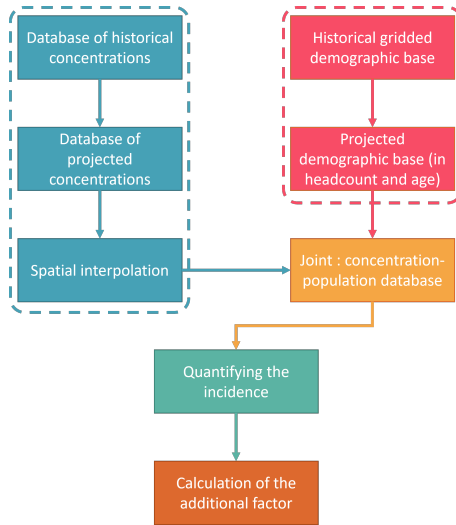


FIGURE 9 – Procedural *Compendium* schematised modelling. All the impulse methods and steps in this schema are implemented in a single Python tool.

The methodology illustrated in *Figure 9* is put into practice, leading to the determination of additional shocks. These are evaluated on the basis of the current estimate of the number of attributable deaths (in the historical situation or baseline), compared with the future estimate of this same quantity (according to a specified horizon and scenario). Given the following variables for the projection year k : F_k , the additive mortality factor, D_0 , the historical number of deaths per year attributable to Ozone pollution (value baseline) and D_k , the number of deaths attributable to Ozone pollution, the additional factor for year k is obtained by performing :

$$F_k = \frac{D_k - D_0}{P_k} \times 100$$

The results for average additional factors at a global granularity are summarised in the table below :

Scenario	2030		2040		2050	
RCP4.5	0,0024	[0,0009;0,004]	0,0051	[0,0019;0,0087]	0,0062	[0,0022;0,0105]
RCP8.5	0,0021	[0,0008;0,0036]	0,0051	[0,0019;0,0087]	0,0067	[0,0024;0,0113]

TABLE 6 – Additional national shocks in % obtained for ozone pollution according to different climate scenarios for the years 2030, 2040 and 2050 - All age groups combined. The results in square brackets represent the values at the lower and upper bounds of the 95 % confidence interval of the factor in **bold**.

It should be noted that the exclusion of fine particles and nitrogen dioxide leads to an underestimation of the results compared with the analysis carried out by the ACPR. However, the methodology for aggregating the impact of all the pollutants has not been communicated, and there is no consensus in the literature on the complex interactions of the cross-effects of each type of pollutant.

Finally, the major advantage of this methodology lies in its ability to provide a detailed assessment of impacts at any point in France. By aggregating the results at different administrative scales, from local to regional, it is possible to gain a better understanding of the spatial variability of risks.

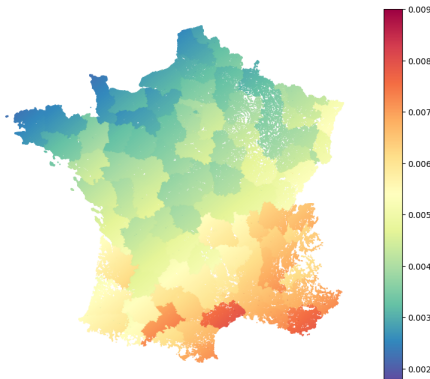


FIGURE 10 – Choropleth map : Zonier - Additional shocks per « *département* » obtained for Ozone pollution according to the SSP5-8.5 climate scenario for the 2050 horizon - All age classes combined - Modelled according to the median values of the exogenous variable.

This illustration shows that the additional impact on mortality is influenced not only by age differences, but also by significant geographical variations. Departments located in the south-east of France show a particularly high exposure to *SOMO35* values, due to the meteorological conditions favourable to its formation. These spatial disparities can be of vital importance to insurance companies, and can be incorporated into the creation of a so-called « *taylor-made* » table, designed to provide an accurate and complete representation of the risk support of a portfolio using geolocalised data.

In conclusion, it should be emphasised that this study makes a significant contribution to the approach to a complex problem linked to the prospective impact of air pollution.

However, certain areas for improvement remain to be explored, in particular a more extensive investigation of the hyperparameter space, the inclusion of the « *cocktail effect* », which is characterised by a potentiation of the risk factor in the event of joint exposure to several pollutants simultaneously, and the quantification of impacts other than mortality, such as repercussions on health, absences from work and the effects of pollution on co-morbidity factors.

Remerciements

Tout d'abord, j'aimerais remercier Monsieur Christophe EBERLE, président fondateur d'Optimind, pour m'avoir permis de réaliser mon stage de fin d'études d'école d'ingénieurs ainsi que cette année d'alternance au sein de son entreprise.

J'adresse également mes remerciements pour la confiance portée en ma candidature à Monsieur Gildas ROBERT, Madame Valérie DEPPE ainsi que Madame Chloé PARFAIT, Partners de la practice AFS (Actuarial & Financial Services) dans laquelle j'ai évolué dans d'excellentes conditions durant cette année.

J'aimerais ensuite, adresser mes remerciements à mes tuteurs en entreprise pour ce mémoire, Messieurs Guillaume BESSON et Victor DEQUATRE pour leur accompagnement infailible, leur expertise et le support qui m'a été prodigué.

Je souhaite également remercier Monsieur Emmanuel BERTHELE, Directeur Recherche et Développement à Optimind, tant pour ses orientations que pour la qualité de ses conseils.

Je souhaiterais ensuite remercier l'ensemble des professeurs que j'ai pu rencontrer lors de mon cursus au sein de l'EURO Institut d'Actuariat, notamment Monsieur Franck VERMET en qualité de directeur, pour le sérieux et la qualité des enseignements qui m'ont été dispensés, mais également Monsieur Pierre AILLOT, tuteur académique de ce mémoire, pour ses précieux conseils, son temps et sa disponibilité.

Je voudrais ensuite remercier l'ensemble des consultants de la practice Actuarial and Financial Services d'Optimind, pour leur accueil chaleureux ainsi que pour leur esprit d'équipe.

Je souhaiterais adresser une mention particulière à Madame Chloé PARFAIT et Monsieur Hugo DIJOURD pour leurs conseils avisés ainsi qu'à mes collègues alternants, Annabel BERARD, Lucas BLANCHETON et Lisa RUEILLAN avec qui j'ai partagé cette année ainsi qu'à Maxime BEILLAN et Ana FERREIRA pour leurs précieux conseils de relecture.

J'aimerais ensuite remercier, l'ensemble des collaborateurs d'Optimind Paris avec qui j'ai pu travailler et échanger, et en particulier mes supérieurs qui ont su me faire confiance durant toute mon alternance, ainsi que toutes les personnes avec qui je me suis lié d'amitié.

J'aimerais enfin remercier ma famille, mes amis et particulièrement mon frère Jules ALLAIRE pour son aide lors de la rédaction de ce mémoire, mais également Thibaut BAYEMI et Clémence LEFEVRE pour leurs encouragements dévoués.

Glossaire

AASQA	<i>Associations Agréées de Surveillance de la Qualité de l'Air indépendantes</i>
ACCMIP	<i>Atmospheric Chemistry and Climate Model Intercomparison Project</i>
ACPR	<i>Autorité de Contrôle Prudentiel et de Résolution</i>
CAMS	<i>Copernicus Atmosphere Monitoring Service</i>
COV	<i>Composés Organiques Volatils</i>
COVNM	<i>Composés Organiques Volatils Non Méthaniques</i>
CRPA	<i>Code des relations entre le public et l'administration</i>
CPATLD	<i>Convention sur la Pollution Atmosphérique Transfrontière à Longue Distance</i>
EIOPA	<i>European Insurance and Occupational Pensions Authority</i>
EQIS	<i>Evaluations Quantitatives d'Impact sur la Santé</i>
Filosofi	<i>Fichier localisé social et fiscal</i>
GEA	<i>Global Energy Assessment</i>
GES	<i>Gaz à Effet de Serre</i>
INERIS	<i>Institut national de l'environnement industriel et des risques</i>
INSEE	<i>Institut National de la Statistique et des Études Économiques</i>
LCSQA	<i>Laboratoire Central de Surveillance de la Qualité de l'Air</i>
LoB	<i>Lines of business</i>
OMS	<i>Organisation Mondiale de la Santé</i>
ONU	<i>Organisation des Nations Unies</i>
ORSA	<i>Own Risk and Solvency Assessment</i>
RCP	<i>Representative Concentration Pathways</i>
RGPD	<i>Règlement Général sur la Protection des Données</i>
SDES	<i>Service de la Donnée et des Études Statistiques</i>
SOMO35	<i>Sum Of Means Over 35 ppb (daily maximum 8-hour)</i>
SSP	<i>Shared Socioeconomic Pathways</i>
RCP	<i>Representative Concentration Pathways</i>
RGPD	<i>Règlement général sur la protection des données</i>

Table des matières

Citation	3
Résumé	4
Abstract	5
Note de synthèse	6
Synthesis note	14
Remerciements	22
Glossaire	23
Table des matières	24
Introduction	33
1 Cadre de l'étude	35
1.1 Assurance et changement climatique	35
1.2 La pollution de l'air	49
1.3 Méthodologies de quantification de l'incidence de la pollution atmosphérique	61
2 Méthodologie de construction d'un zonier de la pollution de l'air sur données historiques	67
2.1 Sources de données utilisées	67
2.2 Interpolation spatiale	78
2.3 <i>Machine Learning</i> : Élection d'un modèle de <i>krigeage</i> - Méthodologie et application	91

2.4	Implémentation et validation du modèle	97
3	Approches de génération et d'utilisation de scénarios prospectifs	106
3.1	Projection de la population en base carroyée	106
3.2	Dynamique prospective des polluants	110
3.3	Forces, limites et améliorations de la méthodologie retenue	118
4	Impact sanitaire prospectif de la pollution de l'air : Résultats	120
4.1	Étude de la matérialité par polluant	120
4.2	<i>Compendium</i> procédural de la modélisation	122
4.3	Résultats : chocs additionnels prospectifs en mortalité	124
4.4	Applications et réutilisations en assurance	132
	Conclusion	135
A	Annexes	138
A.1	Seuils règlementaires par polluant et par origine normative.	138
A.2	Indicateurs de santé liés à la morbidité et la mortalité.	139
A.3	Les données de population : Données au carreau de 1 km ² en France de l'INSEE	141
A.4	Nombre de sites de surveillance de la pollution de fond en France	142
A.5	Cartes choroplèthes	142
A.6	Résultats de facteurs additionnels par tranches d'âges	144

Table des figures

1	Représentation de la densité de population par pixel d'1 km ² selon la base « Revenus, pauvreté et niveau de vie en 2017 - Données carroyées », INSEE	7
2	Synthèse schématisée de la méthodologie adoptée pour la sélection, l'implémentation et la validation du modèle d'interpolation spatiale. Chaque étape ayant été réalisée pour chacun des polluants étudiés (<i>PM</i> _{2.5} , <i>PM</i> ₁₀ , <i>NO</i> ₂ , <i>SOMO</i> ₃₅).	8
3	Schéma procédural de l'évolution conjointe du double volet prospectif de l'étude : pollution/démographie. Le module d'exposition est obtenu finalement par jointure spatiale selon la clé d'identification de chaque pixel. . . .	10
4	<i>Compendium</i> procédural schématisé de la modélisation. L'ensemble des méthodes d'impulsion et des étapes de ce schéma sont implémentées dans un unique outil Python.	11
5	Carte choroplèthe : Zonier - Chocs additionnels (en % de majoration de l'hypothèse initiale de mortalité) par département obtenus pour la pollution à l'Ozone selon le scénario climatique SSP5-8.5 pour l'horizon 2050 - Toutes classes d'âges confondues - Modélisation selon les valeurs médianes de la variable exogène.	12
6	Representation of population density per pixel of 1 km ² according to the « <i>Revenus, pauvreté et niveau de vie en 2017 - Données carroyées</i> » database, INSEE	15
7	Schematic summary of the methodology adopted for the selection, implementation and validation of the spatial interpolation model. Each step was carried out for each of the pollutants studied (<i>PM</i> _{2.5} , <i>PM</i> ₁₀ , <i>NO</i> ₂ , <i>SOMO</i> ₃₅).	16
8	Procedural diagram of the joint evolution of the two prospective aspects of the study : pollution/demography. The exposure module is finally obtained by spatial join according to the identification key of each pixel.	18
9	Procedural <i>Compendium</i> schematised modelling. All the impulse methods and steps in this schema are implemented in a single Python tool.	19
10	Choropleth map : Zonier - Additional shocks per « <i>département</i> » obtained for Ozone pollution according to the SSP5-8.5 climate scenario for the 2050 horizon - All age classes combined - Modelled according to the median values of the exogenous variable.	20

1.1	Schéma récapitulatif de la procédure de traduction des risques climatiques en risques prudentiels traditionnels dans l'ORSA. (<i>Finalyse, Climate change risks in the ORSA</i> [1])	36
1.2	Exemples de correspondance entre les risques climatiques et les risques prudentiels traditionnels. (<i>Finalyse, Climate change risks in the ORSA</i> [1])	37
1.3	Exemple d'une matrice de matérialité à horizon long terme (EIOPA, <i>Application Guidance on Running Climate Change Materiality Assessment and Using Climate Change Scenarios in ORSA</i> , 2023 [2])	39
1.4	Schéma simplifié des étapes d'évaluation de la matérialité d'un risque climatique	40
1.5	Définition des horizons de temps dans l'ORSA (<i>EIOPA 2021a</i> [3])	40
1.6	Définition des horizons de temps dans l'ORSA (<i>EIOPA, 2021a</i> [3])	42
1.7	Tableau récapitulatif des impacts additifs de la pollution atmosphérique sur les décès, les frais de soins et les arrêts de travail par géographie et horizon de projection (<i>Drif et al., 2020</i> [4])	47
1.8	Schéma de la formation d'ozone troposphérique (O_3)	52
1.9	Part des secteurs d'activité dans les émissions anthropiques de NO_x , $COVNM$, PM_{10} et $PM_{2.5}$ en 2021 (<i>Citepa</i> , 2022)	54
1.10	Frise chronologique des principales directives relatives à l'amélioration de la qualité de l'air en Europe	55
1.11	Frise chronologique des principales directives relatives à l'amélioration de la qualité de l'air en France	55
1.12	Évolution des émissions anthropiques de quelques polluants en France métropolitaine (Corse exclue) en indice base 100 (<i>Citepa, 2022</i>)	56
1.13	Évolution des concentrations moyennes annuelles pour les polluants SO_2 , NO_2 , O_3 , PM_{10} et $PM_{2.5}$, en fond urbain, en France métropolitaine (Corse exclue) (<i>Geod'air, 2022</i>)	57
1.14	Évolution de la concentration moyenne d'ozone (O_3) et de la température moyenne annuelle en France en base 100 (période 2000-2020)	60
1.15	Schéma méthodologique de la réalisation d'une EQIS-PA, Santé publique France 2021	63
2.1	Données ré-analysées de concentration moyennes de $PM_{2.5}$ au 1 avril 2018 issues du projet CAMS	69
2.2	Portail GEOD'AIR : Moyenne annuelle de Dioxyde d'Azote ($\mu g/m^3$) en 2013 aux stations de mesures	71
2.3	Représentation de la densité de population par pixel d'1 km^2 selon la base « Revenus, pauvreté et niveau de vie en 2017 - Données carroyées », INSEE	74
2.4	Pyramide des âges : Population en 2050, Scénario Central, INSEE	75

2.5	Évolution temporelle de la part attribuable de la France dans les émissions européennes (EU28) de NO_x	77
2.6	Historique de la température moyenne annuelle en °C en France métropolitaine	78
2.7	Schéma du principe de l'interpolation spatiale	79
2.8	Exemple d'interpolation spatiale par <i>spline</i> cubique réalisée dans un plan en 3 dimensions : vue en perspective.	80
2.9	Principe de l'effet écran : le site s_2 et le site s_7 peuvent potentiellement se masquer l'un l'autre en raison de leur proximité spatiale	81
2.10	Diagramme de Voronoi - Interpolation par plus proche voisin : application sur un champ $\mathcal{D} \subset R^2$ de 50 observations aléatoires en rose (<i>germes</i>) - représentation sur une grille régulière - vue de dessus. On remarque l'effet de crénelage de la méthodologie	82
2.11	Interpolation polynomiale linéaire : application sur un champ $\mathcal{D} \subset R^2$ de 100 observations réparties aléatoirement - représentation par courbes de niveau - vue de dessus. On remarque l'effet pyramidal de la méthodologie	83
2.12	Réduction du poids relatif en fonction de la distance selon plusieurs valeurs de p	84
2.13	Interpolation barycentrique par inverse des distances : application sur un champ $\mathcal{D} \subset R^2$ de 100 observations réparties aléatoirement pour différentes valeurs de p - représentation par courbes de niveau - vue de dessus. On observe que plus grande est la valeur de p plus l'influence aux valeurs les plus proches du point interpolé est grandissante.	85
2.14	<i>Spline</i> d'interpolation (exacte) - cubique : application sur un champ $\mathcal{D} \subset R^2$ de 100 observations réparties aléatoirement - représentation par courbes de niveau - vue de dessus.	86
2.15	Schéma explicatif de la réalisation d'un échantillonnage aléatoire type <i>Train-Test</i>	95
2.16	Schéma explicatif de la réalisation d'un échantillonnage aléatoire type <i>K-Fold</i> pour $k = 3$	96
2.17	Schéma explicatif de la réalisation d'un échantillonnage aléatoire type <i>Leave-One-Out</i>	97
2.18	Synthèse schématisée de la méthodologie adoptée pour la sélection, l'implémentation et la validation du modèle d'interpolation spatiale. Chaque étape ayant été réalisée pour chacun des polluants étudiés ($PM_{2.5}$, PM_{10} , NO_2 , $SOMO35$).	97
2.19	Effet du biais et de la variance sur la précision des prédictions d'un modèle de <i>Machine Learning</i> . Le cercle rose représente la juste valeur à estimer. . .	102
2.20	Indicateurs statistiques calculés à l'aide de la technique de validation croisée <i>Leave-One-Out</i> sur les valeurs moyennes annuelles de différents polluants mesurées et estimées sur les stations de fond pour les années 2013 à 2022. .	103

2.21	Carte choroplèthe : Résultat du module d'interpolation spatiale appliqué au <i>raster</i> de la base carroyée de l'INSEE pour les valeurs du SOMO35 en 2021. Les espaces blanchis sont les pixels en France ne comportant aucun habitant. Ce résultat peut être comparé à celui de l'INERIS via l' <i>Annexe A.2</i> . On observe que le sud-est de la France est assujéti à des températures plus élevées et est donc plus exposé que le reste du pays.	104
3.1	Pyramides des âges pour le département des Hauts-de-Seine (92) en 2018 et en 2070, Scénario Central Insee.	108
3.2	Projection de la température annuelle moyenne en France à horizon 2050 selon les scénarios RCP (CMIP5) et SSP (CMIP6) - Valeur médiane issue d'un ensemble de modèles (<i>World Bank Group, Climate Change Knowledge Portal</i>).	112
3.3	Régressions exponentielles des données historiques d'émissions de NO_x selon différentes profondeurs d'historique - Les profondeurs historiques ici étudiées correspondent aux intervalles 1980-2020, 1990-2020, 2000-2020.	115
3.4	Données d'émissions de NO_x historiques et projetées selon les différents scénarios présentés en 3.2.2 et en 3.2.3. On constate que les émissions sont à tendance baissière, excepté pour le scénario réglementaire D qui n'est pas réaliste.	116
3.5	Synthèse schématisée de la méthodologie adoptée pour le rapatriement des index prospectifs et leur traduction en concentration aux stations.	117
4.1	<i>Compendium</i> procédural schématisé de la modélisation	123
4.2	Projection de la température annuelle moyenne en France à horizon 2050 selon les scénarios RCP (CMIP5) et SSP (CMIP6) - Valeurs correspondant au 90ème centile issue d'un ensemble de modèles (<i>World Bank Group, Climate Change Knowledge Portal</i>).	129
4.3	Carte choroplèthe : Zonier - Chocs additionnels par département obtenus pour la pollution à l'Ozone selon le scénario climatique SSP5-8.5 pour l'horizon 2050 - Toutes classes d'âges confondues - Modélisation selon les valeurs médianes de la variable exogène.	130
A.1	Indicateurs de santé liés à la morbidité et la mortalité quantifiables à l'aide d'AirQ+	140
A.2	Carte choroplèthe des valeurs de <i>SOMO35</i> en France en 2021, INERIS. . .	142
A.3	Légende des valeurs de <i>SOMO35</i> , INERIS. Une valeur de <i>SOMO35</i> s'élevant à 3 000 est égale $3\ 000/365.25 = 8,21$ en concentration moyenne.	143
A.4	Carte choroplèthe : Île de France - Nombre de décès attribuables à la pollution à l'Ozone selon le scénario climatique SSP5-8.5 pour l'horizon 2050 - Toutes classes d'âges confondues - Modélisation selon les valeurs médianes de la variable exogène.	143

Liste des tableaux

1	Facteurs additionnels de la mortalité communiqués par l'ACPR dans ses exercices pilote climatique de 2020 et 2023.	6
2	Résultats obtenus par polluant des métriques pour les modèles optimaux de <i>krigeage</i> sur les échantillons historiques.	9
3	Chocs additionnels nationaux en % obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Toutes classes d'âges confondues. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en gras	12
4	Additional mortality factors communicated by the ACPR in its 2020 and 2023 climate pilot exercises.	14
5	Results obtained by pollutant metric for the optimal models of kriging on historical samples.	17
6	Additional national shocks in % obtained for ozone pollution according to different climate scenarios for the years 2030, 2040 and 2050 - All age groups combined. The results in square brackets represent the values at the lower and upper bounds of the 95 % confidence interval of the factor in bold	20
1.1	Exemple de tableau récapitulatif de la matérialité des risques physiques aigus et chroniques par horizon temporel (EIOPA, <i>Application Guidance on Running Climate Change Materiality Assessment and Using Climate Change Scenarios in ORSA</i> , 2023 [2])	39
1.2	Tableau récapitulatif des métriques d'évaluation par type d'indicateur (EIOPA, 2021 [2])	41
2.1	Nombre de sites de surveillance de la pollution de fond en France par polluant atmosphérique pour les années 2013 à 2022.	71
2.2	Description des variables	72
2.3	Description des variables du jeu de données « <i>National Emission reductions Commitments (NEC) Directive emission inventory data</i> »	76
2.4	Synthèse des méthodes d'interpolation spatiale déterministes	87

2.5	Nombre de stations conservées par polluant (<i>post</i> -retraitements effectués en 2.4.2)	99
2.6	Configurations optimales du modèle de <i>krigeage</i> pour chaque polluant, obtenues à l'issu du <i>Grid Search</i>	100
2.7	Résultats obtenus par polluant des métriques pour les modèles optimaux de <i>krigeage</i> sur les échantillons historiques.	101
3.1	Comparaison de la population entre 2020 et 2050 - Données réelles et données estimées (INSEE) - Les populations sont exprimées en millions.	110
3.2	Incidence qualitative des scénarios climatiques RCP sur la pollution de l'air.	111
3.3	Niveaux d'émissions par polluant en France selon les 4 scénarios de l'INERIS en (Gg=1 000 tonnes) (<i>Rapport du Sénat, 2014 [5]</i>)	115
3.4	Objectifs de réduction d'émissions par polluant fixés pour la France (exprimés en % par rapport à 2005) (<i>Santé Publique France, 2023 [6]</i>)	116
4.1	Synthèse de la matérialité du risque mortalité induit par chaque polluant atmosphérique à différents horizons de temps. Les horizons « Court », « Moyen » et « Long » terme font référence aux horizons de temps de l'ORSA climatique présenté en <i>Figure 1.5</i>	122
4.2	Chocs additionnels nationaux en % obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Toutes classes d'âges confondues. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en gras	125
4.3	Valeurs moyennes nationales de l'indicateur <i>SOMO35</i> et pourcentages de majoration de la valeur historique pour différents scénarios climatiques aux horizons 2030, 2040 et 2050 - Modélisation selon les valeurs médianes de la variable exogène.	125
4.4	Températures et différence à l'historique pour différents scénarios climatiques aux années 2030, 2040 et 2050 - Modélisation selon les valeurs médianes de la variable exogène.	126
4.5	Chocs additionnels nationaux en % obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d'âge : personnes de plus de 65 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en gras	127
4.6	Chocs additionnels nationaux obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Toutes classes d'âges confondues. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en gras - Modélisation selon les valeurs choquées de la variable exogène.	128

4.7	Valeurs moyennes nationales de l'indicateur <i>SOMO35</i> et pourcentages de majoration de la valeur historique pour différents scénarios climatiques aux horizons 2030, 2040 et 2050 - Modélisation selon les valeurs choquées de la variable exogène.	129
4.8	Facteurs additionnels de la mortalité communiqués par l'ACPR dans ses exercices pilote climatique de 2020 et 2023.	131
A.1	Normes de concentration pour NO_2	138
A.2	Normes de concentration pour PM_{10}	138
A.3	Normes de concentration pour $PM_{2.5}$	139
A.4	Normes de concentration pour O_3	139
A.5	Nombre de sites de surveillance de la pollution de fond en France par polluant atmosphérique pour les années 2013 à 2022 (post-retraitements effectués en 2.4.2)	142
A.6	Chocs additionnels nationaux en % obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d'âge : 0 - 10 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en gras	144
A.7	Chocs additionnels nationaux en % obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d'âge : 10 - 25 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en gras	144
A.8	Chocs additionnels nationaux en % obtenus pour la pollution à l'Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d'âge : 10 - 25 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l'intervalle de confiance à 95 % du facteur en gras	145

Introduction

Dans un contexte, où, le changement climatique devient de plus en plus préoccupant, acteurs assurantiels, régulateurs nationaux et européens travaillent collaborativement afin de mieux anticiper et comprendre les risques de demain. Il en résulte la conduite d'études préliminaires novatrices comme la publication de l'*European Insurance and Occupational Pensions Authority* (EIOPA) sur l'importance de la considération et l'utilisation de scénarios de changement climatique au sein de l'*Own Risk and Solvency Assessment* (ORSA), ou encore l'exercice pilote climatique mené par l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) en 2020, reconduit en 2023. Ces textes constituent des références méthodologiques importantes dans l'étude des risques prospectifs induits par le changement climatique.

Par ailleurs, l'assurance de personnes n'est évidemment pas épargnée par l'aggravation en fréquence ou en sévérité des risques entraînés par le changement climatique. Cette aggravation tire principalement son essence de trois aléas distincts que sont les maladies à transmission vectorielle, la pollution de l'air et la surmortalité induite par les épisodes caniculaires extrêmes. Tous trois mettent en péril la santé humaine, induisant des effets de consommation santé plus importants ou encore parfois en provoquant des décès prématurés. Ces aléas doivent par conséquent faire l'objet d'une surveillance de plus en plus accrue.

Cet écrit traitera en particulier de l'un de ces périls au sein du *scope* de l'assurance de personnes ; les impacts de la pollution de l'air. En guise de rappel, selon l'Organisation Mondiale de la Santé (OMS) la pollution de l'air, également appelée pollution atmosphérique, se caractérise par « la contamination de l'environnement intérieur ou extérieur par tout agent chimique, physique ou biologique qui modifie les caractéristiques naturelles de l'atmosphère » (*OMS [7]*).

Historiquement, c'est à partir du 19 novembre 1979 que l'Organisation des Nations Unies (ONU) établit une convention sur la pollution atmosphérique transfrontière à longue distance (CPATLD). Cette date marque le début de la prise de conscience de l'importance que l'Homme doit collectivement porter au suivi et au strict maintien d'une qualité de l'air décente.

En effet, la pollution de l'air est à l'origine de préjudices de sévérités variées allant de la « simple » difficulté respiratoire donnant lieu à une consultation chez un médecin généraliste, à l'hospitalisation, voire au développement d'affections plus graves tels que certains types de cancers parfois à l'origine de cas de décès prématurés. L'agence nationale de santé publique a par ailleurs estimé que la pollution de l'air est responsable de plus de 48 000 morts en France en 2016, représentant un coût sanitaire annuel total de 100 milliards d'eu-

ros (*Santé Publique France, 2023 [8]*). En outre, sur la période 1990-2016, la pollution de l'air constitue le cinquième facteur de risque le plus dangereux pour la santé dans le monde (*The Lancet, 2017 [9]*).

Compte tenu de ces chiffres clés et de ce contexte, il semble nécessaire pour les assureurs commercialisant des contrats en assurance de personnes, d'étudier les impacts de la pollution atmosphérique historique, actuelle et future. L'étude de ce type de risque s'inscrit dans le processus ORSA, issu du second pilier de la norme prudentielle Solvabilité II, au sein duquel l'assureur évalue en interne ses risques et sa solvabilité en pointant les éléments susceptibles de modifier sa situation financière.

Cependant, bien qu'il soit largement répandu dans la pensée commune que l'évolution de la pollution de l'air est intimement liée à celle de l'évolution des gaz à effet de serre, peut-on réellement laisser penser que ce soit systématiquement le cas ? Pollution de l'air et gaz à effet de serre sont, sur certains aspects, totalement dissociables que ce soit en raison de leurs sources d'origine, de leurs comportements chimiques, de leurs tendances d'évolution au cours des dernières années ou encore de leurs effets et conséquences financières.

Aussi, est-ce que le recours à un scénario construit sur les tendances d'évolution des gaz à effet de serre (GES) comme le sont les scénarios *Representative Concentration Pathway* (RCP) ou les scénarios *Shared Socio-economic Pathways* (SSP) actuellement utilisés au sein de l'exercice pilote climatique de l'ACPR, se trouve être le plus pertinent pour une application à la projection de la pollution atmosphérique ?

Cette étude exploratoire a donc pour ambition de répondre à ces besoins et questionnements en tentant d'identifier dans quelle mesure la pollution de l'air peut représenter un risque de matérialité suffisante en assurance de personnes afin d'être considérée au sein de l'ORSA climatique d'une compagnie d'assurance.

Le premier chapitre de ce mémoire a pour objectif de rappeler le contexte de l'étude et de fournir au lecteur les bases théoriques relatives au domaine de l'assurance, du changement climatique et de la pollution de l'air nécessaires à la compréhension de cet écrit.

Ensuite, le chapitre deuxième aborde la présentation des données utilisées et la méthodologie de construction du module exposition couplant estimations de concentrations de différents polluants obtenues par interpolation spatiale à une population par pixel d'1km² en France métropolitaine.

Dans un chapitre troisième, la construction et l'utilisation de différents types de scénarios prospectifs de pollution atmosphérique à horizon 2050 est présentée. Il sera également abordé la méthodologie utilisée pour lier ces scénarios aux concentrations de pollution de l'air à des stations de mesure du réseau national français. En se basant sur des approches existantes et sur leurs résultats, l'écrit mettra en exergue les forces et limites identifiées de la méthodologie en justifiant des résultats obtenus.

Enfin, dans un chapitre quatrième, un *compendium* procédural est réalisé permettant de synthétiser la méthodologie globale de l'étude, puis, les résultats sont présentés et discutés. Enfin des exemples de réutilisation des résultats obtenus sont donnés, et une ouverture sur l'application de la méthodologie à d'autres thématiques actuarielles est réalisée.

Chapitre 1

Contexte de l'étude : Assurance, changement climatique et pollution de l'air

1.1 Assurance et changement climatique

1.1.1 Contexte

À l'ère de profonds changements induits par une activité humaine irraisonnée, maintenue croissante depuis la seconde révolution industrielle, et responsable d'une émission soutenue de gaz à effet de serre rejetés dans l'atmosphère, de nouvelles problématiques climatiques aux conséquences sociales et financières encore incertaines ont vu le jour. Fréquence et intensité des épisodes climatiques extrêmes, inondations, sécheresses, feux de forêts, modification de l'environnement de vie ne sont qu'une infime partie des potentielles incidences aujourd'hui anticipées du changement climatique.

De ces nouvelles problématiques résulte l'apparition de nouveaux risques pesant à la fois sur le système bancaire et sur le système assurantiel. C'est dans ce contexte singulier, soutenu par des publications scientifiques reconnues et de plus en plus alarmantes (comme peuvent être les différents rapports du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC)), que les différents acteurs nationaux et européens prennent de plus en plus part aux discussions et réalisations relatives aux enjeux prospectifs du climat.

Suite à cela, plusieurs travaux préliminaires d'investigation ont été publiés, impliquant à la fois les acteurs réglementaires et les acteurs métiers. Parmi ces réalisations, on retrouve notamment certains travaux de l'EIOPA comme la mise à disposition de CLIMADA, un outil Open-source pour la modélisation et la gestion des risques liés au changement climatique ou encore le guide d'application sur l'évaluation de la matérialité des risques dus au changement climatique au sein du processus ORSA. Par ailleurs le régulateur français, l'ACPR, a également réalisé un exercice pilote climatique mené en 2020 et reconduit cette année, en 2023.

Deux grandes familles de risques sont alors à l'étude au sein de ces travaux. Tout d'abord, il y a les risques dits de « transition ». Ces risques font référence aux préjudices entraînés

par des changements économiques, politiques et sociaux à caractère nécessaires pour faire face au changement climatique. Cela inclut des changements dans les politiques publiques, les technologies émergentes, les réglementations environnementales, la transition vers des sources d'énergie renouvelable, la prise en considération du coût du carbone et la réduction des émissions de gaz à effet de serre.

Les risques de transition peuvent affecter les entreprises et les investisseurs qui ne sont pas prêts à s'adapter à cette nouvelle donne en les exposant à des pertes financières issues par exemple de la dégradation d'actifs et à une réputation dégradée. En conséquence une baisse d'attractivité et du nombre de souscriptions de contrats peut alors être enregistrée pour le cas d'un assureur.

Ensuite, les risques physiques font référence aux conséquences dites « directes » du changement climatique, telles que les tempêtes, les inondations, la sécheresse, la hausse du niveau de la mer et les vagues de chaleur extrêmes. Ces événements climatiques extrêmes peuvent perturber les chaînes d'approvisionnement, endommager les infrastructures et causer tant des dommages aux biens que des dommages aux personnes. Les risques physiques sont donc davantage liés aux effets immédiats du changement climatique sur l'environnement et sur les populations.

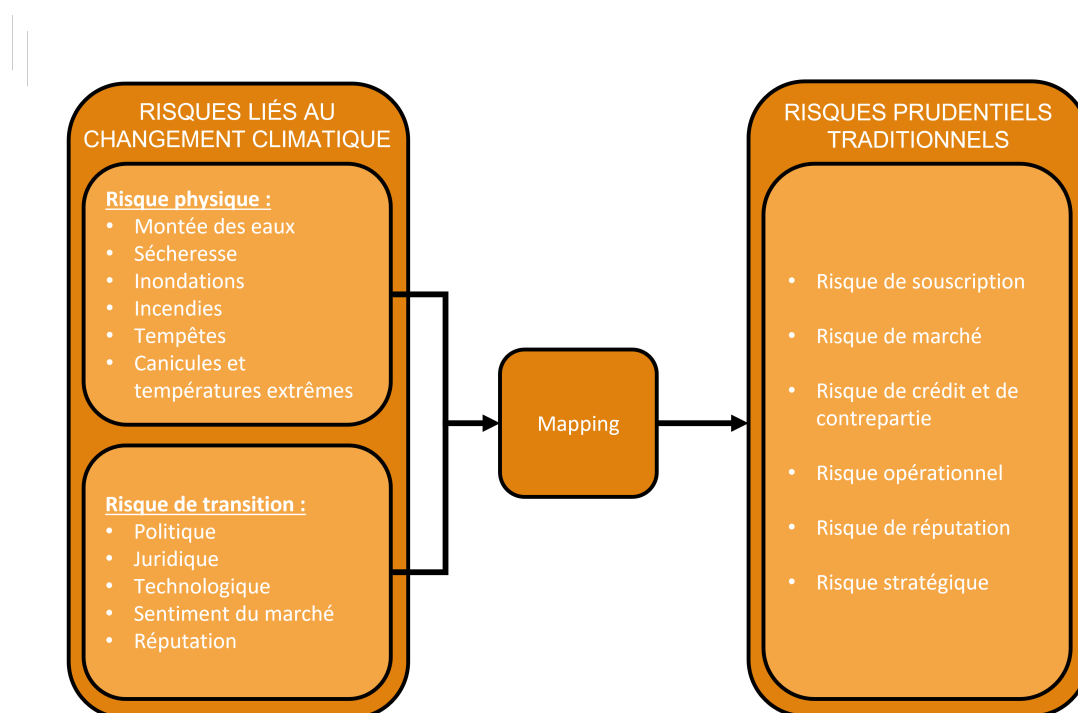


FIGURE 1.1 – Schéma récapitulatif de la procédure de traduction des risques climatiques en risques prudentiels traditionnels dans l'ORSA. (*Finalyse, Climate change risks in the ORSA* [1])

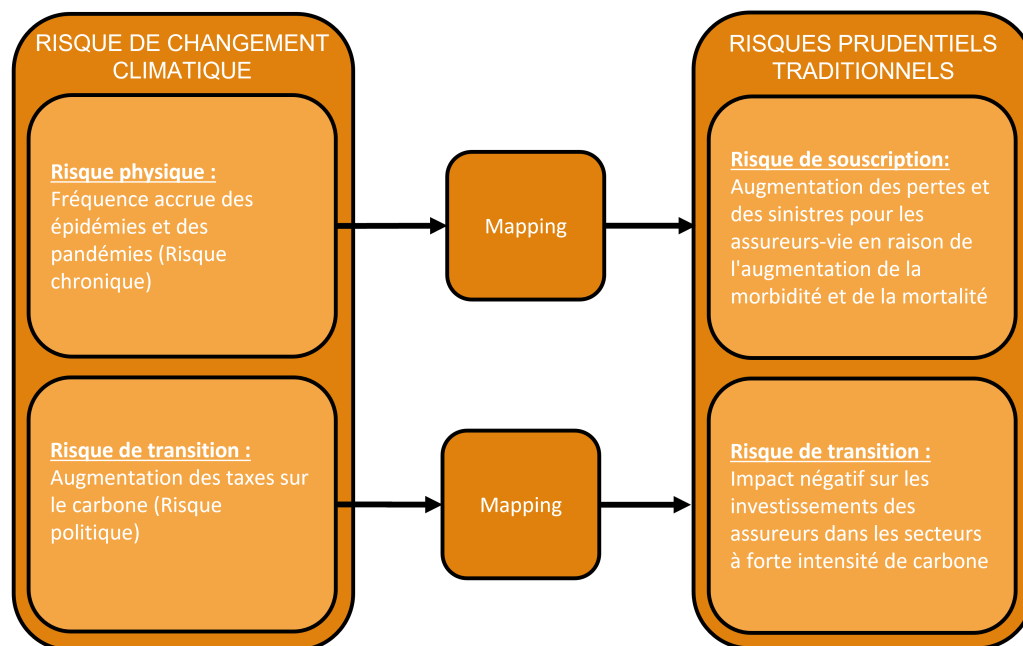


FIGURE 1.2 – Exemples de correspondance entre les risques climatiques et les risques prudentiels traditionnels. (*Finalyse, Climate change risks in the ORSA* [1])

Il s'agira par la suite de présenter le contexte, le contenu, les aboutissants de deux de ces textes, et les mesures dans lesquelles ils seront utiles au sein de cette étude.

1.1.2 Publication EIOPA : évaluation de l'importance relative du changement climatique dans l'ORSA

ORSA et changement climatique

Les entreprises assurantielles ont l'obligation d'évaluer leurs propres risques via le processus de l'ORSA imposé par le pilier 2 du cadre prudentiel Solvabilité 2. Cette évaluation interne de la gestion des risques permet de refléter de la plus fidèle des manières les besoins en solvabilité de l'entreprise, de vérifier le respect des exigences de capital sous-jacentes et enfin de traiter l'importance avec laquelle le profil de risque de l'entreprise s'écarte des hypothèses sur lesquelles ces exigences sont fondées.

Le risque climatique représente ainsi un impact futur potentiel ou certain selon les types de profils à considérer au sein de la gestion globale des risques des entreprises. Compte tenu de cette exposition potentielle ou certaine, les entreprises d'assurance sont donc censées intégrer les risques liés au changement climatique au sein de leur ORSA en analysant l'impact de ces derniers sur leur propre profil de risque ainsi que sur leur niveau d'exposition.

Mise en situation

Un guide d'application pour l'évaluation de l'importance relative de l'utilisation de scénarios de changement climatique dans l'ORSA a été publié le 2 août 2022 par l'EIOPA. Ce guide d'application fait suite à un avis publié en avril 2021 sur la supervision du recours aux scénarios de risques climatiques dans l'ORSA.

Concrètement, il définit les attentes en matière de surveillance concernant l'utilisation de scénarios de changements climatiques par les acteurs assurantiels au sein de leur ORSA.

Cependant, seule une minorité de compagnies évalue les risques liés au changement climatique en ayant recours à l'analyse de différents scénarios. De plus, lorsque ces analyses sont effectuées, elles ne sont généralement projetées qu'à court terme, horizon qui semble inadéquat à celui de la réalisation anticipée des événements liés au changement climatique.

L'EIOPA a donc décidé d'élaborer et de proposer des orientations d'application de scénarios prospectifs au sein de l'ORSA à travers un guide d'application en fournissant des études de cas concrètes sur des entreprises « fictives » mais représentatives du marché de l'assurance. Ce guide d'application a également pour dessein de réduire les coûts de mise en œuvre de l'appréciation de ces risques pour les entreprises d'assurance et particulièrement pour les petites et moyennes entreprises. Par ailleurs, l'EIOPA incite les différents acteurs de la place à continuer d'investiguer et de communiquer sur de nouveaux procédés autres que ceux proposés au sein du guide d'application, ces derniers ne constituant qu'un socle initial méthodologique de convergence de bonnes pratiques.

Composé de trois chapitres, le premier décrit les différentes parties de l'ORSA dans lesquelles les entreprises ont la possibilité d'aborder les risques liés au changement climatique. Le deuxième fournit des indications générales sur l'évaluation de la matérialité et l'utilisation des différents scénarios climatiques. Enfin le troisième chapitre donne des exemples concrets d'application sur des entreprises fictives non-vie et vie de l'évaluation des risques et de la quantification des impacts de la réalisation de différents scénarios climatiques.

Évaluation de la matérialité

La première étape dans la prise en compte des risques induits par le changement climatique au sein de l'ORSA est l'évaluation de la matérialité. Les entreprises sont incitées à identifier les risques significatifs liés au changement climatique dans le cadre de l'exercice de leur activité. Par ailleurs, si le changement climatique n'est pas un risque significatif, ces dernières sont incitées à justifier de comment elles sont parvenues à ce constat (*section 3.14 de l'avis EIOPA 2021a [3]*).

Ci-dessous voici les étapes clés de la description de la matérialité des risques pour une entreprise :

- La définition du contexte commercial : il s'agit de statuer sur quels produits ou portefeuilles l'entreprise est exposée aux risques climatiques et d'identifier quelles activités d'assurance sont concernées, sur quel horizon temporel, et dans quel contexte stratégique.
- La recherche des impacts sur l'entreprise et son business : sont ici investigués les impacts potentiels des risques liés au changement climatique sur l'exposition. Une

distinction peut alors être réalisée entre le risque physique et le risque de transition. L'entreprise interprète ainsi les impacts identifiés sur ses produits au travers de ses états financiers bilan et/ou compte de résultat.

- L'évaluation de la pertinence pour l'entreprise : l'entreprise évalue la matérialité de chaque risque considéré comme étant suffisamment lié au risque de changement climatique sur l'actif et le passif de son bilan. Il est alors pris en compte l'exposition initiale de l'entreprise, l'impact du changement climatique sur l'exposition spécifique et la probabilité de survenance de l'impact identifié. Enfin, l'évaluation de la matérialité du risque peut être représentée dans une matrice de matérialité ou dans un tableau récapitulatif par risque.

Risque	Indicateur	Type de risque climatique
Physique Aigu	Court	Non Matériel
	Moyen	Non Matériel
	Long	Matériel
Physique Chronique	Court	Non Matériel
	Moyen	Non Matériel
	Long	Non Matériel

TABLE 1.1 – Exemple de tableau récapitulatif de la matérialité des risques physiques aigus et chroniques par horizon temporel (EIOPA, *Application Guidance on Running Climate Change Materiality Assessment and Using Climate Change Scenarios in ORSA*, 2023 [2])

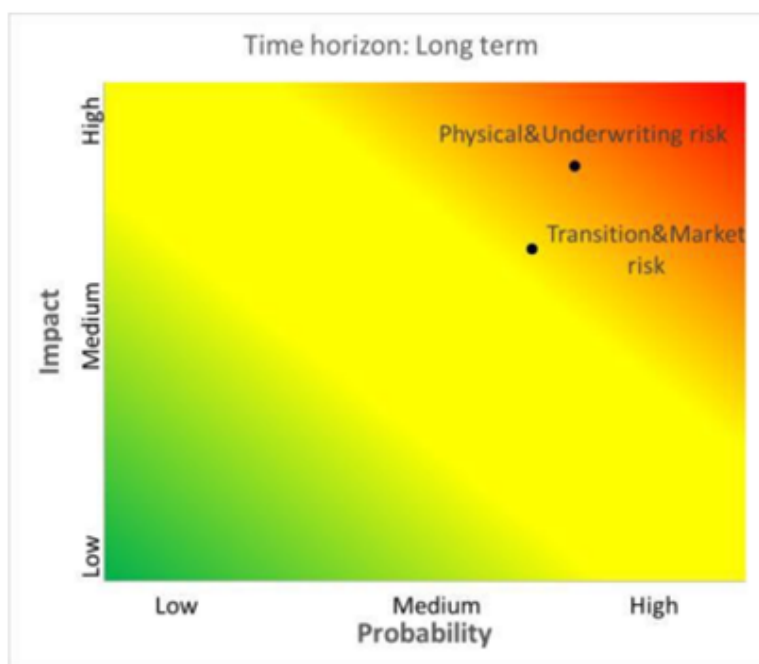


FIGURE 1.3 – Exemple d'une matrice de matérialité à horizon long terme (EIOPA, *Application Guidance on Running Climate Change Materiality Assessment and Using Climate Change Scenarios in ORSA*, 2023 [2])

- Conclusion sur la matérialité : en cas de matérialité suffisante du risque considéré, il est nécessaire de procéder à une évaluation plus approfondie faisant appel à l'uti-

lisation de scénarios climatiques prospectifs. À l'inverse, en cas de matérialité insuffisante du risque, il est toutefois nécessaire d'expliquer pourquoi le risque n'est pas considéré comme significatif au sein de l'ORSA.

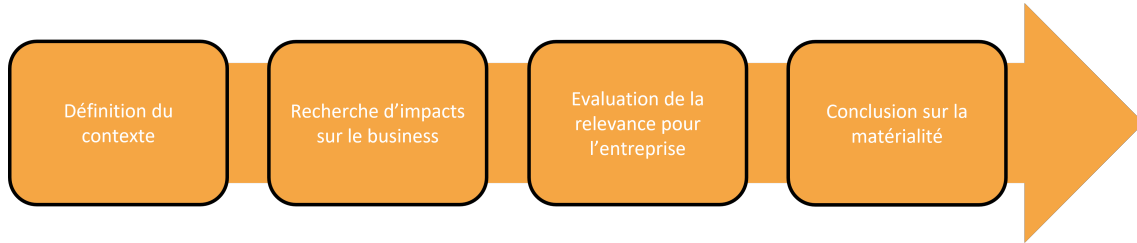


FIGURE 1.4 – Schéma simplifié des étapes d'évaluation de la matérialité d'un risque climatique

Horizons de temps et scénarios

L'avis 2021a de l'EIOPA [3] mentionne que les entreprises doivent évaluer les risques du changement climatique à long terme en utilisant l'analyse de scénarios afin d'informer le département en charge du pilotage stratégique si ce risque impacte significativement ou non les états financiers d'une société.

L'horizon temporel se voit alors dans le cadre de l'étude des risques du changement climatique et de l'ORSA, allongé en comparaison aux horizons utilisés par défaut pour les risques « traditionnels » par les compagnies d'assurance dans l'ORSA. Les définitions des horizons court, moyen et long termes sont alors revus à la hausse.

Il en résulte l'émergence d'un défi qui consiste à concilier la dynamique à très long terme du changement climatique avec la capacité opérationnelle d'évaluer l'impact des risques connexes sur la base du modèle commercial actuel de l'entreprise.

Cela nécessite une nouvelle approche dans l'ORSA pour l'analyse des risques liés au changement climatique. Il s'agit de faire concorder les horizons temporels de l'analyse des risques d'une entreprise avec ceux de ses engagements et particulièrement si ces derniers sont des engagements à long terme.

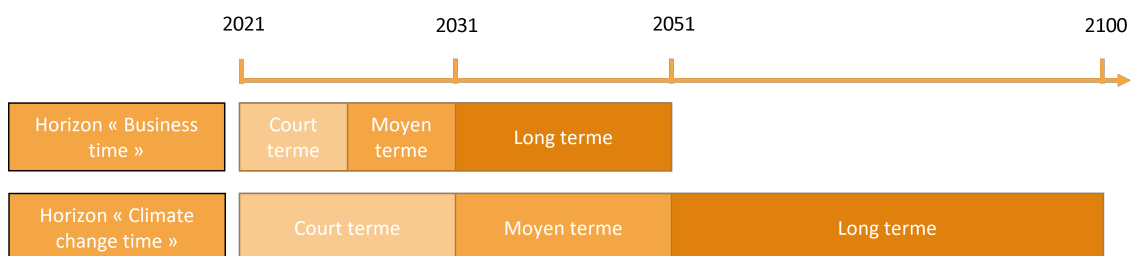


FIGURE 1.5 – Définition des horizons de temps dans l'ORSA (EIOPA 2021a [3])

Sur cette base de temporalité, pour les risques précédemment retenus comme étant matériels, il est alors nécessaire de définir les scénarios qui seront utilisés pour approfondir l'analyse.

Afin de définir ces scénarios, conformément à la section 3.18 de l’avis de l’EIOPA, une approche en 3 étapes est envisagée :

- La définition des scénarios : les entreprises doivent envisager au moins deux scénarios climatiques à long terme, un scénario dans lequel l’augmentation de la température mondiale reste inférieure à 2°C, et de préférence ne dépassant pas 1,5°C, un second dans lequel l’augmentation de la température mondiale augmente de plus de 2°C. D’autres analyses sont également présentées dans le guide d’application avec des horizons temporels plus courts car ils peuvent être tout autant pertinents à étudier pour les entreprises assurantielles.
- La traduction des scénarios définis en risques liés au changement climatique : Il s’agit alors, compte tenu des scénarios retenus en étape précédente, de par exemple identifier les risques liés à l’abandon soudain des combustibles fossiles induisant une baisse de la valeurs des actions exerçant dans ce domaine pour le cas des risques de transition. Cet exemple n’est évidemment pas exhaustifs, il convient alors de tous les identifier.
- La transformation du changement climatique en pertes financières : afin d’interpréter objectivement les incidences des scénarios retenus en étape précédente, il est nécessaire de choisir des indicateurs qui permettront de modéliser les impacts financiers. Ces derniers se classent en plusieurs catégories et se complètent les uns aux autres.

Type d’indicateur	Indicateur	Type de risque climatique
Bilan	Capital de solvabilité requis	Risque physique et de transition
	Excédent de l’actif sur le passif (variation de)	Risque physique et de transition
	Actif sur passif (variation de)	Risque physique et de transition
	Le changement de prix pour chacun des actifs identifiés	Risque de transition
	Variation relative du total des provisions techniques	Risque physique
Profitabilité	Ratio des pertes (S/P)	Risque physique
	Impact global sur le résultat de l’entreprise	Risque physique et de transition
Technique	Pertes brutes/cédées nettes cumulées	Risque physique
	Exposition principale (somme assurée)	Risque physique
	Total des actifs soumis à des risques transitoires	Risque de transition
	Probabilité annuelle de survenance	Risque physique

TABLE 1.2 – Tableau récapitulatif des métriques d’évaluation par type d’indicateur (EIOPA, 2021 [2])

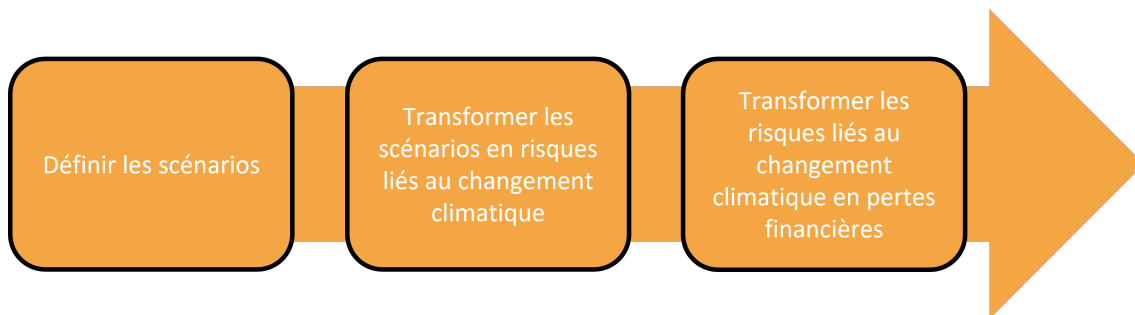


FIGURE 1.6 – Définition des horizons de temps dans l’ORSA (*EIOPA, 2021a [3]*)

Limites de l’études

Comme énoncé précédemment, ce guide d’application définit un socle de convergence de méthodologies et de bonnes pratiques mises à la disposition de tous les acteurs. Il ne se place en aucun cas comme une approche exhaustive à appliquer strictement. Il est même conseillé aux acteurs d’investiguer leurs propres méthodologies et d’identifier, en se basant partiellement ou non sur les directives proposées, des procédés propres à leur portefeuille, exposition et vulnérabilité.

Aussi, l’approche consistant à se limiter à deux scénarios indexés sur la température ne serait-elle pas restrictive pour l’étude de certains risques ? Plus concrètement, les évolutions socio-économiques, les actions politiques et réglementaires ne sont-elles pas prédominantes dans l’étude de certains risques futurs ?

Toutefois, les éléments qui se réfèrent à l’étude de la matérialité et ses étapes de réalisation seront réutilisés au sein de cette étude et serviront d’exemple pour une application à la pollution atmosphérique.

1.1.3 Les exercices pilote climatique de l’ACPR

Présentation générale de l’exercice pilote climatique 2020

Conscient de l’émergence de nouveaux risques ou du renforcement de l’intensité et/ou de la fréquence de certains risques, l’ACPR a mené en 2020 un exercice autour des enjeux liés au changement climatique.

Par le biais de cette étude, le régulateur français avec la participation d’acteurs bancaires et assurantiels sous sa responsabilité, a réalisé pour la première fois une évaluation des risques engendrés par le changement climatique. Plus concrètement, cet exercice a eu pour ambition et pour succès :

1. De parvenir à mobiliser divers acteurs sur la base du volontariat. En quelques chiffres, 15 groupes d’assureurs soit 22 organismes d’assurance et 9 groupes bancaires ont été sollicités durant cet exercice « représentant respectivement 75 % des provisions techniques et des actifs des assureurs et 85 % du total de l’actif des

banques françaises » (ACPR, *Principaux résultats de l'exercice pilote climatique 2020 [10]*).

2. De sensibiliser ces mêmes acteurs à l'émergence de ces nouveaux risques et à l'importance d'une gestion saine et maîtrisée en permettant la mise à disposition de méthodes novatrices. Bien que certaines difficultés méthodologiques aient subsisté compte tenu du caractère inédit de l'exercice, les institutions participantes ont salué les nombreuses avancées dans l'évaluation des risques climatiques. La mobilisation transversale a également permis de croiser les réflexions, de pointer les limites des modèles utilisés, de favoriser la meilleure compréhension des nouveaux risques et *in fine*, de disposer de meilleures conditions pour piloter une activité en connaissance des conséquences du climat.
3. D'évaluer parallèlement la matérialité qualitative et quantitative de tous les risques conséquents au changement climatique, qu'ils soient physiques ou de transition. En d'autres termes cet exercice sert de point catalyseur aux réflexions sur le sujet et favorise l'assignation de moyens financiers et humains pour la réalisation d'études concrètes. Il a également été précisé que cet exercice sert « D'autre part [...] de point de départ pour une introduction croissante de la considération et de l'étude des risques climatiques pour les acteurs de la place française » (ACPR, *Principaux résultats de l'exercice pilote climatique 2020 [10]*).
4. De réaliser une première mesure prospective des risques et vulnérabilités avec 30 ans de profondeur, correspondant à un horizon long-terme dans le *scope* de l'étude des risques traditionnels. Cela correspond à un horizon moyen terme dans le cas de l'étude des risques climatiques.

Celle-ci se basant sur l'hypothèse de bilan dynamique et les impacts constatés. On notera pour résultat de l'étude, une hausse particulièrement significative de la sinistralité se répercutant sur la hausse des primes d'assurance pour couvrir ces pertes envisagées : « Ainsi, sur la base des éléments remis par les assureurs, le coût des sinistres pourrait être multiplié par 5 à 6 dans certains départements français entre 2020 et 2050 ». (ACPR, *Principaux résultats de l'exercice pilote climatique 2020 [10]*).

Cet exercice pilote d'évaluation des risques conséquents au changement climatique se place comme l'un des textes référents dans le secteur bancaire et assurantiel français. Il servira également de texte source pour la suite de cet écrit.

Enfin ayant pour ambition d'être reconduit « régulièrement », à date d'écriture de ce mémoire, le prochain exercice climatique est en cours de réalisation avec une publication envisagée à la fin de l'année 2023.

Reconduction de l'exercice en 2023

À la suite du succès du premier exercice réalisé en 2020, et en adéquation avec l'objectif d'accroître l'analyse des risques climatiques, l'étude a été renouvelée en 2023. Cette reconduction a pour but de compléter le cadre méthodologique du premier exercice en testant notamment la capacité des assureurs à maintenir leurs niveaux de solvabilités face à des dérives de sinistralité induites par les le changement climatique.

En cours, les résultats n'ont pas encore été publiés. En revanche les spécifications et hypothèses de l'exercice ont été communiquées. La réalisation de 2023 reprend en majeure partie les hypothèses du précédent en intégrant de nouvelles approches inédites comme le choix du scénario de référence de hausse globale de la température qui bascule du RCP8.5 au RCP4.5.

D'autres approches sont par ailleurs innovantes comme la combinaison dans chaque scénario des impacts à l'actif et au passif, permettant une analyse conjointe des risques physiques et des risques de transition.

Périls étudiés et méthodologie retenue en assurance de personnes

En plus de ses multiples conséquences sur les événements météorologiques extrêmes cités plus haut, le changement climatique présente un risque pesant également sur la santé humaine. L'exercice suggère ainsi les périls à retenir et propose une méthodologie pour tenter d'appréhender l'incidence supposée sur la santé et la mortalité, périmètre retenu dans le cadre de ce mémoire.

Aussi, la mise en place ou non d'une transition vers une économie décarbonée, la vitesse à laquelle cette dernière s'opèrerait et les conséquences sous-jacentes de la saturation de l'atmosphère aux gaz à effet de serre (sous-entendu le nombre de degrés additionnels du référentiel des accords sur le climat de Paris, COP21), permettent de dessiner différents scénarios prospectifs. Dans les exercices pilote climatique, ces mêmes scénarios entraînent des conséquences directes sur la santé humaine.

Ces derniers se servent ainsi de différents scénarios climatiques pour évaluer l'impact sur différents *Lines of Business* (LoB) relatifs à l'activité d'assurance. En santé-prévoyance, principalement impactée par le risque physique, l'effet du risque climatique est conséquent à une dégradation de l'environnement de vie des assurés qui affecterait leur santé. Ce genre de phénomène peut être étudié à différentes mailles géographiques, pouvant ainsi survenir à l'échelle du pays entier ou rester très local à l'échelle d'une région, d'un département, d'une commune.

Les deux principales sources de sinistralité additionnelle retenues sont l'essor de pandémies issues de maladies à transmission vectorielle et la pollution de l'air en assurance de personnes en France. Concernant l'exercice de 2023, un troisième péril a été ajouté, les épisodes caniculaires.

Dans une démarche de complétude et d'exhaustivité, l'ACPR a combiné son expertise avec celui d'acteurs externes pour certains périls. En guise d'exemple, la Caisse centrale de réassurance (CCR) est intervenue sur le régime catastrophes naturelles (CATNAT) et le courtier AON a fourni des travaux liés aux risques physiques.

L'ACPR a ainsi bénéficié en 2020 de la part d'AON de la réalisation de deux études distinctes relatives à l'essor des pandémies issues de maladies à transmission vectorielle d'une part et les impacts que provoquent la pollution de l'air en assurance de personnes en France d'autre part.

De ces études, l'ACPR a intégré les résultats au sein de l'exercice pilote climatique sous la forme de tableaux recensant des hypothèses d'évolution de tables de mortalité, de frais de santé et d'arrêts de travail, par tranches d'âges, pour différentes zones géographiques et

selon différents horizons temporels.

Conséquences du changement climatique sur la pollution de l'air et impact en assurance de personnes - AON

« *Conséquences du changement climatique sur la pollution de l'air et impact en assurance de personnes* » ([4]) est l'une des deux études réalisées par AON pour le compte de l'ACPR et dont les résultats ont été intégrés au sein des exercices pilote climatique.

Au sein de cette étude, AON fournit une démarche de bout en bout afin d'éclairer les enjeux assurantiels liés à l'évolution de la qualité de l'air dans un contexte de changement climatique. Les particules fines (PM_{10} et $PM_{2.5}$), l'oxyde d'Azote (NO_x) et l'ozone (O_3) font l'objet de projections à horizon 2050, en particulier selon le scénario le plus adverse en matière de hausse anticipée de la température globale, à savoir le RCP8.5. De ces projections sont déduits des impacts en sinistralité additionnelles sur les garanties contenues dans les contrats d'assurances de personnes.

En amont de la partie production de l'étude sont présentés un certain nombre de sources et d'enseignements relatifs à la pollution de l'air. Ces éléments forment un état de l'art de la variation prospective des principaux polluants en France et notamment un tour d'horizon des études déjà réalisées à ce sujet. Pour chaque étude sont évoqués :

- Les objectifs du projet,
- Les hypothèses du projet,
- La zone spatiale et sa résolution,
- Les modèles utilisés,
- Les principaux résultats.

Ci-dessous les principaux points d'attention issus des sources retenues. Ces mêmes points serviront par ailleurs de socle d'hypothèses utilisées lors de la construction des modèles de ce mémoire :

1. La connaissance de la qualité de l'air et l'étude de son évolution future est d'une importance capitale pour les enjeux sanitaires de demain (*Drif et al., 2020 [4]*).
2. Les principaux facteurs qui influent sur la qualité future de l'air sont : le changement climatique, les émissions biogènes et anthropiques et le transport intercontinental de la pollution atmosphérique (*Colette et al. 2013b [11]*). Selon la résolution et l'échelle de l'étude leur ampleur est relativement variable.
3. De ces trois principales origines citées, on peut affirmer que seules les émissions sont les principaux facteurs déterminants à l'horizon 2050 (*Colette et al. 2013b [11]*, *Colette et al. 2013a [12]*). Les variations des seules variables climatiques ont un impact secondaire sur la qualité de l'air (*Lecoeur 2013 [13]*).
4. Pour une prévision correcte des polluants dispersés dans l'atmosphère, il est essentiel d'utiliser un scénario d'émission de précurseurs adéquat qui tienne compte des informations à l'échelle locale et de la législation actuelle qui régit la pollution de l'air (*Likvar et al., 2015 [14]*).
5. Les émissions de polluants atmosphériques en Europe peuvent suivre les scénarios d'émission de polluants atmosphériques *Global Energy Access (GEA)*, qui sont plus adaptés aux études sur la qualité de l'air et comprennent une représentation de

l'ensemble de la législation actuelle sur la qualité de l'air en Europe (*Riahi et al., 2012 [15]*).

6. Un point essentiel pour quantifier l'impact du changement climatique sur la qualité de l'air est de disposer de prévisions avec une résolution spatiale adéquate (*Drif et al., 2020 [4]*).
7. À l'échelle locale, les plus grandes incertitudes sont associées à la résolution du modèle mis en œuvre, en particulier la résolution de l'inventaire des émissions d'entrée, qui est très importante pour les zones urbaines (*Likhvar et al., 2015 [14]*).
8. *Colette et al. (2013a, b) ([12], [11])* ont suggéré que les outils d'évaluation des risques climatiques nécessitaient une résolution plus fine ou égale à un demi-degré (un degré représente environ 111,319 km à l'équateur).
9. En effet, bien que les RCP incluent les émissions de polluants atmosphériques, ils sont conçus exclusivement dans le but d'étudier le forçage radiatif dans le cadre du projet *Atmospheric Chemistry and Climate Model Intercomparison Project* (ACC-MIP) (*Shindell et al. 2012 [16], Young et al. 2012 [17]*). Leur mise en œuvre pour les projections de la qualité de l'air constitue une déviation de leur objectif premier (*Butler et al. 2012 [18], Colette et al. 2012 [19], Fiore et al. 2012 [20]*).

Compte tenu de l'inertie de l'évolution du risque physique, et de l'impossibilité d'agir afin d'obtenir des résultats bénéfiques à court terme même dans le cas d'une transition immédiate vers une économie bas carbone, il a été retenu que l'étude du scénario le plus adverse à savoir le scénario RCP8.5 du GIEC à horizon 2050, serait le seul étudié.

Ce dernier plus communément appelé « *Business as usual* » postule que les émissions actuelles sont prolongées, qu'aucune action de transition vers une économie bas carbone n'est entreprise, et se traduit finalement par une hausse globale de la température de 4°C à l'horizon 2100 (augmentation comprise entre 1,4°C et 2,6°C en 2050).

Finalement, à partir de la synthèse des éléments cités ci-dessus, l'étude présente en résultats les hypothèses d'évolution future de tables de mortalité, de frais de santé et d'arrêts de travail suivantes :

IMPACTS DE LA POLLUTION				
		2021 - 2030	2031-2040	2041-2050
Granularité : Globale	Décès	0,02%	0,02%	0,03%
	Frais de Soins	0,84%	1,25%	1,65%
	Arrêts de Travail	0,07%	0,10%	0,13%
Granularité : Fine				
Bordeaux	Décès	0,01%	0,02%	0,02%
	Frais de Soins	0,68%	1,02%	1,37%
	Arrêts de Travail	0,06%	0,09%	0,13%
Ile de France	Décès	0,02%	0,03%	0,04%
	Frais de Soins	1,09%	1,60%	2,10%
	Arrêts de Travail	0,08%	0,12%	0,17%
Lille	Décès	0,02%	0,03%	0,03%
	Frais de Soins	1,00%	1,47%	1,94%
	Arrêts de Travail	0,08%	0,12%	0,16%
Lyon	Décès	0,02%	0,03%	0,04%
	Frais de Soins	1,09%	1,60%	2,10%
	Arrêts de Travail	0,08%	0,12%	0,17%
Marseille	Décès	0,02%	0,03%	0,04%
	Frais de Soins	1,09%	1,60%	2,10%
	Arrêts de Travail	0,08%	0,12%	0,17%
Montpellier	Décès	0,02%	0,03%	0,03%
	Frais de Soins	0,72%	1,09%	1,47%
	Arrêts de Travail	0,05%	0,08%	0,12%
Nantes	Décès	0,01%	0,02%	0,02%
	Frais de Soins	0,68%	1,01%	1,37%
	Arrêts de Travail	0,05%	0,08%	0,11%
Nice	Décès	0,01%	0,02%	0,02%
	Frais de Soins	0,68%	1,02%	1,37%
	Arrêts de Travail	0,05%	0,08%	0,11%
Strasbourg	Décès	0,01%	0,02%	0,02%
	Frais de Soins	0,68%	1,02%	1,37%
	Arrêts de Travail	0,05%	0,08%	0,11%
Toulouse	Décès	0,01%	0,02%	0,02%
	Frais de Soins	0,68%	1,02%	1,37%
	Arrêts de Travail	0,06%	0,09%	0,13%

FIGURE 1.7 – Tableau récapitulatif des impacts additifs de la pollution atmosphérique sur les décès, les frais de soins et les arrêts de travail par géographie et horizon de projection (*Drif et al., 2020 [4]*)

Soient les variables suivantes :

- Tx_c : taux de mortalité choqué,
- Tx_i : taux de mortalité central,
- F : le facteur additif de mortalité.

On a alors la relation suivante pour la période étudiée :

$$Tx_c = Tx_i + F$$

Le facteur additif correspond à une majoration des taux initiaux. En considérant un facteur

additif de 0,002 % et un taux initial avant choc de 0,03 %, on obtient alors un taux choqué de 0,032 %.

Le facteur Décès correspond au taux de mortalité additionnel qui majore annuellement les hypothèses de mortalité de l'assureur pour chacune des années de projections. En guise d'exemple, considérons le taux flat initial t_i de décès de 0,3 % pour les années 2021 à 2050.

Ainsi, dans le cadre de la majoration de la mortalité induite par la pollution pour la ville de Nantes, où l'on observe ces facteurs additionnels :

- 0,01 % pour l'horizon 2021-2030,
- 0,02 % pour l'horizon 2031-2040,
- 0,02 % pour l'horizon 2041-2050.

On observera les taux choqués suivants pour chaque année de projection :

- 0,31 % pour l'horizon 2021-2030,
- 0,32 % pour l'horizon 2031-2040,
- 0,32 % pour l'horizon 2041-2050.

Le facteur frais de soins correspond à la proportion additionnelle des têtes assurées qui, pour chaque année de projection, vont être amenées à générer des frais de soins supplémentaires dus à la pollution de l'air.

Parmi ce nombre d'assurés qualifiés de cas attribuables, une hypothèse sur la répartition de la nature des frais de soins engagés a été posée. Il a été retenu que 3 % des frais de soins seraient à l'origine de cas d'hospitalisations pour une durée moyenne de 6 jours et 97 % seraient à l'origine de consultations. Concrètement, pour un taux de 1,02 % de frais de soin adossés à un portefeuille de 10 000 assurés, chaque année, en sus de la sinistralité *Best Estimate*, 102 assurés ($10\,000 \times 1,02\%$) généreront des frais de soins additionnels. Parmi ces assurés, 3 ($102 \times 3\%$) seront hospitalisés 6 jours et 99 ($102 \times 97\%$) ne généreront qu'une « simple » consultation. Dans les résultats de l'étude, le facteur de proportion additionnelle communiqué est décliné selon les 3 horizons temporels, de la même manière que pour le facteur décès.

Enfin, le facteur arrêts de travail correspond à la proportion additionnelle des têtes assurées qui, pour chaque année de projection, vont être en arrêt de travail pour une durée moyenne de 6 jours. Un taux de 0,08 % d'arrêt de travail induit pour un portefeuille de 10 000 assurés, chaque année, en sus de la sinistralité *Best Estimate*, 8 arrêts de travail ($10\,000 \times 0,08\%$) pour une durée moyenne de 6 jours.

Limites identifiées des exercices pilotes pour la pollution de l'air

Cet exercice fait cependant apparaître un certain nombre de limites méthodologiques qui seront mises en lumière dans le paragraphe suivant. Considérant les sources énoncées, il est possible de relever :

- Ce point de départ peut être utilisé pour obtenir des premiers résultats mais ne doit nullement être considéré comme étant une approche parfaite. Il doit servir de point de départ aux travaux futurs qui permettront peut-être de mettre en lumière des incohérences ou des approches plus optimales permettant de produire des stress tests climatiques plus en phase avec l'exposition des différents acteurs.

- Le choix du scénario prospectif : Le scénario RCP8.5 est particulièrement adverse en matière de scénario de hausse globale de la température. Cependant peut-on se limiter à l'étude de ce scénario pour l'évolution de la pollution de l'air ? Comme énoncé auparavant l'utilisation des scénarios RCP pour l'étude de la pollution de l'air constitue une déviation de leur utilisation première. De plus l'utilisation d'un unique scénario ne complète pas les recommandations du guide d'application de l'EIOPA qui suggère à minima l'étude de deux scénarios. Dans ce sens l'exercice de 2023 réalise l'étude d'un second scénario le RCP4.5, complétant par la même occasion les recommandations de l'EIOPA.
- Il peut paraître contradictoire d'appliquer des chocs haussiers à chaque nouvelle décennie compte tenu de la tendance historique des concentrations de la plupart des polluants : une étude de la matérialité de chaque polluant est peut-être à considérer avant de les inclure dans le choc à appliquer.
- La maille géographique : les résultats de taux additifs ne couvrent que les zones géographiques les plus urbaines, avec 10 villes parmi les plus importantes en France ainsi qu'un taux à l'échelle nationale. Est-il cependant parfaitement cohérent de négliger la pollution rurale ? En effet un focus sur les chocs en agglomération est réalisé alors qu'une hausse de la pollution à l'ozone en milieu rural est anticipée, « Cette pollution s'observe en général de manière plus intense dans les régions périurbaines et rurales sous le vent des agglomérations. » (*PREV'AIR*, [21]).
- Les chocs publiés en 2020 ne se déclinent pas par tranche d'âge, il est pourtant parfaitement envisageable que certaines populations soient davantage susceptibles d'être fragilisées en raison de leur âge (jeunes enfants ; populations âgées). Cette limite a par ailleurs été corrigée dans l'exercice de 2023.

1.2 La pollution de l'air

D'après une évaluation datant de septembre 2016, 92 % de la population mondiale respire un air ambiant trop pollué (*OMS*, 2016 [22]). Cette pollution de l'air serait par ailleurs « responsable dans le monde d'environ 7 millions de décès en moyenne par an » (*OMS*, 2014 [23]).

Il semble alors nécessaire pour les assureurs de réaliser des travaux permettant de quantifier l'incidence du poids de la pollution de l'air dans la dégradation de la santé afin d'identifier les solutions prudentielles et tarifaires les plus justes et représentatives de leur sinistralité.

En France, la qualité de l'air est aujourd'hui très surveillée et la réglementation progresse dans ce sens. Cette démarche fait suite à des études qui ont mis en lumière le lien entre une exposition à la pollution de l'air, qu'elle soit de court terme ou de long terme, et la dégradation de la santé. Aussi quelles sont les causes de la dégradation de la qualité de l'air ? quelles en sont les conséquences ? et quels sont les leviers d'action à disposition ainsi que leur mesure de mise en œuvre ?

1.2.1 Définitions

Dans un premier temps, il est nécessaire de faire la distinction entre la pollution intérieure et la pollution extérieure. La première fait référence à la dégradation de la qualité de l'air que l'on peut retrouver à l'intérieur des bâtis. La seconde fait référence à la pollution que l'on retrouve dans l'environnement extérieur aux bâtis. Bien que la qualité de l'air en intérieur soit la plupart du temps plus dégradée, elle sera négligée dans cette étude et l'on ne considèrera uniquement la pollution extérieure.

Dans un second temps, il faut différencier la pollution d'origine naturelle de la pollution d'origine humaine également appelée pollution anthropique. Cette dernière est directement issue des activités humaines qui, pour certaines d'entre elles, ont pour conséquence la dégradation de la qualité de l'air via le relâchement dans l'atmosphère de gaz parfois nocifs pour la santé humaine, animale et environnementale. Quant aux émissions naturelles, elles sont caractérisées par une origine indépendante de l'action délibérée de l'homme.

Ensuite on retrouve la pollution transfrontalière. Celle-ci fait référence à la pollution qui « à la suite de l'émission de (ces) polluants dans l'air, les substances se déplacent dans l'air en fonction des différentes conditions météorologiques et particulièrement à l'aide des effets du vent. Aussi il est tout à fait possible de retrouver certaines particules à des milliers de kilomètres de leur source initiale. » (*Ministère de la Transition écologique et de la Cohésion des territoires, Ministère de la Transition énergétique, 2023 [8]*).

Enfin, il est nécessaire de distinguer les polluants dits « primaires » des polluants « secondaires ». Brièvement, les polluants « primaires » font référence aux éléments émis directement dans l'air. Par exemple parmi les polluants primaires figurent les polluants issus du trafic routier ou des différentes industries métallurgiques. Le lien entre la source émettrice et la contamination de l'air se fait donc directement. D'autres part les polluants dits « secondaires » font référence aux polluants issus de transformations chimiques qui se forment en réagissant avec d'autres réactifs présents dans l'air et sous l'effet de conditions particulières (chaleur, rayonnement...).

1.2.2 Les polluants

En France, les concentrations de différents polluants sont suivies quotidiennement en raison du potentiel risque qu'ils représentent sur la santé. Quatre principaux polluants ont été retenus dans le cadre de cette étude. Ils correspondent aux polluants étudiés dans le cadre de l'exercice pilote climatique de 2020. Dans la section suivante il s'agira de les énumérer, en proposant une description succincte de leur origine, qu'elle soit primaire ou secondaire, du fardeau sur la santé humaine qu'ils représentent et des potentielles interactions qu'ils peuvent avoir entre eux.

Les particules ou poussières en suspension (PM)

Il existe deux types de « PM », aussi appelées particules fines, elles sont aujourd'hui mesurées et suivies dans le cadre de la surveillance de la qualité de l'air en France. Ces dernières sont classées par taille.

Tout d'abord, les PM_{10} qui correspondent aux particules de diamètre inférieur à 10 micro-

mètres. Elles sont retenues au niveau du nez et des voies aériennes supérieures. Ensuite, les particules $PM_{2.5}$ font référence aux particules de diamètre inférieur à 2,5 micromètres. Ces dernières sont particulièrement dangereuses du fait de leur taille infime leur permettant de pénétrer profondément dans l'appareil respiratoire de l'homme, jusqu'à passer dans la circulation sanguine.

Ces deux classes de particules fines sont nocives pour la santé car elles sont sources de problèmes respiratoires chez les personnes sensibles et peuvent dans certains cas être responsables de décès prématurés en raison d'affections respiratoires, de maladies cardiovasculaires ou de développement de certains types de cancers.

L'ozone (O_3)

L'ozone est un gaz que l'on retrouve naturellement en grande quantité dans notre atmosphère, celui-ci étant indispensable à la vie terrestre. Naturellement présent entre 12 km et 50 km d'altitude dans la stratosphère, il permet de créer un bouclier contre les rayons ultraviolets provenant du soleil. Toutefois dans la troposphère qui correspond à la couche basse de l'atmosphère allant de la surface du sol à 12 km d'altitude, l'ozone constitue un polluant nocif pour la santé humaine.

L'ozone est un polluant secondaire, ce qui signifie qu'il résulte d'une transformation et n'est donc pas émis directement dans l'atmosphère. Cette transformation est qualifiée de photochimique et fait intervenir différents composés aussi appelés précurseurs tels que les composés organiques volatils (COV), les oxydes d'azote (NOx) et le monoxyde de carbone qui réagissent entre eux en présence de photons de lumières en particulier issus des rayons UV du soleil.

Aussi, « Les épisodes de pollution à l'ozone surviennent principalement durant l'été, lors de situations anticycloniques calmes, ensoleillées et chaudes, avec peu ou pas de vent. Les périodes de canicule sont donc propices à l'apparition de tels épisodes. » [8].

Par ailleurs « sa formation est particulièrement favorisée par des conditions météorologiques spécifiques comme des températures élevées, un fort rayonnement solaire UV et en présence de ses précurseurs en quantité suffisante : oxydes d'azote, composés organiques volatils non méthaniques ($COVNM$), CO et méthane » (*Drif et al., 2020 [4]*).

L'ozone est un polluant particulièrement compliqué à appréhender, son cycle de production/destruction étant complexe et faisant intervenir un certain nombre de paramètres exogènes. Voici ci-dessous un schéma simplifié de sa formation.

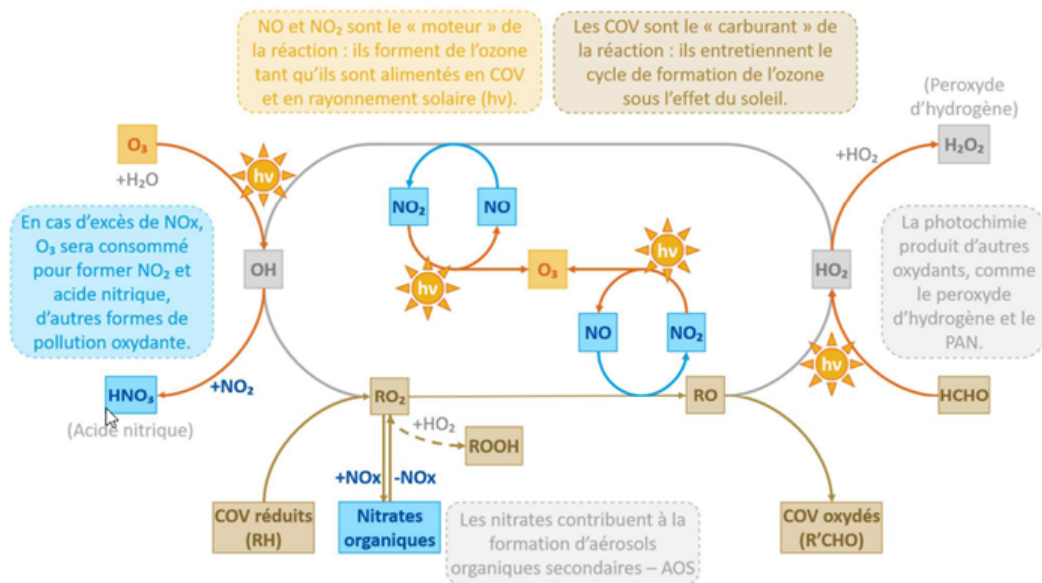


FIGURE 1.8 – Schéma de la formation d'ozone troposphérique (O_3)

Ses impacts sur la santé sont divers, et sont notamment dus à son caractère irritant impactant à la fois l'appareil respiratoire et les yeux. Il est ainsi responsable d'une hausse du taux de mortalité en période de pics de pollution et contribue à l'effet de serre.

Deux indicateurs réglementaires sont généralement utilisés dans les études d'impact de l'incidence de l'ozone sur la santé et l'évaluation basée sur l'exposition à une concentration donnée :

- Les maxima quotidiens moyens de concentration d'ozone troposphérique en période estivale sur la base de moyennes mobiles sur 8 heures,
- Le *SOMO35* défini comme la somme annuelle du maximum journalier de plus de 35 ppb ($70 \mu\text{g}/\text{m}^3$) sur la base de 8 heures. Cet indicateur est particulièrement conçu pour saisir les effets néfastes de l'ozone sur la santé humaine en exposition long terme. Il se définit mathématiquement de la sorte :

$$SOMO35_{\text{non corrigé}} = \sum_i \max \{0, C_i - 35\text{ppb}\}$$

où C_i est la concentration moyenne journalière maximale sur 8 heures et la somme est annuelle, c'est-à-dire effectuée du jour $i = 1$ à 365.25 par année.

Par ailleurs on a la relation $1\text{ppb} = 2\mu\text{g}/\text{m}^3$, ainsi on peut également écrire :

$$SOMO35_{\text{non corrigé}} = \sum_i \max \{0, C_i - 70\mu\text{g}/\text{m}^3\}$$

Étant donné que l'indicateur *SOMO35* est sensible aux valeurs manquantes (c'est-à-dire les jours où les concentrations moyennes journalières maximales sur 8 heures ne sont pas disponibles, incorrectes, ou de couverture temporelle insuffisante). Lorsque ces données journalières sont manquantes, une correction en prenant la couverture

à temps plein (par exemple annuelle) doit être réalisée :

$$\text{SOMO35} = \text{SOMO35}_{\text{non corrigé}} \cdot \frac{N_{\text{total}}}{N_{\text{valide}}}$$

où N_{total} est le nombre de jours dans la période d'intérêt (365 pour une année, 180 pour les mois d'été) et N_{valide} est le nombre de jours avec des valeurs valides. Selon le bureau régional de l'OMS pour l'Europe, les valeurs de SOMO35 se situent habituellement entre 0 et 10 000. On utilise par la suite le SOMO35 pour déduire

des impacts sur la santé en calculant le risque relatif (RR) :

$$RR = \exp \left(\beta \cdot \frac{\text{SOMO35}_{\text{non corrigé}}}{N_{\text{valide}}} \right)$$

Les oxydes d'azote (NOx)

Les oxydes d'azote regroupent eux aussi deux polluants distincts mesurés de manière indépendante, il s'agit du monoxyde d'azote (NO) et du dioxyde d'azote (NO_2). Tous deux sont émis à la suite d'une combustion. Les sources anthropiques de ces types de polluants sont le chauffage au gaz, le trafic routier, la production d'électricité dans le cas du recours aux usines à charbon et les sources naturelles ne sont autres que les éruptions volcaniques, ou encore les sols asséchés qui peuvent émettre du monoxyde d'azote issu du processus biologique de transformation de l'azote des sols.

Une fois parvenu dans l'air, le monoxyde d'azote réagit et forme du dioxyde d'azote, gaz particulièrement irritant pour les bronches et aujourd'hui considéré responsable de crises d'asthme et d'infections pulmonaires. Ce polluant est particulièrement dangereux pour les populations sensibles comme les jeunes enfants et les personnes âgées.

Parallèlement cette classe de polluant possède des effets néfastes sur l'environnement tels que l'acidification des milieux ou encore l'eutrophisation. Ces effets ne seront cependant pas à l'étude dans cet écrit.

Substances polluantes autres

D'autres composants sont aujourd'hui suivis dans le cadre de l'étude de la qualité de l'air en France. On retrouve parmi ces composants le dioxyde de soufre (SO_2), les composés organiques volatils (COV), les métaux lourds (dont Fe , Zn , Ni , As , Cr). Ces polluants présentant un manque de données que ce soit en termes de mesure des concentrations ou en termes de traduction en incidence potentielle sur la santé, ne seront donc pas retenus pour la suite de l'étude.

1.2.3 Les sources de pollution

Les polluants peuvent provenir de différentes sources qu'il s'agira ici de mettre en lumière. En effet, les causes principales de la pollution de l'air primaire sont connues et peuvent être catégorisées par secteur d'activité ou par source naturelle pour un ensemble donné de polluants. Ainsi quatre grandes catégories seront retenues :

1. Les transports (transport routier, aérien, ferroviaire, fluvial) (hors transports internationaux),
2. L'agriculture/Sylviculture,
3. L'industrie (manufacturière, construction, énergie et traitement centralisé des déchets),
4. Résidentiel/Tertiaire.

La quantité de rejet de chacune de ces sources est estimée chaque année par le Centre interprofessionnel technique d'études de la pollution atmosphérique (Citepa), permettant de produire le graphique suivant :

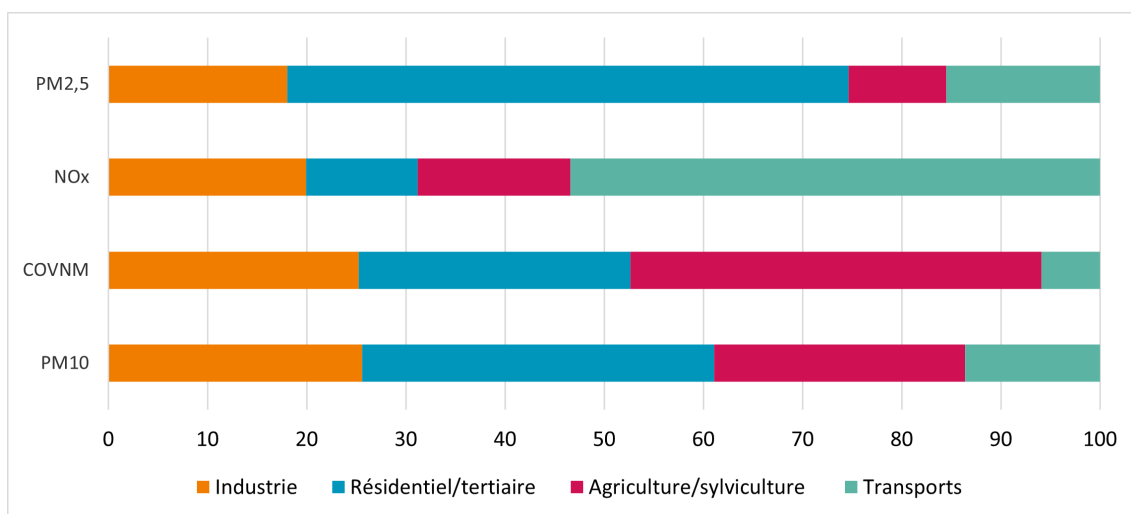


FIGURE 1.9 – Part des secteurs d'activité dans les émissions anthropiques de NO_x , $COVNM$, PM_{10} et $PM_{2,5}$ en 2021 (Citepa, 2022)

Bien que le graphique précédent soit arrêté à l'année civile 2021, on observe globalement que ces proportions sont relativement stables dans le temps depuis les années 2000 en consultant la publication « Les rejets de polluants dans l'air : résultats complémentaires du bilan de la qualité de l'air extérieur en France en 2021 ».

1.2.4 Évolution historique de la pollution de l'air

Évolution de la réglementation

La réglementation en matière de pollution de l'air fait depuis un certain nombre d'années l'objet de la mise en place de lois et d'initiatives de plus en plus restrictives. Ces interventions agissent autant à l'échelle nationale qu'à l'échelle européenne. Ci-dessous voici un bref récapitulatif chronologique des mesures en faveur d'une meilleure qualité de l'air ayant été appliquées au cours des dernières décennies.

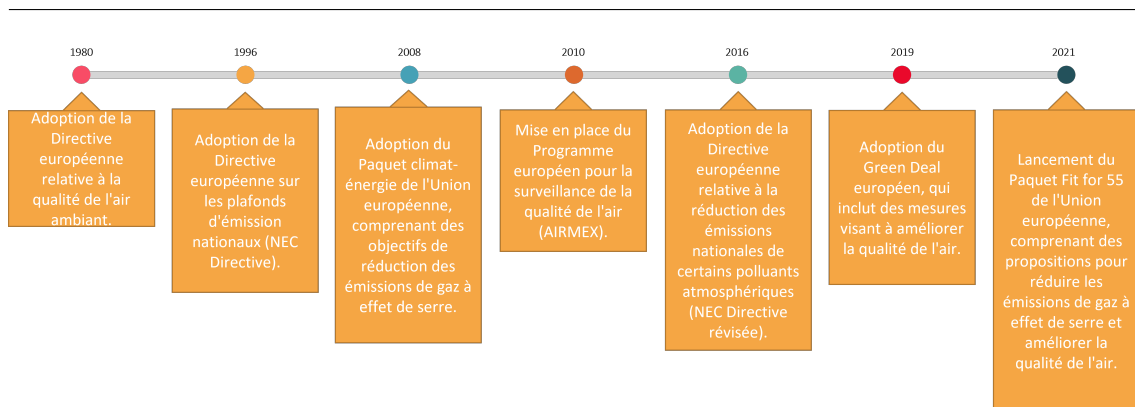


FIGURE 1.10 – Frise chronologique des principales directives relatives à l'amélioration de la qualité de l'air en Europe

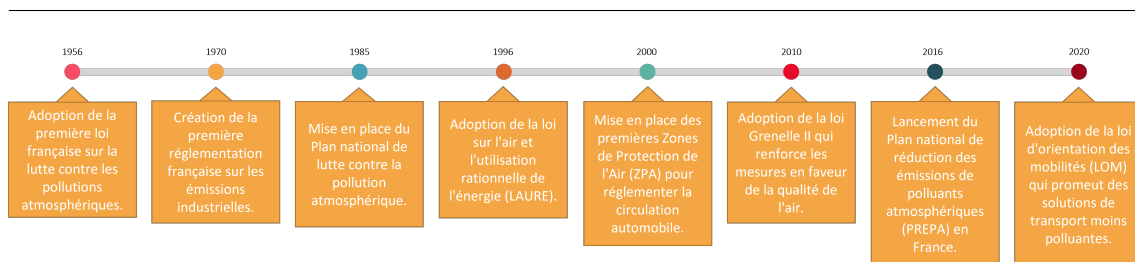


FIGURE 1.11 – Frise chronologique des principales directives relatives à l'amélioration de la qualité de l'air en France

Le Centre Interprofessionnel Technique d'Études de la Pollution Atmosphérique (Citepa) met ainsi à jour « le tableau de bord des engagements sur les polluants », recensant le respect ou non de la France concernant ses objectifs légaux en matière d'émissions.

Évolutions quantitatives historiques

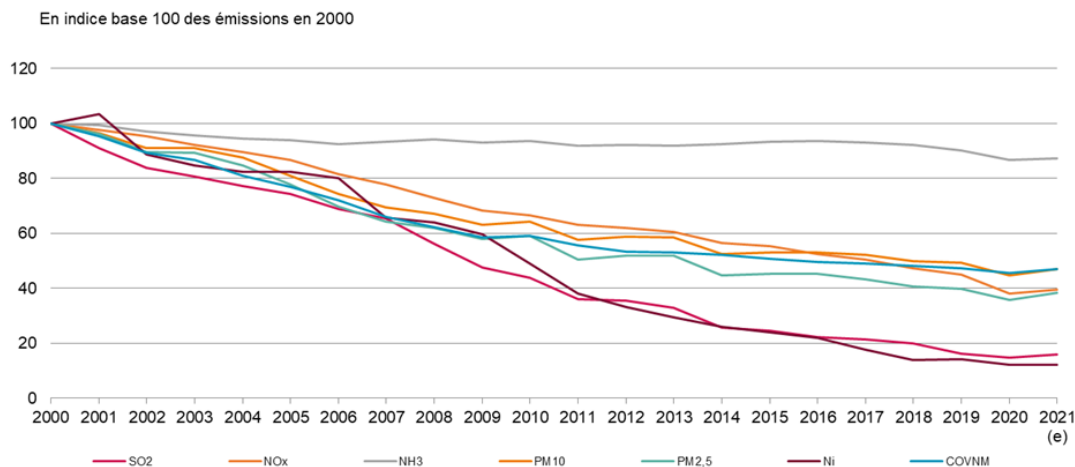


FIGURE 1.12 – Évolution des émissions anthropiques de quelques polluants en France métropolitaine (Corse exclue) en indice base 100 (*Citepa, 2022*)

Sur la base de ce *premier graphique (1.12)*, il est possible de constater que les efforts réglementaires mis en place par les autorités nationales ont permis de réaliser d'immenses progrès dans la baisse des émissions anthropiques internes au périmètre de la France métropolitaine, c'est-à-dire la quantité de polluant issue de l'activité humaine directement rejetée dans l'air. On enregistre ainsi sur la période 2000 à 2021 une baisse des émissions de :

1. - 53 % pour les PM_{10} ,
2. - 62 % pour les $PM_{2,5}$,
3. - 84 % pour le dioxyde de soufre (SO_2),
4. - 60 % pour les oxydes d'azote (NO_x),
5. - 53 % pour les composés organiques volatils non méthaniques ($COVNM$),
6. - 63 % pour le monoxyde de carbone (CO),
7. - 13 % pour l'ammoniac (NH_3).

L'ozone (O_3) ne figurant pas dans ce graphique car étant un polluant secondaire, il ne peut être considéré comme anthropique.

Ces diminutions permettent de mettre en relief la consistance des efforts réglementaires français et l'efficacité des mesures politiques mises en place. Mais comment se répercute cette baisse généralisée des émissions anthropiques, sur le niveau de concentration moyen des polluants étudiés ?

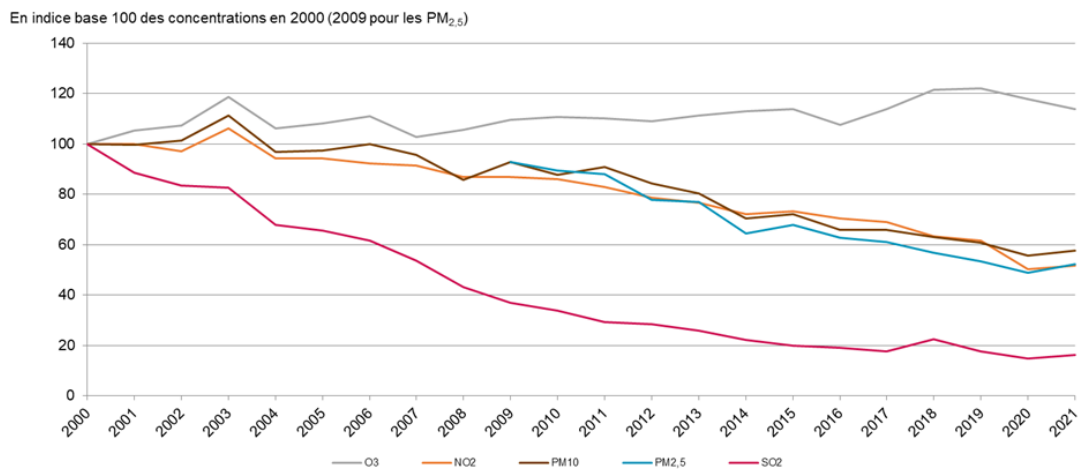


FIGURE 1.13 – Évolution des concentrations moyennes annuelles pour les polluants SO_2 , NO_2 , O_3 , PM_{10} et $PM_{2.5}$, en fond urbain, en France métropolitaine (Corse exclue) (*Geo-d'air*, 2022)

Ce *second graphique (1.13)* nous permet, pour le milieu « fond urbain » (milieu agglomératif), d'enregistrer les variations suivantes de concentration de polluants sur la période 2000 à 2021 :

1. - 48 % pour les $PM_{2.5}$,
2. - 42 % pour les PM_{10} ,
3. - 84 % pour le dioxyde de soufre (SO_2),
4. - 48 % pour les oxydes d'azote (NO_2),
5. +14 % pour l'ozone (O_3), seul polluant enregistrant une hausse de sa concentration sur l'historique retenu.

Bien que les baisses enregistrées en milieu urbain soient moins soutenues que les émissions anthropiques totales enregistrées en Figure 1.12, elles n'en restent pas moins particulièrement importantes. Trois principales raisons peuvent expliquer ces différences de variations.

La première correspond au fait qu'en milieu urbain les polluants sont davantage présents particulièrement pour le dioxyde d'azote (NO_2) qui est un polluant traceur du trafic routier objectivement plus conséquent dans les agglomérations. Il en est de même pour les particules fines issues majoritairement des combustions fossiles, des activités industrielles et des combustions domestiques.

Ensuite, les émissions anthropiques ne traitent par définition, que de la pollution émise par l'homme au sein du périmètre étudié. Cependant les concentrations annuelles moyennes peuvent être perturbées par les émissions des pays voisins, transportées et véhiculées jusqu'en France en raison de conditions météorologiques particulières. On parle alors de pollution transfrontalière.

Enfin, il est finalement nécessaire de faire une distinction entre « immission » et « émission ». On définit l'immission comme étant l'inventaire des concentrations de polluants présents dans l'environnement. Sa mesure est utilisée pour refléter la présence de polluants dans

l'air ambiant et par conséquent, les émissions ne sont pas parfaitement représentatives de la concentration moyenne que l'on peut observer au sein d'un territoire.

Compte tenu de ces premières analyses, il est désormais difficile d'exclure le rôle des politiques règlementaires, qu'elles soient à l'échelle nationale ou européenne dans l'évolution prospective de la qualité de l'air. Ces normes règlementaires et directives ont donc une importance potentielle non négligeable à prendre en considération dans l'élaboration de scénarios prospectifs.

1.2.5 Lien entre émissions et immissions

Afin d'investiguer davantage le lien entre ces deux termes, il s'agira d'étudier d'une part la causalité de ces phénomènes et d'autre part la relation de corrélation.

On définira la causalité dans cet écrit comme étant l'influence par laquelle un événement, un processus, un état ou un objet (une cause) contribue à l'élaboration d'un second événement, processus, état ou objet (un effet) qui sera considéré comme sa conséquence.

Rappelons également que selon *Colette et al., 2013b [11]* et *Colette et al., 2013a [12]* il est possible d'affirmer que seules les émissions anthropiques sont les principaux facteurs déterminants de la concentration moyenne d'un polluant à l'horizon 2050. Les variations des seules variables climatiques ont un impact secondaire sur la qualité de l'air en particulier pour les particules fines ($PM_{2.5}$, PM_{10}) et le dioxyde d'azote (NO_2) (*Lecoœur, 2013 [13]*).

Par conséquent, il est tout à fait raisonnable de penser que les données d'émissions (la cause) soient représentatives des concentrations relevées en France métropolitaine (l'effet). Cependant un biais existe entre ces deux concepts. Pour approfondir ce raisonnement, il s'agira désormais de quantifier la corrélation de ces deux phénomènes pour chacun des polluants étudiés.

Pour rappel, le coefficient de corrélation linéaire ou coefficient de Bravais-Pearson entre deux variables aléatoires réelles X et Y , à variance finie, noté par $\text{Cor}(X, Y)$ ou par r , est défini par :

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Où $\text{Cov}(X, Y)$ désigne la covariance des variables X et Y , et σ_X et σ_Y désignent leurs écarts types.

De manière équivalente :

$$r = \frac{E[(X - E(X))(Y - E(Y))]}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y}$$

où $E[X]$ désigne l'espérance mathématique de la variable X .

En supposant que X soit la variable suivant l'évolution de la concentration d'un polluant en France en base 100 sur la période 2000-2020 (2009-2020 pour les $PM_{2.5}$), et Y suivant l'évolution des émissions en France de ce même polluant¹ sur la même période, on obtient alors :

1. à l'exception des NO_2 dont on a calculé la corrélation selon les émissions de NO_x , seules données à disposition

- $r_{PM_{2.5}} = 0,96$
- $r_{PM_{10}} = 0,87$
- $r_{NO_2} = 0,96$

Les coefficients de corrélation étant tous les trois élevés et proches de 1, on parle alors de corrélation positive forte. Pour la suite de l'étude, on suggérera ainsi que les données d'émissions prospectives seront représentatives de la concentration des polluants cités ci-dessus en France. Ce proxy permettra lors de la constitution des scénarios prospectifs au *Chapitre 3*, d'utiliser les projections d'émissions selon différents scénarios abordés plus tard dans l'écrit.

1.2.6 Impact du changement climatique sur la qualité de l'air

Compte tenu de ces premiers résultats historiques aux tendances particulièrement bassières, il peut être surprenant d'étudier l'incidence de la pollution de l'air. Cependant les impacts de cette dernière sont aujourd'hui encore particulièrement préoccupants. Singulièrement, l'ozone apparaît comme étant le seul polluant dont le niveau de concentration dans l'air n'évolue pas dans le sens attendu par les autorités publiques malgré les efforts mis en place.

Ce résultat est d'autant plus surprenant que les polluants précurseurs de ce gaz sont en baisse (1.12). L'étude des corrélations est donc dans ce cas inexploitable en raisons de corrélations négatives des concentrations d'ozone et des concentrations de ses précurseurs. Il est alors de notre ressort de trouver un phénomène capable d'expliquer la concentration d'ozone en France métropolitaine. Aussi, ce dernier ne serait-il pas le seul polluant influencé par le changement climatique ? Plus généralement peut-on parfaitement dissocier le phénomène de hausse globale de la température de celui de l'évolution de la qualité de l'air en France ?

Il est en effet impossible de rejeter le lien entre ces deux manifestations. L'ozone étant formé lors d'une réaction photochimique faisant intervenir les rayons ultraviolets du soleil et particulièrement favorisé en présence de fortes chaleurs, il est aisé de conclure que pour ce polluant en particulier, la hausse globale de la température anticipée dans les années futures, aura très certainement une incidence néfaste sur la qualité de l'air. C'est ce que nous allons démontrer ci-après.

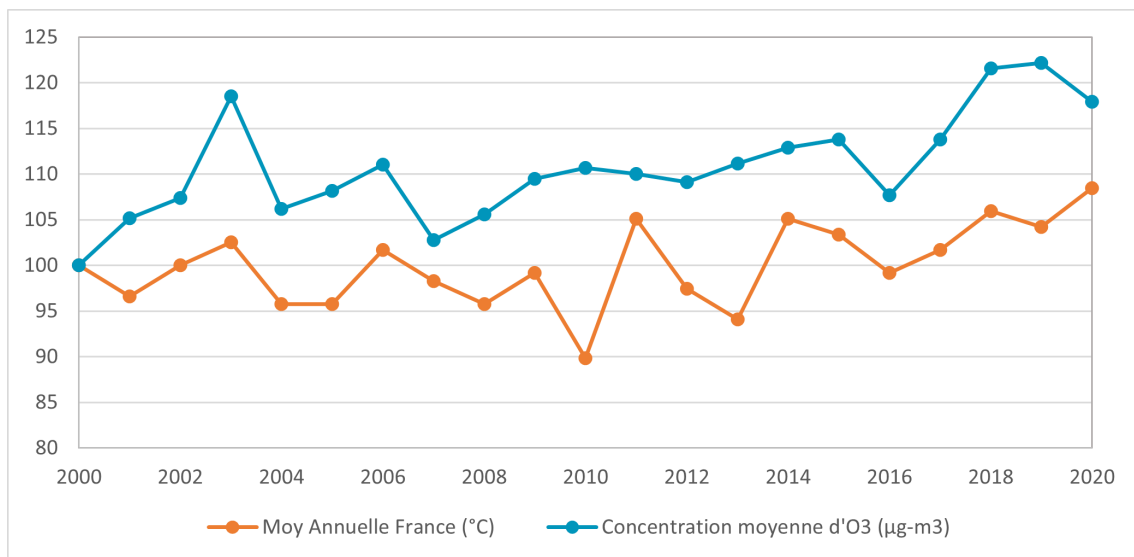
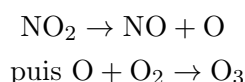


FIGURE 1.14 – Évolution de la concentration moyenne d’ozone (O_3) et de la température moyenne annuelle en France en base 100 (période 2000-2020)

Afin de détailler davantage le lien entre ces deux variables, il s’agira de nouveau d’étudier d’une part la causalité et, d’autre part, la relation de corrélation entre ces deux phénomènes.

L’ozone troposphérique est ainsi principalement issu de la transformation, sous l’effet du rayonnement solaire, des oxydes d’azote (NO_x) et des composés organiques volatils (COV) émis majoritairement par les activités humaines à l’issue d’un procédé chimique appelé photodissociation (*Wikipédia [24]*).

Ainsi, sous l’action de rayonnements solaires de courte longueur d’onde (uv) :



(en présence d’autres oxydants, sinon, NO_2 se reforme : $NO + O_3 \rightarrow NO_2 + O_2$) [24].

Plus généralement l’équation de la production d’ozone troposphérique (mauvais ozone) est souvent vulgarisée comme suit :



La relation de causalité ainsi démontrée par l’équation de production d’ozone troposphérique (effet) réagissant en présence de fortes chaleurs (cause). Les fortes chaleurs ici approximativement les rayons UV de courte longueur d’onde normalement responsables de la réaction photochimique.

En supposant que X soit la variable suivant l’évolution de la concentration en ozone en France en base 100 sur la période 2000-2020, et Y suivant l’évolution de la température moyenne annuelle en France sur la même période, on obtient alors $r = 0,58$.

Le coefficient de corrélation étant particulièrement sensible aux valeurs « déviantes », ce dernier s’élève à $r_{\text{corrigé}} = 0,67$ lorsqu’on retire l’année 2010, année pouvant être

considérée extrême car particulièrement froide. On parle alors de corrélation positive forte dans les deux cas, le coefficient r étant supérieur à 0,5. Cependant ces résultats sont bien en deçà de ceux obtenus en *sous-section précédente* (1.2.5) pour les autres polluants traités.

Pour renforcer la justification du lien qui existe entre ces deux phénomènes, on fera appel à la littérature.

Une étude conduite par *Colette et al., 2015 [25]* renforce le lien établi entre la température et la concentration d’ozone en Europe. En effet, cette ressource étudie la « pénalité climatique », où, les mesures dans lesquelles les perturbations climatiques influent sur les niveaux de concentration en ozone troposphérique malgré la baisse des concentrations de ses polluants précurseurs. Ensuite une seconde étude *Boletti et al., 2020 [26]*, par le biais de l’utilisation de séries temporelles, établit une relation linéaire entre les valeurs maximales quotidiennes d’ozone et la température en la saison estivale (entre mai et septembre). Les tendances sont calculées sur les valeurs moyennes journalières et pour la période 2010-2015. Aussi bien que les échelles temporelles diffèrent avec celle retenue dans cette étude, le lien entre la température et le niveau d’ozone (sous entendu la valeur de l’indicateur *SOMO35*) est bien réel. Il est cependant nécessaire de préciser que la linéarité entre température et ozone ne fait pas consensus au sein de la communauté scientifique.

Pour la suite de l’étude, on suggérera désormais que la température moyenne annuelle sera linéairement représentative de la concentration d’ozone et par conséquent de l’indicateur réglementaire *SOMO35* en France. Ce proxy permettra lors de la constitution des scénarios prospectifs, d’utiliser les projections selon les différents scénarios climatiques RCP, et SSP présentés plus tard dans l’écrit.

En section suivante, la liaison entre la pollution de l’air et une incidence directe sur la santé ou la mortalité de l’Homme est discutée.

1.3 Méthodologies de quantification de l’incidence de la pollution atmosphérique

Il existe aujourd’hui deux grandes approches reconnues dans l’évaluation des impacts de la pollution atmosphérique sur la santé : les études épidémiologiques et les évaluations quantitatives d’impact sur la santé de la pollution de l’air (EQIS-PA) (*ORS, 2022 [27]*).

1.3.1 Les études épidémiologiques

Ces études, en se basant sur des travaux menés depuis plusieurs décennies sur des cohortes nationales d’effectif pouvant dépasser plus de 360 000 personnes suivies (comme peut l’être la cohorte *Escape*), ont permis dans un premier temps d’identifier que lorsque la pollution augmente, la mortalité respiratoire et cardiaque enregistrent une évolution de même tendance avec quelques jours de décalage : on parle alors d’impact à court terme.

D’autre part, elles ont également permis de mettre la lumière sur l’effet long terme qui se traduit par une exposition prolongée à un ou plusieurs polluants. Cette exposition prolongée induit une augmentation de la survenance de certaines maladies comme le sont certains cancers respiratoires ou d’autres maladies comme l’infarctus du myocarde.

Bien que l'impact à court terme soit beaucoup moins significatif que l'impact à long terme, le fardeau sanitaire de la pollution atmosphérique se traduit par l'association additive de ces deux phénomènes.

In fine, ces études permettent sur la base d'un travail toxicologique et épidémiologique complexe, de démontrer l'existence d'un lien causal entre une exposition et une incidence. Cette relation est particulièrement vérifiée pour le cas des polluants atmosphériques retenus, l'impact étant réel, et nul ne pouvant s'y soustraire.

1.3.2 Le risque relatif (*RR*)

Concrètement, le risque relatif estime l'ampleur d'une association entre l'exposition et la maladie en indiquant la probabilité de développer la maladie dans le groupe exposé p_E , par rapport à ceux qui ne sont pas exposés p_U . Cette probabilité est égale à $\frac{p_E}{p_U}$.

Par exemple, si on suppose $RR = 1,06$ (IC 95 % = 1,01-1,12) pour toutes les causes de mortalité avec une augmentation annuelle moyenne de $PM_{2.5}$ de $10 \mu g/m^3$, cela signifie qu'une augmentation annuelle moyenne de $10 \mu g/m^3$ de $PM_{2.5}$ est associée à une augmentation de 6 % des décès toutes causes confondues (avec des intervalles d'incertitude compris entre 1 % et 12 %).

Les risques relatifs dus à la pollution de l'air sont habituellement modélisés avec une fonction log-linéaire :

$$RR = \frac{\exp(\alpha + \beta X)}{\exp(\alpha + \beta X_0)} = \exp[\beta(X - X_0)].$$

Où X désigne la concentration de polluants ($\mu g/m^3$), et X_0 désigne la valeur seuil ou contrefactuelle (par exemple la concentration de fond ou la valeur la plus basse possible ($\mu g/m^3$)). Dans le modèle log-linéaire, β indique la variation du RR pour une variation d'une unité de la concentration X .

Passons désormais à la présentation des évaluations quantitatives d'impact sur la santé de la pollution de l'air (EQIS ou EQIS-PA).

1.3.3 Les évaluations quantitatives d'impact sur la santé de la pollution de l'air (EQIS-PA)

Les études EQIS sont fondées sur des estimations statistiques et permettent de fournir un ordre de grandeur de l'impact de la pollution de l'air sur la santé et la mortalité.

Les EQIS-PA permettent notamment de :

- Calculer l'ampleur du fardeau de la pollution atmosphérique sur la santé
- Quantifier les bénéfices attendus pour la santé des différents scénarios de réduction des niveaux de pollution au sein d'une population donnée en adéquation par exemple avec les seuils réglementaires précédemment cités
- Évaluer les bénéfices potentiels d'une action politique visant à réduire la pollution

- Mesurer les impacts financiers/monétaires de l'exposition à la pollution de l'air. Les estimations ainsi produites et les résultats obtenus servent à accompagner les politiques publiques mais également à légitimer les mesures mises en place avec pour objectif de réduire les niveaux de pollution.

Ces évaluations se déclinent à différentes échelles et sont ainsi réalisées à l'échelle nationale comme à des granularités plus fines comme peut l'être le niveau régional. Leur réalisation suit la méthodologie ci-après :

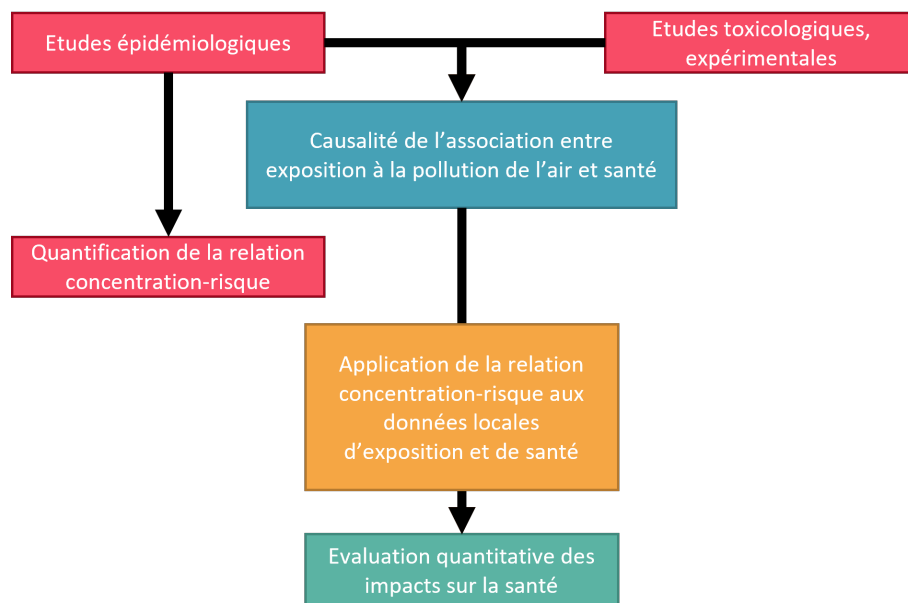


FIGURE 1.15 – Schéma méthodologique de la réalisation d'une EQIS-PA, Santé publique France 2021

Un outil *Open-source* a été créé par l'OMS pour agréger l'ensemble des savoirs

Dans une démarche visant à faciliter l'accès et la diffusion des connaissances, l'Organisation Mondiale de la Santé (OMS) a mis en place un outil *Open-source* appelé AirQ+. Cet outil a été spécialement conçu pour regrouper de manière systématique et centralisée les diverses sources de savoirs disponibles liant pollution de l'air et incidence sur la santé. La sous-section traite davantage de cet outil.

1.3.4 AirQ+

Présentation générale

AirQ+ est un logiciel développé par le bureau régional Europe de l'OMS qui permet de quantifier le fardeau d'une pollution atmosphérique dégradée, sur la santé. Directement mis à disposition de tous en libre accès, AirQ+ inclut différentes méthodologies pour évaluer les effets de l'exposition à long terme (et à court terme) sur la pollution de l'air ambiant. Il permet notamment de travailler sur des polluants tels que les particules fines, le dioxyde d'azote, l'ozone ou encore le « *black carbon* ».

Il est alors possible d'étudier un certain nombre d'affections aiguës ou chroniques, par le biais d'indicateurs de santé liés à la morbidité et la mortalité. Ces derniers sont disponibles en annexe A.2.

Les fondements scientifiques de cet outil se basent sur des études menées dans les pays développés tels que les pays d'Europe et d'Amérique du Nord. Ces études ont quantifié les effets sur la santé des divers polluants énoncés ci-dessus et ont été intégrés dans l'outil. AirQ+ se base ainsi également sur la méthodologie des fonctions réponses utilisées dans le cadre des EQIS en France.

Analyses possibles

Par le biais du logiciel, il est possible de réaliser 3 principales analyses :

- **Une évaluation d'impact** : cette analyse consiste à répondre à la question « Quel serait le changement sur le plan de la santé si le niveau de pollution de l'air diminuait ou augmentait ? ». C'est cette évaluation qui sera réalisée par la suite.
- **Une Charge de morbidité** : cette seconde analyse consiste à répondre à la question « Dans quelle mesure un résultat particulier pour la santé (par exemple la mortalité) est-il attribuable l'exposition actuelle à un polluant atmosphérique ? ».
- **Analyse de risque** : enfin, cette dernière analyse consiste à répondre à la question « Quel est le risque de cancer au cours de la vie à certains polluants atmosphériques pour lesquels il existe un risque unitaire ? ».

Données requises

Le logiciel se présente sous la forme d'un exécutable classique. Après un court instant de chargement, le logiciel invite l'utilisateur à choisir l'une des trois analyses présentées en sous-section précédente 1.3.4.

Pour le cas d'une évaluation d'impact qui est l'analyse qui a été réalisée au sein de l'étude, il est alors demandé à l'utilisateur de compléter les informations suivantes :

- La valeur moyenne de la concentration du polluant choisi sur la zone d'étude (exprimée en $\mu\text{g}/\text{m}^3$ sauf pour l'ozone dont il faut renseigner l'indicateur réglementaire *SOMO35*).
- L'effectif de l'ensemble de la population vivant dans la zone d'étude.
- L'indicateur de santé qu'il souhaite étudier : il est possible de choisir dans une liste déroulante de choix en fonction du polluant étudié. Une nouvelle fois l'ensemble de ces indicateurs sont disponibles en annexe A.2.
- Le taux brut de mortalité également qui est défini comme le nombre de décès pour 100 000 habitants sur une année de la population en milieu d'année.
- La population à risque qui représente le pourcentage de la population exposée. Au sein de l'étude, l'hypothèse retenue considère que la totalité de la population de la zone d'étude est exposée. Il est également parfois possible de ne retenir qu'une catégorie spécifique comme peut l'être le nombre d'individus pour une tranche d'âge spécifique (adultes de 30 ans et plus par exemple.).

- La valeur seuil ou valeur contrefactuelle qui correspond à un niveau de concentration en dessous duquel il est choisi de ne pas dépasser. Comme pour l'étude de Drif et al., 2020 [4], cette valeur seuil sera de $0\mu\text{g}/\text{m}^3$.
- Le risque relatif ou fonction concentration-risque qui traduit la relation entre l'exposition d'une population à la pollution de l'air et ses conséquences sur l'indicateur de santé étudié. Cette notion peut être vulgarisée comme étant la corrélation entre une dose d'exposition et un impact sur la santé ou comme la probabilité de développer une maladie pour les personnes exposées.

Résultats de l'analyse d'impact

Après avoir complété les données requises il est alors possible de lancer un calcul et d'obtenir trois résultats distincts :

La part attribuable : cette valeur correspond à

$$\text{RRX}(\text{expo}, \text{seuil}, \beta) = \exp(\beta \cdot (\text{expo} - \text{seuil}))$$

$$\text{Part attribuable} = \frac{1 \cdot (\text{RRx} - 1)}{1 + 1 \cdot (\text{RRx} - 1)}$$

L'estimation du nombre de cas attribuables : cette valeur correspond

$$\text{Nombre de cas attribuables} = \frac{\text{part_attribuable} \cdot \text{incidence} \cdot \text{population}}{100\ 000}$$

L'estimation du nombre de cas attribuables pour 100 000 personnes à risque : cette valeur correspond

$$\text{Nombre de cas attribuables pour 100 000 personnes} = \text{part_attribuable} \cdot \text{incidence}$$

Méthode d'impulsion : Migration d'AirQ+ sous python

Dans une démarche d'optimisation, et afin de conserver l'analyse à la maille fine envisagée en *Chapitre 2*, l'outil a été réimplémenté sous Python dans le cadre des travaux menés en vue du présent mémoire dans le but d'accélérer le processus quantitatif.

En effet le logiciel AirQ+ impose la réalisation d'une évaluation à la fois pour chaque zone d'étude contraignant l'utilisateur à saisir toutes les données précédemment citées en sous-section 1.3.4 à la main. Cet aspect constitue une limite opérationnelle qui ne concorde pas avec l'étude d'un grand nombre de zones, ce qui sera par la suite retenue et présenté.

C'est notamment pour les raisons citées précédemment qu'il a été décidé de réimplémenter le processus calculatoire de l'outil dans un environnement python. Pour ce faire les fiches méthodiques et la transparence de l'outil AirQ+ ont permis de reproduire la méthodologie. De plus cette migration sous python a permis de construire un algorithme traitant chacune des étapes localement de « bout en bout ». Finalement après analyses des résultats obtenus sous python et sous AirQ+, et la vérification de leur concordance, le modèle a été validé.

Après avoir établi le contexte relatif aux implications assurantielles du changement climatique, exploré la problématique de la pollution atmosphérique et exposé la méthodologie employée pour évaluer ses impacts sur la santé, le prochain chapitre se focalise sur l'élaboration d'une approche novatrice de construction d'un module exposition liant pollution et démographie à une granularité fine.

Chapitre 2

Méthodologie de construction d'un zonier de la pollution de l'air sur données historiques

2.1 Sources de données utilisées

2.1.1 *Open-data*

Pour la réalisation de cette étude, des données *open-data* ont été utilisées. Selon la Commission nationale de l'informatique et des libertés (CNIL), « L'*open-data* désigne un mouvement, né en Grande-Bretagne et aux États-Unis, d'ouverture et de mise à disposition des données produites et collectées par les services publics (administrations, collectivités locales...) » [28].

Les données ainsi mises à disposition de tous peuvent être réutilisées par toute personne qui le souhaite sous certaines conditions fixées par le Code des relations entre le public et l'administration (CRPA) et le Règlement général sur la protection des données (RGPD).

Au sein de cette partie, il s'agira de présenter les typologies de données utilisées et leur source en tentant de présenter quelles ont été les raisons de leur utilisation et les mesures dans lesquelles ces différentes sources interagissent entre elles.

2.1.2 Variables cibles

Dans un premier temps il s'agit de présenter les données qui ont successivement été recueillies, retraitées et utilisées pour la modélisation de l'incidence historique de la pollution de l'air en France métropolitaine.

Les données de pollution de l'air

Il existe aujourd'hui plusieurs sources de données *open-data* relatives à la pollution de l'air en France. Il s'agira ici de les énumérer en présentant leurs avantages et limites, et en

justifiant de la source retenue pour l'étude.

— CAMS

On retrouve tout d'abord les données du projet *Copernicus Atmosphere Services Monitoring* (CAMS) issues du projet européen Copernicus, programme européen d'observation de la Terre, qui s'intéresse à la thématique environnementale de notre planète en offrant des services d'observation de la Terre via des données satellitaires et des données *in situ* (données non spatiales).

Le projet CAMS est chargé de créer et fournir des données et des informations continues sur la composition atmosphérique, notamment en matière de qualité de l'air.

Le jeu de données identifié « *European air quality reanalyses* » fournit des données annuelles ré-analysées de la qualité de l'air pour l'Europe, basées sur des observations provisoires et validées.

Le modèle CAMS produit ainsi des données annuelles de la qualité de l'air à une résolution spatiale de l'ordre de 0,1 degré (soit environ 10 km). La production est actuellement basée sur un ensemble de neuf systèmes d'assimilation de données sur la qualité de l'air à travers l'Europe. Un ensemble médian est ensuite calculé à partir des résultats individuels. Enfin l'écart entre les neuf modèles est alors utilisé pour fournir une estimation de l'incertitude des données fournies.

La ré-analyse des résultats finaux combine les données du modèle avec les observations fournies par l'Agence européenne pour l'environnement (AEE) en un ensemble de données complet et cohérent. D'autres sources d'observations peuvent être également utilisées pour compléter l'assimilation des données *in situ*, comme les données satellitaires.

Les données produites prennent la forme d'un *raster* d'environ 10 km, disponibles en fichiers mensuels contenant des analyses horaires pour les années 2018 à 2021, soit une profondeur de 4 ans. La fréquence de mise à jour est biannuelle, et les fichiers extraits sont au format .NetCDF, pour une liste très exhaustive de polluants, sur la totalité de la surface européenne. Un portail en ligne sur le site de CAMS est mis à disposition pour une extraction des données simplifiée.

Aussi, bien que ces données soient d'apparence appropriées pour l'étude, le volume de données proposé est particulièrement conséquent en raison de la couverture géographique. Ce dernier aspect représente une limite technique conséquente au vu de la puissance de calcul disponible pour la réalisation de ce mémoire. De plus, ce jeu de données ne possède qu'une profondeur historique de 4 ans, point particulièrement discriminant lorsqu'il s'agit de pouvoir construire un modèle robuste qui apprend sur des données passées.

En ce sens, ce jeu de données a été exclu de l'étude bien que ce dernier soit particulièrement pertinent pour l'étude historique de l'impact de la pollution de l'air sur les années couvertes.

CAMS European air quality reanalysis
PM_{2.5} ensemble median, 1 Apr 2018

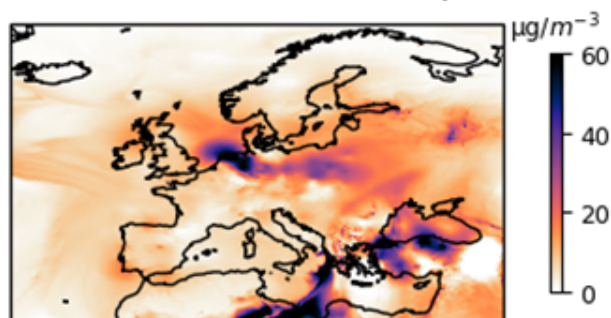


FIGURE 2.1 – Données ré-analysées de concentration moyennes de $PM_{2.5}$ au 1 avril 2018 issues du projet CAMS

- Reconstruction historique de la pollution atmosphérique de fond en France pour la période 2000-2015

Ensuite on retrouve les données « Reconstruction historique de la pollution atmosphérique de fond en France pour la période 2000-2015 ». Ces données couvrent 16 années de concentrations de pollution atmosphérique et d'indicateurs de qualité de l'air en France et sont produites par l'Institut national de l'environnement industriel et des risques (INERIS).

Ces dernières sont produites par le biais de méthodes de *krigeage* qui combinent des mesures de fond de la qualité de l'air ainsi qu'une modélisation réalisée avec le modèle de transport chimique CHIMERE.

Les concentrations à disposition sont horaires et les polluants estimés sont le NO_2 , O_3 , PM_{10} et $PM_{2.5}$ avec une résolution spatiale d'environ 4 kilomètres. Les indicateurs réglementaires nationaux tels que la moyenne annuelle de la concentration dans l'air, le SOMO35 pour l'Ozone sont également calculés à partir de ces données.

Les indicateurs réglementaires sont fournis aux niveaux régional, départemental et municipal sous forme de fichiers .csv pour les différentes années.

Ces données sont particulièrement intéressantes en présentant une profondeur historique de 16 ans. De plus, les indicateurs réglementaires à utiliser pour la quantification de l'incidence de la pollution sont fournis, simplifiant le travail de retraitement pour certains polluants comme l'Ozone. Cependant, les données s'arrêtent en 2015, ce qui pose un problème de cohérence conséquent au manque de continuité des données avec 10 années d'interruption. De plus, le volume des données étant à nouveau particulièrement conséquent, ce dernier a également été exclu.

- GEOD'AIR

La surveillance de la qualité de l'air ambiant en France est confiée à 18 Associations Agréées de Surveillance de la Qualité de l'Air indépendantes (AASQA). Le Laboratoire Central de Surveillance de la Qualité de l'Air (LCSQA) est ainsi l'entité supérieure aux AASQA mandatées par le Ministère en charge de l'Environnement pour développer et gérer la base

nationale de données de qualité de l'air appelée GEOD'AIR.

Le LCSQA a également la charge de l'uniformisation réglementaire des données de la qualité de l'air selon les termes des Directives et Décisions européennes. En effet, ces données sont transmises à l'AEE selon un formalisme bien défini, documenté dans des guides de référence mis à disposition des AASQA. Enfin, l'entité nationale est chargée de fournir des bilans annuels de la qualité de l'air en fournissant des données et cartographies, ces derniers étant produits par le Service de la Donnée et des Etudes Statistiques (SDES).

Ainsi GEOD'AIR agrège l'intégralité des données récoltées par les stations des AAQSA à partir d'analyseurs automatiques. Les données continues sont ensuite agrégées à une résolution horaire selon des critères et des règles strictes (fixées spécifiquement pour le rapportage réglementaire à l'entité européenne). Les données sont disponibles via un portail d'accès qui met aussi à disposition la consultation de cartes interactives, les données relatives aux stations de mesures, un portail d'export avancé des données ainsi qu'une API pour la facilitation de l'extraction des données.

Les données sont disponibles à différentes échelles temporelles (horaire, journalière, mensuelle et annuelle) à partir de points géographiques représentant la position de chacune des 550 stations du réseau de mesures pour les concentrations des polluants atmosphériques suivants :

1. Ozone (O_3),
2. Dioxyde d'azote (NO_2),
3. Dioxyde de soufre (SO_2),
4. Particules de diamètre inférieur à 10 μm (PM_{10}),
5. Particules de diamètre inférieur à 2,5 μm ($PM_{2.5}$),
6. Monoxyde de carbone (CO).

Ces dernières étant des données validées et conformes aux exigences européennes, elles servent à alimenter les modèles communiquant l'indice de l'air journalier publié et mis à la disposition de tous.

C'est finalement ce dernier jeu de données qui a été retenu, fournissant des données uniquement *in situ* avec pour ambition de construire un modèle d'interpolation spatiale indépendant capable de répliquer au mieux les productions et résultats présentés auparavant.

Avec son portail facile d'accès et réactif, il permet aisément d'extraire les données de relevé de pollution à temporalité annuelle, échelle retenue pour l'étude afin de limiter les contraintes calculatoires et la durée des exécutions de code. Enfin les données disponibles sont à l'échelle du point géographique (station), sur une profondeur de 10 ans, de 2013 à 2022.

De plus, ces dernières données donnent la liberté de construction de la grille carroyée pour l'interpolation, sa résolution, sa projection et le choix de la méthodologie. Cet aspect se trouvera particulièrement utile lorsque les données de population seront présentées.

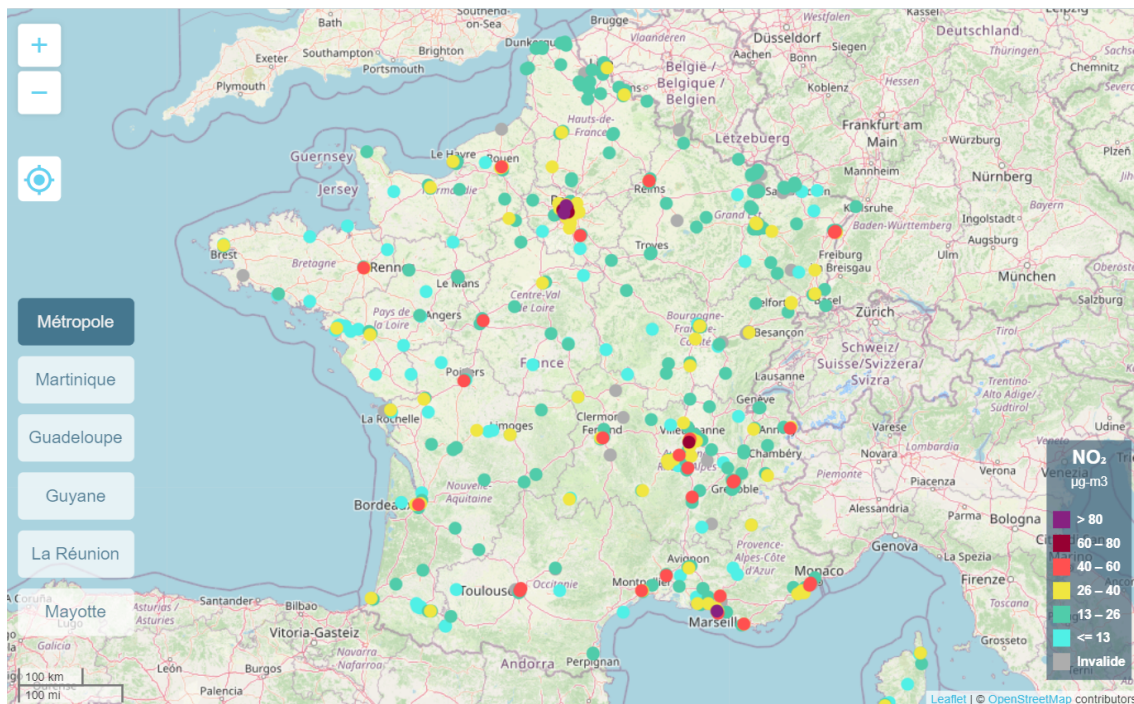


FIGURE 2.2 – Portail GEOD’AIR : Moyenne annuelle de Dioxyde d’Azote ($\mu\text{g}/\text{m}^3$) en 2013 aux stations de mesures

Ont donc été extraites la totalité des stations en France métropolitaine (Corse exclue) présentant une mesure de concentration annuelle pour chacun des polluants que sont NO_2 , O_3 , $\text{PM}_{2.5}$ et PM_{10} . Les mesures annuelles sont alors extraites d’ensembles de données de référence validés.

Pour l’obtention des valeurs de SOMO_{35} , les valeurs journalières historiques du maximum journalier moyen sur 8 heures consécutives par station ont été extraites. Par la suite les données annuelles ont été obtenues en réalisant les calculs présentés en 1.2.2.

Le nombre de sites de relevés de la pollution de fond extraits pour chaque type de polluants atmosphérique et pour chaque année est résumé dans la table suivante.

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
NO_2	460	452	461	456	446	431	417	410	402	396	4 331
O_3	385	363	361	355	347	335	327	319	315	312	3 419
PM_{10}	393	389	387	401	396	391	382	378	377	374	3 868
$\text{PM}_{2.5}$	136	138	134	155	154	157	169	194	209	231	1 677
SOMO_{35}	363	340	337	334	323	310	302	293	288	285	3 175

TABLE 2.1 – Nombre de sites de surveillance de la pollution de fond en France par polluant atmosphérique pour les années 2013 à 2022.

Ci-dessous la liste des variables extraites :

Variable	Type	Description
Date de début	String	Date de début de la mesure
Date de fin	String	Date de fin de la mesure
Organisme	String	Nom de l'organisme responsable des mesures
Code zas	String	Code ZAS (Zone d'Action Simplifiée)
Zas	String	Zone d'Action Simplifiée
Code site	String	Identifiant unique du site
Nom site	String	Nom civil du site
Type d'implantation	String	Type d'implantation du site ("Périurbain", "Rural", ...)
Polluant	String	Nom du polluant mesuré
Type d'influence	String	Type d'influence du polluant (uniquement "Fond")
Réglementaire	String	Indicateur si la mesure est réglementaire ("Oui"/"Non")
Type d'évaluation	String	Type d'évaluation de la mesure ("mesures fixes"/"mesures indicatives")
Type de valeur	String	Type de valeur de la mesure (Moy. annuelle, Max Jour. moy. 8h)
Valeur	Float	Valeur réanalysée de la mesure
Valeur brute	Float	Valeur brute de la mesure
Unité de mesure	String	Unité de mesure de la valeur ($\mu\text{g}/\text{m}^3$)
Taux de saisie	Float	Taux de saisie de la mesure
Couverture temporelle	Float	Couverture temporelle de la mesure
Couverture de données	Float	Couverture de données de la mesure
Code qualité	String	Code de qualité de la mesure (-1/1)
Validité	Integer	Indicateur de validité de la mesure
Latitude	Float	Latitude du site (WGS84)
Longitude	Float	Longitude du site (WGS84)

TABLE 2.2 – Description des variables

— Autres sources

D'autres sources de données ont été identifiées comme AIR PARIF, l'observatoire de la qualité de l'air en Île-de-France, qui fournit cartes et données exhaustives en matière de qualité de l'air mais ne répondant pas à l'ambition de la production de résultats à l'échelle nationale en France métropolitaine.

Ou encore PREV'AIR, plate-forme nationale de prévision de la qualité de l'air, qui est l'une des composantes du dispositif français de surveillance et de gestion de la qualité de l'air, agissant en complément des informations fournies par les réseaux de mesure et d'observation « physiques » gérés par les AASQA.

Cependant les données ne sont pas directement disponibles sur le site. Il est nécessaire de compléter un formulaire de demande de mise à disposition de données numériques de simulation de la qualité de l'air. Il est également expressément mentionné que la fourniture des données numériques est uniquement destinée aux organismes acteurs de la surveillance de la qualité de l'air et/ou utilisateurs de données de qualité de l'air. Aussi les demandes

émanant de particuliers ne sont pas satisfaites.

Il serait particulièrement intéressant de croiser les sources de données produisant des résultats sous forme de *raster* en qualité de données externes et de les utiliser afin de valider la démarche et le modèle d'interpolation spatiale réalisé. Passons désormais à la présentation des données de population utilisées.

Les données de population : Données au carreau de 1 km² en France de l'INSEE

La première base de données utilisée relative à la démographie française se trouve être « Revenus, pauvreté et niveau de vie en 2017 - Données carroyées ». Cette base de données issue du dispositif fichier localisé social et fiscal (Filosofi), fournit des informations socio-économiques pour plus de 30 millions de foyers en France. Ces informations sont par la suite diffusées sur des *rasters* de différentes résolutions et dont la plus petite se trouve être par carreau de 200 m sur 200 m partout en France pour l'année civile 2017.

Ces informations sont pour le compte des institutions publiques très utiles pour, par exemple, connaître les besoins en infrastructures de transport, planifier au mieux les services publics comme la collecte de déchets ou encore dans le processus de quantification des risques.

Ces données carroyées constituent des statistiques locales au niveau géographique le plus fin et permettent aisément de se substituer à l'utilisation d'un portefeuille dans le cadre de l'étude de l'impact de la pollution de l'air en France.

Une nouvelle fois sous contrainte de volume de données et de puissance calculatoire, la base carroyée de 1 km de résolution a été retenue. La liste des 31 variables pour chacun des 374 797 carreaux présents en France métropolitaine est disponible à l'annexe A.3.

Le *raster* produit pour la métropole fait l'objet d'une projection Lambert 93 (EPSG 2154)¹. Toutefois, la grille de carreaux a été produite à partir des données projetées en LAEA (EPSG 3035)², projection utilisée au niveau européen. Les contours des carreaux ainsi obtenus ont ensuite été re-projetés en Lambert 93, projection de référence pour la France métropolitaine.

1. EPSG : European Petroleum Survey Group, un système de référencement pour les coordonnées géographiques.

2. LAEA : Lambert Azimuthal Equal Area, une projection cartographique utilisée à l'échelle européenne.

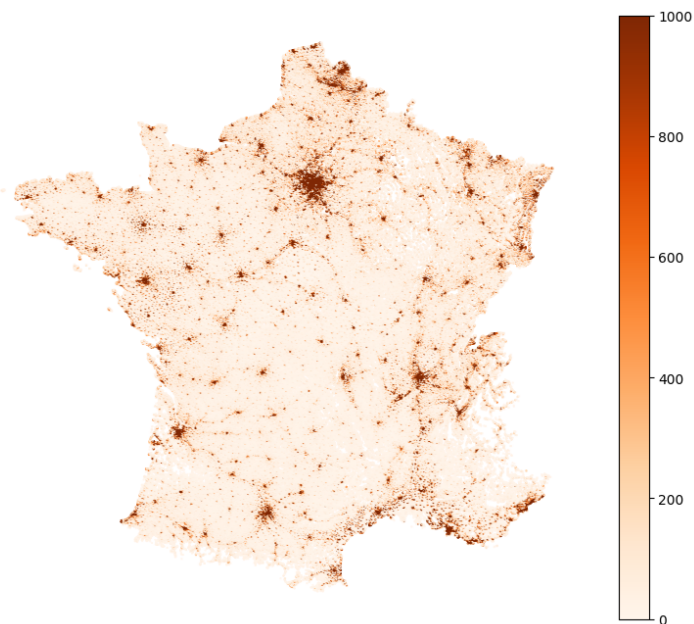


FIGURE 2.3 – Représentation de la densité de population par pixel d'1 km² selon la base « Revenus, pauvreté et niveau de vie en 2017 - Données carroyées », INSEE

Il est possible de consulter ces données via une carte interactive par le biais de deux plateformes distinctes :

- L'espace statistiques nationales de l'INSEE (grille de 1 km de résolution)
- Le géoportail de l'IGN³ (grilles de 200 m et 1 km)

Les données de population : Le scénario central de l'INSEE

Alors que la base de données précédemment présentée constituant la grille de carreaux utilisée dans la modélisation spatiale se trouve arrêtée à l'année civile 2017, l'étude a pour objectif d'étudier l'incidence prospective de la pollution de l'air. Il est alors nécessaire de faire évoluer le module d'exposition au sein de l'étude, en faisant varier annuellement la population française en effectif et en tranche d'âge.

Pour se faire, l'étude s'est appuyée sur le « scénario central de l'INSEE » disponible en libre accès, qui fournit différentes tables détaillée ci-après. On y retrouve par exemple l'onglet « *Structure_Age_DEP* » qui indique la structure par âge quinquennale de la population et l'âge moyen en 2018 et 2070 ainsi que l'onglet « *Population_DEP* » qui indique l'effectif démographique annuel estimé en milliers de personnes par département et année de projection sur la période 2018 à 2070.

Les scénarios de l'INSEE sont construits selon des hypothèses relatives à trois composantes : fécondité, mortalité et migration. Le scénario central constitue celui qui retient les hypothèses centrales pour ces trois composantes et est le scénario le plus souvent retenu pour les modélisations prospectives de la population française.

3. Institut National de l'Information Géographique et Forestière



FIGURE 2.4 – Pyramide des âges : Population en 2050, Scénario Central, INSEE

Enfin, afin de faire concorder les tranches d'âges des deux bases de l'INSEE citées précédemment, quatre tranches d'âges ont été retenues en agrégeant celles qui y sont contenues :

- [0, 10) pour les personnes âgées de 0 à 9 ans
- [10, 25) pour les personnes âgées de 10 à 24 ans
- [25, 65) pour les personnes âgées de 25 à 64 ans
- [65, +100) pour les personnes âgées de 65 ans et plus

2.1.3 Variables exogènes

Émissions historiques

Compte tenu de la synthèse des polluants réalisées en partie 1.2.5 permettant la mise en évidence du lien causal entre émissions et immissions, il a été retenu d'utiliser les données d'émissions historiques et prospectives afin d'expliquer l'évolution des concentrations moyenne des polluants aux stations de mesure. En particulier en raison des fortes relations de causalités et de corrélations qui ont été établies précédemment.

Les données historiques ont été recueillies à partir de la « *National Emission reductions Commitments (NEC) Directive emission inventory data* » qui est la base de données sur les émissions de polluants atmosphériques (ammoniac (NH_3), composés organiques volatils non méthaniques (COVNM), oxydes d'azote (NOx), particules 2.5 ($PM_{2.5}$) et dioxyde de soufre (SO_2)) déclarées chaque année par les États membres à la Commission européenne en vertu de la directive 2016/2284 du Parlement européen et du Conseil relatif à la réduction des émissions nationales de certains polluants atmosphériques.

Ci-après, la table des variables de la base de données « *National Emission reductions Commitments (NEC) Directive emission inventory data* »

Variable	Type	Description
Country_Code	String	Code international du pays
Country	String	Nom du pays
Pollutant_name	String	Nom court du polluant (PM ₁₀ , NMVOC, NO _x , PM _{2.5})
Format_name	String	Classification sectorielle de la NFR19
sector_code	String	Code du secteur émetteur
parent_sector_code	String	Code du secteur parent
sector_name	String	Nom du secteur
Year	Integer	Année de la mesure (1990-2019)
Emissions	Float	Émissions
Unit	String	Unité de mesure des émissions (Gg=1 000 tonnes)
Notations	String	Code de validité de la mesure

TABLE 2.3 – Description des variables du jeu de données « *National Emission reductions Commitments (NEC) Directive emission inventory data* »

Ces données ont par la suite été retraitées et agrégées de telle sorte que les émissions historiques annuelles relatives à la France soient répertoriées par polluant sur la période 1990-2019. N'ont ainsi été retenus que les totaux nationaux additionnant l'ensemble des émissions annuelles recensées, tout secteur confondu. Les mesures n'apparaissant pas pour les années civiles précédant l'an 2000 pour certains pays et apparaissant imprécises, la profondeur historique qui est retenue est la période 2000-2019.

Cette même base de données d'émissions a également permis de déduire la part attribuable de la France dans les émissions européennes, quantité réutilisée lors de la récupération de données issues de scénarios prospectifs généralement communiquant les émissions à l'échelle de l'Europe en *Chapitre 3*.

En effet soient :

- Les émissions françaises pour un polluant donné en année y représentées par $E_{\text{France}}(y)$.
- Le total des émissions européennes pour un polluant donné représentées par $E_{\text{Europe}}(y)$.

Aussi, la part des émissions françaises sur le total des émissions européennes pour l'année y est donnée par :

$$P_{\text{France}}(y) = \frac{E_{\text{France}}(y)}{E_{\text{Europe}}(y)}$$

Pour les années allant de 2000 à 2019 :

$$\begin{aligned}
 P_{\text{France}}(2000) &= \frac{E_{\text{France}}(2000)}{E_{\text{Europe}}(2000)} \\
 P_{\text{France}}(2001) &= \frac{E_{\text{France}}(2001)}{E_{\text{Europe}}(2001)} \\
 &\vdots \\
 P_{\text{France}}(2019) &= \frac{E_{\text{France}}(2019)}{E_{\text{Europe}}(2019)}
 \end{aligned}$$

On obtient alors le graphique suivant pour les émissions de NOx.

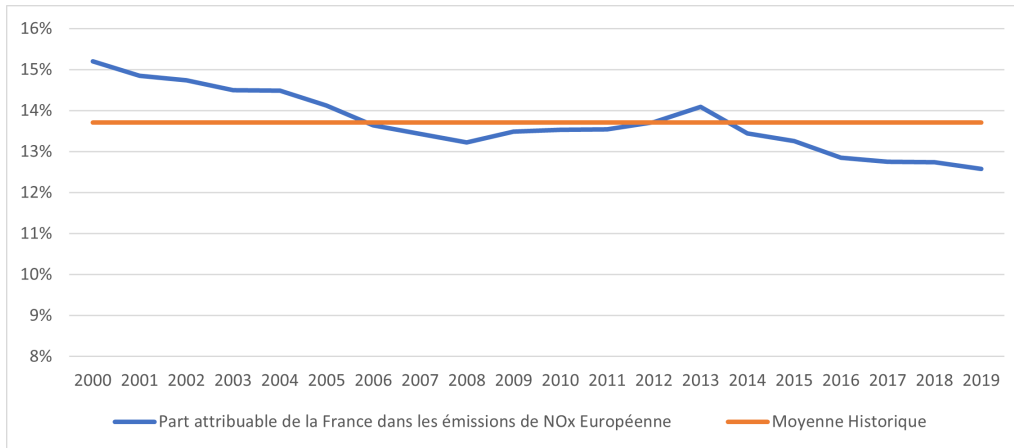


FIGURE 2.5 – Évolution temporelle de la part attribuable de la France dans les émissions européennes (EU28) de NO_x

Après vérification de la quasi-stationnarité de $P_{\text{France}}(y)$ pour chacun des polluants étudiés, l'hypothèse d'utiliser la part attribuable moyenne de la France sur la totalité de l'historique a été retenue.

Enfin, la part attribuable moyenne de la France par rapport aux émissions européennes sur la totalité de l'historique définie par :

$$P_{\text{France_moyen}} = \frac{1}{N} \sum_{y=2000}^{2019} P_{\text{France}}(y)$$

On a alors par exemple pour le cas des NOx ;

$$P_{\text{France_moyen_NOx}} = \frac{1}{N} \sum_{y=2000}^{2019} P_{\text{France_NOx}}(y) = 13,71 \%$$

Quant aux données prospectives, leurs sources et utilisations seront présentées en *Chapitre 3*. Passons désormais aux relevés de températures annuelles historiques, indice d'évolution de la concentration d' O_3 .

Relevés de températures annuelles historiques

Dans une démarche similaire aux données d'émissions, il a été mis en avant le lien moins prononcé mais tout de même significatif entre la concentration d'ozone troposphérique et l'évolution de la température moyenne annuelle en France (1.2.6). De la relation de causalité démontrée et de la corrélation établie entre ces deux mesures, il a été retenu d'utiliser les données de températures historiques et prospectives pour expliquer l'évolution des concentrations moyennes d'ozone aux stations de mesures.

Les données historiques ont été récoltées via « *Temperature in France, 2000-2020* », *International Energy Agency* (IEA). Ces données présentent l'évolution de la température annuelle agrégée pour la France métropolitaine de l'année 2000 à l'année 2019.



FIGURE 2.6 – Historique de la température moyenne annuelle en °C en France métropolitaine

Similairement aux données d'émissions; les données prospectives de température annuelle selon différents scénarios prospectifs seront présentées en *Chapitre 3*.

Il s'agit désormais de présenter la méthodologie développée pour obtenir les concentrations de pollutions en tout pixel de la base carroyée de l'INSEE à partir de données *in situ* de la base GEOD'AIR. Pour ce faire, le concept d'interpolation spatiale est introduit et défini.

2.2 Interpolation spatiale

Dans cette section, un focus sur le concept d'interpolation spatiale est réalisé. Cette méthodologie largement appliquée dans des domaines comme la météorologie ou la géologie, peut trouver des applications en assurance et plus particulièrement encore en raison des conséquences anticipées du changement climatique. Il s'agit dans un premier temps de définir un cadre théorique de cette méthode.

2.2.1 Définition

L'interpolation est la technique mathématique permettant de prédire des valeurs à partir d'un nombre limité de points d'échantillonnage. Cette technique étant particulièrement applicable et utile dans la prédiction de données spatiales; on introduit alors le terme d'interpolation spatiale. Cette dernière permettra dans notre cas de déduire des valeurs inconnues des cellules de la grille de la base de données carroyées de l'INSEE à partir de

points géographiques de valeurs connues correspondant dans notre étude aux relevés de pollution du réseau de stations de mesures.

Il est alors possible d'obtenir des estimations de données continues interpolées comme peuvent l'être l'altitude, les précipitations, les concentrations chimiques ou encore la pollution de l'air.

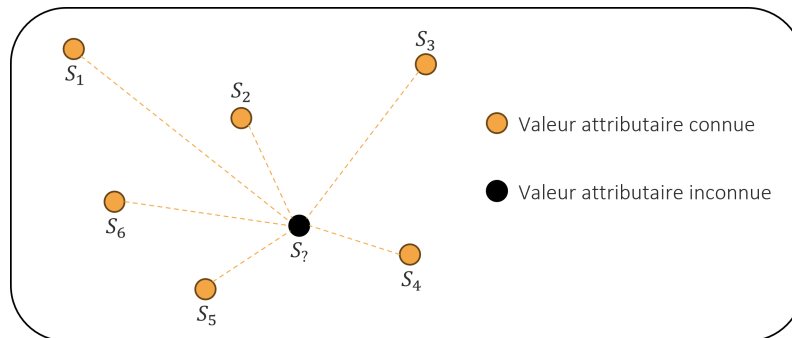


FIGURE 2.7 – Schéma du principe de l'interpolation spatiale

Ensuite, on distingue l'interpolation spatiale exacte de l'interpolation inexacte. En effet, la première fait référence aux modèles qui conserveront les valeurs des points de mesures intactes. L'interpolation inexacte réestime les valeurs des points géographiques connus. Elle permet d'éviter des divergences importantes et marquées dans la surface estimée en sortie du modèle.

Il s'agit par la suite de présenter les différents modèles d'interpolation spatiale ; qu'ils soient déterministes ou stochastiques et leurs notations associées. À l'étude ne seront abordés que les modèles univariés ; c'est-à-dire les modèles qui utilisent une unique variable pour prédire la valeur de la variable à estimer.

2.2.2 Représentations

À partir des données échantillonnées, on rappelle qu'à partir du recours à l'interpolation spatiale, il est possible d'obtenir une valeur approchée en tout point de la représentation régionalisée étudiée.

Les données échantillonnées étant les valeurs des stations de mesures du réseau national de relevés de concentrations des polluants dans l'air, la représentation retenue est par conséquent plane et constitue la base carroyée de l'INSEE. Plus expressément, les points d'observations constituent les points géographiques des stations de mesures et chacun des pixels de 1 km^2 du *raster* projeté en Lambert 93 (EPSG 2154) constitue une cellule du champ à estimer. Les axes du plan spatial étant respectivement la latitude en abscisse et la longitude en ordonnée.

À noter qu'il est tout à fait possible d'ajouter une troisième dimension comme l'altitude selon l'axe Z.

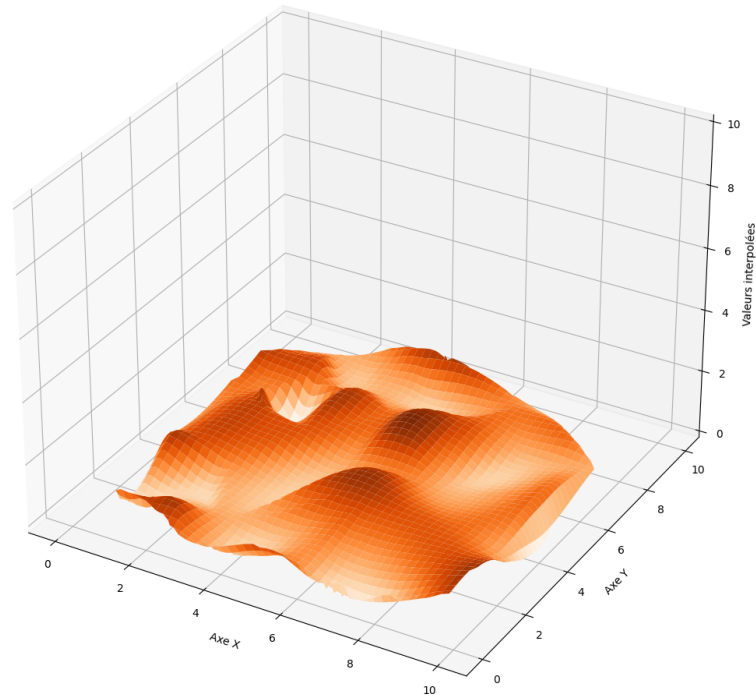


FIGURE 2.8 – Exemple d’interpolation spatiale par *spline* cubique réalisée dans un plan en 3 dimensions : vue en perspective.

2.2.3 Notations

On définit une variable régionalisée comme étant une fonction numérique prenant ses valeurs dans une région limitée, appelée champ. Par la suite, on utilise les notations suivantes :

1. z , la variable régionalisée.
2. \mathcal{D} , le champ de la régionalisation, c’est-à-dire le domaine spatial dans lequel la variable régionalisée est définie. En général, $\mathcal{D} \subset R, R^2$ ou R^3 . L’étude étant réalisée dans R^2
3. $\vec{s} \in \mathcal{D}$, le vecteur de coordonnées (x, y, z) qui indique la position d’un site dans le champ \mathcal{D} (Dans R^2 on a systématiquement $z = 0$).
4. $z(\vec{s})$, la valeur prise par la variable régionalisée z au site $\vec{s} \in \mathcal{D}$.
5. $z(V)$, la valeur moyenne de z sur le domaine $V \subset \mathcal{D}$.
6. n , le nombre de sites où la variable a été mesurée. $z(\vec{s}_1), \dots, z(\vec{s}_n)$, les valeurs prises par z aux sites d’observation $\vec{s}_1 \dots \vec{s}_n$.
7. $\hat{z}(\vec{s}_0)$, une estimation de $z(\vec{s}_0)$ avec $\vec{s}_0 \in \mathcal{D}$.

2.2.4 Présentations des modèles : Les modèles déterministes

Les modèles dits déterministes sont des modèles qui créent des surfaces interpolées à partir de points mesurés en fonction de différents paramètres comme peuvent l'être l'étendue de similitude ou encore le degré de lissage. Ces derniers se divisent en deux groupes : les modèles globaux et locaux.

Une technique d'interpolation spatiale globale permet de calculer les prévisions en capitalisant sur l'information fournie par l'ensemble des données à disposition. Une interpolation locale quant à elle obtiendra une prévision de la valeur à estimer à partir des points mesurés dans les voisinages. Ces points de voisinage sont des zones spatiales plus petites elles-mêmes localisées dans une zone d'étude plus grande (le champ \mathcal{D}).

Enfin on parle d'effet écran lorsqu'une observation de l'échantillonnage est fonction d'une seconde observation lors de l'exécution de l'interpolation. Si l'influence de l'observation dite écrantée est nulle on parle d'écran total, peu élevée on parle d'un écran partiel et modérée pour le cas d'un écran faible.

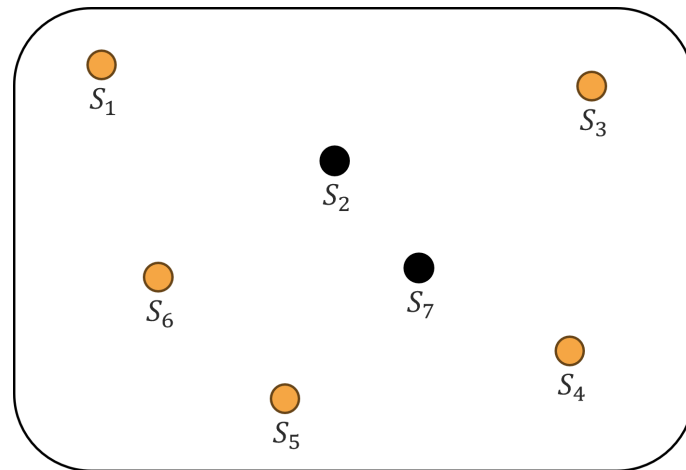


FIGURE 2.9 – Principe de l'effet écran : le site s_2 et le site s_7 peuvent potentiellement se masquer l'un l'autre en raison de leur proximité spatiale

Plus proche voisin (*Nearest Neighbor*)

La méthode d'interpolation spatiale par plus proche voisin est une méthode d'interpolation qui détermine la valeur de la cellule recherchée comme étant la valeur du point d'échantillonnage connu le plus « proche ».

Pour tous les points d'observation de \mathcal{D} , on définit un polygone d'influence tel que chaque point de ce polygone est plus proche du point d'observation que de tout autre site de l'échantillon de valeurs connues :

$$\forall \vec{s} \in \mathcal{P}_i, \forall \vec{s}_j \in \mathcal{D} \setminus \mathcal{P}_i, \|\vec{s}_i - \vec{s}\| \leq \|\vec{s}_j - \vec{s}\|$$

\mathcal{D} est alors partitionné en une famille de polygones convexes. Ces derniers sont nommés « polygone de Thiessen » (ou polygones de Voronoi, cellules de Dirichlet). [29]

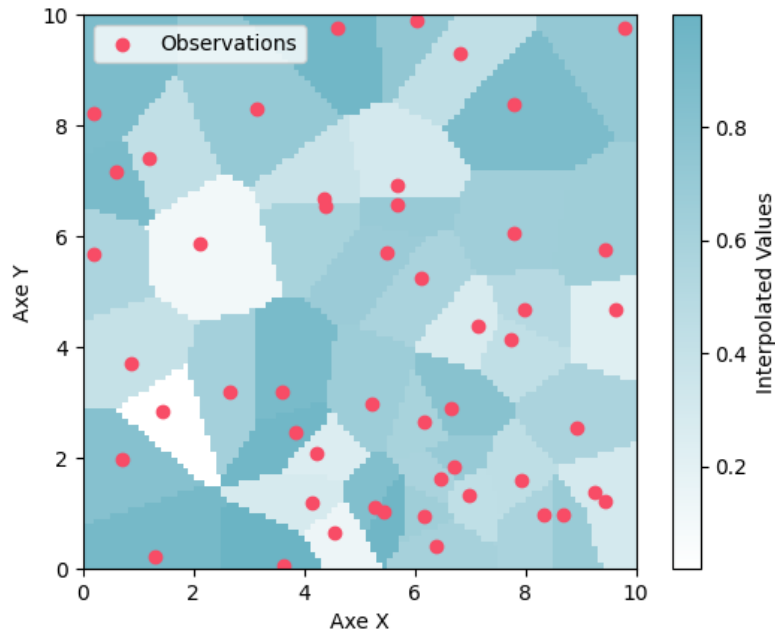


FIGURE 2.10 – Diagramme de Voronoi - Interpolation par plus proche voisin : application sur un champ $\mathcal{D} \subset \mathbb{R}^2$ de 50 observations aléatoires en rose (*germes*) - représentation sur une grille régulière - vue de dessus. On remarque l'effet de crênelage de la méthodologie

Le diagramme de Voronoi ici obtenu en *Figure 2.10* est un découpage du plan spatial en polygones obtenus à partir d'un ensemble discret de points aussi appelés « germes ». Ces germes constituent ici l'échantillon de points aléatoirement générés. Chaque polygone obtenu contient un seul germe et tout point contenu dans le polygone se voit assigner la valeur du germe. Le polygone correspond donc à la surface d'influence spatiale de chaque germe.

Bien que cette méthode soit théoriquement simple et dotée d'une complexité calculatoire assez faible, ses résultats restent discutables et peu recommandés compte tenu du fort effet de crênelage qu'elle génère et de la forte dépendance à la configuration géométrique des points de mesure.

Interpolation linéaire

L'interpolation linéaire est une méthode couramment utilisée pour estimer les valeurs d'une variable régionalisée z à des emplacements non observés à partir des mesures effectuées à des sites d'observation $\vec{s}_1, \vec{s}_2, \dots, \vec{s}_n$.

L'estimation $\hat{z}(\vec{s}_0)$ est obtenue en combinant linéairement les valeurs observées $z(\vec{s}_1), z(\vec{s}_2), \dots, z(\vec{s}_n)$ en fonction de leurs distances relatives par rapport au point \vec{s}_0 . Plus précisément, l'estimation est calculée en utilisant la relation :

$$\hat{z}(\vec{s}_0) = \sum_{i=1}^n w_i(\vec{s}_0) z(\vec{s}_i)$$

où les coefficients $w_i(\vec{s}_0)$ représentent les poids attribués à chaque valeur observée. Ces poids sont généralement déterminés en fonction de la distance entre le point d'estimation \vec{s}_0 et les points d'observation \vec{s}_i . Plus la distance est grande, moins le poids est important.

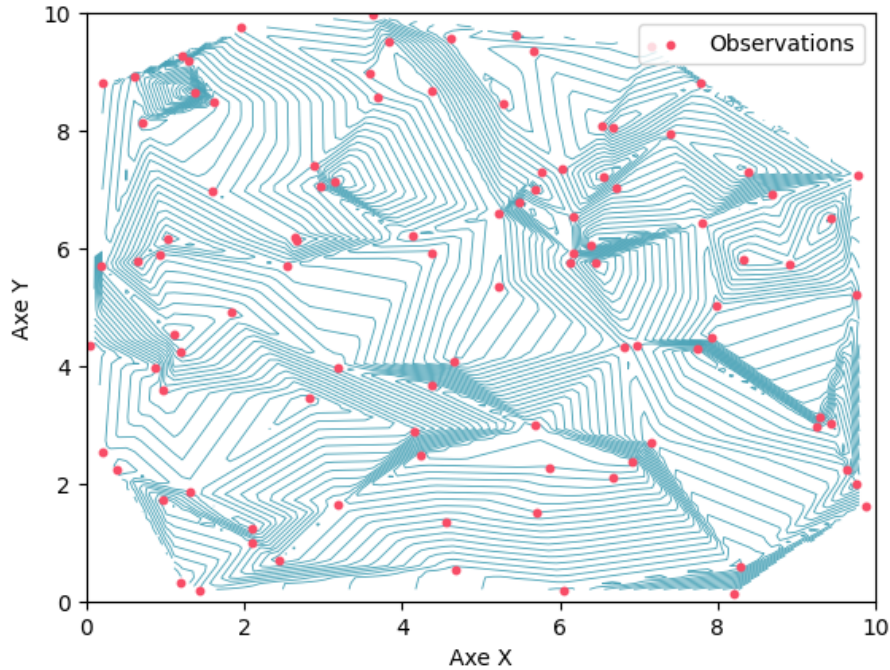


FIGURE 2.11 – Interpolation polynomiale linéaire : application sur un champ $\mathcal{D} \subset \mathbb{R}^2$ de 100 observations réparties aléatoirement - représentation par courbes de niveau - vue de dessus. On remarque l'effet pyramidal de la méthodologie

En résumé, l'interpolation linéaire exploite la relation entre les valeurs observées et leurs distances relatives pour fournir des estimations de la variable régionalisée à des emplacements non observés.

Inverse Distance Weighted (IDW)

La technique d'interpolation par pondération inverse de la distance permet de prédire la valeur d'un point dans l'espace d'étude n'appartenant pas à l'échantillon initial en se servant de la distance spatiale entre ce point et d'autres points de valeur connue.

Aussi la valeur à prédire $z(\vec{s})$ au site $\vec{s} \in \mathcal{D}$ s'obtient en réalisant le calcul suivant :

$$\hat{z}(\vec{s}) = \frac{\sum_{i=0}^n \frac{z(\vec{s}_i)}{\|\vec{s}_i - \vec{s}\|}}{\sum_{i=0}^n \frac{1}{\|\vec{s}_i - \vec{s}\|}}$$

Aussi plus une valeur d'observation est spatialement proche du point à prédire plus elle reçoit un poids plus important dans la prédiction. Au contraire les observations éloignées exerceront une influence relativement moins importante sur la prédiction. On parle de méthode barycentrique.

Il est possible d'introduire un paramètre $p \in \mathbb{R}^+ *$ tel que :

$$\hat{z}(\vec{s}) = \frac{\sum_{i=0}^n \frac{z(\vec{s}_i)}{\|\vec{s}_i - \vec{s}\|^p}}{\sum_{i=0}^n \frac{1}{\|\vec{s}_i - \vec{s}\|^p}}$$

Le paramètre p correspond ici à un paramètre de puissance qui ajuste le poids attribué à chaque élément de l'échantillon. Lorsque p augmente, le poids attribué décroît avec la distance. Ce facteur peut être compris comme un paramètre de lissage de l'interpolation, de la densité et la distribution des échantillons interpolés, et de la distance maximum au-delà de laquelle un échantillon individuel peut influencer les points environnants.

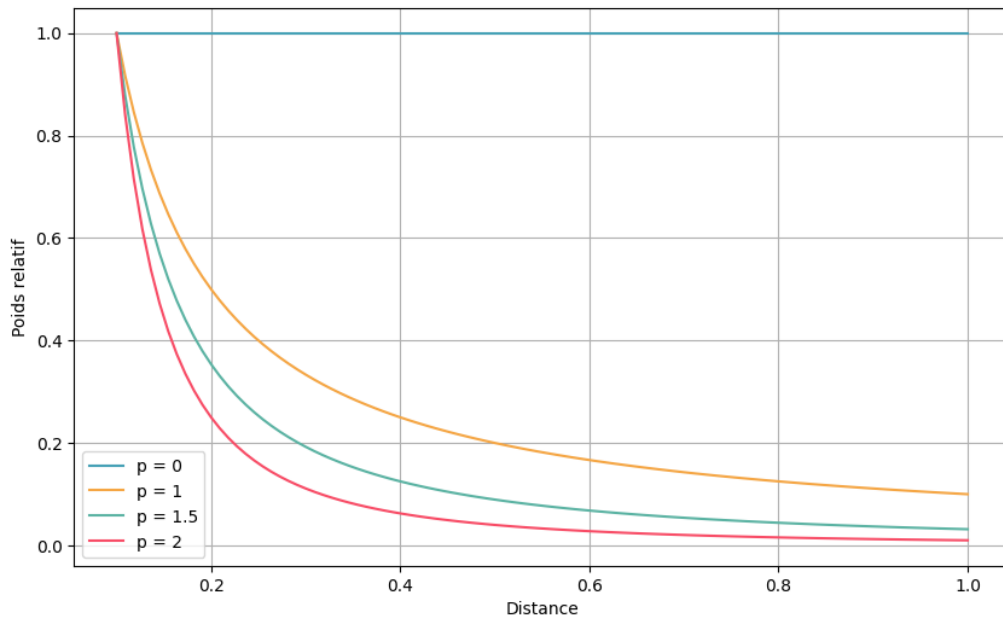


FIGURE 2.12 – Réduction du poids relatif en fonction de la distance selon plusieurs valeurs de p

Pour $0 < p < 1$, en $u(x)$, on observe des sommets lissés autour du point d'interpolation x_k , alors que pour $p > 1$, le pic devient plus pointu. Il est par conséquent nécessaire de bien paramétrer le paramètre p afin que la surface interpolée soit pertinente, fiable et représentative du phénomène continu estimé.

Cette méthodologie est, comme les précédentes, relativement facile à implémenter et présente l'avantage de ne pas être coûteuse en matière de calculs.

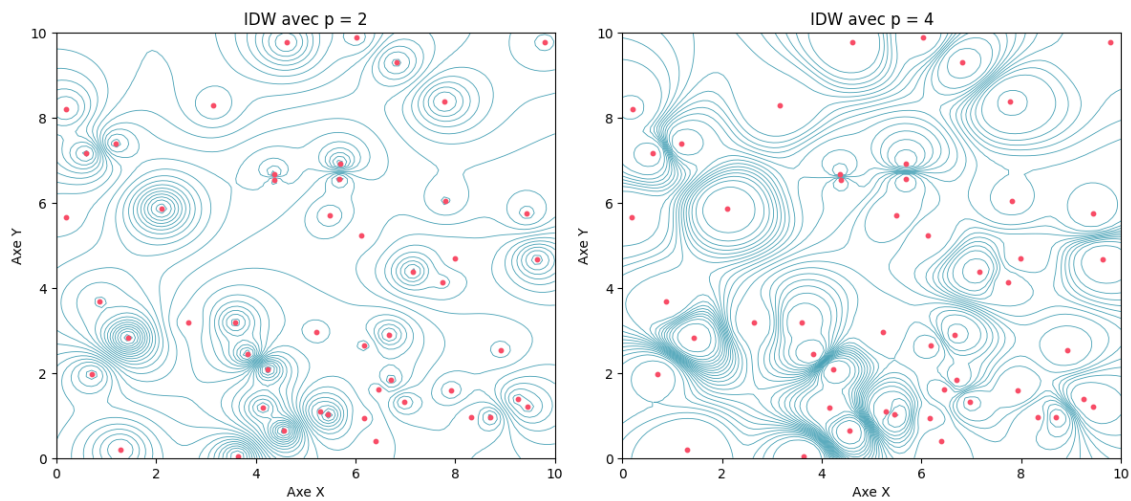


FIGURE 2.13 – Interpolation barycentrique par inverse des distances : application sur un champ $\mathcal{D} \subset \mathbb{R}^2$ de 100 observations réparties aléatoirement pour différentes valeurs de p - représentation par courbes de niveau - vue de dessus. On observe que plus grande est la valeur de p plus l'influence aux valeurs les plus proches du point interpolé est grandissante.

Spline

Dans le domaine de l'analyse numérique, une *spline* est une fonction définie par morceaux par des polynômes (*Wikipédia [30]*). Pour ce qui est de leur usage en interpolation spatiale, elles sont souvent préférées à l'interpolation polynomiale, car elles permettent d'obtenir des résultats similaires en ayant recours à des polynômes de degrés inférieurs.

Les *splines* sont : « une famille de fonctions qui minimisent l'énergie de flexion sous certaines contraintes d'ajustement » [29]. On présentera dans cet écrit deux catégories de fonctions de *spline* :

- Les *splines* d'interpolation : méthodologie d'interpolation exacte
- Les *splines* de lissage : méthodologie d'interpolation inexacte

Les premières consistent en l'obtention d'une fonction la plus lisse possible capable de restituer les valeurs connues de l'échantillon. La fonction $z(\vec{s})$ illustre la surface continue que l'on contraint à passer à chaque point connu. Mathématiquement, dans \mathbb{R}^2 cela revient au problème de minimisation de l'énergie de flexion suivant :

$$\iint \frac{\partial^2 \hat{z}}{\partial x^2} + 2 \frac{\partial^2 \hat{z}}{\partial x \partial y} + \frac{\partial^2 \hat{z}}{\partial y^2} dx dy$$

Avec comme contrainte $\hat{z}(\vec{s}_i) = z(\vec{s}_i)$. On peut montrer que la fonction *spline* solution s'écrit alors sous la forme :

$$\hat{z}(\vec{s}) = \hat{z}(x, y) = a_0 + a_1 x + a_2 y + \sum_{i=1}^n b_i K(\vec{s} - \vec{s}_i)$$

Avec :

$$K(\vec{h}) = \|\vec{h}\|^2 \cdot \ln(\|\vec{h}\|)$$

Les inconnues $a_0, a_1, a_2, b_1, \dots, b_n$ sont solutions du système suivant :

$$\begin{cases} a_0 + a_1x_i + a_2y_i + \sum_{j=1, j \neq i}^n b_j K(\vec{s}_i - \vec{s}_j) = z(\vec{s}_i) \\ \sum_{j=1}^n b_j = 0, \sum_{j=1}^n b_j x_j = 0, \sum_{j=1}^n b_j y_j = 0 \end{cases}$$

En traduisant ce système sous forme matricielle cela revient à résoudre :

$$\begin{pmatrix} 1 & x_1 & y_1 & 0 & \dots & K_{1,n} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & y_n & K_{n,1} & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 1 \\ 0 & 0 & 0 & x_1 & \dots & x_n \\ 0 & 0 & 0 & y_1 & \dots & y_n \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Afin d'obtenir dans ce cas une estimation en un point donné, il est nécessaire de résoudre $n + 3$ équations. Par ailleurs on a l'existence et l'unicité de la solution si les points ne sont pas parfaitement alignés [29].

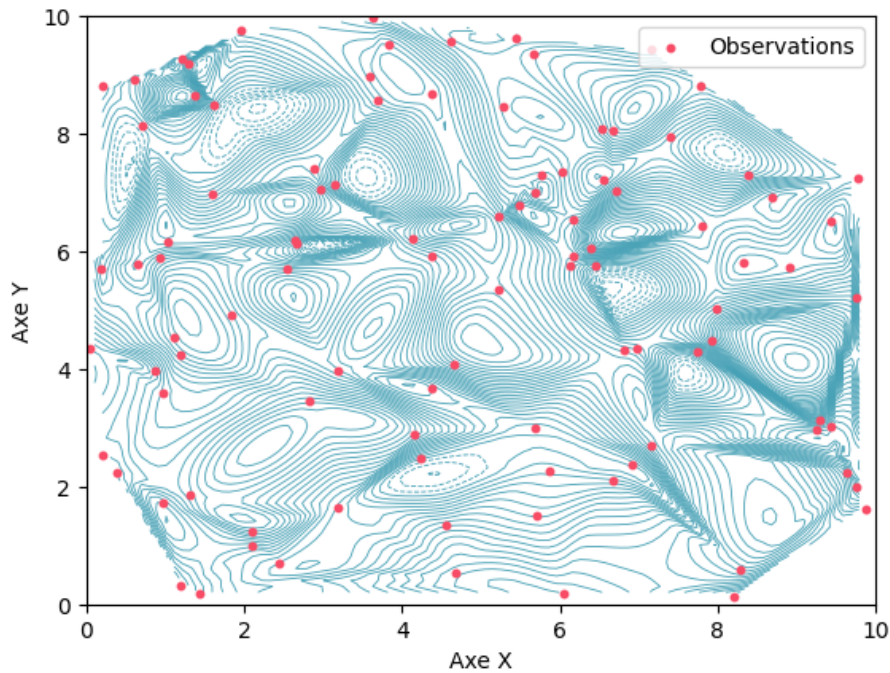


FIGURE 2.14 – *Spline* d'interpolation (exacte) - cubique : application sur un champ $\mathcal{D} \subset \mathbb{R}^2$ de 100 observations réparties aléatoirement - représentation par courbes de niveau - vue de dessus.

Les secondes, consistent en l'obtention d'une fonction la plus lisse possible mais sans la contrainte de restitution de la valeur exacte des points connus ($\hat{z}(\vec{s}_i) = z(\vec{s}_i)$). Mathématiquement cela revient au problème de minimisation de l'énergie de flexion suivant :

$$\sum_{i=1}^n [\hat{z}(\vec{s}_i) - z(\vec{s}_i)]^2 + \rho \iint \frac{\partial^2 \hat{z}}{\partial x^2} + 2 \frac{\partial^2 \hat{z}}{\partial x \partial y} + \frac{\partial^2 \hat{z}}{\partial y^2} dx dy$$

Où ρ fixé a priori est un paramètre strictement positif.

Il est alors possible de montrer que la solution revient à obtenir un cas de *splines* d'interpolation comme précédemment. Finalement, on a le système matriciel suivant :

$$\begin{pmatrix} 1 & x_1 & y_1 & \rho & \dots & K_{1,n} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n & y_n & K_{n,1} & \dots & \rho \\ 0 & 0 & 0 & 1 & \dots & 1 \\ 0 & 0 & 0 & x_1 & \dots & x_n \\ 0 & 0 & 0 & y_1 & \dots & y_n \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

L'estimation nécessite alors à nouveau la résolution de $n + 3$ équations et l'existence et l'unicité de la solution sont à nouveaux assurées lorsque les points ne sont pas alignés.

Synthèse des méthodes déterministes et limites

Nous avons pu parcourir les méthodes d'interpolation spatiales les plus utilisées, mais cependant suite à leurs implémentations pour le cas d'usage de la pollution de l'air, ces méthodes ont présenté un certain nombre de limites qui sont résumées dans la table suivante :

Méthode d'Interpolation	Interpolation Exacte	Effet Écran	Limite(s)
Plus Proche Voisin	Oui	Total	Sensibilité aux données aberrantes, discontinuités
Polynomiale Linéaire	Non	Partiel	Ne préserve pas les tendances complexes
IDW	Oui	Faible	Sensibilité à la valeur du paramètre p
<i>Spline</i> d'interpolation	Oui	Partiel	
<i>Spline</i> de lissage	Non	Faible	Peut sur-interpoler les données

TABLE 2.4 – Synthèse des méthodes d'interpolation spatiale déterministes

Pour conclure, ces méthodes déterministes permettent d'obtenir des résultats esthétiques en ayant recours à une complexité calculatoire relativement faible, mais elles ne sont pas nécessairement précises. En effet, ces méthodes s'appliquent strictement et ne tiennent pas compte de la structure spatiale de la variable à estimer. Cet aspect constitue un frein à la performance du modèle d'interpolation. Cependant il est nécessaire de hautement considérer les sorties du module d'interpolation. Les résultats finaux anticipés de l'étude étant particulièrement dépendants des valeurs des concentrations de polluant estimées en tout point du *raster* de la France métropolitaine.

Par ailleurs, compte tenu des résultats obtenus lors des implémentations respectives des différents modèles déterministes présentés précédemment, il a été décidé de s'orienter sur

d'autres modèles plus complexes. En outre, la littérature et de nombreux travaux considérés comme référents dans le domaine, comme peut l'être *Real et .al 2022 [31]*, nous incitent à investiguer d'autres méthodes. En effet, l'étude « Reconstruction historique de la pollution atmosphérique de fond en France pour la période 2000-2015 » *Real et .al 2022 [31]*, fournit une méthodologie exhaustive de la réalisation d'une interpolation spatiale des stations de mesures de concentrations en France métropolitaine pour les polluants $PM_{2.5}$, PM_{10} , O_3 et NO_2 . Cette méthodologie est « mimétisée » dans la poursuite de ce second chapitre afin de parvenir à la construction du module final d'interpolation.

Il s'agit par la suite de discuter de ces méthodes qui prennent en compte la dépendance spatiale de la variable régionalisée et qui seront par conséquent préférées aux modèles déterministes. Ces modèles diffèrent des modèles déterministes en faisant appel à une dimension probabiliste : ce sont les méthodes stochastiques également appelées méthodes géostatistiques.

2.2.5 Présentation des modèles : Les modèles géostatistiques

Les techniques d'interpolation stochastiques aussi appelées *krigeages* font appel aux propriétés statistiques des points de mesure. Elles permettent notamment de quantifier l'autocorrélation spatiale parmi les mesures en prenant en compte la configuration spatiale des points d'échantillonnage autour de la localisation de la prévision. On parle également de méthodes géostatistiques car sont utilisées simultanément les valeurs observées et l'information de la position pour améliorer l'estimation. Ces méthodes permettent également de récupérer l'information de la précision de chaque estimation.

La particularité d'une méthode géostatistique est qu'elle fait appel à des modèles probabilistes pour approximer le phénomène naturel continu à estimer à l'aide d'un processus aléatoire.

Cependant contrairement aux statistiques classiques les valeurs de l'échantillon ne doivent pas nécessairement vérifier la condition d'indépendance et l'identique distribution. En effet ces hypothèses ne peuvent être vérifiées pour notre contexte de données spatiales.

Rappelons que la notion de variable aléatoire est une fonction renvoyant une valeur résultant d'une expérience aléatoire ; dans notre cas elle permet d'assurer la relation entre un phénomène (la pollution de l'air) et les résultats chiffrés d'une expérience portant sur ce phénomène. Ce résultat est appelé réalisation de la variable aléatoire.

2.2.6 Le *krigeage*

Le *krigeage* est une méthode d'interpolation particulièrement applicable pour des données spatiales. Sa théorie a été développée dans les années 1950 par le mathématicien français G. Matheron en se servant des travaux de l'ingénieur minier sud-africain D. G. Krige qui cherchait à étudier la distribution de minerais à partir d'un échantillon de forages [29]. Les principes et fondements de cette méthodologie sont discutés ci-après.

Principe

Le but du *krigeage* est d'estimer le plus fidèlement la variable régionalisée par une combinaison linéaire des observations retenues au sein de l'échantillon de valeurs réelles. Le critère de précision ici à minimiser est l'erreur quadratique moyenne. Cette méthodologie induit quatre contraintes distinctes qu'il conviendra d'expliciter.

- **Contrainte de linéarité** : La valeur estimée doit être une combinaison linéaire des données.

$$\hat{Z}(\vec{s}_0) = \sum_{i=1}^n \lambda_i(\vec{s}_0) Z(\vec{s}_i)$$

- **Contrainte d'autorisation** : L'espérance et la variance de l'erreur de prévision $\hat{z}(\vec{s}_0) - z(\vec{s}_0)$ doivent obligatoirement exister. Pour s'en assurer, on a alors :

$$\begin{aligned} \hat{Z}(\vec{s}_0) - Z(\vec{s}_0) &= \left[\sum_{i=1}^n \lambda_i(\vec{s}_0) Z(\vec{s}_i) \right] - Z(\vec{s}_0) \\ &= \left[\sum_{i=1}^n \lambda_i(\vec{s}_0) Z(\vec{s}_i) \right] - \left[\sum_{i=1}^n \lambda_i(\vec{s}_0) Z(\vec{s}_0) \right] + \left[\sum_{i=1}^n \lambda_i(\vec{s}_0) Z(\vec{s}_0) \right] - Z(\vec{s}_0) \\ &= \left[\sum_{i=1}^n \lambda_i(\vec{s}_0) [Z(\vec{s}_i) - Z(\vec{s}_0)] \right] + Z(\vec{s}_0) \left[\sum_{i=1}^n \lambda_i(\vec{s}_0) - 1 \right] \end{aligned}$$

L'erreur de prévision est ici combinaison linéaire d'accroissements si et seulement si $\sum_i \lambda_i(\vec{s}_0) = 1$. Dans cet unique cas on a bien l'existence de la variance et de l'espérance sous l'hypothèse de stationnarité sous-jacente.

- **Contrainte de non biais** : L'espérance de l'erreur de prévision doit être nulle. On vérifie donc que l'estimateur de l'erreur de prévision est non biaisé, or d'après l'équation précédente on a :

$$\mathbb{E} \left[\hat{Z}(\vec{s}_0) - Z(\vec{s}_0) \right] = \left[\sum_{i=1}^n \lambda_i(\vec{s}_0) \mathbb{E} [Z(\vec{s}_i) - Z(\vec{s}_0)] \right]$$

- **Contrainte d'optimalité** : Les poids λ_i doivent être estimés de manière à minimiser la variance de l'erreur de prévision $\text{var}[\hat{z}(\vec{s}_0) - z(\vec{s}_0)]$. On se ramène donc à la recherche des extrema du lagrangien suivant :

$$\mathcal{L}(\lambda_1(\vec{s}_0), \dots, \lambda_n(\vec{s}_0), \mu) = \mathbb{E} \left[\left(\sum_{j=1}^n \lambda_j(\vec{s}_0) Z(\vec{s}_j) - Z(\vec{s}_0) \right)^2 \right] + 2\mu \left(1 - \sum_{i=1}^n \lambda_i(\vec{s}_0) \right)$$

Où μ représente un multiplicateur de Lagrange.

Après développement, on peut réécrire le lagrangien sous la forme :

$$\mathcal{L} = - \sum_{i=1}^n \sum_{j=1}^n \lambda_i(\vec{s}_0) \lambda_j(\vec{s}_0) \gamma_{ij} + 2 \sum_{j=1}^n \lambda_j(\vec{s}_0) \gamma_{j0} + 2\mu \left(1 - \sum_{i=1}^n \lambda_i(\vec{s}_0) \right)$$

Où $\gamma_{ij} = \gamma(\vec{s}_i, \vec{s}_j)$

Le calcul des dérivées partielles premières de \mathcal{L} par rapport aux $\lambda_i(\vec{s}_0)$ donne les équations constituant le système *krigeage* :

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \lambda_i(\vec{s}_0)} &= -2 \sum_{i=1}^n \lambda_i(\vec{s}_0) \gamma_{ij} + 2\gamma_{i0} - 2\mu \\ &= 0\end{aligned}$$

Soit, pour tout $i \in [1; n[$,

$$\sum_{j=1}^n \lambda_j(\vec{s}_0) \gamma_{ij} + \mu = \gamma_{i0}$$

Au final, le système à résoudre s'écrit donc :

$$\begin{cases} \sum_{j=1}^n \lambda_j(\vec{s}_0) \gamma_{ij} + \mu = \gamma_{i0} \\ \sum_{j=1}^n \lambda_j(\vec{s}_0) = 1 \end{cases}$$

Ce que l'on peut réécrire sous forme matricielle :

$$\begin{pmatrix} 0 & \gamma_{12} & \gamma_{13} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & 0 & \gamma_{23} & \dots & \gamma_{2n} & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \gamma_{n3} & \dots & 0 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1(\vec{s}_0) \\ \lambda_2(\vec{s}_0) \\ \vdots \\ \lambda_n(\vec{s}_0) \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix}$$

Variogramme

Conventionnellement il est d'usage pour le calibrage d'un modèle de krigeage d'avoir recours à une analyse variographique. Cette étape de la constructions d'un modèle de krigeage consiste en l'exploration des données observées et au calcul des variations des valeurs de la variable cible en fonction des distances aux autre points : on obtient alors le variogramme expérimental. Il est donné par la formule suivante :

$$\hat{\gamma}(\vec{h}) = \frac{1}{2|N(\vec{h})|} \sum_{(i,j) \in N(\vec{h})} [z(\vec{s}_i) - z(\vec{s}_j)]^2$$

Où $N(\vec{h}) = \{(i, j) / \vec{s}_i - \vec{s}_j = \vec{h} / i \neq j\}$ est l'ensemble des paires distinctes et $|N(\vec{h})|$ est le cardinal de $N(\vec{h})$. En pratique, les données n'étant pas régulièrement espacées, on définira l'ensemble $N(\vec{h})$ par $N(\vec{h}) = \{(i, j) / \vec{s}_i - \vec{s}_j \approx \vec{h} / i \neq j\}$. [29]

L'idée est alors de trouver le modèle variographique théorique qui se rapproche le plus du variogramme expérimental. Cela revient à minimiser la quantité :

$$\sum_k [\hat{\gamma}(\vec{r}_k) - \gamma(\vec{r}_k)]^2$$

On rappelle que le variogramme est l'outil qui reflète la structure de la régionalisation de la variable cible. Il en existe différent types, qu'il convient d'introduire. Seuls ceux qui ont

été implémentés sont présentés par la suite, sous contrainte de temps de calcul certains modèles ont été écartés et la liste n'est donc pas exhaustive :

- Modèle sphérique : de portée a et de palier C : linéaire à l'origine.

$$\gamma(r) = \begin{cases} C \left(\frac{3}{2} \frac{r}{a} - \frac{1}{2} \frac{r^3}{a^3} \right) & \text{pour } 0 \leq r \leq a \\ C & \text{pour } r > a \end{cases}$$

- Modèle puissance Modèle puissance d'exposant θ et de facteur d'échelle ω : $\theta \in [0; 2]$. Plus θ est proche de 2 , plus la variable régionalisée est régulière. Plus θ est proche de 0 , plus son comportement est erratique.

$$\gamma(r) = \omega r^\theta$$

- Modèle gaussien de paramètre a et de palier C : parabolique à l'origine.

$$\gamma(r) = C \left(1 - e^{-\frac{r^2}{a^2}} \right)$$

- Modèle *Hole-Effect* : parabolique à l'origine ; il permet de modéliser un variogramme moins stable, présentant des fluctuations autour du palier : il est caractéristique de données plus hétérogènes.

$$\gamma(r) = C \left[1 - \frac{\sin\left(\frac{r}{a}\right)}{\frac{r}{a}} \right]$$

- Modèle linéaire de pente ω : linéaire à l'origine.

$$\gamma(r) = \omega r$$

Cependant, la méthodologie retenue dans le présent mémoire n'est pas celle d'une implémentation d'un modèle de krigeage par analyse variographique classique. En l'absence de connaissances relatives à la structure spatiale du phénomène continu à estimer et sous contrainte de temps, il a été convenu d'utiliser une méthodologie type *Machine Learning* afin d'automatiser le calibrage, la sélection et la validation du modèle d'interpolation spatiale pour chaque polluant étudié. Cette méthodologie est développée en section suivante.

2.3 *Machine Learning* : Élection d'un modèle de *krigeage* - Méthodologie et application

Après avoir présenté les ressources théoriques mathématiques des modèles géostatistiques, il s'agit par la suite de discuter de la méthodologie adoptée pour la sélection, le paramétrage, l'entraînement, la validation et l'implémentation du module final d'interpolation spatiale.

Cette méthodologie appelée hyperparamétrage est une approche basée sur la théorie du *Machine Learning* qui consiste en l'automatisation de l'apprentissage supervisé d'un modèle. La nature supervisée de cette approche réside dans le fait que sont utilisées les observations réelles des stations de mesure pour entraîner et ajuster le modèle de *krigeage*.

Enfin, rappelons que pour le cas d'usage de la pollution de l'air, une seule variable régionalisée est à estimer pour chacun des polluants (respectivement les concentrations moyennes annuelles des particules fines, de dioxyde d'azote et la valeur de l'indicateur *SOMO35* pour l'ozone). Il s'agit donc d'une approche dite univariée où seules les données aux stations de l'échantillon retraité permettent d'expliquer la variable cible en tout pixel du *raster*.

Sur la base de ces rappels, il est désormais possible d'introduire la démarche d'élection du meilleur modèle d'interpolation spatiale par procédure d'hyperparamétrage.

2.3.1 Hyperparamétrage du *Krigeage*

L'hyperparamétrage d'un modèle de *machine learning* se définit comme la procédure utilisée pour la détermination des paramètres optimaux. Il consiste en l'ajustement automatique des paramètres du modèle qui ne sont pas directement appris à partir des données mais qui influencent la façon dont le modèle est construit et fonctionne (Diserbeau, 2019 [32]).

Les hyperparamètres eux, se réfèrent aux paramètres externes qui ont un impact sur les résultats du processus d'apprentissage. Ils sont définis préalablement et jouent un rôle crucial dans la configuration du processus d'optimisation. Dans le cas du *krigeage* sont considérés comme hyperparamètres :

- Variogramme : Le choix du modèle de variogramme (sphérique, gaussien, exponentiel, etc.) ainsi que les paramètres associés.
- Fonction de covariance : Le choix de la fonction de covariance spatiale (par exemple, la distance euclidienne, la distance de Manhattan, etc.).
- Type de modèle : Le type de modèle de *krigeage* à utiliser (ordinaire, universel, simple, etc.).
- Paramètres du modèle : Les paramètres spécifiques au type de modèle choisi, tels que les coefficients de dérive, les termes de tendance, etc.
- Paramètres de régularisation : Les paramètres de régularisation pour éviter le surajustement dans les modèles de *krigeage*.
- Méthode d'estimation : La méthode utilisée pour estimer les paramètres du modèle de *krigeage* (par exemple méthode des moindres carrés).
- Modèle de dérive : Le modèle de dérive spatiale utilisé pour capturer les tendances dans les données.

La procédure d'hyperparamétrage a recours à plusieurs concepts comme peuvent l'être l'échantillonnage de l'espace des hyperparamètres ; les métriques d'évaluation ou encore les méthodes de validation croisée qu'il convient de définir.

2.3.2 Échantillonnage de l'espace des hyperparamètres

Sur la base des paramètres à déclarer d'un modèle de *krigeage* présentés ci-dessus, la première étape de l'hyperparamétrage consiste en la définition de l'espace de recherche de paramètres que l'on souhaite assigner au modèle pour *in fine* tester et évaluer ses performances traduisant de sa capacité de généralisation de l'information.

Si l'on considère les paramètres d'un modèle de *krigeage* comme étant des variables, alors les paramètres à tester sont les modalités de chaque variable contenues dans un vecteur \mathbf{m} , où $\mathbf{m} = (m_1, m_2, \dots, m_n)$. L'espace des hyperparamètres \mathcal{H} constitue ainsi le produit cartésien de ces vecteurs de modalités, défini comme :

$$\mathcal{H} = \mathbf{m}_1 \times \mathbf{m}_2 \times \dots \times \mathbf{m}_n$$

où \mathbf{m}_i représente l'ensemble des modalités possibles pour le i -ème paramètre.

L'étape suivante consiste à choisir le moyen de parcourir l'espace \mathcal{H} des hyperparamètres.

Échantillonnage par grille (*Grid Search*)

Soit $\mathbf{h} \in \mathcal{H}$ une combinaison des modalités pour chaque paramètre \mathbf{m}_i , l'échantillonnage par grille ou *Grid Search* consiste à tester chaque élément \mathbf{h} de \mathcal{H} . L'objectif est de trouver la combinaison d'hyperparamètres qui produit les meilleures performances selon un critère spécifique appelé métrique d'évaluation (2.3.3). Le résultat de cette procédure n'est autre que le jeu de paramètres optimal du modèle de *krigeage* pour les données étudiées.

Échantillonnage aléatoire (*Random Search*)

La procédure de *Grid Search* bien qu'exhaustive est particulièrement coûteuse en temps de calcul à mesure que $|\mathcal{H}|$ ou la quantité de données traitées augmentent.

Aussi plutôt que d'explorer systématiquement l'espace complet des hyperparamètres comme c'est le cas pour le *Grid Search*, l'échantillonnage aléatoire ou *Random Search* explore un sous-ensemble aléatoire de cet espace. Dans cette procédure on spécifie un nombre \mathbf{n} de combinaisons de paramètres \mathbf{h} à tester tel que $\mathbf{n} < |\mathcal{H}|$ et on sélectionne aléatoirement \mathbf{n} configurations de paramètres de $\mathbf{h} \in \mathcal{H}$. Cette approche présente l'avantage de réduire le coût computationnel, tout en maintenant la capacité de trouver des configurations performantes d'hyperparamètres. En revanche cette procédure présente la limite de ne pas systématiquement renvoyer le meilleur jeu de paramètres en retournant un optimum local.

Échantillonnage bayésien

Basé sur une approche probabiliste, l'échantillonnage bayésien des hyperparamètres est une méthodologie puissante qui intègre l'incertitude liée à chacun des paramètres permettant de fournir des informations complémentaires au décisionnaire pour la sélection des valeurs optimales des hyperparamètres.

Brièvement, on définit une distribution de probabilité sur les hyperparamètres en amont de la procédure censée refléter la connaissance que l'on a des valeurs possibles de paramètres avant d'observer les données. Dans un second temps, en utilisant les données à

disposition et le modèle de *krigeage*, on met à jour cette distribution pour obtenir une distribution *a posteriori* des hyperparamètres. Finalement, cette distribution *a posteriori* fournit l'incertitude sur les valeurs des hyperparamètres et les valeurs des paramètres les plus probables.

2.3.3 Les métriques d'évaluation

Lors d'une procédure d'hyperparamétrage il est nécessaire de spécifier l'objectif à optimiser. Il convient alors de choisir une métrique d'évaluation qu'on cherchera en fonction du choix réalisé à maximiser ou à minimiser. Voici une liste des métriques usuellement utilisées et leur objectif respectif (Maximisation ou Minimisation) :

- Erreur Quadratique Moyenne (MSE - *Mean Squared Error*) (Minimisation) :

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{z}(\vec{s}_i) - z(\vec{s}_i))^2$$

- Racine de l'Erreur Quadratique Moyenne (RMSE - *Root Mean Squared Error*) (Minimisation) :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{z}(\vec{s}_i) - z(\vec{s}_i))^2}$$

- Erreur Absolue Moyenne (MAE - *Mean Absolute Error*) (Minimisation) :

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{z}(\vec{s}_i) - z(\vec{s}_i)|$$

- Coefficient de Détermination (R^2 , *R-squared*) (Maximisation) :

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{z}(\vec{s}_i) - z(\vec{s}_i))^2}{\sum_{i=1}^n (z(\vec{s}_i) - z(V))^2}$$

- Erreur Moyenne Relative (MER - *Mean Error Ratio*) (Minimisation) :

$$MER = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{z}(\vec{s}_i) - z(\vec{s}_i)}{z(\vec{s}_i)} \right| \times 100 \%$$

2.3.4 Validation croisée

La validation croisée est une technique fondamentale en apprentissage automatique et en statistiques pour évaluer les performances et la généralisation d'un modèle. Cette méthode vise à estimer comment un modèle se comportera sur de nouvelles données non vues, en simulant la séparation entre les données d'entraînement et les données de *test*. L'idée centrale de la validation croisée est de diviser le jeu de données disponible en plusieurs ensembles de manière à ce que chaque sous-ensemble soit utilisé à la fois pour l'entraînement et la validation.

La validation croisée offre l'avantage de fournir une estimation plus réaliste de la capacité de généralisation du modèle qu'un simple entraînement *train-test*, car elle évite le risque de sur-ajustement à un seul ensemble de données de *test*.

De plus, elle permet de maximiser l'utilisation des données pour l'apprentissage et l'évaluation. Cependant, la validation croisée peut prendre plus de temps, car le modèle est entraîné et évalué plusieurs fois.

Pour conclure, il s'agit une étape cruciale pour sélectionner les hyperparamètres optimaux et évaluer la performance du modèle avant son déploiement sur de nouvelles données.

Principe de jeu d'entraînement et de jeu de *Test* (*train-test*)

Le concept de jeu d'entraînement et de jeu de *Test* est une étape clé en apprentissage automatique. Cette étape fondamentale est conçue pour évaluer les performances du modèle entraîné et sa capacité de généralisation des données qu'on lui soumet. L'objectif à terme est de simuler la manière dont le modèle se comporte lorsque de nouvelles données qui ne sont pas intervenues lors de la phase d'entraînement lui sont soumises.

Le jeu d'entraînement appelé « *Train* » est utilisé pour former le modèle en ajustant ses paramètres aux données disponibles. Il s'agit de la partie du jeu de données sur laquelle le modèle apprend les relations entre les variables d'entrée et les cibles de sortie. Une fois que le modèle est entraîné, il est important de l'évaluer sur un ensemble distinct appelé jeu de « *Test* » qui ne fait pas partie du jeu d'entraînement.

Le jeu de *Test* est utilisé pour évaluer la performance du modèle sur des données qu'il n'a jamais rencontrées auparavant. Il permet de mesurer à quel point le modèle généralise bien à de nouvelles observations. En comparant les prédictions du modèle sur le jeu de *Test* avec les valeurs réelles, on peut obtenir des mesures appelées métriques d'évaluation telles que l'erreur moyenne, l'erreur quadratique moyenne ou la précision.

Il est en effet important de veiller à ce que le jeu de *test* soit indépendant du jeu d'entraînement afin d'obtenir une évaluation représentative de la performance réelle du modèle. En pratique, la proportion entre les jeux d'entraînement et de *test* peut varier en fonction de la taille totale du jeu de données et des besoins spécifiques du problème mais couramment on utilise les proportions 80 % de données d'entraînement pour 20 % de données de *test*.

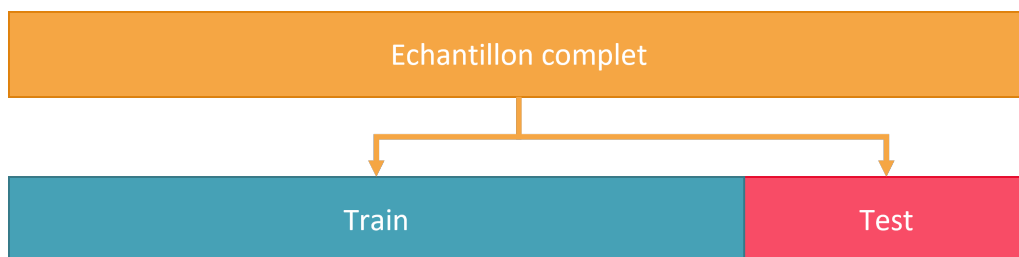


FIGURE 2.15 – Schéma explicatif de la réalisation d'un échantillonnage aléatoire type *Train-Test*.

K-Folds

Une approche courante de la validation croisée est la validation croisée *k-folds*. Dans cette méthode, le jeu de données est divisé en k sous-ensembles (ou « *folds* ») de taille égale. Le modèle est ensuite entraîné sur $k - 1$ sous-ensembles et évalué sur l'ensemble restant.

Ce processus est répété k fois, chaque fois en utilisant un *fold* différent comme ensemble de validation. Cela induit que le temps de calcul sera multiplié par un facteur k par rapport à une procédure *Train-Test* classique car le modèle est testé sur chaque sous ensemble. Les performances du modèle sont le résultat de la moyenne de la métrique d'évaluation choisie sur les k évaluations pour obtenir une mesure de la performance globale du modèle.

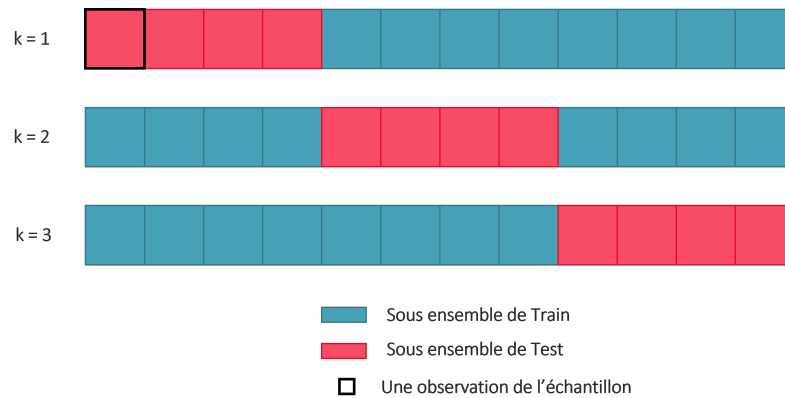


FIGURE 2.16 – Schéma explicatif de la réalisation d'un échantillonnage aléatoire type *K-Fold* pour $k = 3$.

Leave-One-Out

En se basant fondamentalement sur la même démarche que la procédure de *K-Folds*, la méthodologie de *Leave-One-Out* présente la particularité extrême de la création et l'évaluation d'un modèle pour chaque exemple de l'ensemble de données d'apprentissage. Plus expressément, supposons que nous disposons de l'ensemble de données \mathcal{Z} , la validation croisée *Leave-One-Out* consiste à itérer sur chaque élément de l'ensemble de données \mathcal{Z} , en laissant un élément hors de l'ensemble d'apprentissage à chaque itération, et en utilisant ce dernier comme ensemble de *test*. En reprenant les notations introduites en sous section 2.2.3, on a :

- À chaque itération i , l'élément \vec{s}_i est retiré de l'ensemble de données, et un modèle de *krigeage* est formé en utilisant les autres sites d'observation comme ensemble d'apprentissage.
- Ensuite, ce modèle est utilisé pour estimer la valeur de z au site \vec{s}_i exclu.
- Ce processus est répété pour chaque élément de l'ensemble de données \mathcal{Z} , utilisant chaque site d'observation comme point de validation.

Finalement, une série d'estimations $\hat{z}(\vec{s}_i)$ est obtenue pour chaque site d'observation \vec{s}_i . La performance du modèle est évaluée en comparant ces estimations aux valeurs réelles de z aux sites d'observation et en calculant la métrique d'évaluation retenue pour chaque observation.

Cependant, il est important de noter que cette méthode peut être coûteuse en matière de calcul, surtout si le nombre d'éléments dans l'ensemble de données \mathcal{Z} est grand.

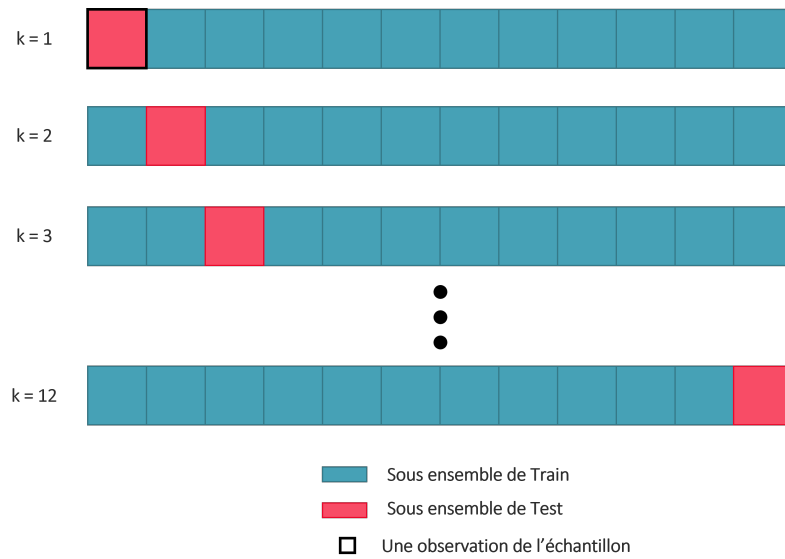


FIGURE 2.17 – Schéma explicatif de la réalisation d'un échantillonnage aléatoire type *Leave-One-Out*.

Passons désormais à l'application réelles sur les données de pollution de l'air.

2.4 Implémentation et validation du modèle

2.4.1 Synthèse de la démarche

Afin de construire le module d'interpolation spatiale capable d'estimer le plus fidèlement possible les différentes concentrations de polluants à l'étude, la méthodologie a été résumée en *Figure 2.18*.

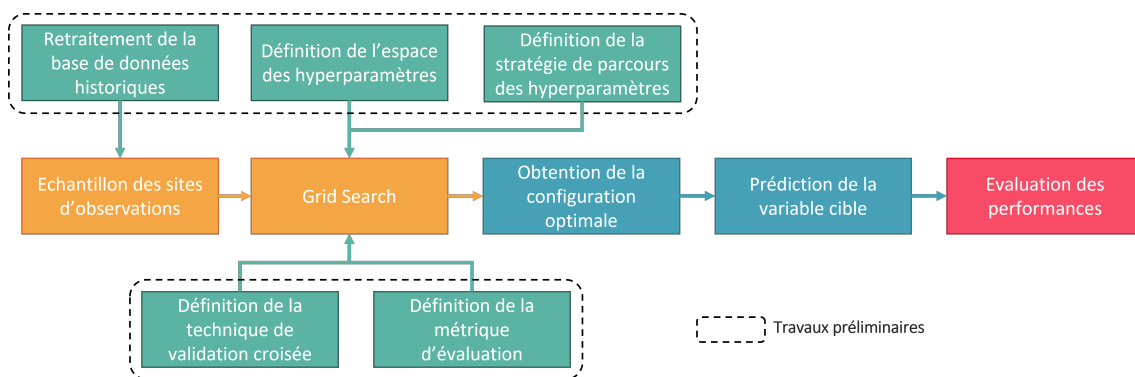


FIGURE 2.18 – Synthèse schématisée de la méthodologie adoptée pour la sélection, l'implémentation et la validation du modèle d'interpolation spatiale. Chaque étape ayant été réalisée pour chacun des polluants étudiés ($PM_{2.5}$, PM_{10} , NO_2 , $SOMO35$).

Passons désormais à l'explication respective des étapes présentées ci-dessus en commençant par les travaux préliminaires à l'implémentation de l'hyperparamétrage.

2.4.2 Travaux préliminaires

1. **Retraitements de la base historique** : Suite à l'étape exploratoire des données extraites de la base *GEOD'AIR*, il a été constaté que certaines données étaient invalides ou présentaient des limites à la réalisation de l'étude.

Certains retraitements ont donc été appliqués comme la suppression des stations de mesures qui présentent une couverture temporelle annuelle insuffisante (inférieure à 70 %). Une couverture temporelle correspond en temporalité annuelle au nombre de jours où l'on a mesuré des données valides, cohérentes et exploitables sur le nombre total de jours d'une année (cette information étant communiquée pour chaque station).

Ensuite, les stations de relevés autres que les stations de type influence de fond ont été supprimées. Il s'agit des stations de type industrielles et de trafic qui sont davantage placées dans des lieux dont la concentration est influencée par une activité humaine soutenue résultant de mesures de concentration locales particulièrement élevées. Ces dernières constituant des valeurs extrêmes, sont difficiles à estimer et représentent donc une limite dans un système de représentation cartographique à l'échelle nationale.

Les stations dont la variable validité était égale à -1 traduisant une mesure inexploitable, ont également été supprimées.

Enfin, les valeurs annuelles de concentrations par polluant ont été moyennées sur la totalité de l'historique disponible. Par ailleurs les stations de mesure qui, sur la profondeur historique de 10 ans, ne présentaient pas un minimum de 5 mesures ont été supprimées de la base.

Finalement, tous retraitements réalisés, on obtient l'échantillon final des sites d'observations. Ce dernier se compose des effectifs de stations de mesures par polluant suivants :

Polluant	Nombre de stations
NO_2	261
O_3	298
PM_{10}	224
$PM_{2.5}$	101
$SOMO35$	301

TABLE 2.5 – Nombre de stations conservées par polluant (*post*-retraitements effectués en 2.4.2)

Passons désormais à la construction de l'espace des hyperparamètres testés.

2. **Définition de l'espace des hyperparamètres** : Voici la liste des paramètres qui ont été définis et testés dans l'étude :

- **Méthode** (m_1) : *Ordinary, Universal*
- **Modèle de Variogramme** (m_2) : *Linear, Power, Gaussian, Spherical, Hole Effect*
- **Nombre de Lag** (m_3) : 4, 6, 8, 10
- **Poids** (m_4) : *True, False*

On rappelle que l'espace des hyperparamètres \mathcal{H} est le résultat du produit cartésien des vecteurs de modalités m_i pour le i -ème paramètre tel que :

$$\mathcal{H} = \mathbf{m}_1 \times \mathbf{m}_2 \times \mathbf{m}_3 \times \mathbf{m}_4$$

où \mathbf{m}_i représente l'ensemble des modalités possibles pour le i -ème paramètre.

Finalement on a donc : $|\mathcal{H}| = 2 \cdot 5 \cdot 4 \cdot 2 = 80$ combinaisons de paramètres h

3. **Définition de la stratégie de parcours de l'espace des hyperparamètres** : Considérant la grille de paramètres testée énoncée en 2., c'est la méthodologie de *Grid Search* qui a été retenue compte tenu du nombre peu élevé ($2 \times 5 \times 4 \times 2 = 80$) de jeux de paramètres h différents.

4. **Définition de la technique de validation croisée** : Pour le cas de l'interpolation spatiale, c'est la méthode singulière de validation croisée par *Leave-One-Out* qui est majoritairement retenue. En effet la méthodologie de *Real et .al 2022 [31]*, prescrit une validation croisée type *Leave-One-Out* dans le cadre de l'hyperparamétrage d'un modèle de *kriigeage*.

Bien que, comme énoncé en 2.3.4, cette technique soit très coûteuse en terme de temps de calcul (le modèle étant exécuté le même nombre de fois que le nombre d'observations que comporte l'échantillon étudié), le nombre de stations *post*-retraitement ne dépassant pas 301 sites d'observations comme le suggère la *Table 2.5*, cette méthodologie s'est avérée réalisable.

5. **Définition de la métrique d'évaluation** : Pour la métrique d'évaluation, la RMSE a été privilégiée et utilisée pour évaluer les performances des modèles pour chaque jeu de paramètres testé. La RMSE étant une mesure d'erreur, la procédure de *GridSearch* a cherché à retenir le jeu de paramètres qui la minimise.

En effet la RMSE est considérée comme une mesure standard dans les applications de *Machine Learning*, pour les avantages suivants :

- Stabilité : c’est une mesure considérée comme stable car la racine carrée appliquée au carré des erreurs permet de normaliser les valeurs.
- Interprétation plus intuitive : elle est exprimée dans la même unité de mesure que les données exploitées
- Mesure plus informative : elle prend en compte à la fois la variance des erreurs de prédiction et la précision.

2.4.3 Implémentation de l’hyperparamétrage : Résultats des configurations optimales

Suite à ces travaux préliminaires, la procédure de *Grid Search* a été implémentée et réalisée à l’aide de la librairie *scikit-learn* du langage de programmation *Python*. Ainsi sur l’échantillon retraité des sites d’observations à valeurs moyennes de concentration historique la procédure a retourné les configurations optimales suivantes :

Polluant	Type de modèle	Nombre de <i>lags</i>	Modèle du <i>variogram</i>	Pondération
<i>PM_{2.5}</i>	<i>Ordinary</i>	6	<i>Spherical</i>	Oui
<i>PM₁₀</i>	<i>Universal</i>	6	<i>Spherical</i>	Oui
<i>NO₂</i>	<i>Universal</i>	8	<i>Power</i>	Non
<i>SOMO35</i>	<i>Universal</i>	8	<i>Power</i>	Non

TABLE 2.6 – Configurations optimales du modèle de *krigeage* pour chaque polluant, obtenues à l’issue du *Grid Search*.

Il s’agira désormais d’analyser plus profondément la consistance de ces résultats via l’étude des métriques classiques d’analyse de performance des modèles de *Machine Learning*.

2.4.4 Prédiction et évaluation des performances

Afin de poursuivre, un code a été réalisé afin d’implémenter une interpolation spatiale via les configurations optimales obtenues en 2.4.3 , de relancer une procédure type *Leave-One-Out* et enfin, d’obtenir pour chaque station d’observation \vec{s}_i une estimation $\hat{z}(\vec{s}_i)$ de la concentration du polluant étudié, afin de la comparer à sa valeur réelle $z(\vec{s}_i)$. Ainsi cette procédure a été itérée sur chaque polluant retenu et ont été obtenus des vecteurs de valeurs prédites.

Par ailleurs, on rappelle également les équations des métriques étudiées telles que :

1. **Coefficient de détermination (r^2)** : indique la proportion de la variance totale des valeurs réelles expliquée par le modèle.

$$r^2 = 1 - \frac{\sum (z(\vec{s}_i) - \hat{z}(\vec{s}_i))^2}{\sum (z(\vec{s}_i) - z(V))^2}$$

2. **Corrélation** : indique à quel point les prédictions et les vraies valeurs sont corrélées

linéairement.

$$\text{Corrélation} = \frac{\sum(z(\vec{s}_i) - z(V))(\hat{z}(\vec{s}_i) - \bar{\hat{z}})}{\sqrt{\sum(z(\vec{s}_i) - z(V))^2 \sum(\hat{z}(\vec{s}_i) - \bar{\hat{z}})^2}}$$

3. **Erreur Absolue Moyenne (MAE)** : quantifie la différence moyenne entre les prédictions du modèle et les vraies valeurs.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |z(\vec{s}_i) - \hat{z}(\vec{s}_i)|$$

4. **Biais** : indique si les prédictions sont en surestimation ou sous-estimation des valeurs réelles.

$$\text{Biais} = \frac{1}{n} \sum_{i=1}^n (z(\vec{s}_i) - \hat{z}(\vec{s}_i))$$

5. **Variance** : indique la dispersion des erreurs autour de la valeur réelle.

$$\text{Variance} = \frac{1}{n} \sum_{i=1}^n (z(\vec{s}_i) - \hat{z}(\vec{s}_i) - \text{Biais})^2$$

6. **Racine Carrée de l'Erreur Quadratique Moyenne (RMSE)** : donne une idée de l'écart typique entre les prédictions et les vraies valeurs.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z(\vec{s}_i) - \hat{z}(\vec{s}_i))^2}$$

Afin de qualifier la pertinence des configurations optimales des modèles de *krigeage* obtenus en 2.4.3, les vecteurs de valeurs prédites ont été mis en phase avec les valeurs réelles via le calcul des métriques décrites ci-dessus. Les résultats sont reportés dans la table suivante :

Polluant	Val. Moyenne	r ²	Corrélation	MAE	Biais	Variance	RMSE
NO ₂	16,22	0,43	0,68	3,84	0,07	25,91	5,09
PM _{2.5}	10,63	0,14	0,37	1,46	0,03	3,81	1,95
PM ₁₀	16,95	0,38	0,62	1,59	0,02	4,87	2,21
SOMO35	12,28	0,71	0,84	1,56	-0,01	4,69	2,17

TABLE 2.7 – Résultats obtenus par polluant des métriques pour les modèles optimaux de *krigeage* sur les échantillons historiques.

En théorie, lorsque l'ajustement d'un modèle aux données est amélioré, le biais, qui représente l'écart entre les prédictions du modèle et les valeurs réelles, tend à diminuer.

Cependant, cette réduction du biais s'accompagne généralement d'une augmentation de la variance. En effet, un modèle fortement ajusté sera contraint d'effectuer des ajustements plus prononcés pour s'aligner au mieux aux données d'apprentissage.

Cette adaptation propre aux données d'apprentissage peut rendre le modèle instable lorsqu'il est appliqué à un échantillon de validation distinct. En conséquence, un modèle à variance élevée peut perdre en capacité de généralisation, car il peut avoir du mal à faire des prédictions cohérentes et précises sur de nouvelles données.

Lorsqu'un modèle présente un biais significatif mais une variance faible, on le qualifie de « sous-apprentissage ». En revanche, un modèle avec un biais faible mais une variance élevée est considéré comme étant en situation de « sur-apprentissage ». On introduit le concept de « compromis biais-variance ». Le compromis idéal est atteint dans le cas d'un biais et d'une variance faible.

Ce dilemme classique en apprentissage automatique doit donc être utilisé pour évaluer la pertinence d'un modèle de *Machine Learning*.

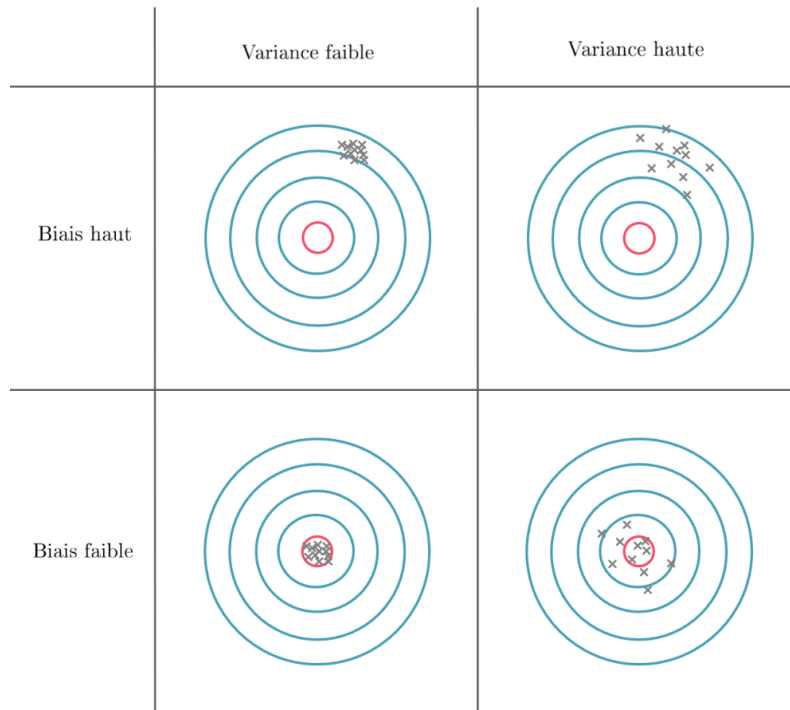


FIGURE 2.19 – Effet du biais et de la variance sur la précision des prédictions d'un modèle de *Machine Learning*. Le cercle rose représente la juste valeur à estimer.

Ainsi lorsqu'on se réfère à la *Table 2.7*, il est possible, par polluant de considérer :

- PM_{10} : Pour les PM_{10} on obtient un biais positif de 2 % mais une variance élevée plaçant notre modèle en situation de sur-apprentissage. La RMSE reportée à la valeur moyenne des points de mesure nous donne une erreur typique de 28 %. La corrélation des prédictions et des valeurs réelles est relativement importante (0,62) signifiant que lorsque les vraies valeurs augmentent, les prédictions ont tendance à augmenter également. En revanche le coefficient r^2 censé être proche de 1 ne s'élève qu'à 0,38 traduisant d'une adéquation faible aux données. Enfin la MAE est acceptable avec une valeur de 1,59.

En synthèse, le modèle est acceptable en l'état mais mériterait d'être ajusté davantage afin de mieux s'adapter aux données réelles, en cherchant à réduire le RMSE tout en maintenant une corrélation élevée et un biais faible.

- $PM_{2.5}$: Pour le cas des particules fines type $PM_{2.5}$ on obtient un biais proche de 0 et positif mais supérieur au biais obtenu pour les PM_{10} . Cela signifie qu'à nouveau, le modèle a une tendance à surestimer les valeurs réelles. La variance est quant à

elle particulièrement élevée. On obtient une valeur de RMSE égale à 1,95 ce qui, une fois reporté à la moyenne des observations en première colonne, représente 18 % d'écart typique entre la valeur réelle et la valeur prédite. La corrélation affichée est la plus faible de tous les modèles, ce qui signifie que le modèle n'arrive pas à capter la relation linéaire entre ces deux variables.

On peut tenter d'expliquer le manque de précision de ce modèle par le nombre de stations de relevés qui s'élève à 101. En effet les particules fines inférieures à $25\mu\text{m}$ n'ayant été mesuré qu'à compter de l'année 2009 en France. Le modèle de *krigeage* étant dépendant de la structure spatiale des points de mesure, la répartition géographique de ces stations joue un rôle dans la qualité des résultats d'interpolation. La mesure de concentration des PM_{10} étant d'avantage développé et comme le suggère *Real et .al 2022 [31]*, il aurait été tout à fait pertinent d'utiliser les stations de mesure de PM_{10} pour expliquer les valeurs de concentration des PM_{10} via l'élaboration d'un modèle de *co-krigeage* et en utilisant la relation du taux de particules $PM_{2.5}$ contenues dans le taux de PM_{10} fourni par l'*OMS, 2016 [33]*.

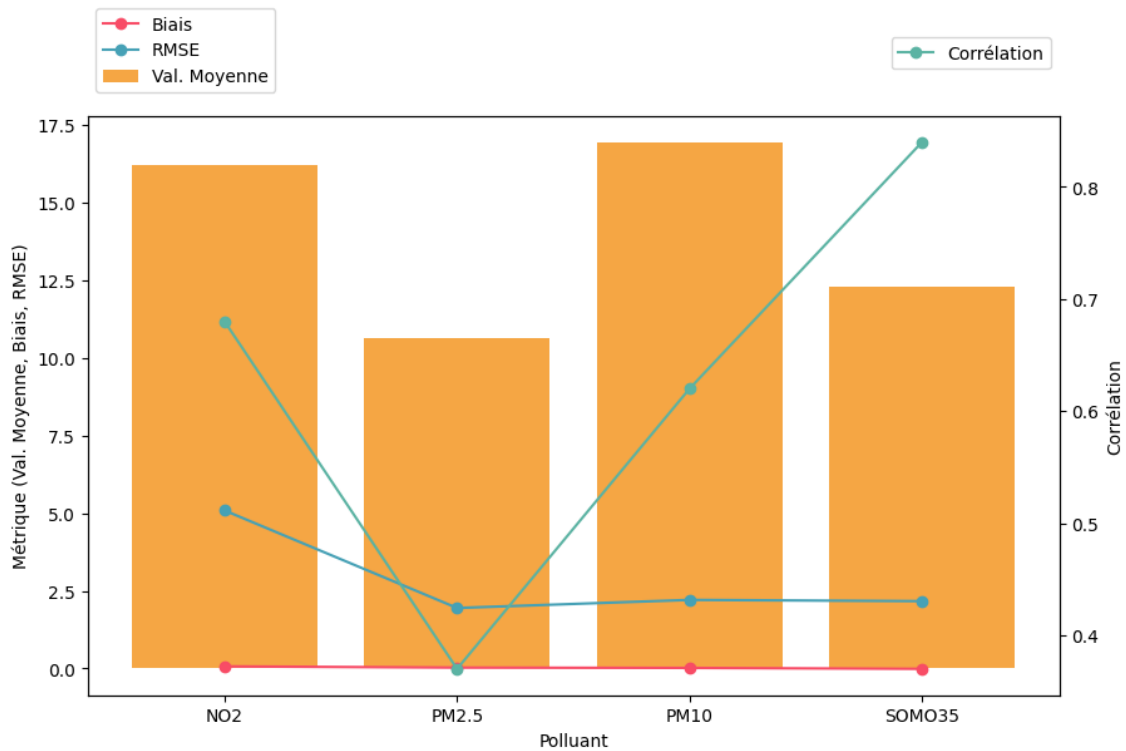


FIGURE 2.20 – Indicateurs statistiques calculés à l'aide de la technique de validation croisée *Leave-One-Out* sur les valeurs moyennes annuelles de différents polluants mesurés et estimés sur les stations de fond pour les années 2013 à 2022.

- NO_2 : Pour le cas du dioxyde d'azote, on obtient une nouvelle fois un biais relativement faible de 7 %, et une corrélation élevée légèrement inférieure à 70 %. En revanche la variance explose avec une valeur proche de 26. La RMSE une fois reportée à la valeur moyenne des observations est égale à 31% (*Table 2.7*).

En résumé, pour le polluant NO_2 , le modèle montre une certaine adéquation glo-

bale aux données, mais avec une marge d'erreur significative (MAE élevée) et une variance assez élevée. Bien que la corrélation soit forte, la variance et la MAE soulignent la nécessité d'une meilleure précision des prédictions pour une utilisation plus fiable.

Ces résultats peuvent s'expliquer par le caractère très localisé de ce polluant, qui se retrouve majoritairement en marge des grands axes routiers français. Le modèle ainsi utilisé n'est peut-être pas le plus adapté.

- O_3 (SOMO35) : Les résultats du modèle développé pour la mesure réglementaire SOMO35 sont meilleurs que pour les polluants précédents. En effet on obtient un indice r^2 égal à 0,71 et une corrélation de 0,84 traduisant d'un modèle restituant fidèlement les valeurs testées. La variance est, quant à elle, également élevée et la RMSE reportée à la valeur moyenne des observations représente une erreur typique de l'ordre de 17 %.

En résumé, pour le polluant SOMO35, le modèle montre une adéquation élevée aux données (r^2 et corrélation élevés) avec une précision relative (MAE faible), un biais faible et une stabilité dans les prédictions (variance et RMSE). Ces résultats indiquent que le modèle est capable de prédire de manière précise les valeurs de SOMO35 en tout point de la France métropolitaine.

Ci-dessous un exemple d'interpolation spatiale obtenue pour la mesure de SOMO35 en 2021 pour la France métropolitaine (Corse exclue) :

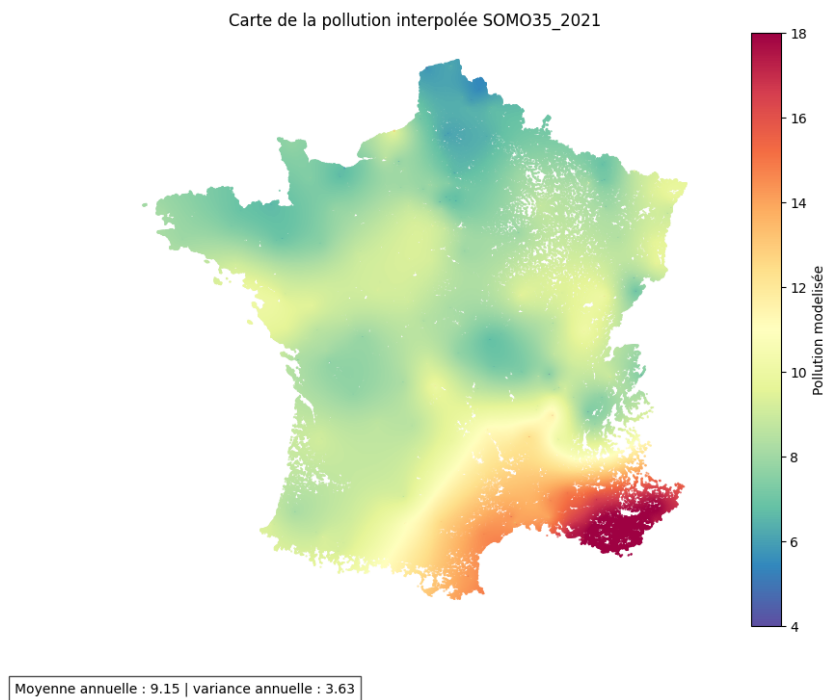


FIGURE 2.21 – Carte choroplèthe : Résultat du module d'interpolation spatiale appliqué au *raster* de la base carroyée de l'INSEE pour les valeurs du SOMO35 en 2021. Les espaces blanchis sont les pixels en France ne comportant aucun habitant. Ce résultat peut être comparé à celui de l'INERIS via l'Annexe A.2. On observe que le sud-est de la France est assujetti à des températures plus élevées et est donc plus exposé que le reste du pays.

2.4.5 Limites identifiées de l’approche

Comme vu précédemment, le *Machine Learning* a été utilisé dans cette étude dans le but de calibrer les modèles de *krigeages*. Bien que la démarche trouve ses racines de sources littéraires, les résultats restent objectivement améliorables *via* l’exploration de d’autres approches, méthodes et modèles.

Pour commencer durant cette étude seules les stations de mesures en France ont été utilisées pour estimer la pollution. Il aurait été tout à fait possible d’utiliser des stations de mesures des pays limitrophes. Ces données auraient pu améliorer la qualité des estimations aux extrémités de la base carroyée.

Ensuite, il serait intéressant durant la procédure d’hyperparamétrage et particulièrement durant le processus de *Grid Search* d’utiliser une ou plusieurs métriques d’évaluation différentes de la RMSE. Un modèle externe aurait pu être utilisé pour contraindre l’hyperparamétrage à améliorer la minimisation du couple biais-variance.

Un autre levier d’action pourrait être un élargissement de l’espace des hyperparamètres testés. Sous contrainte de temps et de puissance calculatoire, la grille testée s’est vu diminuée. En outre, l’implémentation de la méthodologie de parcours de l’espace des hyperparamètres par approche bayésienne aurait pu potentiellement fournir des informations supplémentaires pour le choix final des configurations optimales.

Ensuite certains modèles de *krigeage* n’ont volontairement pas été abordés sous contrainte de temps : le *krigeage* bayésien ou encore le *co-krigeage*.

Par ailleurs, il serait intéressant d’utiliser plus finement les résultats obtenus. Par exemple analyser les résultats en dissociant les types de station (rurale, périurbaine, urbaine, etc.).

Enfin il serait tout à fait pertinent de faire appel à d’autres données *Open Data* et l’*Open Source*, pour espérer trouver des variables exogènes capables d’expliquer plus précisément une concentration de pollution à un pixel donné. Le nombre d’individus de chaque pixel fourni par la base carroyée de l’INSEE en est un bon exemple.

Ainsi, dans ce chapitre nous avons pu présenter l’ensemble des sources de données utilisées, comprendre comment à partir de données historiques de concentrations il a été possible de construire un modèle d’interpolation spatiale capable d’estimer la valeur d’une concentration, et toute la méthodologie sous-jacente par approche de *Machine Learning* pour y parvenir.

Chapitre 3

Approches de génération et d'utilisation de scénarios prospectifs

Dans ce troisième chapitre, il s'agira de traiter du caractère prospectif de l'étude, tant sur les données que sur la méthodologie adoptée pour projeter les polluants dans le temps à l'horizon 2050. Dans un premier temps, le chapitre présentera la projection du module démographique de l'étude en France métropolitaine. Dans un second temps, il traitera de la projection du module pollution de l'air aux stations de mesure selon différents scénarios. Enfin les forces et limites de la méthodologie retenue seront abordées.

3.1 Projection de la population en base carroyée

Pour rappel, le module démographique de l'étude repose sur la base de population fournie par l'INSEE : « Revenus, pauvreté et niveau de vie en 2017 - Données carroyées ». Cette base carroyée constitue un recensement de la population française estimée en tout carreau d'1 km² en France métropolitaine pour l'année civile 2017. Seulement l'étude ayant pour ambition de quantifier les impacts de la pollution de l'air à un horizon long terme, il est donc obligatoire de faire évoluer cette base démographique au fil des années de projection.

C'est cette démarche qui sera présentée au sein de cette section, à commencer par l'évolution en effectifs de la base carroyée.

3.1.1 Projection en effectifs

Alors que la base de données carroyées fournit un nombre d'individus par pixel en France ainsi que le nombre d'individus par tranche d'âges définies au sein de la table des variables en *Annexe A.3*, la seconde base de données de l'INSEE concernant le scénario central présenté en *2.1.2* introduit une dimension future avec une projection à long terme de la démographie française.

L'INSEE fournit ainsi un fichier Excel appelé « projections_scenario_central » (*Insee Omphale, 2022*) dans lequel plusieurs tables sont fournies. La première à laquelle nous allons nous intéresser se nomme « Population_DEP ».

Dans cette table de projection (la plus récente), l'INSEE communique le nombre d'habitants (en milliers de personnes) estimé par département et par année de projection (2018-2070). Par raisonnement matriciel, on définit M une matrice 101×53 tel que $M(i, j)$ est le nombre d'habitants (en milliers) selon l'INSEE pour un département i et une année j spécifique :

$$M = \begin{pmatrix} M(1, 1) & M(1, 2) & \dots & M(1, 53) \\ M(2, 1) & M(2, 2) & \dots & M(2, 53) \\ \vdots & \vdots & \ddots & \vdots \\ M(101, 1) & M(101, 2) & \dots & M(101, 53) \end{pmatrix}$$

De cette première matrice définie, il est alors possible de calculer $\frac{M(i, j+1)}{M(i, j)}$ la matrice 101×52 des coefficients de passage d'une population d'une année j à une année $j + 1$ pour un département i . On a alors :

$$\frac{M(i, j+1)}{M(i, j)} = \begin{pmatrix} \frac{M(1, 2019)}{M(1, 2018)} & \frac{M(1, 2020)}{M(1, 2019)} & \dots & \frac{M(1, 2070)}{M(1, 2069)} \\ \frac{M(2, 2019)}{M(2, 2018)} & \frac{M(2, 2020)}{M(2, 2019)} & \dots & \frac{M(2, 2070)}{M(2, 2069)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{M(101, 2019)}{M(101, 2018)} & \frac{M(101, 2020)}{M(101, 2019)} & \dots & \frac{M(101, 2070)}{M(101, 2069)} \end{pmatrix}$$

Enfin, afin de projeter la population d'un département i de l'année initiale 2018 à une année j on réalise l'opération suivante :

$$C(i, j) = \prod_{k=2018}^j \frac{M(i, k+1)}{M(i, k)}$$

En généralisant, soit la matrice C (101×53) des coefficients multiplicatifs à appliquer à la population d'un département i arrêtée à l'année 2018, afin d'obtenir sa population projetée à l'année j :

$$C = \begin{pmatrix} C(1, 1) & C(1, 2) & \dots & C(1, 53) \\ C(2, 1) & C(2, 2) & \dots & C(2, 53) \\ \vdots & \vdots & \ddots & \vdots \\ C(101, 1) & C(101, 2) & \dots & C(101, 53) \end{pmatrix}$$

tel que :

$$C = \begin{pmatrix} \prod_{k=2018}^{2018} \frac{M(1, k+1)}{M(1, k)} & \prod_{k=2018}^{2019} \frac{M(1, k+1)}{M(1, k)} & \dots & \prod_{k=2018}^{2070} \frac{M(1, k+1)}{M(1, k)} \\ \prod_{k=2018}^{2018} \frac{M(2, k+1)}{M(2, k)} & \prod_{k=2018}^{2019} \frac{M(2, k+1)}{M(2, k)} & \dots & \prod_{k=2018}^{2070} \frac{M(2, k+1)}{M(2, k)} \\ \vdots & \vdots & \ddots & \vdots \\ \prod_{k=2018}^{2018} \frac{M(101, k+1)}{M(101, k)} & \prod_{k=2018}^{2019} \frac{M(101, k+1)}{M(101, k)} & \dots & \prod_{k=2018}^{2070} \frac{M(101, k+1)}{M(101, k)} \end{pmatrix}$$

Pour finir, afin de faire évoluer la base de données de pixels en France métropolitaine, et ne disposant pas de données plus fines d'évolution démographique par territoire administratif,

l'hypothèse forte suivante a été appliquée : les pixels suivent la tendance démographique du département auquel leur centroïde est rattaché.

Ainsi à chaque année de projection étudiée, les pixels ont vu leur nombre d'individus évoluer à partir des coefficients de la matrice C . Par ailleurs, la base de données carroyée de l'INSEE étant arrêtée à l'année 2017, les coefficients de passage $\frac{M(i,2019)}{M(i,2018)}$ lui ont été appliqués afin d'obtenir l'effectif démographique par département en 2018, ne disposant pas de l'information $\frac{M(i,2018)}{M(i,2017)}$.

3.1.2 Projection en tranche d'âge

Dans un second temps, il faut considérer le vieillissement de la population de la base carroyée. La seconde table « Structure_Age_DEP » du fichier « projections_scenario_central » (*Insee Omphale, 2022*) permet de prendre en compte ce phénomène.

Cette seconde table fournit par département la structure d'âge des années 2018 et 2070 uniquement. Plus explicitement, cette table donne le pourcentage de chaque classe d'âge quinquennale par département en 2018 et en 2070.

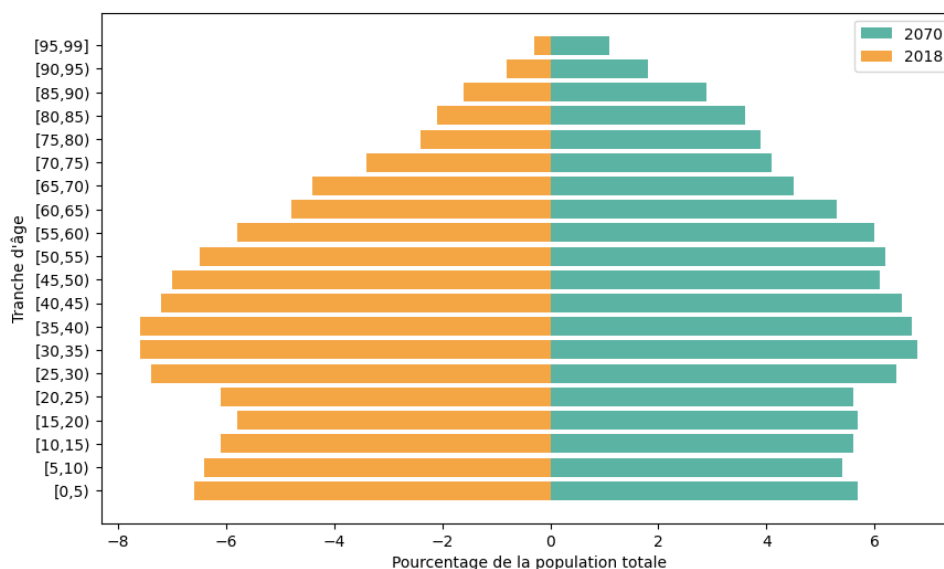


FIGURE 3.1 – Pyramides des âges pour le département des Hauts-de-Seine (92) en 2018 et en 2070, Scénario Central Insee.

Aussi, un outil Python a été construit pour vieillir la démographie française en concordance avec les chiffres communiqués par l'INSEE.

Ce dernier vise à projeter les structures par tranche d'âge pour chaque département en utilisant des coefficients de passage entre tranches d'âge. Pour ce faire, la structure par âge entre les années 2018 et 2070 a été interpolée linéairement afin d'obtenir le coefficient d'évolution des tranches d'âges annuelles par département.

Aussi soit la matrice A_{2018} (101×20) tel que $A_{2018}(i, j)$ représente le pourcentage de la démographie pour la tranche d'âge j et le département i pour l'année 2018.

$$A_{2018} = \begin{pmatrix} A_{2018}(1, 1) & A_{2018}(1, 2) & \dots & A_{2018}(1, 20) \\ A_{2018}(2, 1) & A_{2018}(2, 2) & \dots & A_{2018}(2, 20) \\ \vdots & \vdots & \ddots & \vdots \\ A_{2018}(i, 1) & A_{2018}(i, 2) & \dots & A_{2018}(i, 20) \\ \vdots & \vdots & \ddots & \vdots \\ A_{2018}(101, 1) & A_{2018}(101, 2) & \dots & A_{2018}(101, 20) \end{pmatrix}$$

Pareillement, soit la matrice A_{2070} (101×20) tel que $A_{2070}(i, j)$ représente le pourcentage de la démographie pour la tranche d'âge j et le département i pour l'année 2070.

$$A_{2070} = \begin{pmatrix} A_{2070}(1, 1) & A_{2070}(1, 2) & \dots & A_{2070}(1, 20) \\ A_{2070}(2, 1) & A_{2070}(2, 2) & \dots & A_{2070}(2, 20) \\ \vdots & \vdots & \ddots & \vdots \\ A_{2070}(i, 1) & A_{2070}(i, 2) & \dots & A_{2070}(i, 20) \\ \vdots & \vdots & \ddots & \vdots \\ A_{2070}(101, 1) & A_{2070}(101, 2) & \dots & A_{2070}(101, 20) \end{pmatrix}$$

Il a donc été calculé le coefficient d'évolution linéaire pour chaque département et chaque classe d'âge. Ce dernier s'obtient en réalisant :

$$D = \frac{A_{2070} - A_{2018}}{2070 - 2018}$$

tel que D la matrice (101×20) et $D(i, j)$ indique la variation annuelle moyenne pour le département i et la tranche d'âge j entre les années 2018 et 2070.

Finalement il est possible d'obtenir la structure d'âge par année et par département à n'importe quelle année de projection k en effectuant le calcul suivant :

$$A_k(i, j) = A_{2018}(i, j) + (k - 2018) \cdot D(i, j)$$

Enfin, les tranches d'âges ont été agrégées pour les faire concorder avec celles de la base carroyée (2.1.2). Comme pour l'évolution de l'effectif de la population, on pose l'hypothèse que chaque pixel suit les tendances du département auquel il est associé. Cette évolution des tranches d'âges a été appliquée à la base carroyée en multipliant le nombre d'individus total du pixel par les proportions obtenues. Une vérification a été appliquée si les valeurs d'individus par tranche d'âge projetées sont positives. Si certaines valeurs sont négatives, une réallocation est effectuée pour s'assurer que toutes les valeurs soient positives.

Suite au développement de la méthodologie de projection démographique, il s'agira de vérifier la précision des résultats.

3.1.3 Cohérence des résultats

En agrégeant les résultats projetés issus des méthodologies énoncées précédemment on obtient la table suivante pour l'année 2020 (début de projection de l'étude) et l'année

ultime de projection, 2050. Une comparaison est alors réalisée avec les chiffres réels de l'INSEE.

Année	Tranche d'âge	Pop. INSEE	%	Pop. Outil	%	Différence INSEE
2020	0-9	7,8	12 %	8,2	13 %	1,48 %
	10-24	12,2	18 %	10,1	16 %	-2,10 %
	25-64	33,6	50 %	32,1	51 %	1,21 %
	65,+100	13,7	20 %	12,4	20 %	-0,59 %
	Total	67,3	100 %	62,9	100 %	0,00 %
2050	0-9	7,2	10 %	7,6	12 %	1,32 %
	10-24	11,2	16 %	9,5	15 %	-1,45 %
	25-64	31,9	46 %	31,1	48 %	2,27 %
	65,+100	18,9	27 %	16,2	25 %	-2,14 %
	Total	69,2	100 %	64,4	100 %	0,00 %

TABLE 3.1 – Comparaison de la population entre 2020 et 2050 - Données réelles et données estimées (INSEE) - Les populations sont exprimées en millions.

On constate que de légers écarts se créent, notamment en raison des périmètres différents (les statistiques INSEE incluent les départements d'outre-mer et la Corse), et également en raison des chiffres qui diffèrent entre la base carroyée (données estimées) et les données réelles de l'INSEE. Par ailleurs, l'estimation par pixel est probablement susceptible de générer des écarts infimes qui, une fois agrégés deviennent perceptibles. Les proportions par tranche d'âge sont en revanche relativement bien conservées avec un écart maximum de +2,27 %.

Abordons désormais la section relative à l'étude de la dynamique prospective anticipée des polluants.

3.2 Dynamique prospective des polluants

3.2.1 Approches dans la littérature

Les scénarios de changement climatique explorent les conséquences possibles des activités humaines sur le système climatique en fonction de modifications dans les systèmes socio-économiques. Cependant, bien qu'ils soient utiles dans de nombreux cas d'usage, sont-ils pour autant les plus représentatifs du futur de la pollution de l'air en France ?

Les scénarios RCP

Les scénarios *Representative Concentration Pathways* (RCPs) élaborés par le GIEC et utilisés dans son 5ème rapport sur le climat, sont des scénarios prospectifs à proprement parler et sont aujourd'hui les principaux retenus dans la plupart des études. Ils sont nommés en fonction du niveau cible de forçage radiatif à l'horizon 2100, c'est-à-dire en fonction de leur niveau de perturbation de l'équilibre du système climatique global par des facteurs externes.

Les estimations de forçage radiatif sont basées sur le forçage des gaz à effet de serre et d'autres agents permettant de décrire quatre trajectoires (2.6, 4.5, 6.0 et 8.0W/m²). Ces

dernières couvrent ainsi une grande gamme de scénarios projetés jusqu'en 2100 traduisant différentes tendances de niveaux de concentration de gaz à effet de serre qui pourraient se produire dans le futur.

Les quatre RCP sélectionnés ont été considérés comme représentatifs de la littérature et comprennent un scénario d'atténuation conduisant à un niveau de forçage très faible (RCP2.6), deux scénarios de stabilisation moyenne (RCP4.5/RCP6) et un scénario d'émissions très élevées (RCP8.5). Aucune description socio-économique n'est volontairement jointe à ces scénarios, toutefois en voici une traduction :

- RCP 8.5 : Scénario à fortes émissions correspondant à un avenir sans changement de politique pour réduire les émissions, et caractérisé par une forte augmentation des GES conduisant à des concentrations élevées dans l'atmosphère (scénario le plus adverse). C'est ce premier qui a été utilisé dans l'étude de Drif et al.,2020 [4].
- RCP 6.0 : Scénario d'émissions élevées à intermédiaires dans lequel les émissions de GES atteignent un pic vers 2060 puis se réduisent progressivement jusqu'à la fin du siècle.
- RCP 4.5 : Scénario intermédiaire caractérisé par une réduction des émissions ambitieuse. Les émissions de GES augmentent encore légèrement avant de commencer à diminuer en 2040. Malgré son caractère ambitieux, il faut noter que ce scénario est toujours en deçà des objectifs fixés par les accords de Paris malgré une réduction projetée des émissions mondiales de 50 % d'ici 2080.
- RCP 2.6 : Seul scénario du GIEC conforme à la limite de 1,5-2°C fixée par les accords de Paris. Ce scénario RCP ambitieux se traduit par une réduction des émissions des GES atteignant un pic à l'horizon 2020, puis qui déclineraient selon une trajectoire linéaire pour devenir nettes négatives en 2100.

Bien que ces scénarios ne soient pas conçus en première intention pour décrire l'évolution prospective des émissions de polluants atmosphériques qu'il faut distinguer des GES, voici une brève description qualitative des tendances anticipées par scénario :

Composante du scénario	RCP2.6	RCP4.5	RCP6	RCP8.5
Gaz à effet de serre (GES)	Très faible	Atténuation moyenne-faible Niveau de référence très bas	Niveau de référence moyen Atténuation élevée	Niveau de référence élevé
Pollution de l'air	Moyen-Faible	Moyen	Moyen	Moyen-Élevé

TABLE 3.2 – Incidence qualitative des scénarios climatiques RCP sur la pollution de l'air.

Toutefois, ces scénarios ont uniquement été utilisés pour expliquer l'évolution prospective de la concentration d'ozone, ces scénarios étant particulièrement adapté pour traduire d'une hausse globale ou locale de la température moyenne annuelle.

Les scénarios SSP

En plus des scénarios RCP précédemment cités, les « *Shared Socioeconomic Pathways* » (SSP) ont été développés pour modéliser la manière dont les facteurs socio-économiques pourraient évoluer au cours du prochain siècle. Les SSP examinent globalement cinq façons

différentes dont le monde pourrait évoluer en l’absence de politique climatique et comment différents niveaux d’atténuation du changement climatique pourraient être atteints lorsque les objectifs d’atténuation des RCP sont combinés aux SSP. Ces scénarios sont la dernière génération de scénarios créés par le GIEC et sont utilisés dans son 6ème rapport sur le climat. Ci-dessous les 5 scénarios « SSP » et leur signification :

- SSP1 : *Sustainability – Taking the Green Road* (faibles défis à l’atténuation et à l’adaptation)
- SSP2 : *Middle of the Road* (défis moyens en matière d’atténuation et d’adaptation)
- SSP3 : *Regional Rivalry – A Rocky Road* (défis élevés en matière d’atténuation et d’adaptation)
- SSP4 : *Inequality – A Road Divided* (Faibles défis pour l’atténuation, défis élevés pour l’adaptation)
- SSP5 : *Fossil-fueled Development – Taking the Highway* (Défis élevés pour l’atténuation, faibles défis pour l’adaptation)

Ces deux types de scénarios distincts développés par le GIEC sont cependant complémentaires. En effet, les RCP fixent des trajectoires pour les concentrations de gaz à effet de serre et de ce fait, fixent l’ampleur du réchauffement climatique, alors que les SSP définissent le cadre dans lequel les réductions d’émissions seront ou ne se seront pas réalisées. Les données d’élévation de la température selon ces différents scénarios climatiques ont été récupérées via la plateforme « *Climate Change Knowledge Portal for Development Practitioners and Policy Makers* ». Ce portail développé par *World Bank Group* donne la possibilité d’extraire des données de températures annuelles prospectives par pays et selon les différents types de scénarios introduits ci-dessus. Les températures annuelles médianes en France issues d’un ensemble de modèles pour les années 2015 à 2100 selon différents scénarios ont donc été extraites. Ces modélisations (CMIP5 et CMIP6¹) sont représentées dans la figure suivante :

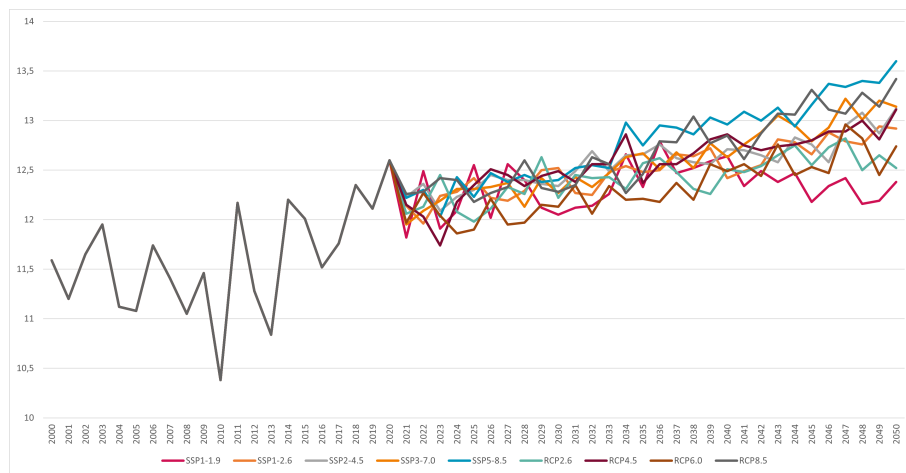


FIGURE 3.2 – Projection de la température annuelle moyenne en France à horizon 2050 selon les scénarios RCP (CMIP5) et SSP (CMIP6) - Valeur médiane issue d’un ensemble de modèles (*World Bank Group, Climate Change Knowledge Portal*).

1. CMIP5 : Coupled Model Intercomparison Project Phase 5, initiative d’intercomparaison des modèles climatiques pour comprendre les changements climatiques. CMIP6 étant la continuation de l’initiative CMIP5 pour améliorer les modèles et les projections climatiques.

Les scénarios GEA

Poursuivons sur la présentation des scénarios « *Global Energy Access* » (GEA). En effet comme mentionné dans les travaux de *Drif et al., 2020 [4]*, les émissions de polluants atmosphériques en Europe peuvent suivre les scénarios d'émission GEA, qui sont plus adaptés aux études sur la qualité de l'air et comprennent une représentation de l'ensemble de la législation actuelle sur la qualité de l'air en Europe (Riahi et al., 2012 [15]). Ce sont donc ces scénarios qui ont été retenus pour la récupération de données projetées des émissions des polluants suivants : $PM_{2.5}$, PM_{10} , et NO_2 .

Les scénarios GEA comprennent essentiellement un seul scénario normatif de la transition vers un système énergétique durable. Dans ce scénario, des voies alternatives sont développées, qui décrivent les transformations vers des objectifs normatifs pour l'accès à l'énergie, les impacts environnementaux de la conversion et de l'utilisation de l'énergie, et la sécurité énergétique.

Par ailleurs, toutes les trajectoires modélisées stabilisent l'augmentation de la température moyenne mondiale à 2°C au-dessus des niveaux pré-industriels.

Les scénarios GEA se déclinent au travers de 3 composantes :

1. Le niveau de demande énergétique :
 - *GEA-Efficiency* (Faible demande)
 - *GEA-Supply* (Forte demande)
 - *GEA-Mix* (Demande intermédiaire)
 - *GEA-Counterfactual* (Aucune action réalisée)
2. Les carburants et technologies de transport majoritaires :
 - Transports conventionnels (Carburants liquides)
 - Transports avancés (Électrique, hydrogène)
3. La diversité du portefeuille d'options du côté de l'offre énergétique :
 - Portefeuille complet (Toutes les options)
 - Portefeuille restreint (Exclusion appliquées, exemple : pas de nucléaire.)

Il existe comme pour les types de scénarios précédemment cités, une base de données en libre accès visant à explorer les résultats issus des hypothèses des voies de transformation empruntées par les scénarios GEA (*Public GEA Scenario Database, Science for Global Insight*). Ces résultats sous-tendent la grande majorité des travaux compris au sein du chapitre 17 de l'étude *Global Energy Assessment Toward a Sustainable Future, 2012 [34]*.

Cette base en ligne met à disposition les données d'émissions projetées de différents polluants à horizon 2100 en Europe. Ont donc été utilisés les parts attribuables moyennes historiques de la France dans les émissions européennes énoncées en 2.1.3, afin de ne conserver que les émissions prospectives propres au territoire français.

3.2.2 Approche par régression historique

Afin de vérifier la cohérence des scénarios GEA retenus, une régression des données historiques d'émission a été effectuée. On appelle régression la méthode statistique utilisée pour modéliser la relation entre une variable dépendante (ou réponse) et une ou plusieurs

variables indépendantes (ou prédicteurs). L'objectif principal de la régression dans ce cas d'étude est de comprendre et de quantifier la relation historique liant les années et les niveaux d'émissions afin de prédire les émissions futures.

Pour ce faire, la régression exponentielle a été retenue étant particulièrement appropriée notamment pour ses propriétés de :

- **Positivité** : La fonction exponentielle est toujours strictement positive. Cela signifie que pour toute valeur réelle de la variable indépendante, la valeur de la fonction exponentielle est positive. Mathématiquement, cela s'exprime comme suit : $f(x) = e^x > 0$ pour tout $x \in \mathbb{R}$
- **Décroissance** (pour les exponentielles de base e^{-x}) : La fonction exponentielle e^{-x} est strictement décroissante. Cela signifie que si $x_1 < x_2$, alors $e^{-x_1} > e^{-x_2}$. En d'autres termes, lorsque la variable indépendante augmente, la valeur de la fonction exponentielle e^{-x} diminue.

Ce type de régression est alors particulièrement adapté aux situations où les émissions historiques diminuent, ce qui comme abordé en 1.2.4 se produit historiquement en raison des mesures de réduction des émissions ou des réglementations environnementales instaurées.

Rappelons la forme générale de l'équation de régression exponentielle :

$$y = a \cdot e^{bx}$$

où :

- y représente la variable dépendante, dans ce cas les émissions de polluants,
- x est la variable indépendante, généralement le temps (années dans ce contexte),
- a est le coefficient d'échelle, déterminant l'ordonnée à l'origine de la courbe exponentielle
- b est le coefficient d'indice, contrôlant la pente de la courbe exponentielle. b est ici strictement négatif, les émissions étant en baisse au cours du temps.

Par ailleurs, l'équation de minimisation du coefficient de détermination r^2 dans le cas d'une régression exponentielle est basée sur la différence entre les valeurs observées y_i et les valeurs prédites par le modèle exponentiel \hat{y}_i . Ainsi, en calculant le coefficient de détermination r^2 , on obtient la mesure de la proportion de la variance totale de la variable dépendante y qui est expliquée par le modèle.

On rappelle l'équation générale du coefficient de détermination (r^2) dans une régression exponentielle :

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

- n est le nombre d'observations,
- y_i sont les valeurs observées de la variable dépendante,
- \hat{y}_i sont les valeurs prédites par le modèle exponentiel pour chaque observation,
- \bar{y} est la moyenne des valeurs observées de la variable dépendante.

L'objectif de la régression exponentielle est de trouver les valeurs optimales des coefficients

(a et b) dans l'équation $y = a \cdot e^{bx}$ de manière à maximiser la valeur de r^2 , c'est-à-dire à minimiser la somme des carrés des résidus. Plusieurs régressions ont été effectuées avec différents niveaux de profondeurs historiques. Ont été retenus les profondeurs qui maximisaient r^2 . Les résultats des régressions sont les suivants :

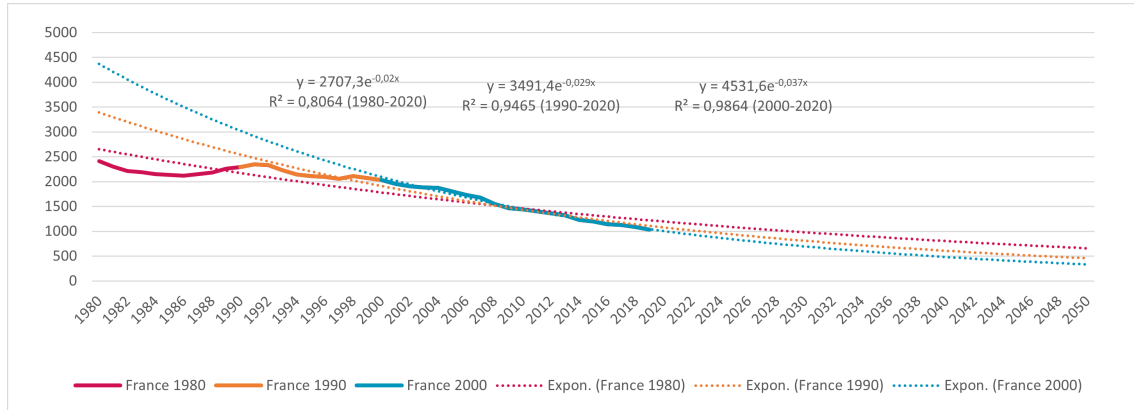


FIGURE 3.3 – Régressions exponentielles des données historiques d'émissions de NO_x selon différentes profondeurs d'historique - Les profondeurs historiques ici étudiées correspondent aux intervalles 1980-2020, 1990-2020, 2000-2020.

3.2.3 Approches réglementaires

Finalement on peut introduire d'autres types de scénarios qui sont produits par les autorités administratives publiques que l'on appellera scénarios réglementaires. Dans un premier temps on présentera les scénarios d'émissions produits par l'INERIS. Ceux-ci se présentent ainsi :

- Scénario A : les émissions des états membres de l'Union européenne en 2030 sont similaires à celles de 2005.
- Scénario B : les États membres appliquent en 2030 la réglementation en vigueur actuellement.
- Scénario C : la France respecte en 2030 les objectifs proposés par la Commission européenne pour la révision de la directive NEC, tandis que les 27 autres Etats membres restent aux niveaux d'émissions correspondant à la réglementation en vigueur.
- Scénario D : tous les Etats membres respectent, en 2030, les nouveaux objectifs NEC.

Ces scénarios se traduisent par les niveaux d'émissions de $PM_{2.5}$ et de NO_x suivants :

Polluant	Scé. A	Scé. B	Scé. C	Scé. D
$PM_{2.5}$	271	169	140,7	140,7
NO_x	1 351,4	440,9	401,3	401,3

TABLE 3.3 – Niveaux d'émissions par polluant en France selon les 4 scénarios de l'INERIS en (Gg=1 000 tonnes) (*Rapport du Sénat, 2014 [5]*)

Enfin on retrouve les scénarios des politiques publiques de réduction de la pollution de l'air du gouvernement français. Ces derniers sont construits comme un pourcentage de baisse des émissions par rapport à l'année de référence, 2005.

Polluant	À horizon 2020	À horizon 2030
SO_2	-55 %	-77 %
NO_x	-50 %	-69 %
$COVNM$	-43 %	-52 %
NH_3	-4 %	-13 %
$PM_{2.5}$	-27 %	-57 %

TABLE 3.4 – Objectifs de réduction d'émissions par polluant fixés pour la France (exprimés en % par rapport à 2005) (*Santé Publique France, 2023 [6]*)

La totalité de ces scénarios ont été extraits ou créé pour chaque polluant. Les scénarios d'émissions précédemment énoncés sont résumés dans la figure suivante sur l'horizon temporel 2020-2050 pour le polluant NO_x :



FIGURE 3.4 – Données d'émissions de NO_x historiques et projetées selon les différents scénarios présentés en 3.2.2 et en 3.2.3. On constate que les émissions sont à tendance baissière, excepté pour le scénario réglementaire D qui n'est pas réaliste.

3.2.4 Traduction des émissions en immissions aux stations

Sur le fondement de la relation supposée linéaire proposée en 1.2.5 entre les données d'émission et les données d'immission pour le cas des polluants NO_x , PM_{10} , $PM_{2.5}$, et, entre les données annuelles de température et les données de concentration d' O_3 , cette sous-section traite de la méthodologie adoptée pour faire évoluer les concentrations aux stations du réseau de mesure en France. Cette méthodologie est synthétisée dans la figure suivante. Elle sera davantage expliquée ensuite.

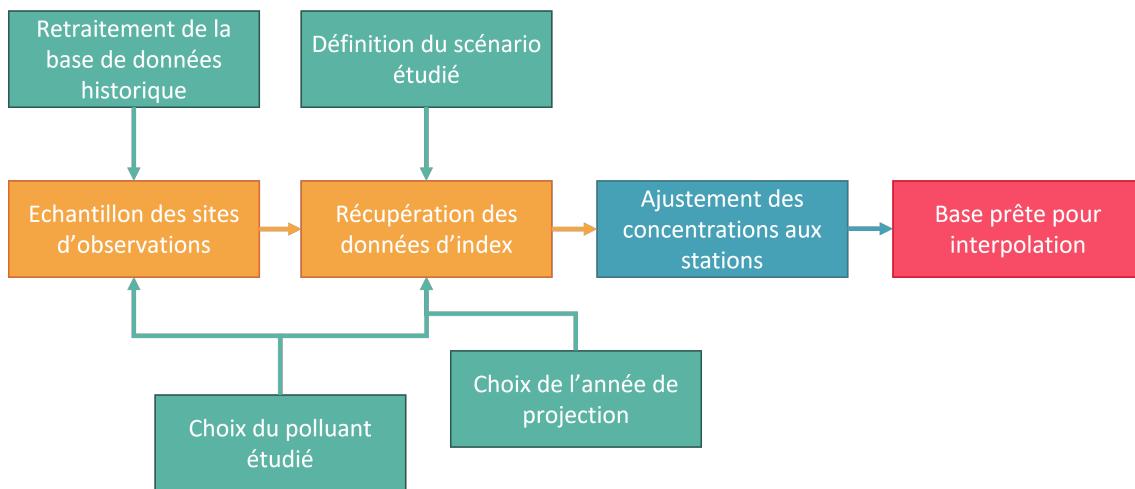


FIGURE 3.5 – Synthèse schématisée de la méthodologie adoptée pour le rapatriement des indices prospectifs et leur traduction en concentration aux stations.

La première étape constitue la récupération et préparation de la base des stations de mesures. Dans la *Figure 3.5*, cela correspond à la case « Échantillon des sites d'observations ». Il s'agit des mêmes retraitements de la base historique qu'effectués lors des *Travaux préliminaires* en 2.4.2. On obtient donc pour un polluant donné, les mesures moyennées des stations de relevé de concentration propres au polluant étudié.

Ensuite, on définit l'année de projection et le scénario que l'on souhaite analyser pour le polluant désigné. Par un jeu de tables préparées en amont, il est alors possible d'obtenir les valeurs d'émission (ou de température pour le cas de l' $O_3(SOMO35)$) à l'année de projection choisie et, sur l'historique 2013-2022.

On parlera alors d'index lors qu'on désignera le résultat de l'opération suivante :

$$I(p, n) = \frac{E(p, n)}{E(p, 0)}$$

tel que :

- p un polluant parmi NO_x , PM_{10} , $PM_{2.5}$.
- n une année de projection.
- $E(p, n)$ représente la valeur d'émission pour le polluant p et l'année n .
- $E(p, 0)$ représente la valeur moyenne des émissions historiques (2013-2022) pour le polluant p .
- $I(p, n)$ représente la valeur de l'index.

ou, pour le cas du *SOMO35* :

$$I(SOMO35, n) = \frac{T(SOMO35, n)}{T(SOMO35, 0)}$$

tel que :

- n une année de projection.
- $T(SOMO35, n)$ représente la valeur de la température annuelle en France à l'année n .
- $T(SOMO35, 0)$ représente la valeur moyenne des températures annuelles historiques en France (2013-2022).
- $I(SOMO35, n)$ représente la valeur de l'index.

Cet index est en fait le coefficient multiplicateur appliqué aux valeurs des concentrations des stations de la base historique des stations de mesures.

Soit C_p le vecteur des mesures de concentration historiques des stations de relevés pour le polluant p . Lorsqu'on multiplie ce vecteur par l'index $I(p, n)$ pour l'année n , on obtient le vecteur $C_p \times I(p, n)$, le vecteur des mesures de concentrations projetées à l'année n pour le polluant p .

De cette manière on fait évoluer linéairement les concentrations (données d'immission) proportionnellement à l'évolution de la variable exogène explicative retenue pour le polluant. Le résultat constitue la base des concentrations pour un polluant et une année donnée selon un scénario prospectif défini. Il ne reste alors qu'à utiliser le modèle d'interpolation spatiale optimal du polluant (2.6) appliqué à $C_p \times I(p, n)$ pour obtenir une valeur de la variable régionalisée en tout point du *raster* de la base carroyée.

3.3 Forces, limites et améliorations de la méthodologie retenue

L'approche prospective tire sa consistance des sources de données utilisées et des sources littéraires citées pour justifier la démarche appliquée. En effet, les données prospectives sont issues de scénarios validés par la communauté scientifique climatique. Par ailleurs les relations émissions-immissions sont tout à fait cohérentes à l'horizon 2050 comme le souligne *Colette et al. 2013b [12]*, *Colette et al. 2013b [11]* et *Lecoeur 2013 [13]* en 1.1.3.

Cependant des limites peuvent être identifiées dans l'approche retenue, à commencer par le format historique du modèle. En effet peut-on réellement raisonner en moyenne historique pour la base de données des concentrations et également pour celles des émissions.

Ensuite ne bénéficiant pas de données par code commune, l'hypothèse retenue de faire évoluer la base démographique carroyée en effectif et en âge selon les tendances administratives départementales est, quant à elle, également réductrice. Plus généralement, il nous est impossible de prévoir à horizon long terme le comportement démographique d'un pixel d'1 km² en France.

Il est également nécessaire de nuancer la relation linéaire qui existe entre émissions et immissions. Les résultats des corrélations, pour la plupart très élevés, ne sont pas tous du même ordre. L'Ozone (SOMO35) dont la valeur de corrélation à la température s'élève à 0,58 (1.2.5) constitue la mesure la plus basse. Cette simplification ayant été choisie en raison de la complexité que représente ce polluant dont la création dépend de nombreux paramètres externes. Les modèles utilisés par les organismes de suivi de la pollution de l'air sont par conséquent beaucoup plus complexes en prenant en compte les émissions de

précurseurs et les grandeurs météorologiques. L'analyse annuelle permet de simplifier la méthodologie et renforce la crédibilité de l'approche d'indexation linéaire à une variable exogène.

Suite à la restitution détaillée du caractère prospectif de l'étude au sein de ce chapitre troisième traitant des différents scénarios, données et méthodologies utilisées, il s'agira dans le dernier chapitre de présenter les résultats et de les analyser.

Chapitre 4

Impact sanitaire prospectif de la pollution de l'air : Résultats

Au sein de ce dernier chapitre, il s'agira de traiter de la convergence de toutes les méthodes et données présentées précédemment, et d'analyser les aboutissants de cette étude en mettant en exergue les forces, limites et éventuelles améliorations à apporter.

4.1 Étude de la matérialité par polluant

Au sein de cette section est réalisée une étude de la matérialité telle que décrite dans la publication de l'EIOPA : « *Application guidance on climate change risk scenarios in the ORSA* » (EIOPA, 2021 [2]). Il s'agit donc de mettre en exergue les risques que peuvent représenter les polluants étudiés dans le cadre de l'exercice de l'activité d'un assureur. Par ailleurs, si le polluant ne constitue pas un risque significatif, il s'agit alors de justifier de ce constat (*section 3.14 de l'avis EIOPA 2021a [3]*). Pour y parvenir, reprenons les étapes présentées en *sous-section 1.1.2*.

4.1.1 Définition du contexte, identification des risques et impacts sur le business

Il a été démontré que la pollution atmosphérique présente des effets néfastes sur la santé, et que nulle personne en France ne peut s'y soustraire. Hausse de la consommation santé, séjours hospitaliers, ou dans certains cas, décès prématurés, sont partie des effets de cette pollution. Pour rappel, la pollution extérieure représente 9 % de la mortalité (toutes causes confondues) en France (*Santé Publique France, 2023 [8]*). Ce phénomène doit donc être particulièrement suivi pour les acteurs commercialisant des contrats d'assurance de personnes. Ci-dessous la liste des branches d'assurance issues de *l'article R321-1 du Code des assurances [35]* impactées par l'incidence de la pollution de l'air :

- n°1 : Accidents
- n°2 : Maladie
- n°20 : Vie-Décès

Par conséquent, tout contrat contenant une garantie liée à l'une de ces trois branches peut

se voir impacté par les risques sous-jacents à la pollution de l'air. Les conséquences ciblent donc le passif d'acteurs tels que les assureurs, les instituts de prévoyance ou encore les mutuelles, via une dérive de leur sinistralité.

Pour un cas simplifié de primes/sinistres constants à chaque exercice projeté dans le temps, cette dérive induisant une hausse de la sinistralité anticipée, entraîne une baisse du résultat de l'entreprise. Il convient alors pour cette dernière, en connaissance du niveau de dérive, de réaliser (ou non) des actions de pilotage diverses.

Parmi ces actions il est possible pour l'acteur d'ajuster sa stratégie de souscription, à titre d'exemple, se désengageant d'une géographie, en retravaillant sa segmentation pour adapter son profil de risque ou encore en modifiant les conditions des garanties de ses contrats particulièrement exposés. Un autre levier d'action consiste en la modification de ses modèles de tarification avec par exemple l'application d'une surprime pour le cas de contrats fortement exposés, ou la mise en place d'une hausse globale de la prime généralisée à l'ensemble du portefeuille en faisant appel à la mutualisation de tous les assurés.

Tous ces enjeux légitiment l'intérêt et la nécessité d'introduire l'étude prospective de la pollution de l'air et ses impacts afin de quantifier le plus fidèlement possible les dérives de sinistralité associées. On notera en particulier l'exposition des contrats type santé-prévoyance individuelle ou collective.

Cependant, étudiés indépendamment, les polluants ne présentent pas les mêmes impacts prospectifs, plus particulièrement pour le cas d'une étude en lien avec le changement climatique.

4.1.2 Évaluation de la relevance du risque par polluant

Dans cette sous-section, il s'agit d'évaluer la pertinence de l'étude de chacun des polluants pris un à un et de considérer la nécessité de les projeter via l'élaboration de scénarios.

Aussi, en raisonnant uniquement en matière de risques climatiques, ne ressort qu'un seul polluant : l'Ozone. En effet, il s'agit du seul polluant représentant un risque directement impacté par le changement climatique, son processus de création étant intimement lié aux périodes de fortes chaleur (1.2.6) amenées à croître en fréquence et en intensité à l'avenir. Il est donc tout à fait fondé de poursuivre vers une analyse quantitative des impacts liés à ce polluant.

Les autres polluant à savoir les particules fines et le dioxyde d'azote peuvent eux-aussi être liés au changement climatique dans une moindre mesure, cependant, comme évoqué en 1.2.4, ces polluants sont davantage assujettis aux effets induits par les politiques réglementaires françaises et européennes. En effet, ces deux familles de polluants sont des phénomènes locaux et en dehors de perturbations naturelles comme les éruptions volcaniques ou le déplacement de particules fines en provenance du Sahara, les émissions internes sont représentatives des concentrations relevées aux stations de mesures du réseau national de surveillance de la qualité de l'air (1.2.4).

Par conséquent, compte tenu des tendances baissières anticipées tous scénarios confondus réalisés en *section 3.2* pour ces deux familles de polluants, il est attendu une baisse de la mortalité. Ces polluants ne génèrent donc peu ou pas de mortalité additionnelle et peuvent donc être écartés de l'étude quantitative.

En effet, la baisse du nombre de cas attribuables entraîne des facteurs négligeables car proches de 0 ou bien négatifs. Cela signifie que les efforts des réglementations mises en places par les autorités publiques françaises et européennes permettent, pour ces polluants, de gagner en espérance de vie.

4.1.3 Conclusion sur matérialité

Dans cette étape de l'évaluation de la matérialité, il s'agit de réaliser l'arbitrage final en ayant recours aux étapes précédentes afin de guider la prise de décision. Pour ce faire, la matérialité des risques est résumée dans la table suivante :

Risque	Polluant	Horizon de temps	Matérialité
Mortalité	$PM_{2.5}$	Court	Modérée
		Moyen	Faible
		Long	Faible
	PM_{10}	Court	Modérée
		Moyen	Faible
		Long	Faible
	NO_2	Court	Modérée
		Moyen	Faible
		Long	Faible
	O_3	Court	Modérée
		Moyen	Modérée
		Long	Forte

TABLE 4.1 – Synthèse de la matérialité du risque mortalité induit par chaque polluant atmosphérique à différents horizons de temps. Les horizons « Court », « Moyen » et « Long » terme font référence aux horizons de temps de l'ORSA climatique présenté en *Figure 1.5*.

Dans un soucis de complétude, l'analyse quantitative a tout de même été réalisée pour l'ensemble des polluants étudiés bien que considérés non matériels.

4.2 Compendium procédural de la modélisation

Au sein de cette section, il s'agit de synthétiser l'entièreté de la méthodologie modélisatrice développée dans le cadre de la quantification de l'incidence de pollution de l'air et le calcul des chocs de mortalité additionnels que cette dernière représente dans le futur. L'ensemble de ces travaux ayant été centralisés dans un unique *Notebook* en langage de programmation Python.

Ainsi, à l'aide des éléments développés dans les chapitres précédents, une projection de la base des stations de mesure retraitées aux valeurs de concentrations historiques moyennées est réalisée à l'horizon souhaité. Les index de la variable exogène explicative sont alors utilisés pour faire évoluer les concentrations d'un polluant aux stations dans le temps. Par exemple, à l'année 2030, pour un scénario climatique donné, la température moyenne annuelle projetée est utilisée pour majorer les valeurs de concentration historiques proportionnellement au ratio $\frac{T_{2030}}{T_{baseline}}$ où $T_{baseline}$ est la température moyenne sur l'historique

2013-2022 (3.2.4). Enfin, l'interpolation spatiale à partir des mesures majorées est réalisée par le biais du modèle de *krigeage* et des hyperparamètres obtenus en *Table 2.6* pour le polluant étudié. Cette première étape fait référence aux cases bleues de la *Figure 4.1*.

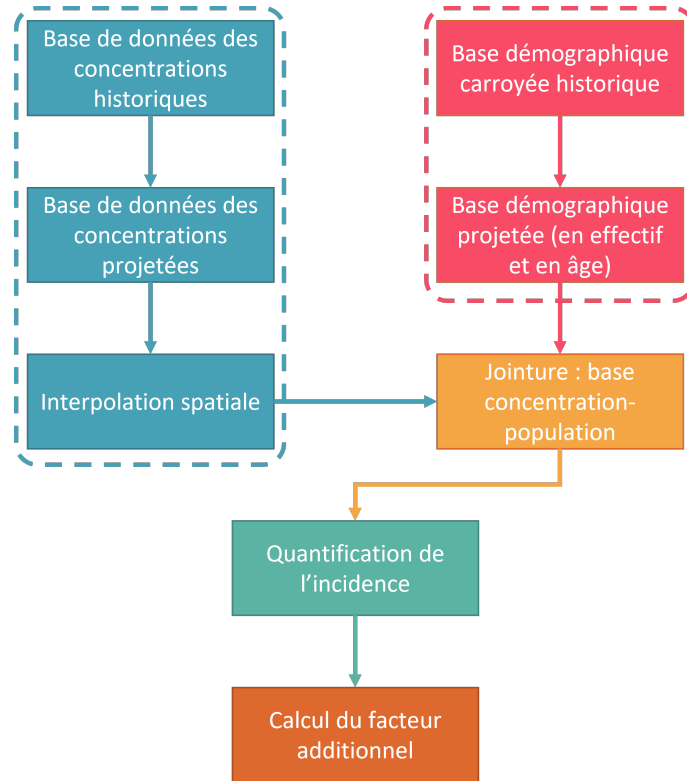


FIGURE 4.1 – *Compendium* procédural schématisé de la modélisation

Pour les étapes représentées par les cases roses de cette figure, est réalisé la projection démographique en âge et en effectif de la base carroyée selon la procédure détaillée en *Section 3.1*.

L'interpolation spatiale et la projection démographique étant produites sur le même *raster*, il est alors possible de joindre les deux bases projetées par le biais de l'identifiant unique de chaque pixel de la base carroyée, la valeur de la variable « $Idcar_{1km}$ ». Ce croisement permet d'obtenir une nouvelle base de données mettant en phase population et niveau d'exposition à un polluant et une année donnée.

À la suite de cette jointure, la quantification du nombre de cas attribuables est effectuée pour chacun des pixels de la base carroyée selon la formule énoncée dans le paragraphe relatif aux *Résultats de l'analyse d'impact* en sous section 1.3.4. Le nombre de cas attribuables de chacun des pixels est alors sommé pour la géographie souhaitée. C'est ce dernier résultat qui est utilisé pour l'obtention *in fine* des facteurs additionnels de mortalité.

En section suivante l'étude quantitative est conduite via l'exécution du code accomplissant chaque étape de la méthodologie développée en *Figure 4.1* et les chocs additionnels de mortalité induits par l'Ozone en France sont calculés.

4.3 Résultats : chocs additionnels prospectifs en mortalité

4.3.1 Notion de choc additionnel

Les résultats de ce mémoire s'inspirent de ceux obtenus dans l'étude de *Drif et al., 2020 [4]*, à savoir la communication de tables de taux de mortalité additionnel majorant annuellement les hypothèses initiales d'un assureur.

En utilisant les valeurs moyennées historiques de *SOMO35* présentées en 2.4.2, et la méthodologie des EQIS réutilisée au sein de l'outil AirQ+, on obtient un nombre moyen historique de décès attribuables à la pollution à l'Ozone. Ce dernier est alors appelé valeur *baseline*, ou valeur de référence.

De même, le nombre de décès à la pollution à l'Ozone est obtenu pour chacun des scénarios conduits impactant les valeurs de *SOMO35* à différents horizons de temps.

On calcule alors les taux de mortalité additionnels (ou facteurs additionnels) en réalisant la différence relative du nombre de décès attribuables à la pollution à l'Ozone à chaque année de projection par rapport à la situation historique. On rapporte ensuite ce premier résultat (le nombre de décès additionnels) à la démographie exposée à l'année de projection considérée.

Soient les variables suivantes :

- F_k : le facteur additif de mortalité à l'année de projection k ,
- D_0 : le nombre de décès historique par an attribuables à la pollution à l'Ozone (valeur *baseline*),
- D_k : le nombre de décès attribuables à la pollution à l'Ozone à l'année de projection k ,
- P_k : la population exposée à l'année de projection k .

On a alors la relation suivante pour l'année k étudiée :

$$F_k = \frac{D_k - D_0}{P_k} \times 100$$

Ces facteurs ont été calculés suivant la méthodologie énoncée pour chaque scénario climatique aux horizons 2030, 2040, et 2050. Les résultats sont présentés dans la section suivante.

4.3.2 Tables de résultats

Résultats granularité globale

Scénario	2030		2040		2050	
RCP2.6	0,002	[0,0007 ; 0,0034]	0,0046	[0,0017 ; 0,0078]	0,0053	[0,0019 ; 0,0089]
RCP4.5	0,0024	[0,0009 ; 0,004]	0,0051	[0,0019 ; 0,0087]	0,0062	[0,0022 ; 0,0105]
RCP6.0	0,0019	[0,0007 ; 0,0032]	0,0046	[0,0017 ; 0,0078]	0,0056	[0,002 ; 0,0095]
RCP8.5	0,0021	[0,0008 ; 0,0036]	0,0051	[0,0019 ; 0,0087]	0,0067	[0,0024 ; 0,0113]
SSP1-1.9	0,0018	[0,0006 ; 0,0031]	0,0048	[0,0017 ; 0,0082]	0,0051	[0,0018 ; 0,0086]
SSP1-2.6	0,0024	[0,0009 ; 0,0041]	0,0045	[0,0016 ; 0,0076]	0,0059	[0,0021 ; 0,01]
SSP2-4.5	0,0022	[0,0008 ; 0,0037]	0,0049	[0,0018 ; 0,0083]	0,0062	[0,0022 ; 0,0105]
SSP3-7.0	0,0021	[0,0008 ; 0,0035]	0,0048	[0,0017 ; 0,0081]	0,0062	[0,0023 ; 0,0105]
SSP5-8.5	0,0023	[0,0008 ; 0,0038]	0,0053	[0,0019 ; 0,009]	0,0069	[0,0025 ; 0,0117]

TABLE 4.2 – Chocs additionnels nationaux en % obtenus pour la pollution à l’Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Toutes classes d’âges confondues. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l’intervalle de confiance à 95 % du facteur en **gras**.

On constate naturellement que plus le scénario climatique est adverse plus son facteur associé est élevé, en particulier pour les scénarios RCP8.5 et SSP5-8.5. Ensuite, on remarque que plus l’horizon de projection est élevé plus le choc associé s’élève lui aussi.

En effet, en faisant référence à la *Figure 3.2*, une hausse de la température moyenne annuelle en France est anticipée quel que soit le scénario à l’étude. Les effets des réductions d’émissions induits par une transition même accélérée vers une société décarbonée ne prendront effet qu’au delà de 2050, année ultime considérée dans cette application. Cependant les chocs obtenus sont relativement bas avec des valeurs de l’ordre de 1×10^{-3} %.

Ci-dessous, voici les résultats obtenus des valeurs de l’indicateur réglementaire *SOMO35* selon les différents scénarios et selon les différents horizons :

Scénario	SOMO35			Majoration de l’historique (en %)		
	2030	2040	2050	2030	2040	2050
HIST	4 301	4 301	4 301	0 %	0 %	0 %
RCP2.6	4 392	4 496	4 499	2 %	5 %	5 %
RCP4.5	4 489	4 622	4 711	4 %	7 %	10 %
RCP6.0	4 359	4 489	4 579	1 %	4 %	6 %
RCP8.5	4 413	4 618	4 823	3 %	7 %	12 %
SSP1-1.9	4 331	4 543	4 449	1 %	6 %	3 %
SSP1-2.6	4 499	4 464	4 643	5 %	4 %	8 %
SSP2-4.5	4 435	4 568	4 715	3 %	6 %	10 %
SSP3-7.0	4 410	4 539	4 722	3 %	6 %	10 %
SSP5-8.5	4 456	4 658	4 888	4 %	8 %	14 %

TABLE 4.3 – Valeurs moyennes nationales de l’indicateur *SOMO35* et pourcentages de majoration de la valeur historique pour différents scénarios climatiques aux horizons 2030, 2040 et 2050 - Modélisation selon les valeurs médianes de la variable exogène.

Ces valeurs de *SOMO35* sont obtenues pour les températures moyennes annuelles suivantes :

Scénario	Températures			Différence à l'historique		
	2030	2040	2050	2030	2040	2050
HIST	11,97	11,97	11,97	0,00	0,00	0,00
RCP2.6	12,22	12,51	12,52	0,25	0,54	0,55
RCP4.5	12,49	12,86	13,11	0,52	0,89	1,14
RCP6.0	12,13	12,49	12,74	0,16	0,52	0,77
RCP8.5	12,28	12,85	13,42	0,31	0,88	1,45
SSP1-1.9	12,05	12,64	12,38	0,08	0,67	0,41
SSP1-2.6	12,52	12,42	12,92	0,55	0,45	0,95
SSP2-4.5	12,34	12,71	13,12	0,37	0,74	1,15
SSP3-7.0	12,27	12,63	13,14	0,30	0,66	1,17
SSP5-8.5	12,40	12,96	13,60	0,43	0,99	1,63

TABLE 4.4 – Températures et différence à l'historique pour différents scénarios climatiques aux années 2030, 2040 et 2050 - Modélisation selon les valeurs médianes de la variable exogène.

Résultats granularité globale - différenciation par classes d'âges

Pour tenter d'approfondir l'analyse et fournir des facteurs plus fins, ces derniers ont été calculés pour ces mêmes scénarios et horizons mais pour différentes classes d'âges (2.1.2).

- [0, 10) pour les personnes âgées de 0 à 9 ans,
- [10, 25) pour les personnes âgées de 10 à 24 ans,
- [25, 65) pour les personnes âgées de 25 à 64 ans,
- [65, +100) pour les personnes âgées de 65 ans et plus.

Les résultats obtenus pour les 3 premières classes d'âges sont presque nuls tous scénarios et tous horizons confondus, il seront donc considérés négligeables. Les résultats sont tout de mêmes communiqués en annexe A.6.

En revanche, pour la dernière classe d'âges qui correspond aux personnes de plus de 65 ans on obtient des facteurs additifs jusqu'à 3 fois plus élevés que ceux présentés précédemment. Voici les résultats obtenus :

Scenario	2030		2040		2050	
RCP2.6	0,003	[0,0011 ; 0,0052]	0,0126	[0,0046 ; 0,0214]	0,0177	[0,0064 ; 0,0299]
RCP4.5	0,0044	[0,0016 ; 0,0074]	0,0145	[0,0052 ; 0,0245]	0,0208	[0,0075 ; 0,0352]
RCP6.0	0,0026	[0,0009 ; 0,0044]	0,0125	[0,0045 ; 0,0212]	0,0188	[0,0068 ; 0,0319]
RCP8.5	0,0033	[0,0012 ; 0,0057]	0,0144	[0,0052 ; 0,0244]	0,0225	[0,0081 ; 0,038]
SSP1-1.9	0,0022	[0,0008 ; 0,0037]	0,0133	[0,0048 ; 0,0225]	0,0169	[0,0061 ; 0,0287]
SSP1-2.6	0,0045	[0,0016 ; 0,0077]	0,0122	[0,0044 ; 0,0206]	0,0198	[0,0072 ; 0,0335]
SSP2-4.5	0,0036	[0,0013 ; 0,0062]	0,0137	[0,0049 ; 0,0231]	0,0209	[0,0075 ; 0,0353]
SSP3-7.0	0,0033	[0,0012 ; 0,0056]	0,0133	[0,0048 ; 0,0224]	0,021	[0,0076 ; 0,0355]
SSP5-8.5	0,0039	[0,0014 ; 0,0067]	0,015	[0,0054 ; 0,0253]	0,0234	[0,0085 ; 0,0396]

TABLE 4.5 – Chocs additionnels nationaux en % obtenus pour la pollution à l’Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d’âge : personnes de plus de 65 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l’intervalle de confiance à 95 % du facteur en **gras**.

En effet, la classe d’âge représentée présente un quotient de mortalité beaucoup plus élevé que pour les autres classes d’âges. De plus, un vieillissement de la population est anticipée en France selon le scénario central de l’INSEE (2.1.2). Naturellement la proportion de décès imputables à une exposition à l’Ozone augmente et cet effet se répercute sur les résultats des facteurs additionnels calculés ci-dessus en *Table 4.5*.

Les mêmes tendances que pour la table tous âges confondus sont observables. En effet le facteur additionnel tend à augmenter à mesure que l’horizon s’allonge et pareillement pour l’augmentation du préjudice imputable aux scénarios climatiques. Les résultats sont pour cette classe d’âge majoritairement de l’ordre de $1 \times 10^{-2}\%$, soit dans certains cas, presque 10 fois plus élevés que pour la table précédente.

De ce constat, il est pertinent d’explorer les répercussions potentielles de cette réalité sur les garanties d’assurance spécifiques. Plus précisément, les produits d’assurance tels que les polices d’assurance-vie et d’assurance décès pourraient être directement touchées. Par conséquent, les primes d’assurance pour ces produits pourraient nécessiter des ajustements afin de refléter plus précisément les risques liés à la mortalité accrue dans cette catégorie d’âge. Une analyse plus approfondie de l’impact spécifique sur chaque type de garantie serait essentielle pour évaluer les implications financières et actuarielles de manière plus détaillée.

Sensibilité à la variable exogène : la température

Afin de confronter les résultats du modèle, une sensibilité a été menée sur la variable exogène responsable des variations des valeurs de l’indicateur SOMO35 à savoir la température prospective. Ont ainsi été utilisées, non plus les données de températures annuelles médianes en France issues de l’agrégation d’un ensemble de modèles de projection comme énoncé en 3.2, mais les données correspondant aux 90ème centile des températures modélisées. Les données de températures sont donc majorées par rapport à la réalisation précédente, et cela permet par conséquent de quantifier l’impact de cette hausse sur les résultats de chocs additionnels nationaux. Une comparaison est alors réalisée avec les résultats obtenus en *Table A.8*.

En reprenant la même démarche que précédemment, on obtient avec les nouvelles données

de températures projetées choquées, la table suivante :

Scénario	2030	2040	2050
RCP2.6	0,0029 [0,0011 ;0,005]	0,0059 [0,0021 ;0,0099]	0,0065 [0,0024 ;0,011]
RCP4.5	0,003 [0,0011 ;0,0051]	0,0061 [0,0022 ;0,0103]	0,0073 [0,0026 ;0,0123]
RCP6.0	0,0024 [0,0009 ;0,0041]	0,0056 [0,002 ;0,0094]	0,0065 [0,0023 ;0,011]
RCP8.5	0,003 [0,0011 ;0,0051]	0,0062 [0,0022 ;0,0105]	0,0077 [0,0028 ;0,013]
SSP1-1.9	0,0032 [0,0012 ;0,0054]	0,0058 [0,0021 ;0,0098]	0,006 [0,0022 ;0,0101]
SSP1-2.6	0,0034 [0,0012 ;0,0057]	0,0059 [0,0021 ;0,01]	0,0073 [0,0026 ;0,0123]
SSP2-4.5	0,0033 [0,0012 ;0,0055]	0,0062 [0,0023 ;0,0105]	0,0074 [0,0027 ;0,0125]
SSP3-7.0	0,0035 [0,0013 ;0,0059]	0,0066 [0,0024 ;0,0112]	0,0079 [0,0028 ;0,0133]
SSP5-8.5	0,0037 [0,0013 ;0,0063]	0,0069 [0,0025 ;0,0117]	0,0088 [0,0032 ;0,0148]

TABLE 4.6 – Chocs additionnels nationaux obtenus pour la pollution à l’Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Toutes classes d’âges confondues. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l’intervalle de confiance à 95 % du facteur en **gras** - Modélisation selon les valeurs choquées de la variable exogène.

Les chocs se voient alors naturellement majorés en comparaison aux résultats de la *Table A.8*. Ainsi pour une augmentation moyenne de 5 % de la valeur de l’indicateur réglementaire *SOMO35* induite par une augmentation moyenne de 6 % de la variable exogène, on observe une augmentation moyenne de 21 % du facteur additionnel tous scénarios et tous horizons confondus.

Ce résultat s’explique une nouvelle fois par le vieillissement de la population française, ce qui la rend d’autant plus exposé à la pollution à l’Ozone et à ses effets sur la mortalité prématurée.

Dans l’utilisation des valeurs de températures choquées, l’horizon court terme (2030) est celui qui fait face à des majorations des facteurs additionnels les plus élevées. En effet les températures semblent augmenter plus rapidement à court terme comme le montre la figure suivante :

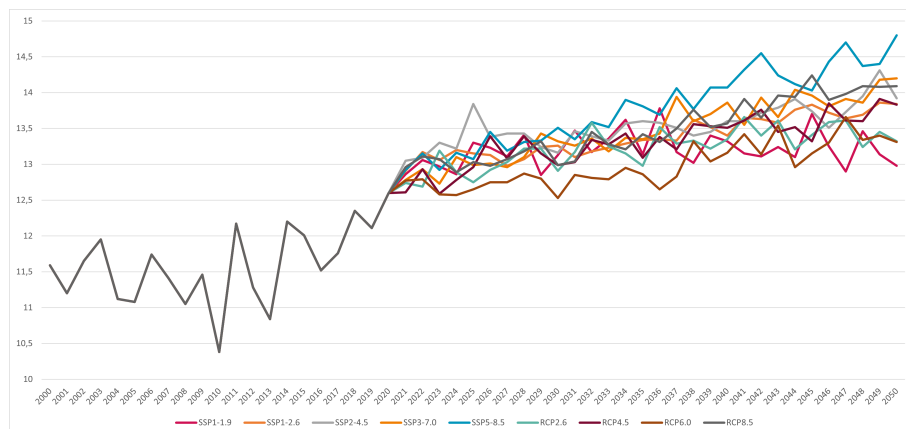


FIGURE 4.2 – Projection de la température annuelle moyenne en France à horizon 2050 selon les scénarios RCP (CMIP5) et SSP (CMIP6) - Valeurs correspondant au 90ème centile issue d'un ensemble de modèles (*World Bank Group, Climate Change Knowledge Portal*).

Cette figure montre également qu'au delà d'une augmentation plus accélérée de la température moyenne annuelle à horizon 2030, tous les scénarios présentent à l'horizon ultime (2050) des températures annuelles moyennes plus élevées qu'en *Figure 3.2*. Les valeurs de SOMO35 dans cette seconde réalisation sont donc plus élevées comme le traduit la table suivante :

Scénario	SOMO35			Majoration à l'historique (en %)		
	2030	2040	2050	2030	2040	2050
HIST	4 301	4 301	4 301	0 %	0 %	0 %
RCP2.6	4 640	4 798	4 787	8 %	12 %	11 %
RCP4.5	4 668	4 855	4 970	9 %	13 %	16 %
RCP6.0	4 503	4 729	4 783	5 %	10 %	11 %
RCP8.5	4 668	4 877	5 064	9 %	13 %	18 %
SSP1-1.9	4 719	4 787	4 665	10 %	11 %	8 %
SSP1-2.6	4 765	4 816	4 974	11 %	12 %	16 %
SSP2-4.5	4 729	4 888	5 003	10 %	14 %	16 %
SSP3-7.0	4 787	4 981	5 103	11 %	16 %	19 %
SSP5-8.5	4 855	5 056	5 319	13 %	18 %	24 %

TABLE 4.7 – Valeurs moyennes nationales de l'indicateur *SOMO35* et pourcentages de majoration de la valeur historique pour différents scénarios climatiques aux horizons 2030, 2040 et 2050 - Modélisation selon les valeurs choquées de la variable exogène.

Le maximum de *SOMO35* atteint pour les valeurs médianes de températures s'élevant à 4 888 (*Table 4.3*) pour le scénario *SSP5 – 8.5* à l'horizon 2050 est largement dépassé pour les valeurs choquées de températures avec une valeur de 5 319 pour le même scénario et le même horizon. Par conséquent les facteurs additionnels se retrouvent tous majorés, ce qui est également le cas pour les résultats de la tranche d'âge des personnes de plus de 65 ans.

4.3.3 Résultats granularité fine : Maille départementale

En capitalisant sur la méthodologie retenue dans le cadre de ce mémoire, il est possible de descendre en granularité afin de fournir non plus des tables de facteurs additionnels nationales mais des tables propres à une géographie donnée. Afin d'illustrer ces propos, un zonier fournissant les chocs additionnels par département a été réalisé et est présenté dans la figure suivante :

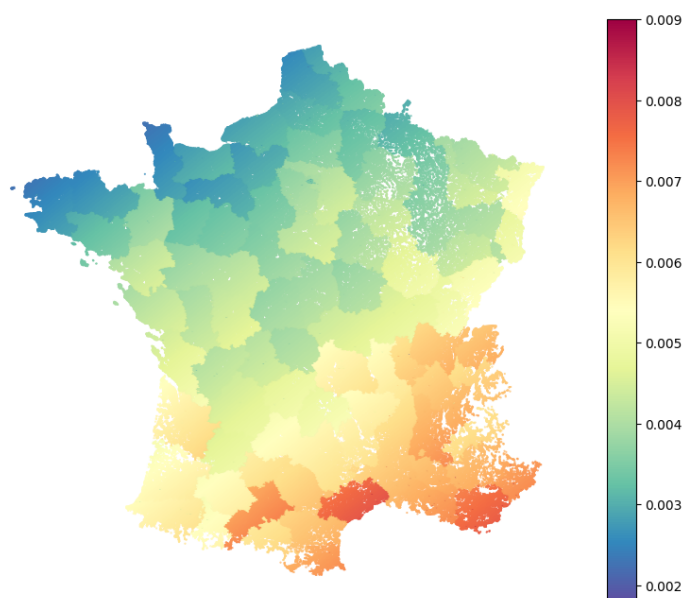


FIGURE 4.3 – Carte choroplèthe : Zonier - Chocs additionnels par département obtenus pour la pollution à l'Ozone selon le scénario climatique SSP5-8.5 pour l'horizon 2050 - Toutes classes d'âges confondues - Modélisation selon les valeurs médianes de la variable exogène.

La représentation choisie en *Figure 4.3* étant le découpage administratif départemental mais il est tout à fait possible de descendre encore en granularité jusqu'au niveau pixel. C'est ce qui est illustré pour l'estimation du nombre de décès attribuables à la pollution à l'ozone pour la ville de Paris selon scénario le climatique SSP5-8.5 à horizon 2050 en *Annexe A.4*.

4.3.4 Comparaison des résultats

Pour conclure cette section, les aboutissants de cette étude sont comparés aux résultats d'études externes réalisant des travaux similaires : les exercices pilote climatique de l'ACPR et un autre mémoire d'actuariat traitant de l'impact du changement climatique sur l'assurance de personnes.

Exercices pilote climatique de l'ACPR

Au sein des exercices pilote climatique, l'ACPR a bénéficié de travaux de quantification des facteurs additionnels qu'induit la pollution atmosphérique sur différents périls de l'as-

surance de personnes de la part d'AON (1.7). Les hypothèses relatives à la mortalité sont résumées dans la table suivante :

Granularité	2023 - RCP4.5			2020 - RCP8.5		
	2024 - 2030	2031-2040	2041-2050	2021 - 2030	2031-2040	2041-2050
France	0,02 %	0,02 %	0,02 %	0,02 %	0,02 %	0,03 %
Bordeaux	0,01 %	0,01 %	0,02 %	0,01 %	0,02 %	0,02 %
Île de France	0,02 %	0,03 %	0,03 %	0,02 %	0,03 %	0,04 %
Lille	0,02 %	0,02 %	0,03 %	0,02 %	0,03 %	0,03 %
Lyon	0,02 %	0,03 %	0,03 %	0,02 %	0,03 %	0,04 %
Marseille	0,02 %	0,03 %	0,03 %	0,02 %	0,03 %	0,04 %
Montpellier	0,02 %	0,02 %	0,03 %	0,02 %	0,03 %	0,03 %
Nantes	0,01 %	0,01 %	0,02 %	0,01 %	0,02 %	0,02 %
Nice	0,01 %	0,01 %	0,02 %	0,01 %	0,02 %	0,02 %
Strasbourg	0,01 %	0,01 %	0,02 %	0,01 %	0,02 %	0,02 %
Toulouse	0,01 %	0,01 %	0,02 %	0,01 %	0,02 %	0,02 %

TABLE 4.8 – Facteurs additionnels de la mortalité communiqués par l'ACPR dans ses exercices pilote climatique de 2020 et 2023.

Il est possible de constater que les résultats de l'ACPR sont beaucoup plus élevés que ceux obtenus au sein de cette étude en *Table A.8*. Ces résultats s'apparentent d'avantage aux facteurs additionnels obtenus pour les personnes âgées de plus de 65 ans en *Table 4.5*.

En effet, dans son étude AON a choisi de considérer l'ensemble des polluants $PM_{2,5}$, PM_{10} , NO_2 et O_3 ce qui n'est pas le cas de ce mémoire compte tenu de l'évaluation de la matérialité réalisée précédemment.

En outre, la méthodologie de calcul des facteurs additionnels agrégeant les polluants n'est pas communiquée et il est impossible de simplement sommer les facteurs additionnels entre eux. En effet, les sommer générerait une erreur d'interprétation car les effets des polluants sur la santé ne sont pas indépendants.

On parle d'effet cocktail lorsque plusieurs polluants atmosphériques sont présents simultanément dans l'environnement. Plutôt que de considérer les effets de chaque polluant individuellement, l'effet cocktail prend en compte les interactions et les synergies entre ces polluants ce qui peut aggraver les effets sanitaires.

Généralement lors de la réalisation d'EQIS, les effets des particules fines et de l'Ozone sont sommés indépendamment. Mais aucune méthodologie n'a été développée pour quantifier convenablement les effets de tous les polluants agrégés.

De plus, les facteurs additionnels obtenus dans le cadre des exercices considèrent à la fois les effets de la pollution à long terme et les effets à court terme ce qui n'est pas le cas de ce mémoire qui se focalise uniquement sur les effets à long terme compte tenu de la résolution temporelle annuelle adoptée dans la modélisation.

Mémoire d'actuariat : « *Climate change impact on Health and protection* », Elsa SUFRIN

Ce mémoire d'actuariat est le premier de son registre à traiter du risque de mortalité lié à la pollution en France. Il présente une méthodologie analogue pour la quantification des

effets de la pollution de l'air sur la mortalité en France et sur un portefeuille. En se basant également sur une relation linéaire liant température et valeur de *SOMO35*, ce mémoire d'actuariat quantifie les facteurs additionnels qui majorent les hypothèses classiques de mortalité de l'assureur AXA. Par la même occasion, ce mémoire produit des résultats pour l'entièreté de la population française.

Les résultats obtenus présentés précédemment sont similaires à ceux obtenus dans le cadre de ce mémoire d'actuariat. Cependant les méthodologies diffèrent et les données imputées également.

Ce mémoire obtient pour une augmentation de 1.5 °C un choc additionnel de 0,0086 % contre 0,0069 % pour le scénario SSP-8.5 non choqué à horizon 2050 qui représente une hausse de 1.6 °C (*Table 4.4*).

Le choc obtenu minore celui du mémoire présenté car la hausse de la température n'est pas traduite par une augmentation uniforme de l'exposition de la population française, hypothèse retenue par « *Climate change impact on Health and protection* » (*Sufrin, 2022 [36]*).

En effet, les valeurs de *SOMO35* évoluent selon la dépendance spatiale captée par le modèle de *krigeage* et l'augmentation n'est pas appliquée uniformément à l'ensemble du territoire mais proportionnellement aux valeurs des stations de mesures historiques. Plus explicitement, les stations dont les valeurs de *SOMO35* sont élevées subiront de plus fortes majorations que les stations à plus faibles valeurs.

La région la plus assujettie à une augmentation des valeurs de *SOMO35* étant le Sud-Est de la France (*Figure 2.21*), et, la population de cette région n'étant pas la plus densément peuplée en comparaison par exemple à l'Île de France, le facteur additionnel sera par conséquent moins important que si l'on a recours à une augmentation uniforme à toute l'échelle du territoire.

La méthodologie présentée précédemment permet donc de plus fidèlement distribuer et restituer les concentrations d'Ozone (et par extension les valeurs de *SOMO35*) en se substituant de l'hypothèse de distribution géographique uniforme de l'exposition, par l'introduction d'une considération de la répartition spatiale des valeurs de concentration basée sur l'expérience passée.

4.4 Applications et réutilisations en assurance

Cette section vise à mettre en avant des utilisations opérationnelles potentielles des travaux réalisés et résultats obtenus. Ces applications constituent par ailleurs une première identification de travaux futurs pouvant faire l'objet de développements complémentaires.

4.4.1 Méthode d'impulsion : table *taylor-made*

Bien que l'utilisation première du *raster* de la base carroyée de l'INSEE ait été de permettre la diffusion de la variable continue régionalisée correspondant à la mesure de concentration d'un polluant donné en s'affranchissant des frontières usuelles telles que les zones administratives, une autre application particulièrement intéressante peut-être explorée.

L'utilisation d'une grille de pixel d'1 km² et l'estimation du phénomène continu de la pollution de l'air en France métropolitaine offrent ainsi la possibilité à n'importe quel assureur le souhaitant, de pouvoir créer sa propre table de facteurs additionnels. En effet, une application de la méthodologie développée dans le cadre de ce mémoire consisterait à obtenir une table *taylor-made* des facteurs additionnels propres à l'exposition d'un portefeuille. En fournissant une base de données géolocalisées d'un portefeuille d'assurés il est alors possible de pouvoir estimer l'incidence de la pollution de l'air prospective de ce dernier.

Cette méthode permet de :

- Renforcer et affiner les chocs standards communiqués par le régulateur en justifiant de la consistance des résultats obtenus. Une évaluation de la matérialité peut par ailleurs argumenter en faveur d'une table de facteurs additionnels minorant celle du régulateur.
- Mettre en lumière une exposition sous-estimée par les facteurs standards communiqués par le régulateur à l'exemple d'un portefeuille concentré dans une zone à forte exposition et/ou présentant une proportion élevée d'assurés considérés comme population à risque.
- Explorer des scénarios climatiques divers non étudiés par le régulateur, si l'on considère qu'aucun des deux scénarios communiqués par l'ACPR n'est susceptible de se réaliser en privilégiant un autre.

L'assureur dispose alors de davantage d'outils pour explorer les conséquences prospectives de la pollution de l'air sur ses sommes assurées.

4.4.2 ORSA climatique

Une autre application des résultats de ce mémoire consiste en la conduction d'un ORSA climatique intégrant les hypothèses de facteurs additionnels obtenus.

La réalisation de l'ORSA climatique permettrait alors de lier les chocs obtenus à l'évaluation de l'impact potentiel du changement climatique sur la solvabilité et la gestion des risques d'un assureur. L'assureur pourrait alors selon les différents scénarios et horizons, évaluer les mesures dans lesquelles les variations climatiques influencent la fréquence et la sévérité des sinistres liés à la mortalité, permettant de l'informer sur sa sensibilité au risque de pollution.

Aussi, en fonction des horizons considérés, les résultats de chocs obtenus peuvent conduire l'assureur à faire évoluer son business-mix ou sa politique de souscription (soit en termes de positionnement géographique, soit en termes de majorations appliquées à certaines tranches d'âge).

4.4.3 Application analogue : facteurs additionnels en santé et en arrêts de travail

L'étude visant à quantifier la mortalité additionnelle résultant de la pollution de l'air ainsi que sa méthodologie présentent un potentiel de réutilisation significatif dans d'autres contextes d'évaluation des risques, pour les garanties santé et arrêts de travail. La méthodologie développée pour estimer les effets de la pollution de l'air sur la mortalité peut être adaptée pour évaluer les facteurs additionnels liés à d'autres indicateurs de santé, tels

que les taux de morbidité et la prévalence de maladies spécifiques. De plus, cette approche pourrait être appliquée à la modélisation des conséquences de la pollution de l'air sur les taux d'arrêt de travail, en identifiant les groupes de population les plus touchés. Une nouvelle comparaison avec les hypothèses des exercices pilote climatique serait également pertinent.

La conduite d'évaluations similaires des facteurs additionnels en santé et en arrêt de travail par une réutilisation intelligente de l'approche présentée dans ce mémoire pourrait contribuer à une meilleure compréhension des conséquences et impacts de la pollution atmosphérique historique et prospective en assurances de personnes.

4.4.4 Autres

Enfin, la méthodologie développée dans le cadre de cette étude pourrait tout à fait être appliquée à d'autres usages que la pollution atmosphérique. En effet, le principe de diffusion d'une variable continue par le recours à des techniques d'interpolation spatiale, qu'elles soient déterministes ou stochastiques, pourrait être réutilisé par exemple pour :

- La diffusion de phénomènes météorologiques : application à l'assurance paramétrique,
- L'étude de la structure spatiale de la sinistralité d'un portefeuille : construction de zoniers et tarification géographique,
- L'estimation des dégâts survenus dans des zones non observées pour la survenance d'aléas types CATNAT (catastrophe naturelles).

Finalement, l'utilisation d'une représentation gridée permet de s'affranchir des mailles usuelles (généralement les frontières administratives) afin d'estimer plus finement la variable cible et *in fine*, le facteur de risque. Cette méthodologie permet notamment d'éviter les disparités géographiques et les incohérences parfois générées aux zones frontalières d'un zonier par exemple.

Conclusion

Étant donné l'étude conduite, il s'agit de rappeler les deux objectifs essentiels du présent mémoire. D'une part, il s'est attaché à explorer les défis contemporains liés à la pollution atmosphérique prospective et à comprendre ses retombées sur la qualité de l'environnement de vie des assurés et par extension sur le secteur assurantiel. D'autre part, il visait à élaborer une approche innovante à fine granularité afin d'évaluer l'impact de la pollution atmosphérique sur la mortalité, en générant des résultats capables de compléter les hypothèses présentées par l'ACPR au sein de ses exercices pilotes climatiques.

Qualifiant la démarche entreprise, le point de départ de cette étude consistait en la mise en place du cadre de recherche. Celui-ci englobe une présentation globale et interdisciplinaire des enjeux liés à la résilience du marché de l'assurance face aux risques climatiques émergents, ainsi qu'une présentation de textes référents dans l'étude des risques climatiques ; les exercices pilote climatique et le guide d'application de l'utilisation de scénarios climatiques dans l'ORSA. En outre, une base théorique relative à la pollution atmosphérique ainsi que les méthodologies actuellement en usage pour quantifier l'impact de cette dernière sur la santé sont présentées.

Indispensable pour l'évaluation des risques sanitaires liés à la pollution atmosphérique, la mise en relation du couple population-concentration sur un territoire donné se révèle cruciale. Les données de pollution sont obtenues à partir de l'*Open-data* via la base GEO-D'AIR. Pour rappel, cette base recense les niveaux de concentrations de différents polluants ($PM_{2.5}$, PM_{10} , NO_2 et O_3) aux stations de mesures du réseau national de la surveillance de la qualité de l'air en France. Ces données exhaustives ont fait l'objet d'un processus de retraitement approprié parmi la conversion des valeurs d'ozone troposphérique en valeurs de l'indicateur réglementaire *SOMO35*. Simultanément, pour les données démographiques, une première approche dite d'impulsion est présentée, impliquant l'utilisation de la base carroyée de l'INSEE. Cette source fournit une démographie exhaustive et détaillée à la maille carreau de 1km^2 , permettant de s'affranchir des frontières administratives traditionnelles. Elle fournit également une projection spatiale sous forme de grille, répondant ainsi aux exigences de maillage pour l'estimation de la pollution de l'air.

Pour l'estimation des données de pollution atmosphérique, une seconde méthode novatrice est employée : l'interpolation spatiale. Cette approche, basée sur les points échantillonnés issus des stations de mesure retraitées de la base GEO-D'AIR, permet d'estimer de manière continue la valeur d'une variable, la mesure de concentration, à travers tout l'espace régionalisé. Une automatisation du paramétrage du krigeage a été effectuée à travers une approche d'apprentissage automatique type *Grid Search* pour l'obtention de la configuration optimale de chaque modèle pour chacun des polluants étudiés. Cette automa-

tisation permet d'explorer différentes configurations de paramètres sans avoir besoin d'une expertise approfondie en géostatistique ou de connaissances particulières sur la structure spatiale d'un polluant. Cette démarche améliore l'efficacité et la précision du processus d'estimation, ces derniers étant par ailleurs quantifiés aux travers de différentes métriques permettant de finalement valider les modèles ainsi obtenus.

Le présent mémoire ayant des objectifs de résultats prospectifs, un module projetant le couple démographie-pollution a été réalisé. Dans sa première composante, il projette la distribution démographique en effectif de population et en structure d'âge, en s'appuyant sur les hypothèses départementales issues du scénario central élaboré par l'INSEE. Chaque pixel de la base carroyée évoluant selon les tendances du département auquel il appartient. La seconde composante du module, en se basant sur la littérature et une évaluation causalité-corrélation, intègre des variables exogènes, visant à expliquer les concentrations de polluants mesurées par les stations de surveillance jusqu'à l'horizon 2050. La variation de ces variables exogènes caractérise le risque en faisant évoluer dans le futur les niveaux de concentrations des stations selon différents scénarios climatiques et réglementaires établis. Cette double approche prospective offre ainsi un aperçu dynamique et prévisionnel de l'interaction complexe entre les facteurs démographiques et environnementaux et établit le module exposition de l'étude.

Compte tenu des projections par polluant obtenues, l'analyse de la matérialité des risques est introduite. Consistant en une évaluation systématique de l'incidence potentielle des niveaux projetés de pollution atmosphérique sur les paramètres clés de santé humaine, et sur le constat d'une baisse anticipée du nombre de décès attribuables aux polluants $PM_{2.5}$, PM_{10} et NO_2 , ces derniers ont été écartés de l'étude quantitative.

Le calcul des impacts supplémentaires sur la mortalité a été réalisé spécifiquement pour l'ozone troposphérique, étant donné que ses facteurs de risque sont susceptibles d'augmenter en réponse à l'élévation de la température globale en France, à laquelle ce polluant est directement lié. Ces facteurs additionnels ont été déterminés en fonction de différentes tranches d'âge, de variations géographiques et de scénarios climatiques à des échéances temporelles de 2030, 2040 et 2050. Les résultats centraux, toutes classes d'âges confondues, obtenus sous-estiment les hypothèses des exercices pilotes climatiques, à l'exception du groupe d'âge des personnes âgées de plus de 65 ans. Ce résultat met en évidence la résilience imposée par le régulateur aux acteurs de la place à travers les hypothèses de l'exercice pilote. Par ailleurs, en exploitant la haute résolution de cette étude, une segmentation par département a été établie, permettant de capturer les variations spatiales fines des concentrations de pollution atmosphérique à travers le territoire français.

En somme, ce mémoire propose une méthodologie novatrice, exploitant des méthodes d'impulsion susceptibles d'approfondir la compréhension des risques futurs associés à la pollution atmosphérique. Cependant, il convient de remettre les travaux effectués en perspective avec les limites qu'ils présentent. En effet, l'estimation de la pollution en tout point du territoire est conditionnée par la répartition des points géographiques échantillonnés, à savoir les stations. De plus, les hypothèses relatives à l'évolution temporelle de la base démographique sont fortes. En effet, il est absolument impossible de prédire le comportement de chaque carreau de 1 km^2 en France, les évolutions démographiques nationales faisant intervenir davantage de composantes telles que la natalité ou les mouvements de population. Enfin, il aurait été tout à fait pertinent d'approfondir l'étude à travers une

exploration plus large de l'espace des hyperparamètres testés pour les modèles de krigeage, ou encore en tenant compte des effets croisés que peut induire une exposition simultanée à plusieurs polluants.

Enfin, d'autres résultats sont envisageables en réutilisant la méthodologie proposée dans le présent mémoire, tels que l'obtention de chocs additionnels en santé et en arrêts de travail, ou encore l'application de la quantification des chocs additionnels à un portefeuille géolocalisé. Plus généralement l'approche d'interpolation spatiale peut être extrapolée à la quantification de facteurs de risques à continuité spatiale autres que la pollution de l'air, comme peuvent être les variables météorologiques.

Annexe A

Annexes

A.1 Seuils réglementaires par polluant et par origine normative.

Tableau des normes pour le dioxyde d'azote (NO_2)	Mesure de concentration	Origine de la norme	Unité de mesure
Objectif de qualité	$40 \mu g/m^3$	FR	en moyenne annuelle
Valeurs limites pour la protection de la santé humaine	$200 \mu g/m^3$	UE	en moyenne horaire à ne pas dépasser plus de 18 heures par an
	$40 \mu g/m^3$	UE	en moyenne annuelle
Seuil d'information et recommandation	$200 \mu g/m^3$	FR	en moyenne horaire
Seuils d'alerte	$400 \mu g/m^3$	UE	moyenne horaire pendant 3 heures consécutives

TABLE A.1 – Normes de concentration pour NO_2

Tableau des normes pour les particules (PM_{10})	Mesure de concentration	Origine de la norme	Unité de mesure
Objectif de qualité	$30 \mu g/m^3$	FR	en moyenne annuelle
Valeurs limites pour la protection de la santé humaine	$50 \mu g/m^3$	UE	en moyenne journalière à ne pas dépasser plus de 35 jours par an
	$40 \mu g/m^3$	UE	en moyenne annuelle
Seuil d'information et recommandation	$50 \mu g/m^3$	FR	en moyenne journalière
Seuils d'alerte	$80 \mu g/m^3$	UE	en moyenne journalière

TABLE A.2 – Normes de concentration pour PM_{10}

Tableau des normes pour les particules ($PM_{2.5}$)	Mesure de concentration	Origine de la norme	Unité de mesure
Objectif de qualité	$10 \mu g/m^3$	FR	en moyenne annuelle
Valeur cible pour la protection de la santé humaine	$20 \mu g/m^3$	FR	en moyenne annuelle
Valeur limite 2015 pour la protection de la santé humaine	$25 \mu g/m^3$	UE	en moyenne annuelle

TABLE A.3 – Normes de concentration pour $PM_{2.5}$

Tableau des normes pour l’ozone (O_3)	Mesure de concentration	Origine de la norme	Unité de mesure
Objectif de qualité	$120 \mu g/m^3$	-	pour le maximum journalier de la moyenne sur 8 heures par an
Valeur cible pour la protection de la santé humaine	$120 \mu g/m^3$	-	maximum journalier de la moyenne sur 8 heures à ne pas dépasser plus de 25 jours par an (en moyenne sur 3 ans)
Seuil d’information et de recommandation	$180 \mu g/m^3$	-	en moyenne horaire
Seuils d’alerte	$240 \mu g/m^3$	-	en moyenne horaire
Seuils d’alerte nécessitant la mise en oeuvre progressive de mesures d’urgence	1 ^{er} seuil : $240 \mu g/m^3$	-	moyenne horaire pendant 3 heures consécutives
	2 ^e seuil : $300 \mu g/m^3$	-	moyenne horaire pendant 3 heures consécutives
	3 ^e seuil : $360 \mu g/m^3$	-	en moyenne horaire

TABLE A.4 – Normes de concentration pour O_3

A.2 Indicateurs de santé liés à la morbidité et la mortalité.

Indicateurs sanitaires	Codes CIM-10	Indicateurs d'exposition									
		Pollution de l'air ambiant								Pollution de l'air intérieur	
		Long terme					Court terme			Long terme	
		PM _{2.5}	PM ₁₀	NO ₂	O ₃	CS	PM _{2.5}	PM ₁₀	NO ₂	O ₃	Combustible solide
Mortalité, toutes causes (naturelles)		x		x		x		x		x	
Mortalité, affections aiguës des voies respiratoires inférieures chez les enfants de 0 à 4 ans	J10-J22	x									x
Mortalité, BPCO (adultes 30+)	J40-J44, J47	x									
Mortalité, BPCO (femmes 30+)	J40-J44, J47										x
Mortalité, BPCO (hommes 30+)	J40-J44, J47										x
Mortalité, cardiopathies ischémiques (adultes 25+)	I20-I25	x									
Mortalité, cardiopathies ischémiques (femmes 25+)	I20-I25										x
Mortalité, cardiopathies ischémiques (hommes 25+)	I20-I25										x
Mortalité, cancer du poumon (adultes 30+)	C33-C34, D02.1-D02.2, D38.1	x									
Mortalité, cancer du poumon (femmes 30+)	C33-C34, D02.1-D02.2, D38.1										x
Mortalité, cancer du poumon (hommes 30+)	C33-C34, D02.1-D02.2, D38.1										x
Mortalité, accident vasculaire cérébral (adultes 25+)	I60-I63, I65-I67, I69.0-I69.3	x									
Mortalité, accident vasculaire cérébral (femmes 25+)	I60-I63, I65-I67, I69.0-I69.3										x
Mortalité, accident vasculaire cérébral (hommes 25+)	I60-I63, I65-I67, I69.0-I69.3										x
Mortalité, maladies de l'appareil respiratoire	J00-J99				x					x	
Mortalité, maladies cardiovasculaires	I00-I99									x	
Mortalité post néonatale, toutes causes			x								
Prévalence de la bronchite chez l'enfant			x								
Prévalence des symptômes de bronchite chez les enfants asthmatiques âgés de 5 à 14 ans				x							
Incidence de la bronchite chronique chez l'adulte			x								
Incidence des symptômes de l'asthme chez les enfants asthmatiques							x				
Hospitalisations pour causes cardiovasculaires (y compris AVC)						x					
Hospitalisations pour causes cardiovasculaires (sauf AVC)										x	
Hospitalisations pour causes respiratoires						x		x		x	
Perte de jours de travail, population en âge de travailler uniquement						x					
Jours d'activité restreinte						x					
Jours d'activité restreinte mineure										x	

Acronymes: Classification Internationale des Maladies (CIM), Accident Vasculaire Cérébral (AVC), Broncho Pneumopathie Chronique Obstructive (BPCO), Particules en suspension dans l'air d'un diamètre inférieur à 2,5 micromètres (PM_{2.5}), Particules en suspension dans l'air d'un diamètre inférieur à 10 micromètres (PM₁₀), Dioxyde d'azote (NO₂), Ozone (O₃), Carbone suie (CS)

FIGURE A.1 – Indicateurs de santé liés à la morbidité et la mortalité quantifiables à l'aide d'AirQ+

A.3 Les données de population : Données au carreau de 1 km² en France de l'INSEE

La liste des 31 variables pour chacun des 374 797 carreaux présents en France métropolitaine est présentée ci-dessous :

- Idcar_1km : Identifiant Inspire du carreau de 1 km
- I_est_1km : Vaut 1 si le carreau est imputé par une valeur approchée, 0 ou 2 sinon.
- lcog_geo : Code officiel géographique au 1er janvier 2021 de la ou des commune(s) dans laquelle (lesquelles) se trouve le carreau.
- Ind : Nombre d'individus
- Men : Nombre de ménages
- Men_pauv : Nombre de ménages pauvres
- Men_1ind : Nombre de ménages d'un seul individu
- Men_5ind : Nombre de ménages de 5 individus ou plus
- Men_prop : Nombre de ménages propriétaires
- Men_fmp : Nombre de ménages monoparentaux
- Ind_snv : Somme des niveaux de vie winsorisés des individus
- Men_surf : Somme de la surface des logements du carreau
- Men_coll : Nombre de ménages en logements collectifs
- Men_mais : Nombre de ménages en maison
- Log_av45 : Nombre de logements construits avant 1945
- Log_45/70 : Nombre de logements construits entre 1945 et 1969
- Log_70/90 : Nombre de logements construits entre 1970 et 1989
- Log_ap90 : Nombre de logements construits depuis 1990
- Log_inc : Nombre de logements dont la date de construction est inconnue
- Log_soc : Nombre de logements sociaux
- Ind_0/3 : Nombre d'individus de 0 à 3 ans
- Ind_4/5 : Nombre d'individus de 4 à 5 ans
- Ind_6/10 : Nombre d'individus de 6 à 10 ans
- Ind_11/17 : Nombre d'individus de 11 à 17 ans
- Ind_18/24 : Nombre d'individus de 18 à 24 ans
- Ind_25/39 : Nombre d'individus de 25 à 39 ans
- Ind_40/54 : Nombre d'individus de 40 à 54 ans
- Ind_55/64 : Nombre d'individus de 55 à 64 ans
- Ind_65/79 : Nombre d'individus de 65 à 79 ans
- Ind_80p : Nombre d'individus de 80 ans ou plus
- Ind_inc : Nombre d'individus dont l'âge est inconnu

A.4 Nombre de sites de surveillance de la pollution de fond en France

	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
NO2	274	268	275	267	250	249	244	234	232	230	2523
O3	322	310	309	302	288	282	279	270	272	268	2902
PM10	214	219	211	209	217	220	216	211	211	192	2120
PM2.5	76	84	82	88	95	97	99	108	122	128	979
SOMO35	324	314	316	311	291	289	285	279	280	278	2967

TABLE A.5 – Nombre de sites de surveillance de la pollution de fond en France par polluant atmosphérique pour les années 2013 à 2022 (post-retraitements effectués en 2.4.2)

A.5 Cartes choroplèthes

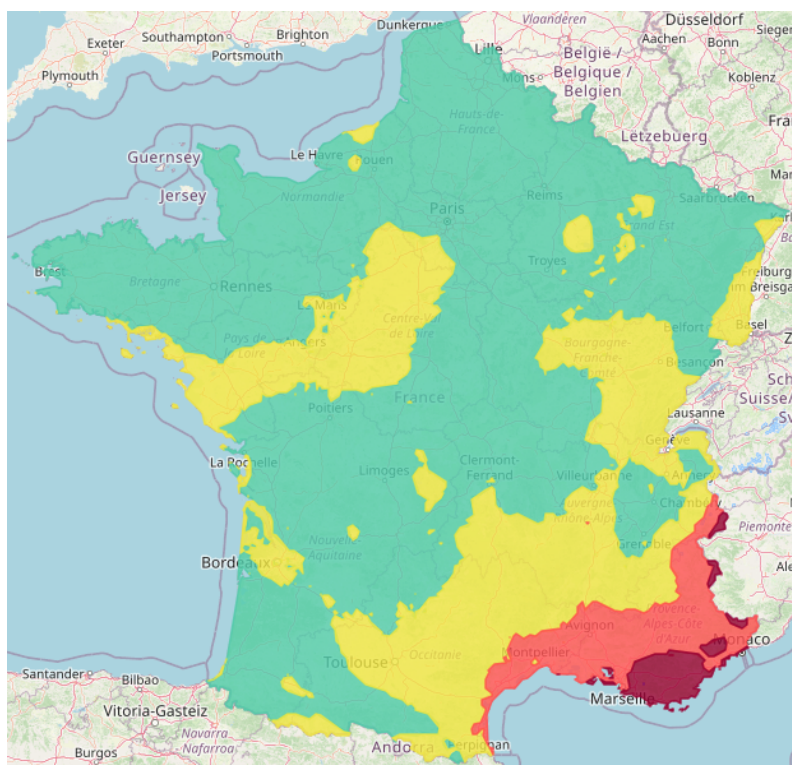


FIGURE A.2 – Carte choroplèthe des valeurs de *SOMO35* en France en 2021, INERIS.

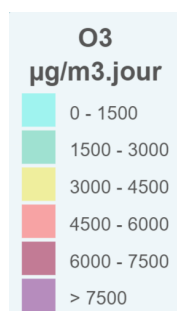


FIGURE A.3 – Légende des valeurs de *SOMO35*, INERIS. Une valeur de *SOMO35* s'élevant à 3 000 est égale $3\,000/365.25 = 8,21$ en concentration moyenne.

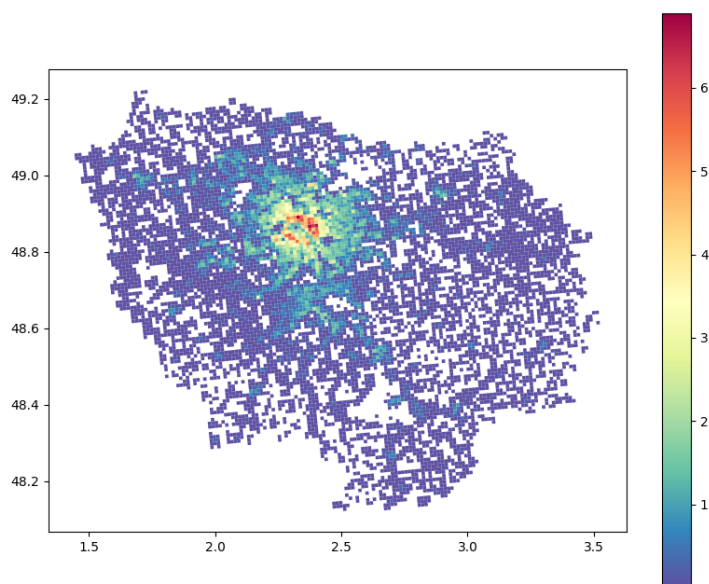


FIGURE A.4 – Carte choroplèthe : Île de France - Nombre de décès attribuables à la pollution à l'Ozone selon le scénario climatique SSP5-8.5 pour l'horizon 2050 - Toutes classes d'âges confondues - Modélisation selon les valeurs médianes de la variable exogène.

A.6 Résultats de facteurs additionnels par tranches d'âges

Scénario	2030	2040	2050
HIST	0 [0;0]	-0,0001 [0;-0,0001]	-0,0001 [-0,0001;-0,0002]
RCP2.6	0 [0;0]	0 [0;0]	-0,0001 [0;-0,0002]
RCP4.5	0 [0;0,0001]	0 [0;0]	-0,0001 [0;-0,0001]
RCP6.0	0 [0;0]	0 [0;0]	-0,0001 [0;-0,0002]
RCP8.5	0 [0;0]	0 [0;0]	-0,0001 [0;-0,0001]
SSP1-1.9	0 [0;0]	0 [0;0]	-0,0001 [0;-0,0002]
SSP1-2.6	0 [0;0,0001]	0 [0;0]	-0,0001 [0;-0,0002]
SSP2-4.5	0 [0;0]	0 [0;0]	-0,0001 [0;-0,0001]
SSP3-7.0	0 [0;0]	0 [0;0]	-0,0001 [0;-0,0001]
SSP5-8.5	0 [0;0,0001]	0 [0;0]	-0,0001 [0;-0,0001]

TABLE A.6 – Chocs additionnels nationaux en % obtenus pour la pollution à l’Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d’âge : 0 - 10 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l’intervalle de confiance à 95 % du facteur en **gras**.

Scénario	2030	2040	2050
HIST	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
RCP2.6	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
RCP4.5	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
RCP6.0	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
RCP8.5	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
SSP1-1.9	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
SSP1-2.6	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
SSP2-4.5	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
SSP3-7.0	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]
SSP5-8.5	-0,0001 [0;-0,0001]	-0,0001 [0;-0,0002]	-0,0002 [-0,0001;-0,0003]

TABLE A.7 – Chocs additionnels nationaux en % obtenus pour la pollution à l’Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d’âge : 10 - 25 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l’intervalle de confiance à 95 % du facteur en **gras**.

Scénario	2030		2040		2050	
HIST	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0002 ; 0,0001]
RCP2.6	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0002 ; 0,0001]
RCP4.5	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0001 ; 0,0002]
RCP6.0	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0001 ; 0,0001]
RCP8.5	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0001 ; 0,0002]
SSP1-1.9	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0002 ; 0,0001]
SSP1-2.6	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0001 ; 0,0001]
SSP2-4.5	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0001 ; 0,0002]
SSP3-7.0	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0001 ; 0,0002]
SSP5-8.5	0,0001	[-0,0001 ; 0,0003]	0	[-0,0001 ; 0,0002]	0	[-0,0001 ; 0,0002]

TABLE A.8 – Chocs additionnels nationaux en % obtenus pour la pollution à l’Ozone selon différents scénarios climatiques pour les années 2030, 2040 et 2050 - Classe d’âge : 10 - 25 ans. Les résultats entre crochets représentent les valeurs aux bornes inférieure et supérieure de l’intervalle de confiance à 95 % du facteur en **gras**.

Bibliographie

- [1] Finalyse. Climate change risks in the orsa.
- [2] European Insurance and Occupational Pensions Authority (EIOPA). Application guidance on running climate change materiality assessment and using climate change scenarios in orsa. Technical report, EIOPA, 2021.
- [3] EIOPA. Opinion on the supervision of the use of climate change risk scenarios in orsa. 2021a.
- [4] Yannick Drif, Palmira Messina, and Pierre Valade. Conséquences du changement climatique sur la pollution de l'air et impact en assurance de personnes. 2020.
- [5] SÉNAT. Rapport fait au nom de la commission d'enquête sur le coût économique et financier de la pollution de l'air, 2015. Enregistré à la Présidence du Sénat le 8 juillet 2015. Dépôt publié au Journal Officiel – Édition des Lois et Décrets du 9 juillet 2015. Rapport fait au nom de la commission d'enquête (1) sur le coût économique et financier de la pollution de l'air, Président M. Jean-François HUSSON, Rapporteure Mme Leila AÏCHI, Sénateurs. Tome I : Rapport.
- [6] Ministère de la Transition écologique. Politiques publiques pour réduire la pollution de l'air, 2023.
- [7] Organisation Mondiale de la Santé (OMS). Pollution atmosphérique. Onglet "Vue d'ensemble".
- [8] Ministère de la Transition énergétique Ministère de la Transition écologique et de la Cohésion des territoires. Pollution de l'air : origines, situation et impacts, 2023. Date de publication le 22 février 2023.
- [9] Emmanuela Gakidou, Ashkan Afshin, Amanuel Alemu Abajobir, Kalkidan Hassen Abate, Cristiana Abbafati, Kaja M Abbas, Foad Abd-Allah, Abdishakur M Abdulle, Semaw Ferede Abera, Victor Aboyans, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016 : a systematic analysis for the global burden of disease study 2016. *The Lancet*, 390(10100) :1345–1422, 2017.
- [10] Autorité de Contrôle Prudentiel et de Résolution (ACPR). Les principaux résultats de l'exercice pilote climatique 2020. *Analyses et synthèses*, (n°122), 2021.
- [11] A Colette, S Schucht, B Bessagnet, et al. Evaluation des stratégies de lutte contre la pollution de l'air à longue distance dans le contexte du changement climatique. *Programme PRIMEQUAL*, 2, 2013.
- [12] Augustin Colette, Bertrand Bessagnet, Robert Vautard, Sophie Szopa, Shilpa Rao, Simone Schucht, Zbigniew Klimont, Laurent Menut, Gaëlle Clain, Frédéric Meleux, et al. European atmosphere in 2050, a regional air quality and climate perspective under cmip5 scenarios. *Atmospheric Chemistry and Physics*, 13(15) :7451–7471, 2013.

- [13] Eve Lecoecur. *Influence de l'évolution climatique sur la qualité de l'air en Europe*. PhD thesis, Université Paris-Est, 2013.
- [14] Victoria N Likhvar, Mathilde Pascal, Konstantinos Markakis, Augustin Colette, Didier Hauglustaine, Myrto Valari, Zbigniew Klimont, Sylvia Medina, and Patrick Kinney. A multi-scale health impact assessment of air pollution over the 21st century. *Science of the Total Environment*, 514 :439–449, 2015.
- [15] Keywan Riahi, Frank Dentener, Dolf Gielen, Arnulf Grubler, Jessica Jewell, Zbigniew Klimont, Volker Krey, DL McCollum, Shonali Pachauri, Shilpa Rao, et al. Energy pathways for sustainable development. 2012.
- [16] Drew Todd Shindell, J-F Lamarque, Michael Schulz, M Flanner, C Jiao, M Chin, PJ Young, Yunha H Lee, L Rotstayn, N Mahowald, et al. Radiative forcing in the accmip historical and future climate simulations. *Atmospheric Chemistry and Physics*, 13(6) :2939–2974, 2013.
- [17] PJ Young, AT Archibald, KW Bowman, J-F Lamarque, V Naik, DS Stevenson, S Tilmes, A Voulgarakis, O Wild, D Bergmann, et al. Pre-industrial to end 21st century projections of tropospheric ozone from the atmospheric chemistry and climate model intercomparison project (accmip). *Atmospheric Chemistry & Physics Discussions*, 12(8), 2012.
- [18] Tim M Butler, ZS Stock, MR Russo, HAC Denier Van Der Gon, and Mark G Lawrence. Megacity ozone air quality under four alternative future scenarios. *Atmospheric Chemistry and Physics*, 12(10) :4413–4428, 2012.
- [19] A Colette, R Koelemeijer, G Mellios, S Schucht, JC Péré, C Kouridis, B Bessagnet, H Eerens, K Van Velze, and L Rouil. Cobenefits of climate and air pollution regulations. *The context of the European Commission Roadmap for moving to a low carbon economy in, 2050*, 2012.
- [20] DA Hauglustaine, F Hourdin, L Jourdain, M-A Filiberti, S Walters, J-F Lamarque, and Elisabeth A Holland. Interactive chemistry in the laboratoire de météorologie dynamique general circulation model : Description and background tropospheric chemistry evaluation. *Journal of geophysical research : Atmospheres*, 109(D4), 2004.
- [21] PREV’AIR. Origine et sources de pollution.
- [22] Organisation mondiale de la Santé (OMS). Communiqué de presse : L’oms publie les estimations nationales de l’exposition à la pollution de l’air et les effets sur la santé. septembre 2016. Genève.
- [23] Organisation mondiale de la Santé (OMS). Communiqué de presse : 7 millions de décès prématurés sont liés à la pollution de l’air chaque année. mars 2014. Genève.
- [24] Wikipédia. Ozone troposphérique. https://fr.wikipedia.org/wiki/Ozone_troposph%C3%A9rique, 2023. Consulté le 17 mai 2023.
- [25] Augustin Colette, Camilla Andersson, Alexander Baklanov, Bertrand Bessagnet, Jørgen Brandt, Jesper H Christensen, Ruth Doherty, Magnuz Engardt, Camilla Geels, Christos Giannakopoulos, et al. Is the ozone climate penalty robust in europe? *Environmental Research Letters*, 10(8) :084015, 2015.
- [26] Eirini Boleti, Christoph Hueglin, Stuart K Grange, André SH Prévôt, and Satoshi Takahama. Temporal and spatial analysis of ozone concentrations in europe based on timescale decomposition and a multi-clustering approach. *Atmospheric Chemistry and Physics*, 20(14) :9051–9066, 2020.

- [27] S Host, T Cardot, A Saunal, V Gherzi, and F Joly. Mortalité attribuable à la pollution atmosphérique en île-de-france. quelle évolution depuis 10 ans et quels bénéfices d’une amélioration de la qualité de l’air dans les territoires. *Observatoire régional de santé Île-de-France*, 2022.
- [28] Commission nationale de l’informatique et des libertés (CNIL). Définition de l’open data.
- [29] Pierre Bosser. Interpolation spatiale. 2011.
- [30] Wikipedia. Spline, 2023.
- [31] E. Real, F. Couvidat, A. Ung, L. Malherbe, B. Raux, A. Gressent, and A. Colette. Historical reconstruction of background air pollution over france for 2000–2015. *Earth System Science Data*, 14(5) :2419–2443, 2022.
- [32] Yann Diserbeau. Proposition d’accélérateurs pour la mise en place d’une couverture indicielle des risques météo-sensibles. 2019.
- [33] WHO Ambient Air Quality Database. Online Database, 2016. Accessed on [Date].
- [34] Keywan Riahi, Frank Dentener, Dolf Gielen, Arnulf Grubler, Jessica Jewell, Zbigniew Klimont, Volker Krey, David McCollum, Shonali Pachauri, Shilpa Rao, and et al. *Energy Pathways for Sustainable Development*, page 1205–1306. Cambridge University Press, 2012.
- [35] Article r321-1 du code des assurances. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032177314/. Consulté en ligne.
- [36] Eisa Sufrin. Climate change impact on health and protection, 03 2022.
- [37] Detlef P Van Vuuren, Jae Edmonds, Mikiko Kainuma, Keywan Riahi, Allison Thomson, Kathy Hibbard, George C Hurtt, Tom Kram, Volker Krey, Jean-Francois Lamarque, et al. The representative concentration pathways : an overview. *Climatic change*, 109 :5–31, 2011.
- [38] EIOPA. Methodological paper on inclusion of climate change in the nat cat calibration. 2021b.
- [39] Keywan Riahi, Shilpa Rao, Volker Krey, Cheolhung Cho, Vadim Chirkov, Guenther Fischer, Georg Kindermann, Nebojsa Nakicenovic, and Peter Rafaj. Rcp 8.5—a scenario of comparatively high greenhouse gas emissions. *Climatic change*, 109 :33–57, 2011.
- [40] Airq+ glossary. <https://www.airqplus.eu/glossary/>, 2021. Version de Février 2021.