

**Master 2 Econométrie et Statistique,
parcours Econométrie Appliquée**

Rapport de stage de fin d'étude

Modélisation du prix de l'Immobilier à partir de données géographiques et structurelles



Noa Le Roux

20 août 2025

Sous la direction de:
Anthony Dupiellet,
Olivier Darne

Remerciements

Je souhaite exprimer ma profonde gratitude aux professeurs et intervenants du Master ECAP (Économétrie, Statistiques) de l'IAE de Nantes (Nantes Université), grâce auxquels j'ai pu acquérir de solides compétences dans de nombreux domaines des statistiques et de l'économétrie, constituant un socle large et structuré de savoir-faire. Leurs enseignements sont précieux, et je suis convaincu que je continuerai à m'y référer tout au long de ma carrière.

Je tiens également à remercier mes camarades de promotion, avec qui nous nous sommes mutuellement soutenus et aidés tout au long de ces deux années. Les échanges et réflexions collectives que nous avons partagés se sont révélés extrêmement enrichissants, permettant d'approfondir et d'affiner la qualité de nos analyses économétriques.

Je remercie également chaleureusement l'équipe ESG du Crédit Mutuel Arkéa, qui m'a réservé un accueil bienveillant et m'a accompagné tout au long de mon stage, en partageant avec moi leur expertise dans le domaine de l'ESG.

Enfin, j'adresse mes remerciements les plus sincères à mon tuteur de stage, Ahtony Dupiellet, pour la qualité de son encadrement durant ces six mois. Il m'a non seulement transmis de nombreuses connaissances sur le risque physique, mais m'a également accordé sa confiance quant aux méthodes statistiques envisageables, me permettant ainsi d'explorer différentes pistes de solutions.

Liste des sigles

ESG : Critères environnementaux, sociaux et de gouvernance

FGR : Fonction Gestion des Risques

CRCI : Conseil d'administration, assisté du Comité des Risques et de Contrôle Interne

PRISM : Physical Risks Internal Scoring Model

GRETA : Global Risk ESG Topics Assessment

NGFS : Network for Greening the Financial System

PDEF : Programme de Données Extra-Financières

ABEI : Arkéa Banque Entreprises Institutionnels

CMA : Crédit Mutuel Arkea

CNCM : Confédération Nationale du Crédit Mutuel

CMAF : Crédit Mutuel Alliance Federal

DPE : Diagnostic de Performance Énergétique

PPV1 : 1 plus proche voisin

PPV3 : 3 plus proches voisins

LR : Likelihood Ratio

AIC : Akaike information criterion

SAR : Spatial Autoregressive Model

SEM : Spatial Error Model

SDM : Spatial Durbin Model

SLX : Spatial Lag of X Model

RF : Random Forest

PDP : Partial Dependence Plot

Sommaire

I RAPPORT DE STAGE	5
1 Introduction	5
1.1 Environnement de l'entreprise	6
1.2 Le risque ESG	9
2 Présentation de l'outil PRISM	12
2.1 Risques Aigus	13
2.2 Risques Chroniques	19
3 Contexte du stage	23
3.1 Objectifs du stage	23
3.2 Gestion du temps	30
4 Bilan	32
II ÉTUDE STATISTIQUE	33
1 Introduction	34
2 Analyse exploratoire	35
2.1 Évolution des ventes	35
2.2 Prix des biens (€)	36
2.3 Étiquettes DPE	41
2.4 Âge des bâtiments (année)	43
2.5 Surface Habitable (m^2)	45
2.6 Surface de Terrain (m^2)	47
2.7 Autres variables	47
2.8 Matrice des corrélations	48
3 Méthodologie économétrique	50
4 Modélisation spatiale	51
4.1 Mesure de l'autocorrélation spatiale globale	51
4.2 Sélection des modèles	53
4.3 Interprétation des modèles	57
5 Modélisation machine-learning	71
6 Conclusion	75
7 Annexes	83

Première partie

RAPPORT DE STAGE

1 Introduction

Dans le cadre de mon cursus au sein du Master Économétrie Appliquée et Statistiques (ECAP) de l'IAE de Nantes, j'ai eu l'opportunité de manipuler des jeux de données issus de thématiques variées telles que l'économie, la finance, la santé, ainsi qu'un domaine qui a particulièrement retenu mon attention : l'environnement, et plus spécifiquement les enjeux climatiques. Cette orientation thématique s'est concrétisée lors d'un premier stage de quatre mois à l'Ifremer, au cours duquel j'ai travaillé sur une problématique environnementale en lien avec l'économie, à travers la modélisation statistique des pêcheries en mer Méditerranée. Cette expérience ayant été particulièrement enrichissante, j'ai souhaité approfondir cette voie en m'orientant vers une approche davantage centrée sur le risque climatique, en cohérence avec mes aspirations professionnelles.

C'est dans ce contexte que j'ai intégré le Crédit Mutuel Arkéa pour un stage de six mois, de mars à août 2025, au sein du département en charge de l'analyse des risques climatiques : l'équipe ESG (*Environnement, Social et Gouvernance*). Ma mission principale a porté sur la dimension cartographique de l'outil PRISM (*Physical Risk Internal Scoring Model*), qui vise à évaluer le risque physique climatique (aigu et chronique) sur l'ensemble du territoire métropolitain. L'objectif de PRISM v3 est de mesurer l'exposition des clients de la banque aux aléas climatiques, en prenant en compte une large gamme de phénomènes : événements extrêmes ponctuels (inondations, tempêtes, feux de forêts, vagues de chaleur et de froid, gels) ainsi que des évolutions plus progressives (sécheresses, mouvements de terrain, variations des températures et des précipitations, élévation du niveau de la mer).

En parallèle de cette mission principale, j'ai été amené à contribuer à plusieurs projets complémentaires, notamment :

- La mise en œuvre d'une analyse en composantes principales spatialisée (GWPCA) afin de construire des pondérations quantitatives entre les sous-indicateurs de chaque aléa utilisés dans l'outil PRISM ;
- Le développement d'une [application R-Shiny](#) (*mot de passe : 123*) permettant d'accéder aux données DRIAS à l'échelle du code postal et de la commune, offrant ainsi un diagnostic de risque climatique à l'adresse ;
- Une revue critique des travaux de la NGFS (Network for Greening the Financial System) portant sur l'impact des risques physiques sur les secteurs économiques en France ;
- La réalisation d'une régression hédonique visant à estimer l'effet de la performance énergétique des logements (DPE) sur les prix immobiliers en Bretagne

- Le calcul du pourcentage de perte attribuable aux aléas climatiques, déterminé par le ratio entre la valeur du sinistre climatique et la valeur initiale du bien.

1.1 Environnement de l'entreprise

1.1.1 Présentation du Crédit Mutuel Arkéa

Le Crédit Mutuel Arkéa (CMA) est un groupe de bancassurance coopératif et mutualiste français, dont le siège social est situé au Relecq-Kerhuon, dans le Finistère. Il constitue l'une des quatre caisses fédérales du Crédit Mutuel et est la seconde en importance après le Crédit Mutuel Alliance Fédérale. Ses racines remontent au mouvement coopératif initié en Allemagne par Frédéric-Guillaume Raiffeisen en 1849, avec l'apparition des premières caisses rurales de Crédit Mutuel en France dès 1884. Cet héritage fonde un modèle financier solidaire et ancré localement.

L'histoire du groupe prend forme avec la création de l'Office central de Landerneau en 1911, qui coordonnait diverses coopératives. En 1926, la Caisse régionale de Bretagne de crédit agricole mutuel (CB-CAM) est fondée, menant en 1970 à la naissance officielle du Crédit Mutuel de Bretagne (CMB), qui s'est ouvert aux ouvriers, notamment ceux de l'arsenal de Brest. En 1996, le Crédit Mutuel du Sud-Ouest (CMSO) rejoint le CMB, et le groupe Crédit Mutuel Arkéa est officiellement créé en 2002, regroupant initialement les fédérations du CMB, du CMSO et du Crédit Mutuel Massif Central (ce dernier ayant depuis rejoint Crédit Mutuel Alliance Fédérale en 2020). Le groupe est aujourd'hui constitué des fédérations du Crédit Mutuel de Bretagne et du Sud-Ouest, ainsi que d'une quarantaine de filiales spécialisées, incluant des acteurs majeurs comme Suravenir (assurance-vie) et Fortuneo Banque (banque en ligne).

En tant que banque coopérative et mutualiste, le Crédit Mutuel Arkéa n'est pas coté en Bourse et appartient directement à ses sociétaires. Ces derniers sont à la fois actionnaires et clients, détenant le capital social sous forme de parts sociales et participant aux décisions stratégiques lors des assemblées générales. Le conseil d'administration est présidé par Julien Carmona depuis 2021, et le comité exécutif par Hélène Bernicot depuis 2020. Le Crédit Mutuel Arkéa est un pionnier de la finance responsable en France : il a été le premier groupe bancaire français à adopter une raison d'être en mai 2020, soulignant son engagement pour l'impact environnemental et territorial. Cette démarche a été renforcée en mai 2022, lorsqu'il est devenu la première banque française à adopter le statut d'entreprise à mission, intégrant des objectifs sociaux et environnementaux à long terme dans ses statuts. Le groupe pilote le plan stratégique "Transitions 2024", axé sur la performance durable et la transition environnementale.

En 2024, le groupe comptait 11 564 collaborateurs et plus de 5,4 millions de sociétaires et clients. Ses fonds propres part du Groupe s'élevaient à 9,9 milliards d'euros, et son total de bilan atteignait 198,4 milliards d'euros. Le résultat net part du Groupe s'est établi à 395 millions d'euros. La production de crédits pour la transition environnementale a augmenté de plus de 10% en 2024, atteignant environ 1,1 milliard d'euros, soit deux fois plus en trois ans. La collecte brute d'épargne en faveur de la transition environnementale a également progressé

de plus de 60% par rapport à 2023, dépassant 1,1 milliard d'euros. Après une période de désaccord concernant son autonomie vis-à-vis de la Confédération Nationale du Crédit Mutuel (CNCM), un protocole d'accord politique pour un Crédit Mutuel uni et pluriel a été adopté à l'unanimité le 2 mai 2023. Cet accord, également validé par le Conseil d'administration de la CNCM le 3 mai 2023, vise à préserver la cohésion du groupe tout en reconnaissant pleinement le rôle prudentiel de la CNCM et en renforçant l'autonomie des fédérations et des groupes régionaux.

Le groupe se distingue par un ancrage territorial fort, notamment en Bretagne et Nouvelle-Aquitaine, étant l'un des rares établissements bancaires français à conserver l'ensemble de ses centres de décision en région. Il promeut activement l'usage de la langue bretonne dans ses services (chèquiers, DAB, charte "Ya d'ar brezhoneg") et est un partenaire majeur du mécénat culturel (Le Quartz, Vieilles Charrues) et du sport (équipes cyclistes Arkéa-B B Hotels, Stade Brestois 29, Arkéa Arena).

1.1.2 Présentation du pôle des risques transverses

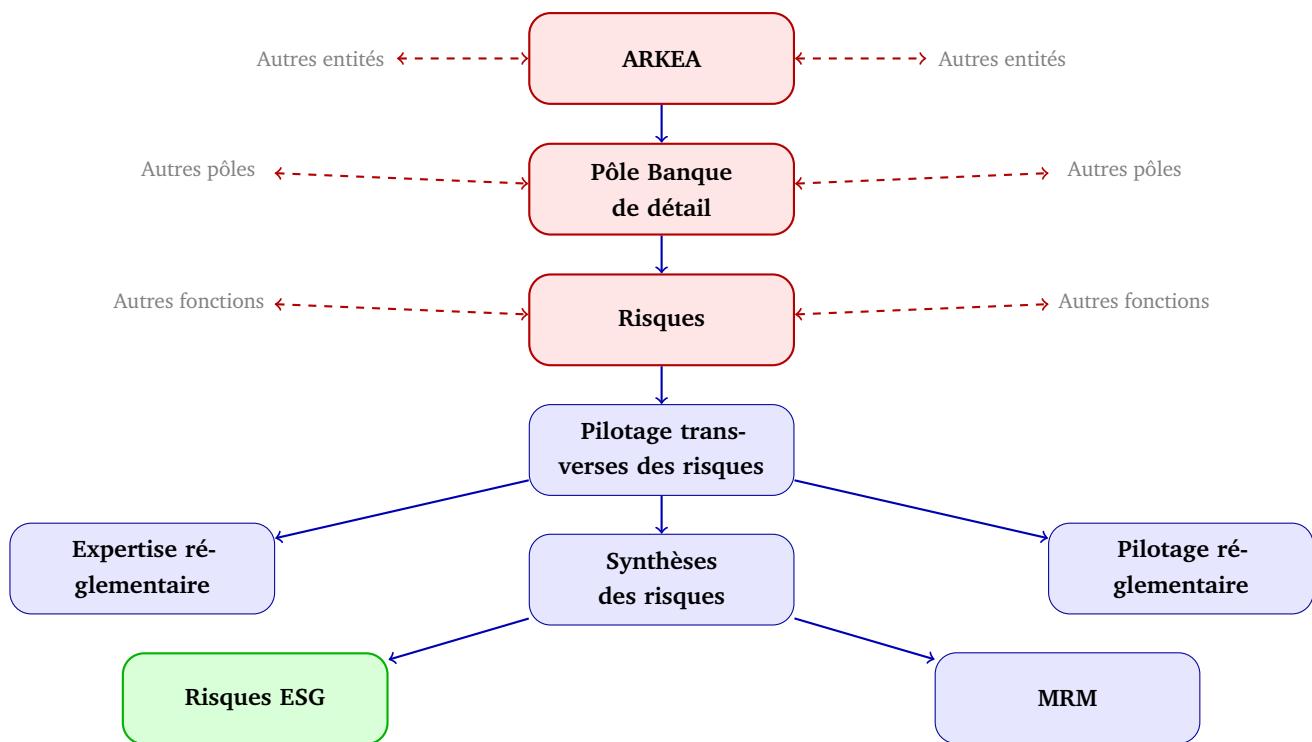


FIGURE 1 – Organigramme fonctionnel - Focus sur les Risques

La Direction du Pilotage Transverse des Risques au sein de Crédit Mutuel Arkéa est une entité clé chargée d'assurer le pilotage consolidé et la coordination globale de la Fonction Gestion des Risques (FGR) pour l'ensemble du groupe. Sa mission principale est de traiter les risques émergents et transverses, tels que les risques climatiques et ESG, qui, par leur nature matricielle, impactent différentes catégories de risques bancaires. Pour ce faire, cette direction a mis en place une organisation transverse visant à anticiper la transmission de ces risques à d'autres domaines

et à harmoniser les dispositifs de gestion à travers le groupe.

Ce pôle joue un rôle essentiel dans l'accompagnement des filiales du groupe, dont 21 ont été spécifiquement accompagnées en 2022, dans la mise en œuvre et le suivi des risques climatiques, ainsi que dans l'adaptation des dispositifs définis au niveau du groupe à leurs contextes spécifiques. La Direction du Pilotage Transverse des Risques collabore activement avec d'autres directions dédiées à la gestion des risques, comme la Direction du Risque de Crédit et de Contrepartie, pour élaborer des feuilles de route communes, partager des visions cibles et définir les chantiers nécessaires à l'intégration des risques ESG dans leurs activités respectives. Le service Expertise Réglementaire, rattaché à ce pôle, assure une veille réglementaire continue sur l'ensemble des typologies de risque et fournit une expertise aux interlocuteurs Risques, y compris sur les problématiques de contrôle interne et de rémunérations.

La Direction du Pilotage Transverse des Risques est également responsable du suivi consolidé de la gestion des risques climatiques pour l'ensemble du groupe, via un tableau de bord des risques climatiques trimestriel. Ce tableau de bord informe la gouvernance des risques encourus et du niveau de risque pris au niveau du groupe, tout en déclinant les engagements stratégiques, comme la stratégie climat et les politiques sectorielles.

1.1.3 Présentation de la partie synthèse des risques

Au sein de la Direction du Pilotage Transverse des Risques, le Département Synthèse des Risques est spécifiquement chargé de la gestion des risques émergents, incluant explicitement les risques environnementaux et climatiques. Ce département a été le point d'ancrage de cette activité depuis juin 2019 et est également en charge de la coordination de la Fonction Gestion des Risques (FGR), ainsi que du suivi consolidé de l'ensemble des risques du groupe, notamment par le biais de tableaux de bord et de rapports.

Face à l'intensification des exigences réglementaires et à l'élargissement progressif du périmètre aux risques Environnementaux, Sociaux et de Gouvernance (ESG), une évolution significative a eu lieu en 2022 avec la création d'un service dédié "Risques ESG" au sein de ce département. Ce service a vu ses effectifs renforcés pour atteindre 7 personnes en 2024, soulignant l'importance stratégique accordée à ces enjeux. Les principales missions du service "Risques ESG" consistent à identifier, mesurer et gérer les risques ESG, à diffuser la culture des risques associée, et à assurer le suivi des engagements ESG internes ainsi que des attentes des superviseurs. En raison du caractère intrinsèquement transverse et matriciel des risques ESG, ce service collabore étroitement avec les autres Directions de la Direction des Risques pour garantir une approche cohérente et intégrée.

1.2 Le risque ESG

1.2.1 Définitions et caractéristiques

Les risques ESG constituent une catégorie de risques qui va au-delà de l'analyse financière traditionnelle. Ils sont caractérisés par leur nature polymorphe et matricielle, ce qui signifie qu'ils ne se matérialisent pas directement mais se transmettent à d'autres catégories de risques existantes, telles que le risque de crédit, le risque opérationnel, le risque de marché, le risque de liquidité, le risque de réputation et le risque juridique/de responsabilité. Le Crédit Mutuel Arkéa reconnaît ces canaux de transmission, par exemple, une inondation peut endommager des infrastructures informatiques (risque opérationnel), ou la banque peut être tenue responsable des impacts du changement climatique (risque juridique et de réputation). Les risques physiques et de transition, bien que distincts, sont interdépendants ; par exemple, des politiques de décarbonation peuvent augmenter les risques de transition à court terme mais diminuer les risques physiques à long terme.

Ces risques présentent des caractéristiques uniques qui les distinguent des risques financiers classiques. Ils sont souvent dépourvus d'historique de crises climatiques passées, rendant leur modélisation plus complexe. Leur temporalité est plus longue, avec une matérialisation des risques pouvant s'étendre sur 30 à 50 ans, contrairement aux risques classiques dont les impacts sont généralement visibles sur 3 à 5 ans, ce qui pose le défi de la "tragédie des horizons" [14]. De plus, ces risques sont universels, impactant l'ensemble de la planète bien qu'à des degrés divers selon les géographies. Enfin, ils sont considérés comme "certains" quant à leur survenance mais "incertains" quant à leur ampleur, rendant leur évaluation et leur gestion particulièrement délicates.

1.2.2 Intégration dans le Cadre d'Appétence au Risque et la Gouvernance

Le Crédit Mutuel Arkéa a explicitement intégré les risques ESG dans son cadre d'appétence au risque et sa cartographie des risques, considérant cela comme un outil fondamental de pilotage. Cette intégration vise à assurer une adéquation entre la prise de risque et les engagements du groupe, notamment sa Raison d'Être et sa Stratégie Climat, tout en maîtrisant son empreinte carbone.

Le Conseil d'administration, assisté du Comité des Risques et de Contrôle Interne (CRCI) et du Comité Stratégie et Responsabilité Sociétale, valide les orientations stratégiques et assure la surveillance des risques, intégrant les enjeux ESG à leur charte. Des indicateurs quantitatifs ont été progressivement ajoutés au cadre d'appétence depuis 2021/2022 pour le suivi de ces risques, incluant les expositions sectorielles (charbon, pétrole/gaz), les émissions de CO₂ directes et les contreparties avec des notations ESG dégradées. Une politique interne de gestion des risques environnementaux a été rédigée pour organiser, identifier, mesurer, encadrer, surveiller, contrôler et communiquer sur ces risques.

1.2.3 Outils d'Identification et de Mesure

Pour identifier et évaluer les risques ESG, le Crédit Mutuel Arkéa a développé et mis en œuvre des outils et méthodologies spécifiques. Le référentiel sectoriel des risques climatiques (physiques et de transition) permet d'estimer l'exposition d'une entreprise en fonction de son secteur d'activité, en se basant sur des études d'organismes reconnus et les émissions de CO₂. En complément, l'outil interne **PRISM** (*Physical Risk Internal Scoring Model*, voir section 2) évalue l'exposition géographique aux risques physiques aigus (inondations, sécheresses, tempêtes) et chroniques (hausse des températures, modification des précipitations, élévation du niveau de la mer) au niveau du code postal en France métropolitaine, s'appuyant sur des bases de données scientifiques nationales et internationales. Un travail similaire est en cours pour les risques liés à la perte de biodiversité, avec un référentiel sectoriel et une approche géographique.

Pour l'évaluation individualisée des risques ESG, l'outil **GRETA** (*Global Risk ESG Topics Assessment*) a été déployé pour les grandes entreprises emprunteuses (+ 10M€ de CA). Ce questionnaire collecte des données ESG pour déterminer une cotation des risques ESG, qui intègre une cotation sectorielle ajustée des pratiques de l'entreprise, et vise à évaluer les mesures d'atténuation mises en place par les contreparties. La cotation ESG de GRETA est intégrée dans le processus d'octroi de crédit. L'analyse par scénarios (SPICE) est une autre des composantes de la gestion des risques, utilisant des scénarios du GIEC et du NGFS (comme Net Zero 2050, Delayed Transition) pour projeter les impacts économiques et financiers à différents horizons (jusqu'à 2050), permettant au groupe de développer sa propre méthodologie de stress test climatique interne.

1.2.4 Suivi, Reporting et Diffusion de la Culture du Risque

Le Crédit Mutuel Arkéa assure le suivi et le pilotage de ses risques ESG au moyen d'un tableau de bord trimestriel, présenté à la gouvernance, qui mesure l'état d'exposition du groupe et l'avancement des objectifs. Enrichi d'informations sur les politiques sectorielles, les expositions aux risques physiques et de transition, les émissions de CO₂ et les émetteurs à notation ESG dégradée, cet outil complète les publications externes du groupe, telles que le rapport TCFD et le rapport Pilier 3, conformes aux exigences réglementaires européennes en matière d'informations qualitatives et quantitatives sur les risques ESG. La gestion des données est centralisée par le Programme de Données Extra-Financières (PDEF), lancé en 2021, avec un effort constant d'amélioration de la collecte et de la structuration, et la diffusion d'une culture ESG est activement encouragée par des formations régulières destinées aux administrateurs, dont une administratrice référente ESG-climat, aux collaborateurs et aux correspondants Risques ESG dans les filiales.

Pionnier dans le secteur bancaire, le Crédit Mutuel Arkéa a également développé dès 2020, en dehors de toute contrainte réglementaire, une méthodologie innovante pour monétiser ses impacts extra-financiers, devenant la première banque française à mesurer en euros sa performance globale, combinant résultats financiers et extra-financiers. Cette approche, d'abord mise en œuvre au sein d'ABEI, s'est progressivement élargie à un périmètre couvrant, en 2024, 81% des financements, 95% des achats externes et 98% des effectifs. Elle repose sur une vingtaine d'indicateurs

extra-financiers (*émissions de GES, consommation d'eau, emplois soutenus, accès au logement ou aux soins*) permettant de quantifier les impacts environnementaux et socio-économiques positifs et négatifs. En 2024, chaque million d'euros de financements ou d'investissements générait en moyenne 109 000 € d'impact global positif, pour une performance extra-financière totale de 9,3 milliards d'euros, stable par rapport à 2023, avec une amélioration des indicateurs environnementaux malgré une dégradation des indicateurs socio-économiques liée au contexte économique.

2 Présentation de l'outil PRISM

PRISM v3 est un outil interne développé par le service risques ESG du Crédit Mutuel Arkéa dans le but de cartographier l'exposition aux risques climatiques physiques par axe géographique. Son ambition est d'intégrer des aléas aigus et chroniques, ainsi que des aspects historiques et prospectifs, afin de prévenir et de protéger les clients du Crédit Mutuel Arkéa de ces risques. Cet outil vise à évaluer le niveau de risque physique au niveau le plus fin, par coordonnées GPS.

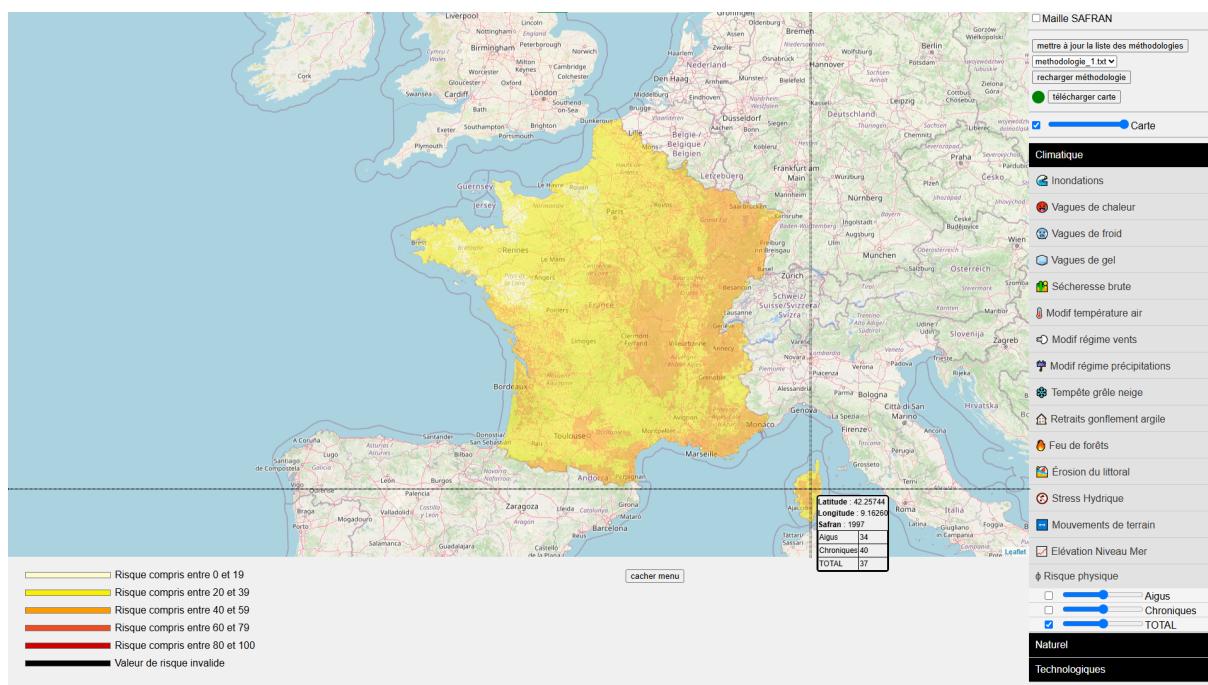


FIGURE 2 – Capture d'écran de l'outil PRISM simulation v.1.1.2

Les risques physiques se réfèrent aux effets financiers du changement climatique, tels que la multiplication des événements climatiques extrêmes et les modifications progressives du climat, ainsi qu'à la dégradation de l'environnement (pollution de l'air, de l'eau, des sols, stress hydrique, perte de biodiversité, déforestation). Dans sa première version, l'évaluation des risques physiques de PRISM v3 se concentre uniquement sur les effets du changement climatique.

Les risques physiques sont classés en deux catégories principales :

- **Les risques aigus** découlent d'événements extrêmes ponctuels. Ces aléas incluent les inondations, les sécheresses (météorologiques et RGA), les tempêtes, la grêle et la neige (TGN), les vagues de chaleur, les vagues de froid, les vagues de gel, et les glissements de terrain. Des événements comme les cyclones, les ouragans et les typhons sont également considérés comme des risques aigus.
- **Les risques chroniques** résultent de changements graduels du climat. Parmi ceux-ci, on trouve la variation des températures de l'air, la variation des régimes de précipitations, la hausse du niveau de la mer, l'érosion du trait de côte, l'évolution des régimes des vents, et

le stress hydrique.

Dans la méthodologie de scoring de PRISM v3, les aléas sont divisés en deux catégories pour une meilleure flexibilité et pour éviter le lissage du score total :

- **Les aléas majeurs** sont des aléas présents de manière généralisée en France et ayant un impact climatique significatif sur de nombreux domaines (assurance, immobilier, agriculture, etc). Pour les risques aigus, les aléas majeurs identifiés sont les inondations, la sécheresse RGA, et les tempêtes-grêle-neige (TGN). Pour les risques chroniques, il s'agit de la modification des pluies, de la modification des températures et du stress hydrique. Les scores des aléas majeurs sont équipondérés dans le calcul du score global.
- **Les aléas mineurs** sont des aléas non homogènes géographiquement sur la France et/ou ayant des impacts moindres en termes d'exposition. Pour ces aléas mineurs (vagues de chaleur, sécheresse brute, feux de forêts, vagues de froid, vagues de gel, glissements de terrain pour les risques aigus ; et modification des vents, érosion du littoral, élévation du niveau de la mer pour les risques chroniques), leur score est calculé comme la valeur maximale de l'ensemble de leurs scores sur un point géographique donné. Cette approche permet de prendre en compte l'aléa le plus fort sans introduire de complexités cartographiques ou lisser le score des aléas majeurs.

2.1 Risques Aigus

Les risques physiques aigus résultent d'événements climatiques extrêmes et ponctuels. Le score global des risques aigus constitue une somme équipondérée des scores des aléas aigus majeurs et du score maximal des aléas aigus mineurs.

2.1.1 Inondation (Aléa majeur)

Une inondation se caractérise par une submersion temporaire d'un espace par de l'eau liquide, qu'elle soit d'origine naturelle ou artificielle. Ce phénomène englobe la remontée de nappes phréatiques, le ruissellement, le débordement de cours d'eau et la submersion marine. Les inondations génèrent des destructions significatives de biens immobiliers et mobiliers, occasionnent des blessures, provoquent la coupure de réseaux de communication et entraînent d'importantes perturbations économiques et sociales. En France, le coût des catastrophes naturelles liées aux inondations a été estimé entre 7,7 et 25,1 milliards d'euros sur la période 1995-2019 [6].

TABLE 1 – Scoring de l'aléa Inondation

Sous-Indicateur	Source	Règle de scoring
Pondération 40%		
Zones sujettes aux remontées de nappes	BRGM – "RemNappe"	Score 100 pour les zones sujettes aux débordements de nappes ou aux inondations de caves. Score 0 si aucune inondation de nappe ni de cave ou si l'adresse est hors de la zone couverte par la base
Zones sujettes aux inondations par ruissellement	BRGM – "Zonages Inondation - Rapportage 2020" (TRI ruissellement)	Score 100 pour une forte probabilité d'occurrence. Score 50 pour une probabilité moyenne d'occurrence. Score 30 pour une faible probabilité d'occurrence. Score 0 si l'adresse est hors de la couverture de la base
Zones sujettes aux inondations par débordement de cours d'eau	BRGM – "Zonages Inondation - Rapportage 2020" (TRI débordement)	Score 100 pour une forte probabilité d'occurrence. Score 50 pour une probabilité moyenne d'occurrence. Score 30 pour une faible probabilité d'occurrence. Score 0 si l'adresse est hors de la couverture de la base
Zones sujettes aux submersions marines	BRGM – "Zonages Inondation - Rapportage 2020" (TRI submersion marine)	Score 100 pour une forte probabilité d'occurrence. Score 50 pour une probabilité moyenne d'occurrence. Score 30 pour une faible probabilité d'occurrence. Score 0 si l'adresse est hors de la couverture de la base
Pondération 20%		
Écart des précipitations quotidiennes extrêmes	DRIAS – Météo France	Score 100 pour un écart de +5 mm ou plus. Score proportionnel entre 0 et 100 pour des variations de 0 à +5 mm. Score 0 pour des variations négatives ou nulles
Pondération 20%		
Nombre d'arrêtés Catastrophe Naturelle (CatNat) par commune	Ministère de la transition écologique – "GAS-PAR"	Score 0 : si le nombre de sinistres observés est égal à 0. Score 100 : pour le nombre maximum de sinistres enregistré au niveau national (ex : 34 à Nice). Le score est calculé par proportionnalité : (Nombre de sinistres sur un code commune / Max de sinistres de la base) × 100
Pondération 20%		
Taux d'imperméabilité des sols	Copernicus – "Imperviousness Density 2018" ¹³	Le score de 0 à 100 correspond directement au pourcentage d'imperméabilisation du sol

2.1.2 Retrait Gonflement des Argiles (Aléa majeur)

Le retrait-gonflement des argiles correspond à un mouvement du sol et du sous-sol provoqué par l’alternance entre la saturation en eau et la sécheresse des sols argileux, entraînant leur rétraction ou leur gonflement. Ce phénomène affecte directement la stabilité des fondations des constructions humaines et peut rendre certaines zones inhabitables tout en générant des pertes de production et de revenus significatives. Le coût des catastrophes naturelles liées au retrait-gonflement des argiles en France a été évalué à 12,3 milliards d'euros entre 1989 et 2018 [4].

TABLE 2 – Scoring de l'aléa Retrait Gonflement des Argiles

Sous-Indicateur	Source	Règle de scoring
Pondération 50%		
Zones sujettes au RGA	BRGM	Score 100 pour aléa fort, Score 50 pour aléa moyen, Score 30 pour aléa faible, Score 0 pour aléa nul
Pondération 25%		
Arrêtés CatNat sécheresse	Ministère de la transition écologique – "GAS-PAR"	Score proportionnel au nombre de sinistres (maximum 50 à Bassillac)
Pondération 12,5%		
Jours de sécheresse SSWI	Météo France	Score = (Nb jours SWI < -0,4) / 365 × 100
Pondération 12,5%		
Jours de pluie	Météo France	Score = (Nb jours pluie > 1mm) / 365 × 100

2.1.3 Tempête-Grêle-Neige (Aléa majeur)

L'aléa tempête-grêle-neige constitue un agrégat de phénomènes atmosphériques caractérisés par des vents violents atteignant ou dépassant la force 10 sur l'échelle de Beaufort, des chutes de grêle avec des grêlons de plus de 5 mm de diamètre, et des dégâts occasionnés par le poids ou l'humidité de la neige. Ces phénomènes génèrent des coûts considérables pour les assureurs, estimés à 4,67 milliards d'euros en 2022 en France [13], et peuvent entraîner des dégâts importants sur les infrastructures ainsi qu'un grand nombre de sinistres déclarés.

TABLE 3 – Scoring de l'aléa Tempête-Grêle-Neige

Sous-Indicateur	Source	Règle de scoring
Pondération 50%		
Fréquence arrêtés CatNat	Ministère de la transition écologique – "GAS-PAR"	Score proportionnel à la fréquence départementale (maximum 21,1% dans les Landes)
Pondération 50%		
Vents forts Q98	Météo France	Score basé sur l'échelle de Beaufort : Score 100 pour niveau 9+ (20,56 m/s)

2.1.4 Feux de Forêts (Aléa mineur)

Les feux de forêts se définissent comme une combustion non contrôlée se développant dans les forêts ou les milieux naturels tels que les landes, garrigues ou maquis. Les facteurs de prédisposition comprennent le type de végétation et sa teneur en eau, les conditions climatiques caractérisées par la sécheresse et le vent, ainsi que l'occupation du territoire incluant les activités humaines et l'interface forêt-habitat. Ces incendies entraînent des pertes forestières importantes, réduisent la valeur marchande des bois, menacent la biodiversité et causent des destructions de biens accompagnées de perturbations économiques significatives.

TABLE 4 – Scoring de l'aléa Feux de Forêts

Sous-Indicateur	Source	Règle de scoring
Pondération 50%		
Périmètre zones forestières	Base de données forestières	Score 100 pour forêts de feuillus/conifères, Score 90 pour forêts mélangées, Score 80 pour feuillus/mutations, Score 70 pour landes/broussailles
Pondération 25%		
IFM historique	Météo France	Score = Nb jours > 40 × 5 (maximum 100 pour 20+ jours)
Pondération 25%		
IFM prospectif	Météo France	Même logique que IFM historique

2.1.5 Glissements de terrain (Aléa mineur)

Les glissements de terrain correspondent à des déplacements naturels des sols et sous-sols, pouvant être lents ou brutaux, causés ou aggravés par l'activité humaine telle que la modification du sol ou l'exploitation minière, ou par des facteurs naturels incluant la composition du sol, l'inclinaison, les variations d'humidité, l'eau, la gravité et les tremblements de terre. Ces phénomènes englobent l'érosion, les glissements proprement dits, les éboulements, les coulées de boue, les effondrements et les affaissements. Ils provoquent la destruction de biens, occasionnent

des blessures, bloquent les infrastructures de transport, polluent les terrains et interrompent les services. Le coût des mouvements de terrain en France a été estimé à 700 millions d'euros sur la période 1995-2019 [5].

TABLE 5 – Scoring de l'aléa Glissements de terrain

Type de mouvement	Zone	Règle de scoring par distance
Pondération 100%		
Glissement	Massif	Score 100 (500m), Score 50 (1000m)
Glissement	Hors massif	Score 100 (250m), Score 50 (500m)
Chute de blocs/Éboulement	Massif	Score 100 (250m), Score 50 (500m), Score 25 (750m)
Chute de blocs/Éboulement	Hors massif	Score 100 (100m), Score 50 (150m), Score 25 (250m)
Coulée	Massif	Score 100 (500m), Score 70 (1000m), Score 25 (1500m)
Coulée	Hors massif	Score 100 (250m), Score 50 (500m)
Effondrement/Affaissement	Toutes zones	Score 100 (250m), Score 30 (500m)
Érosion de berges	Toutes zones	Score 100 (100m)
Cavité	Toutes zones	Score 100 (250m), Score 30 (500m)

2.1.6 Sécheresse (Aléa mineur)

La sécheresse se caractérise par un déficit en eau survenant sur une période plus ou moins longue dans un territoire donné. Elle peut revêtir différentes formes : météorologique par l'absence de précipitations, agricole par l'appauvrissement du sol en eau, ou hydrologique par l'assèchement des nappes phréatiques et des cours d'eau. Cette sécheresse dite "brute" se distingue du retrait-gonflement des argiles bien qu'elle y soit également liée. Elle provoque une baisse des rendements agricoles, une dégradation des sols et des végétaux, et peut générer d'autres risques connexes tels que le stress hydrique, les mouvements de terrain, les incendies et paradoxalement les inondations.

TABLE 6 – Scoring de l'aléa Sécheresse

Sous-Indicateur	Source	Règle de scoring
Pondération 50%		
Période de sécheresse SPI	Météo France	Score = 1 - (1 - Période de sécheresse) ¹⁰ 100
Pondération 50%		
Jours sécheresse profonde	Météo France	Score = (Nb jours sécheresse) / 365 × 100

2.1.7 Vague de chaleur (Aléa mineur)

Une vague de chaleur se définit comme un épisode de plusieurs jours consécutifs où la température maximale excède la normale de plus de 5°C. Météo France considère qu'un événement de vague de chaleur se produit lorsque la température dépasse 25,3°C pendant une journée ou 23,4°C pendant au moins trois jours consécutifs [15]. Ces phénomènes entraînent des conséquences catastrophiques sur la santé humaine et génèrent des impacts économiques majeurs, notamment par la perte de récoltes agricoles et les problèmes de refroidissement des centrales nucléaires.

TABLE 7 – Scoring de l'aléa Vague de chaleur

Sous-Indicateur	Source	Règle de scoring
Pondération 25%		
Jours vague chaleur historique	Météo France	Score 100 pour 30+ jours, intervalles de 6 jours
Pondération 25%		
Jours vague chaleur prospectif	Météo France	Même logique que historique
Pondération 50%		
Nuits tropicales prospectif	Météo France	Score proportionnel aux nuits >20°C

2.1.8 Vague de froid (Aléa mineur)

Une vague de froid correspond à une période durant laquelle la température minimale demeure inférieure de plus de 5°C à la normale pendant au moins cinq jours consécutifs [9]. Ces épisodes provoquent un stress métabolique et augmentent la sensibilité aux maladies chez les populations humaines. Ils entraînent également une augmentation de la demande énergétique pour le chauffage ainsi qu'un ralentissement ou une perte de production agricole. Une valeur de dix jours ou plus de vague de froid est considérée comme élevée.

TABLE 8 – Scoring de l'aléa Vague de froid

Sous-Indicateur	Source	Règle de scoring
Pondération 50%		
Jours vague froid historique	Météo France	Score 100 pour 10+ jours, intervalles de 2 jours
Pondération 50%		
Jours vague froid prospectif	Météo France	Même logique que historique

2.1.9 Vague de gel (Aléa mineur)

Le phénomène de gel se caractérise par la solidification de l'eau liquide lorsque la température devient inférieure ou égale à 0°C [9], pouvant se manifester par du givre ou de la gelée. Les gelées tardives survenant entre mars et mai font l'objet d'une attention particulière en raison de leur caractère destructeur pour les cultures. Ces épisodes affectent la santé humaine par le stress

métabolique et l'augmentation de la sensibilité aux maladies, impactent le secteur énergétique par l'accroissement de la demande de chauffage, perturbent les transports par la formation de verglas, et compromettent l'agriculture par le ralentissement, le retard ou la perte de récoltes, notamment par la "gelée noire" particulièrement meurtrière.

TABLE 9 – Scoring de l'aléa Vague de gel

Sous-Indicateur	Source	Règle de scoring
Pondération 50%		
Jours gel tardif historique	Météo France	Score 100 pour 10+ jours, intervalles de 2 jours
Pondération 50%		
Jours gel tardif prospectif	Météo France	Même logique que historique

2.2 Risques Chroniques

Les risques physiques chroniques découlent de changements graduels et progressifs du climat. Le score global des risques chroniques constitue une somme équipondérée des scores des aléas chroniques majeurs et du score maximal des aléas chroniques mineurs.

2.2.1 Variation des températures (Aléa majeur)

La variation des températures correspond à l'écart de température moyenne journalière exprimé en degrés Celsius entre une période de référence s'étendant de 1976 à 2005 et une période de projection couvrant 2021 à 2050 [9]. Ces variations témoignent des changements climatiques progressifs et constituent un indicateur fondamental de l'évolution du climat régional.

TABLE 10 – Scoring de l'aléa Variation des températures

Niveau de risque	Score	Variation de température
Pondération 100%		
Très faible	0-20	0 à 1,02°C
Faible	21-40	1,02 à 1,14°C
Moyen	41-60	1,14 à 1,22°C
Important	61-80	1,22 à 1,30°C
Très important	81-100	1,30°C

2.2.2 Variation des précipitations (Aléa majeur)

La variation des précipitations se définit comme l'écart des précipitations journalières moyennes exprimé en millimètres entre une période de référence de 1976 à 2005 et une période de projection de 2021 à 2050 [9]. Les variations positives et négatives sont prises en compte car elles représentent toutes deux un déséquilibre climatique significatif par rapport aux conditions de référence.

TABLE 11 – Scoring de l'aléa Variation des précipitations

Niveau de risque	Score	Variation de précipitations
Pondération 100%		
Très faible	0-20	Augmentation 0 à 0,10mm ou diminution 0 à -0,02mm
Faible	21-40	Augmentation 0,10 à 0,16mm ou diminution -0,02 à -0,06mm
Moyen	41-60	Augmentation 0,16 à 0,22mm ou diminution -0,06 à -0,10mm
Important	61-80	Augmentation 0,22 à 0,30mm ou diminution -0,10 à -0,16mm
Très important	81-100	Augmentation 0,30mm ou diminution -0,16mm

2.2.3 Stress hydrique (Aléa majeur)

Le stress hydrique fait référence à la disponibilité de l'eau dans une région donnée et constitue un indicateur crucial des tensions sur les ressources hydriques. Il mesure le rapport entre la demande en eau et les ressources disponibles, permettant d'évaluer la pression exercée sur les écosystèmes aquatiques et la capacité d'approvisionnement en eau des territoires.

TABLE 12 – Scoring de l'aléa Stress hydrique

Niveau de risque	Score	Stress hydrique
Pondération 100%		
Très faible	0-20	< 10%
Faible	21-40	10-20%
Moyen	41-60	20-40%
Important	61-80	40-80%
Très important	81-100	> 80%

2.2.4 Érosion du trait de côte (Aléa mineur)

L'érosion du trait de côte se caractérise par le taux d'érosion du littoral exprimé en mètres par an. Ce phénomène résulte de l'action combinée des vagues, des marées, des courants et des conditions météorologiques sur les côtes, entraînant un recul progressif du littoral. L'érosion côtière menace les infrastructures, les habitations et les écosystèmes littoraux, constituant un enjeu majeur d'aménagement du territoire.

TABLE 13 – Scoring de l'aléa Érosion du trait de côte

Niveau de risque	Score	Taux d'érosion
Pondération 100%		
Très faible	0-20	Communes non littorales
Faible	21-40	Érosion limitée
Moyen	41-60	Érosion modérée
Important	61-80	Érosion significative
Très important	81-100	0,2476 à 9,0656 m/an

2.2.5 Évolution des régimes des vents (Aléa mineur)

L'évolution des régimes des vents correspond à l'écart du nombre de jours de vent fort, défini comme une vitesse quotidienne du vent dépassant le 98ème centile, entre une période de référence de 1976 à 2005 et une période de projection de 2021 à 2050 [9]. Cette évolution témoigne des modifications des patterns atmosphériques et peut avoir des implications importantes pour les activités humaines, notamment l'énergie éolienne, les transports et l'agriculture.

TABLE 14 – Scoring de l'aléa Évolution des régimes des vents

Niveau de risque	Score	Variation (jours)
Pondération 100%		
Très faible	0-20	Augmentation 0 à 0,74 ou diminution 0 à -0,3
Faible	21-40	Augmentation 0,74 à 1,36 ou diminution -0,3 à -0,67
Moyen	41-60	Augmentation 1,36 à 1,94 ou diminution -0,67 à -1,01
Important	61-80	Augmentation 1,94 à 2,78 ou diminution -1,01 à -1,47
Très important	81-100	Augmentation 2,78 ou diminution -1,47

2.2.6 Hausse du niveau de la mer (Aléa mineur)

La hausse du niveau de la mer correspond à la projection de l'augmentation du niveau marin exprimée en centimètres. Ce phénomène résulte principalement de la dilatation thermique des océans due au réchauffement climatique et de la fonte des glaciers et des calottes glaciaires. L'élévation du niveau marin menace les zones côtières par l'inondation permanente ou temporaire, l'érosion accélérée et la salinisation des sols et des nappes phréatiques.

TABLE 15 – Scoring de l'aléa Hausse du niveau de la mer

Niveau de risque	Score	Hausse projetée
Pondération 100%		
Très faible	0-20	Communes non littorales
Faible	21-40	0 à 26 cm
Moyen	41-60	26 à 27 cm
Important	61-80	27 à 29 cm
Très important	81-100	29 à 41 cm

3 Contexte du stage

3.1 Objectifs du stage

L'objectif principal de ce stage était de contribuer à l'analyse cartographique du risque climatique physique à travers l'outil PRISM V3.

Mon rôle consistait à apporter une approche statistique et quantitative à l'outil PRISM, notamment en mobilisant les compétences acquises en économétrie spatiale et en SIG via QGIS. À mon arrivée, le développement de PRISM était temporairement suspendu pour des raisons stratégiques, ce qui a orienté mes travaux vers une phase exploratoire. Cela m'a permis de faire preuve d'initiative tout au long du stage.

Par ailleurs, j'ai assisté de manière régulière à des réunions (5 à 10 par semaine), couvrant des sujets variés :

- Présentation des rôles au sein de l'équipe ESG
- Formation sur l'outil PRISM et les données climatiques utilisées
- Échanges interbancaires sur les approches de modélisation du risque climatique
- Discussions sur les méthodologies de scoring géographique et sectoriel du risque physique

3.1.1 Réflexion sur la pondération des indicateurs climatiques

Une mission importante fut de questionner la méthode de pondération des sous-indicateurs utilisés dans les indices de risque PRISM. En effet, ceux-ci étaient initialement définis par expertise, sans base quantitative solide. Par exemple, dans le cadre du risque d'inondation, la pondération du "taux d'imperméabilisation des sols" était fixée à 20%, sans justification statistique.

De plus, le fait que les sous-indicateurs soient exprimés dans des unités hétérogènes (mm de précipitations, nombre d'arrêtés CATNAT, probabilités, etc.), bien que normalisés sur une échelle de 0 à 100, pose un problème d'interprétation de la mesure d'exposition.

Pour remédier à cette limite, j'ai exploré la littérature scientifique afin d'identifier une méthode de pondération objectivée. L'approche la plus pertinente rencontrée est celle de l'Analyse en Composantes Principales spatialisée (GWPCA)[\[11\]](#) [\[16\]](#) [\[17\]](#), permettant de générer des pondérations basées sur la variance expliquée par les composantes principales, indépendamment des unités des variables.

Il a été nécessaire d'harmoniser les données selon une unité spatiale commune, en l'occurrence le code postal. Pour ce faire, chaque indicateur a été recalculé à l'échelle des codes postaux en fonction de la superficie de leur intersection avec les polygones d'origine des indicateurs. Concrètement, cela consiste à estimer la surface du code postal incluse dans le polygone de chaque indicateur, puis à pondérer la valeur de l'indicateur en proportion de cette surface (*voir annexe 30*).

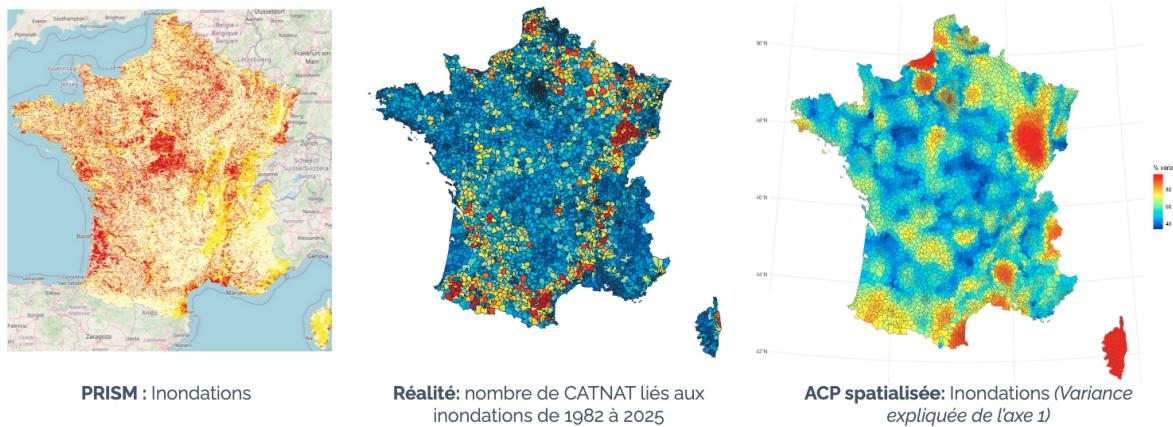


FIGURE 3 – Comparaison entre les différentes méthodes pour l'aléa inondation

J'ai ainsi appliqué cette méthode sur un ensemble d'indicateurs, comparant les résultats obtenus à ceux de PRISM (voir Figure 3 et Annexe 31). A noter qu'il a fallu convertir les données au même format spatial, ici au code postal, via le script python dont nous avons parlé précédemment.

3.1.2 Développement d'une application interactive R-Shiny

Une autre mission majeure a été le développement d'une application interactive à partir des données climatiques issues de la plateforme DRIAS. L'objectif initial était la création de plus de 150 cartes, mais cette approche s'avérant peu ergonomique, j'ai proposé [une solution dynamique et centralisée sous R-Shiny](#) (*mot de passe d'accès : 123*).

Étapes de conception :

1. Collecte des données [DRIAS](#), selon plusieurs **scénarios RCP** (2.6, 4.5, 8.5) et **horizons temporels** (référence, H1, H2, H3)
2. Conversion des données format grille SAFRAN (8x8km) au format code postal, via une pondération en fonction de l'air du code postal figurant dans la grille. ([annexe 30](#))
3. Documentation des indicateurs selon les définitions de la DRIAS [9]
4. Création de l'application et intégration des différentes fonctionnalités
5. [Mise en ligne de l'application](#)

Après présentation de la version 1, qui était encore une maquette, il a été décidé d'intégrer différentes fonctionnalités pour la version 2.

Fonctionnalités intégrées dans la version 2 :

1. Possibilité de visualiser les données au format départemental
2. Export des données au format départemental ou communal

3. Extraction de données en Excel selon les paramètres sélectionnés
4. Téléchargement des cartes au format PDF avec titres dynamiques (les titres varient en fonction de l'indicateur, horizon et scénario sélectionnés)
5. Représentation des variations en pourcentage par rapport à la période de référence
6. Intégration de nouveaux indicateurs

Enfin, après présentation de la version 2, d'autres fonctionnalités ont été intégrées dans la version 3, qui est la version finale.

Fonctionnalités additionnelles intégrées dans la version 3 :

1. **Géolocalisation par adresse via l'API de la BAN [8]**, permettant à l'utilisateur de rentrer et trouver son adresse directement dans l'application
2. **Diagnostic local personnalisé** : représentation graphique du climat à l'adresse, comparaison à la moyenne nationale, selon différents scénarios et horizons (*inspiré par ClimaDiag, voir Figure 6*)
3. **Sécurisation par mot de passe**

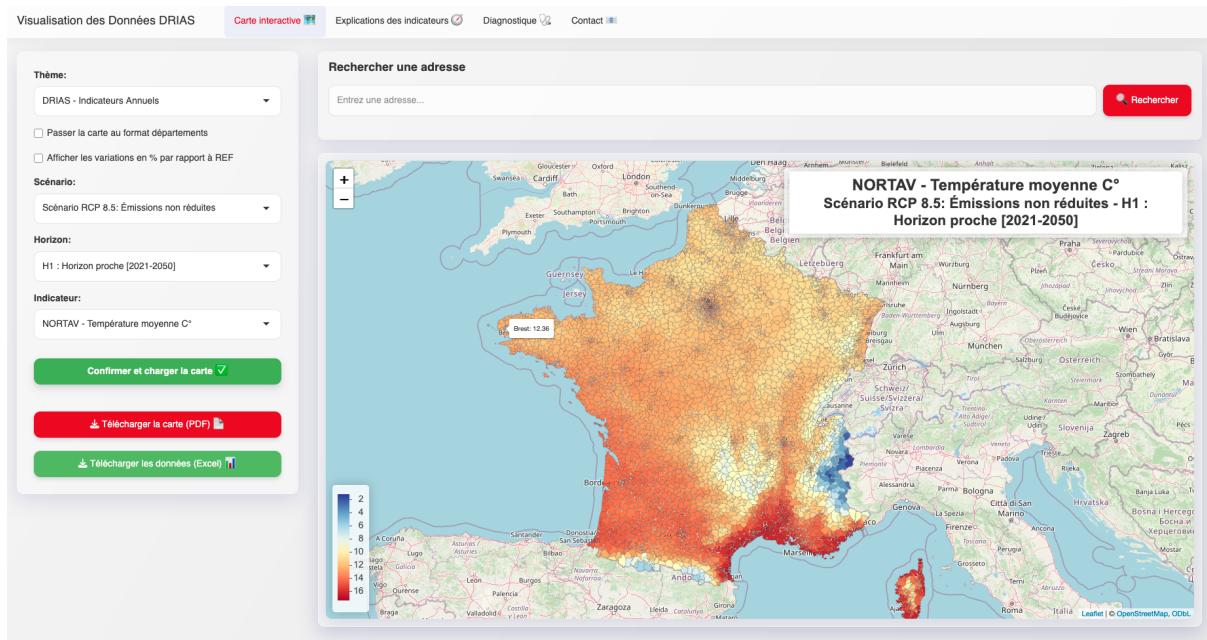


FIGURE 4 – Capture d'écran de l'application en ligne

La figure 4 illustre l'interface visuelle de l'application. Celle-ci se structure autour d'une carte centrale, accompagnée, sur la partie droite de l'écran, de modules de sélection permettant à l'utilisateur de choisir les indicateurs climatiques, les horizons temporels et les scénarios climatique mais aussi de télécharger la carte au format PDF et les données au format Excel. Dans la partie supérieure de l'interface, plusieurs onglets sont accessibles, dont un dédié à l'explication des

indicateurs utilisés.



FIGURE 5 – Capture d'écran de l'application en ligne

L'utilisateur a la possibilité de saisir une adresse (voir figure 5) en cliquant sur le bouton "Chercher". Cette action interroge l'API de la Base Adresse Nationale (BAN) qui localise l'adresse saisie. Une fois l'adresse identifiée, elle est positionnée sur la carte. L'utilisateur peut alors accéder au diagnostic climatique en cliquant sur "Voir le diagnostic climatique", ce qui redirige vers l'onglet "Diagnostique". Dans cet onglet, il est possible de télécharger le diagnostic au format PDF (figure 6).

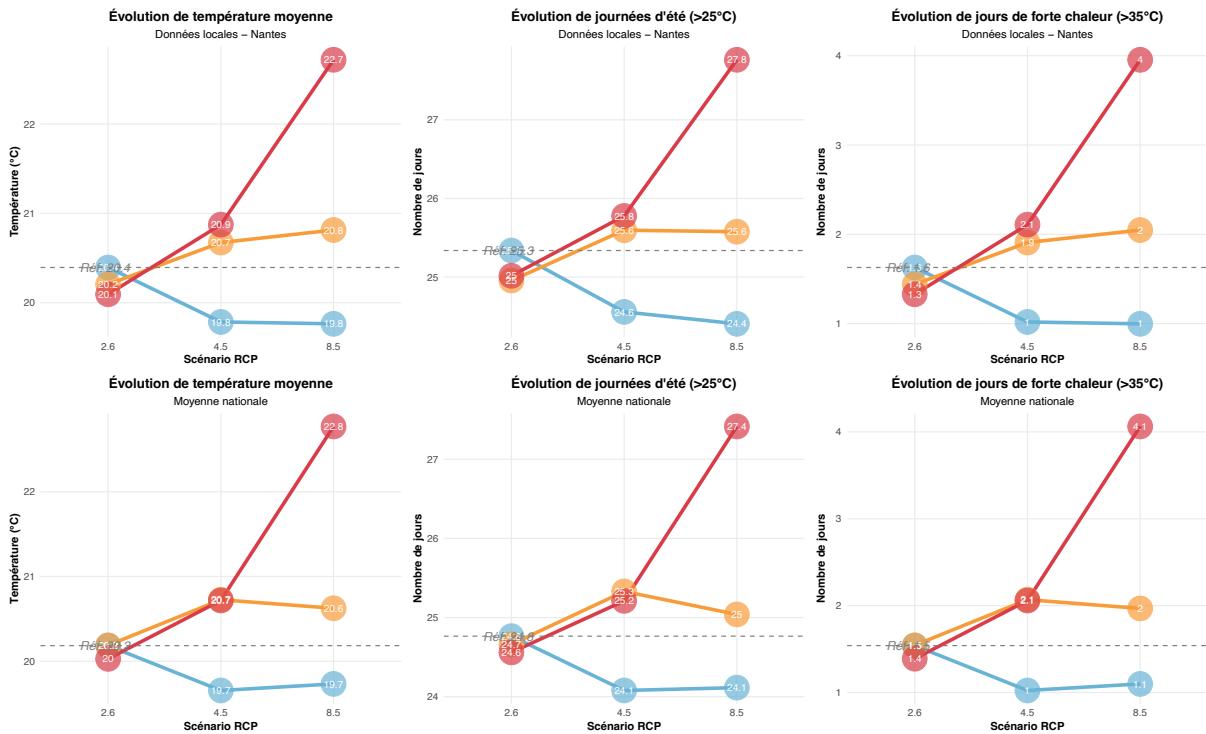


FIGURE 6 – Capture d'écran de l'application en ligne

Les graphiques présentés dans la figure 6 affichent les indicateurs climatiques issus de la base DRIAS, à l'échelle du code postal correspondant à l'adresse saisie, et selon différents scénarios RCP. Les graphiques situés dans la partie supérieure représentent les données spécifiques à l'adresse saisie (niveau code postal), tandis que ceux situés en bas de page correspondent aux moyennes nationales, permettant ainsi une mise en perspective des résultats locaux à l'échelle du territoire français.

L'application a été très bien reçue par l'équipe ESG, qui l'a utilisée dans différents contextes internes. Des discussions ont eu lieu sur sa sécurisation et son intégration au réseau Arkéa, mais sans aboutissement technique final.

3.1.3 Étude sur les travaux de la NGFS

Dans le cadre de la construction d'un score sectoriel du risque climatique, j'ai également travaillé sur les scénarios de la NGFS (Network for Greening the Financial System), notamment ceux publiés en mai 2025. Ces scénarios, bien que très complets, s'apparentent à une modélisation "boîte noire", rendant leur appropriation complexe.

J'ai donc analysé la [documentation technique](#) dans le but de produire des schémas explicatifs pour en faciliter la compréhension, tant pour moi que pour l'équipe (annexes 32 et 33). Ce travail a permis de proposer une intégration des scénarios NGFS dans la construction du score sectoriel, notamment dans une logique de stress-testing bancaire, en cohérence avec les objectifs d'Arkéa.

Cette mission m'a conduit à solliciter directement la NGFS pour clarifier certains points méthodologiques.

3.1.4 Modélisation économétrique : Impact du DPE sur le prix immobilier

Enfin, j'ai mené une étude économétrique visant à quantifier l'impact de l'étiquette DPE (Diagnostic de Performance Énergétique) sur le prix des biens immobiliers, uniquement en Bretagne, à la demande d'Arkéa.

À partir de la [BDNB \(Base de Données Nationale des Bâtiments\)](#), enrichie des données de la [DVF \(Demande de valeurs foncières\)](#) et de la base [DPE](#), j'ai modélisé la valeur foncière d'un bien (Y) en fonction de plusieurs variables explicatives (X_i) :

- **Caractéristiques structurelles** : nombre de pièces, d'étages, surfaces, année de construction
- **Données socio-démographiques** : taux de propriétaires, logements vacants, revenu médian communal
- **Accessibilité géographique** : distance aux espaces naturels (bleus, verts) et aux nuisances (espaces "rouges")
- **Exposition aux risques naturels** (nombre de sinistres CATNAT)

Deux modèles distincts ont été élaborés : un pour les maisons et un autre pour les appartements. Les résultats ont été présentés graphiquement et réservés à la région Bretagne, ce qui représentait déjà un volume important de données, (voir annexes [34](#), [35](#), [36](#), [37](#)).

Les résultats présentés en annexe mettent en évidence l'effet du DPE sur la valeur des biens immobiliers, en comparaison avec l'étiquette de référence **D**. Les écarts estimés sont les suivants :

- **Étiquette A** : +0,2% pour les maisons et +10,3% pour les appartements ;
- **Étiquette B** : +8,1% pour les maisons et +17,6% pour les appartements ;
- **Étiquette C** : +3,2% pour les maisons et +11,6% pour les appartements ;
- **Étiquette E** : -7,6% pour les maisons et -4,3% pour les appartements ;
- **Étiquette F** : -15,8% pour les maisons et -8,2% pour les appartements ;
- **Étiquette G** : -24,8% pour les maisons et -14,6% pour les appartements.

3.1.5 Ratios de perte par aléas climatique

Dans le prolongement des travaux menés sur l'immobilier, j'ai procédé, à la demande de l'équipe, à la fusion de la base de données des sinistres de Suravenir Assurances avec la base immobilière du CMA. Cette intégration a permis de calculer des ratios de pertes associés aux différents aléas climatiques. Le ratio est défini comme :

$$\left(\frac{\text{montant du sinistre pour un bien}}{\text{valeur du bien}} \right) \times 100$$

L'application de cette méthode à l'ensemble du parc immobilier répertorié dans la base du CMA permet d'estimer un ratio moyen par type d'aléa, avec une distinction entre maisons individuelles et appartements. Ainsi, il ressort par exemple qu'en moyenne, une inondation engendre des dommages correspondant à environ 5% de la valeur du bien concerné (*voir table 16*).

TABLE 16 – Ratios de sinistres pour les inondations (en pourcentage de la valeur du bien)

	Moyenne	1er quartile	Médiane	3ème quartile	Maximum	<i>n</i>
Global	5,44%	0,45%	1,23%	4,10%	136,40%	944
Maisons	6,04%	0,49%	1,33%	4,44%	136,40%	693
Appartements	3,78%	0,32%	0,96%	2,94%	107,78%	251

À la suite de cette étape, il a été décidé d'intégrer les quartiles issus de l'outil PRISM dans l'analyse. L'objectif était d'identifier, parmi les adresses appartenant à un quartile donné du tableau 16, la répartition des niveaux de risque 0, 1, 2, 3 et 4 selon la classification de PRISM. Les résultats correspondants sont présentés dans le tableau 17.

TABLE 17 – Ratios de sinistres pour les inondations par niveau de risque PRISM (en pourcentage de la valeur du bien)

Catégorie	Moyenne	1er quartile	Médiane	3ème quartile	Maximum
Global	5,44%	0,45%	1,23%	4,10%	136,40%
Risque PRISM 0	-	2,96%	2,55%	3,40%	0,84%
Risque PRISM 1	-	2,96%	2,55%	4,60%	7,62%
Risque PRISM 2	-	31,50%	35,31%	37,02%	29,66%
Risque PRISM 3	-	11,86%	20,85%	17,87%	17,79%
Risque PRISM 4	-	50,84%	38,72%	37,01%	44,06%

La lecture du tableau 17 peut se faire de la manière suivante :

- Parmi les adresses composants les ratios entre le minimum et le 1er quartile, 2.96% présentent un risque PRISM de niveau 0, 31.35% un risque PRISM de niveau 2 et 50.84% un risque PRISM de niveau 4.
- Parmi les adresses composants les ratios composant les ratios entre le 1er quartile et la médiane, 20.85% présentent un risque PRISM de niveau 3 tandis que 38.72% présentent un risque PRISM de niveau 4, etc...

Enfin il a été de décider de regarder les ratios moyens par niveau de risque PRISM, afin de voir s'ils sont significatifs :

TABLE 18 – Analyse des ratios de sinistres et charges par niveau de risque PRISM pour les inondations

Niveau de risque PRISM	Ratio (%)			Charge (€)			n
	Méd.	Moy.	Max	Méd.	Moy.	Max	
0	0.95	2.17	14.6	2042	2654	10 237	23
1	2.57	4.80	28.0	2736	7264	42 013	42
2	1.23	4.29	100.0	1386	2437	12 992	314
3	1.36	4.83	136.0	1766	3780	33 913	161
4	1.11	4.88	115.0	1427	3962	64 410	402

La lecture du tableau 18 peut se faire de la manière suivante :

- Parmis les adresses en risque PRISM de niveau 0, le ratio (montant du sinistre / valeur du bien) est en moyenne de 2,17% avec une charge moyenne de 2654€.
- Parmis les adresses en risque PRISM de niveau 4, le ratio (montant du sinistre / valeur du bien) est en moyenne de 4,88% avec une charge moyenne de 3962€.

Cette analyse a été établie sur trois aléas : inondations, sécheresses et tempêtes.

3.2 Gestion du temps

Il est difficile d'établir un emploi du temps hebdomadaire précis, dans la mesure où les semaines de stage étaient relativement hétérogènes en termes de charge et de nature des tâches. La figure 7 présente une vision synthétique, sous la forme d'un emploi du temps mensuel couvrant l'ensemble des six mois de stage. Ce graphique met en évidence l'évolution des principales missions menées, en omettant volontairement certaines tâches secondaires afin de préserver la lisibilité de l'ensemble.

De manière générale, la répartition du temps de travail a été bien équilibrée. Même durant les périodes moins actives du projet PRISM, la liberté accordée par mon tuteur pour explorer des pistes de travail complémentaires m'a permis de maintenir un rythme constant et de faire progresser la réflexion globale du projet.

Par ailleurs, durant les congés de mon tuteur en Juillet, j'ai assumé la responsabilité de représenter l'équipe lors de certaines réunions, notamment celles portant sur la construction du score géographique en collaboration avec la CNCM. À cette occasion, il m'a été confié la tâche de proposer de nouveaux indicateurs, de sélectionner des bases de données pertinentes et de suggérer des pondérations adaptées en fonction des besoins identifiés, assurant ainsi la continuité des travaux en son absence.

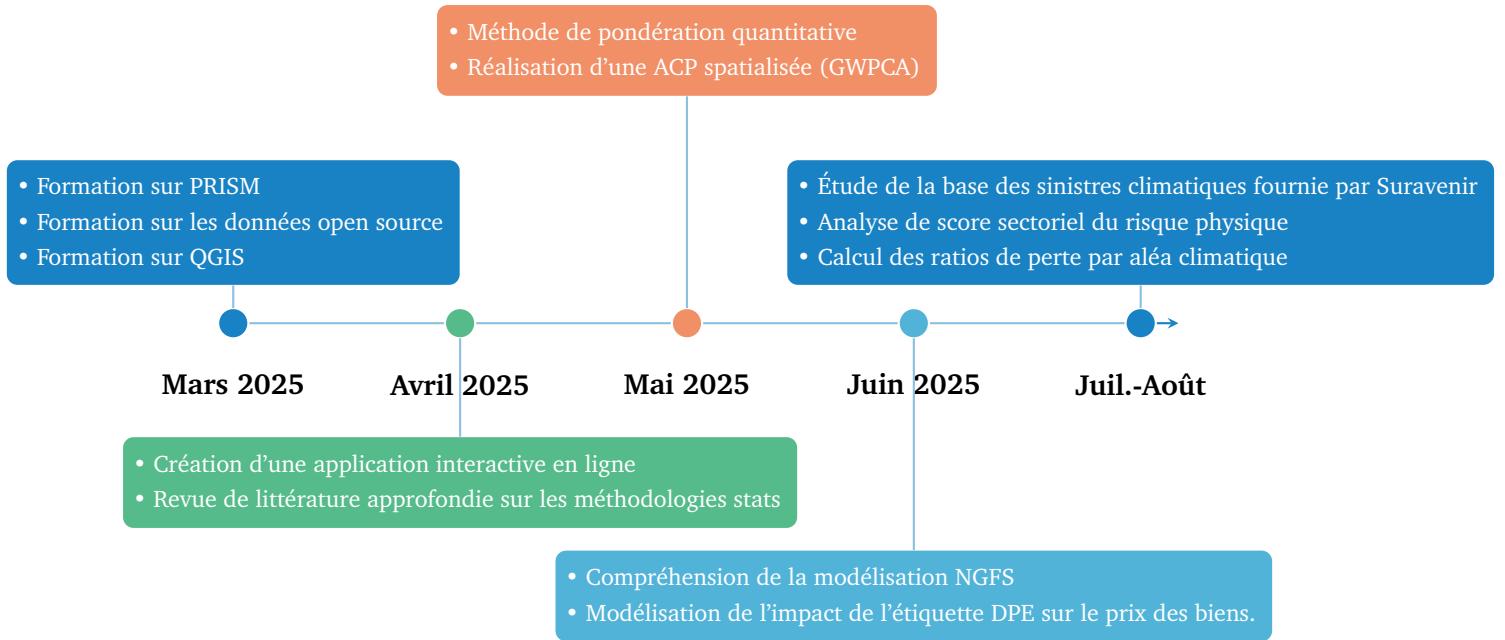


FIGURE 7 – Calendrier mensuel du stage (Mars à Aout 2025)

4 Bilan

Ce stage m'a offert l'opportunité de renforcer mes compétences dans de nombreux domaines, en particulier dans l'analyse des aléas climatiques et la manipulation de données spatiales.

Tout d'abord, j'ai pu approfondir mes connaissances en matière de modélisation des aléas naturels. Cela m'a conduit à explorer les différentes bases de données open source disponibles, à en comprendre le fonctionnement, et à réfléchir à la manière de modéliser ces aléas dans une perspective pertinente pour le secteur bancaire. J'ai ainsi étudié les scénarios climatiques envisageables, les horizons temporels adaptés, ainsi que des notions clés telles que les périodes de retour. Cette démarche m'a également permis de mieux appréhender les différentes typologies d'aléas (*différencier les inondations par remontée de nappe, débordement de cours d'eau, ruissellement, submersion marine, etc.*) et d'acquérir une capacité d'analyse permettant de produire des indicateurs pertinents à partir de ces informations.

La familiarisation avec les bases de données open source, qu'elles relèvent du domaine environnemental, climatique, immobilier ou autre, constitue un acquis précieux, dont je suis convaincu qu'il me sera utile dans mes futures expériences professionnelles.

Par ailleurs, une large part de mon travail a porté sur des données géographiques. L'utilisation du système d'information géographique QGIS a été centrale dans cette démarche. Bien que j'aie déjà été introduit à l'analyse spatiale grâce à des cours d'économétrie spatiale, ce stage m'a permis de monter en compétence de manière significative, tant dans l'exploitation des données spatiales via QGIS que dans leur traitement en langage Python.

J'ai également été sensibilisé aux enjeux de la finance durable à travers une introduction au fonctionnement des critères ESG, ainsi qu'aux rapports et réflexions associés à chacune de ces dimensions au sein de l'équipe.

Je tiens à souligner l'enthousiasme et la disponibilité de mon tuteur, qui m'a accordé une grande autonomie pour explorer des pistes de travail au cours des différents projets. Cette liberté m'a permis de consolider et d'élargir mes connaissances de manière continue.

Enfin, cette expérience m'a confronté aux réalités du monde de l'entreprise, où les données disponibles sont rarement parfaites. Bien que j'aie été préparé à cela par une précédente expérience de stage à l'Ifremer, j'ai à nouveau constaté la nécessité d'adopter des approches pragmatiques, parfois éloignées du cadre académique strict. Cela suppose de savoir adapter ses méthodes, voire de transgresser certaines règles statistiques, afin de faire avancer les projets de manière opérationnelle. Cette flexibilité, bien qu'exigeante, me semble constituer une compétence essentielle dans un contexte professionnel.

Deuxième partie

ÉTUDE STATISTIQUE

Cadre de l'étude

Après avoir examiné l'impact de l'étiquette DPE (Diagnostic de Performance Énergétique) ainsi que d'autres caractéristiques structurelles, socio-démographiques et géographiques sur le prix de vente des maisons et appartements en Bretagne (*voir section 3.1.4*), nous cherchons désormais à affiner notre analyse. En effet, dans l'étude initiale, nous avions recours à un modèle de régression linéaire multiple relativement simple, dans lequel la dimension géographique était prise en compte par l'ajout d'un effet fixe au niveau départemental.

Compte tenu de la taille importante du jeu de données initial (environ 68 000 observations), la mise en œuvre de modèles d'économétrie spatiale s'est avérée techniquement inenvisageable. En effet, ces modèles impliquent notamment des opérations lourdes en calcul, telles que l'inversion de matrices de grande dimension, ce qui engendre des contraintes significatives en termes de temps de traitement et de ressources informatiques. Afin de contourner ces limitations et de permettre une modélisation spatiale rigoureuse, l'analyse a été restreinte à un sous-échantillon plus maniable, en se concentrant exclusivement sur le périmètre de la ville de Brest.

Dans ce contexte, nous menons une analyse en économétrie spatiale, afin de comparer les résultats obtenus à ceux précédemment observés à l'échelle de la Bretagne. Cette étape sera suivie par l'application de techniques de modélisation issues du Machine Learning, combinées à des méthodes d'explicabilité. L'objectif est de mieux comprendre les déterminants du prix de vente à la fois de manière globale (tendances générales) et individuelle (cas par cas). Cette approche exploratoire vise à poser les bases d'une analyse plus étendue, susceptible d'alimenter les réflexions stratégiques du Crédit Mutuel Arkéa.

Marché immobilier breton, Diagnostic de performance énergétique (DPE), Prix de l'immobilier, Ville de Brest, Économétrie spatiale, Modèle SEM, SAR, SLX, SDM, Machine-Learning, Explicabilité

1 Introduction

Le marché immobilier breton, à l'instar du reste du territoire national, connaît une phase d'ajustement depuis l'été 2022, caractérisée par une baisse des volumes de transactions de 19% et une contraction des prix, de 3% pour les maisons et de 4% pour les appartements, en l'espace d'une année [3]. Dans ce contexte de rééquilibrage, l'influence du Diagnostic de Performance Énergétique (DPE) est devenue un critère structurant, redéfinissant les attentes des acquéreurs et la valeur des biens. Un logement est qualifié de "passoire thermique" lorsque son DPE est classé F ou G, signifiant une isolation insuffisante et des systèmes de chauffage inefficaces, entraînant une déperdition significative de chaleur, une consommation d'énergie élevée et une empreinte carbone accrue. Ces logements subissent une décote moyenne de 15% sur leur prix de vente, soit environ -452€ /m² par rapport aux biens classés D, et une marge de négociation plus élevée, atteignant 5,9%. Depuis juillet 2021, les prix des logements bien notés (A à D) ont augmenté de 3,4%, tandis que ceux des biens F ou G ont chuté de 3,5%, et leur délai de vente est en moyenne plus long de cinq jours. Toutefois, ces biens énergivores suscitent un intérêt paradoxalement élevé en raison de leurs prix plus accessibles, générant 67% de contacts supplémentaires pour les annonces de logements classés G par rapport aux biens de classe D [18].

La loi Climat et Résilience est un moteur essentiel de cette transformation du marché, introduisant des échéances progressives d'interdiction de location des logements les moins performants. Les propriétaires de logements classés G+ (consommation supérieure à 450 kWh/m²/an) ne peuvent plus conclure de nouveaux baux depuis janvier 2023, et cette interdiction s'est étendue à tous les logements classés G depuis le 1er janvier 2025. Les logements classés F seront concernés en 2028 et ceux classés E en 2034. En France, 600 000 logements classés G sont actuellement loués et directement affectés par cette interdiction de nouveaux baux [12]. Bien que le renouvellement de bail reste possible, le locataire est en droit d'exiger des travaux de mise en conformité. Cette législation a déjà entraîné une baisse significative de l'offre de passoires thermiques à la location, avec une diminution de près de 64% pour cent depuis janvier 2021. Face à ces contraintes, de nombreux propriétaires retirent leurs biens du marché locatif pour les mettre en vente, les laisser vacants, ou les transformer en locations saisonnières, une catégorie non encore soumise aux mêmes interdictions.

Étudier cette dynamique s'avère essentiel pour comprendre comment les réglementations environnementales, notamment celles liées à la performance énergétique des logements, transforment le fonctionnement du marché immobilier, influencent les comportements d'achat et de vente, et modifient l'offre disponible, en particulier dans un contexte de tension structurelle. Une telle analyse permet au Crédit Mutuel Arkéa (CMA) d'anticiper les évolutions du marché, d'adapter ses politiques de financement immobilier et de mieux évaluer les risques associés aux biens dévalorisés ou moins conformes aux exigences réglementaires. Ces évolutions ont un impact direct sur le coût du risque, et donc sur la rentabilité de la banque. En intégrant ces dimensions, la banque renforce sa capacité à proposer des solutions de crédit cohérentes avec les enjeux économiques, sociaux et environnementaux, tout en accompagnant au mieux ses clients dans la transition énergétique du parc immobilier.

Cette étude s'inscrit dans le cadre d'une demande formulée par l'actuel directeur des risques du groupe Arkéa, Benoît Genest. Celui-ci souhaitait évaluer l'impact de l'étiquette du Diagnostic de Performance Énergétique (DPE) sur les prix de vente des biens immobiliers, en particulier depuis l'entrée en vigueur de la réglementation DPE 2025. Étant donné que le portefeuille immobilier d'Arkéa est majoritairement situé en Bretagne, il a été jugé pertinent de concentrer l'analyse sur cette région.

L'analyse portant sur la Bretagne étant terminée (*voir annexes 34, 35, 36, 37*), il a été décidé, afin d'approfondir l'approche méthodologique, de recourir à des modèles d'économétrie spatiale, mieux adaptés à la nature géographique des données. Toutefois, en raison du temps de calcul nécessaires à ces modèles sur des grands jeux de données (*notamment à cause des inversions de matrices*), cette approche sera limitée à la ville de Brest, qui présente un volume de données suffisant pour garantir la robustesse des résultats. Dans un second temps, des modèles d'apprentissage automatique seront utilisés. Ceux-ci permettront d'enrichir l'analyse en offrant des perspectives à la fois globales et locales sur les déterminants du prix des logements, notamment à travers leurs capacités **d'interprétabilité locale**.

2 Analyse exploratoire

2.1 Évolution des ventes

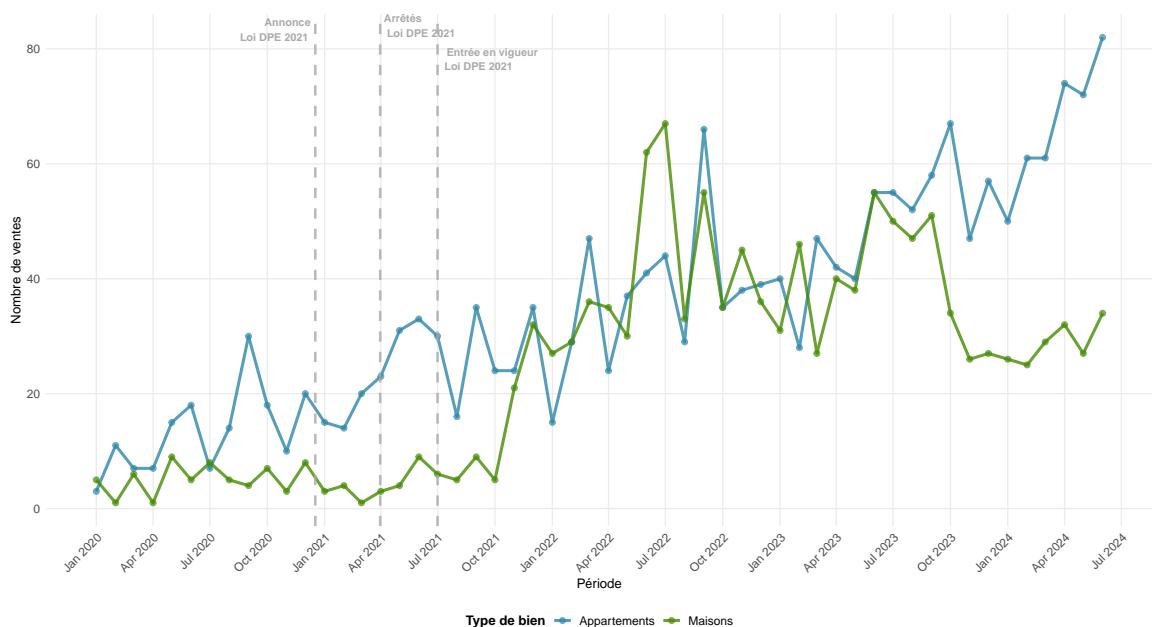


FIGURE 8 – Évolution des ventes immobilières à Brest par type de bien

La figure 8 illustre l'évolution temporelle du nombre de ventes immobilières à Brest, en distinguant les appartements et les maisons. On observe une relative stabilité du volume de transactions jusqu'à l'annonce et surtout l'entrée en vigueur de la loi DPE 2021. Cette dynamique suggère un effet d'anticipation ou d'ajustement du marché face à la réglementation, particulièrement marqué pour les maisons. Cette baisse des ventes peut être interprétée comme une réaction du marché à l'incertitude ou à la contrainte posée par l'interdiction progressive de mise en location des biens classés F et G.

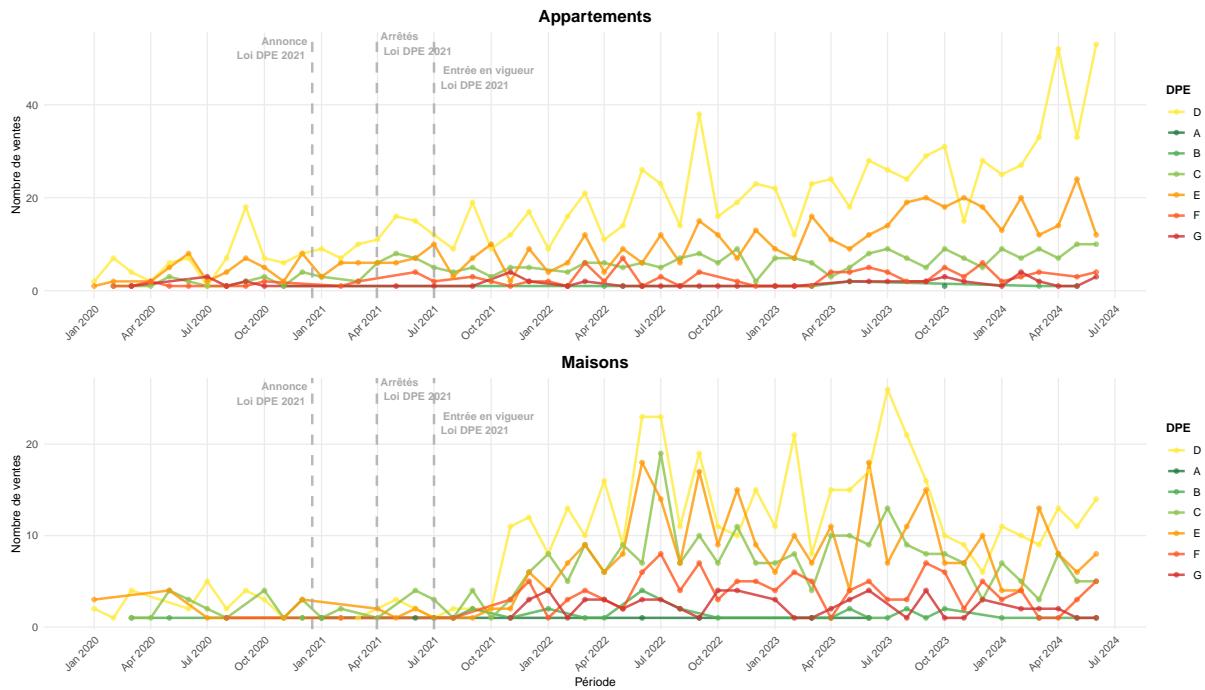


FIGURE 9 – Évolution des ventes immobilières à Brest par classe DPE et type de bien

La figure 9 détaille l'évolution du nombre de ventes par classe DPE, séparément pour les appartements et les maisons. On remarque une forte prédominance des classes D et E dans les transactions, reflet d'un parc immobilier ancien. Les biens classés F et G, bien que présents, sont progressivement moins échangés, en particulier après les arrêtés et la mise en œuvre de la loi DPE 2021. Cette diminution pourrait indiquer une désaffection du marché pour les logements énergivores, probablement en raison des restrictions réglementaires imminentes et du coût associé à leur rénovation.

2.2 Prix des biens (€)

Les figures 10 et 11 cartographient les prix des logements par secteur. Les zones les plus centrales et proches du littoral présentent les valeurs les plus élevées, suggérant un effet positif de la localisation sur les prix. À l'inverse, les zones périphériques ou excentrées affichent des prix nettement plus bas. Cette spatialisation est cohérente avec les dynamiques classiques d'un marché urbain, où l'accessibilité et l'attractivité locale jouent un rôle déterminant.

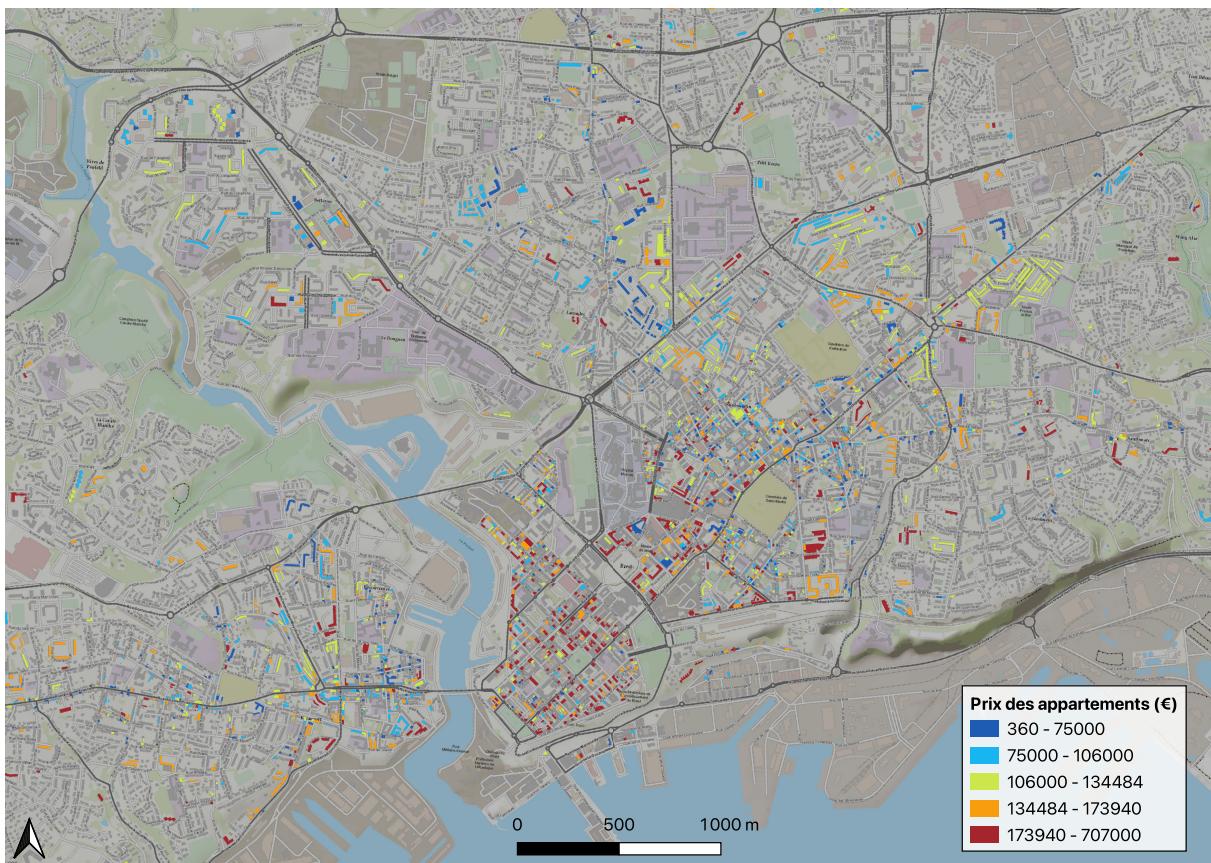


FIGURE 10 – Cartographie des prix des appartements à Brest

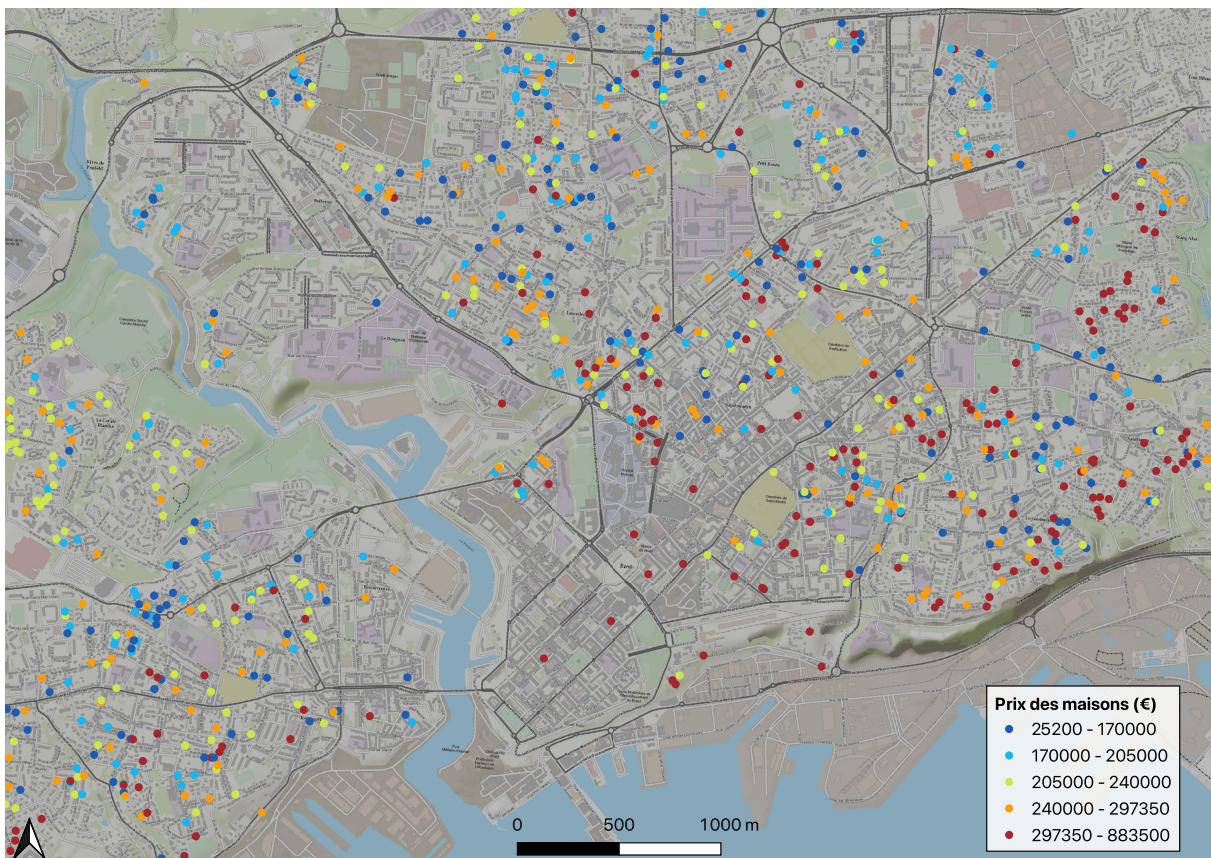


FIGURE 11 – Cartographie des prix des maisons à Brest

La figure 12 trace l'évolution des prix moyens des appartements et maisons à Brest. On y observe une tendance haussière continue depuis 2020, avec une accélération marquée à partir de 2021. Malgré une légère stabilisation récente, les prix n'ont pas connu de correction majeure. Ce mouvement pourrait refléter une tension sur le marché de l'offre, conjuguée à un effet d'ajustement suite à la réforme DPE.

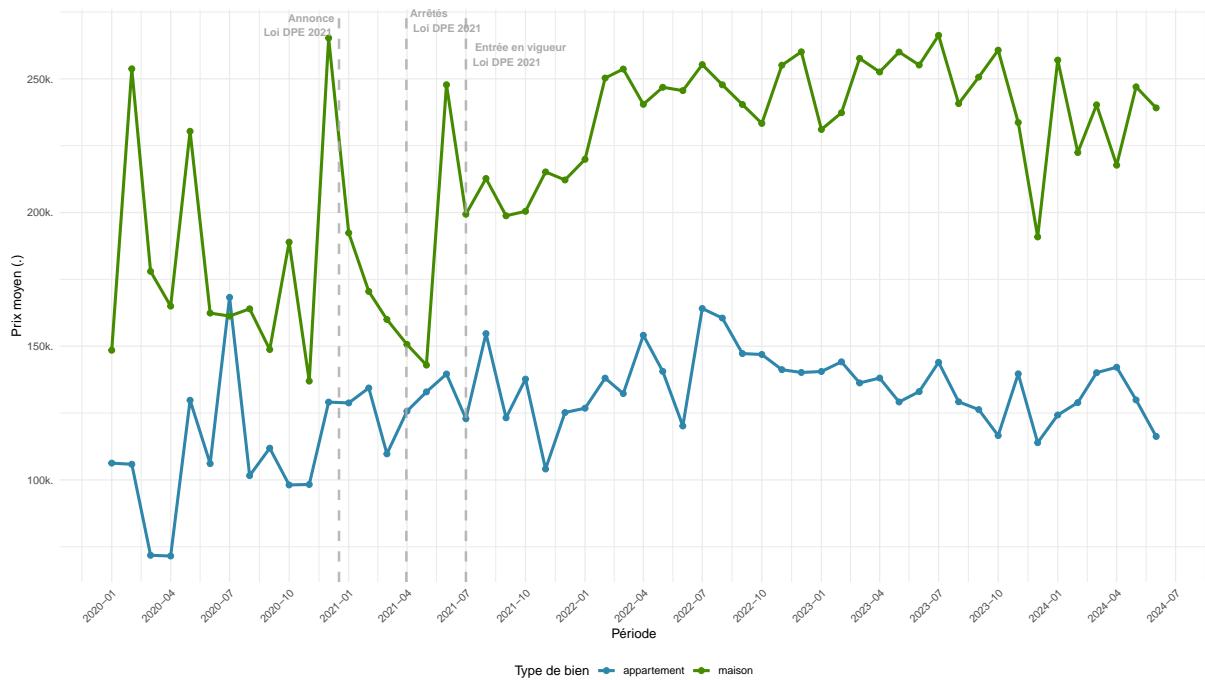


FIGURE 12 – Évolution des prix à Brest par type de biens

La figure 13 met en évidence des écarts de prix significatifs entre les classes DPE. Les logements classés A à C sont nettement plus chers en moyenne que ceux classés E à G, avec une hiérarchie relativement linéaire. Cette tendance suggère l'existence d'une prime énergétique, où l'efficacité thermique perçue est valorisée par le marché, en cohérence avec les politiques publiques incitatives.

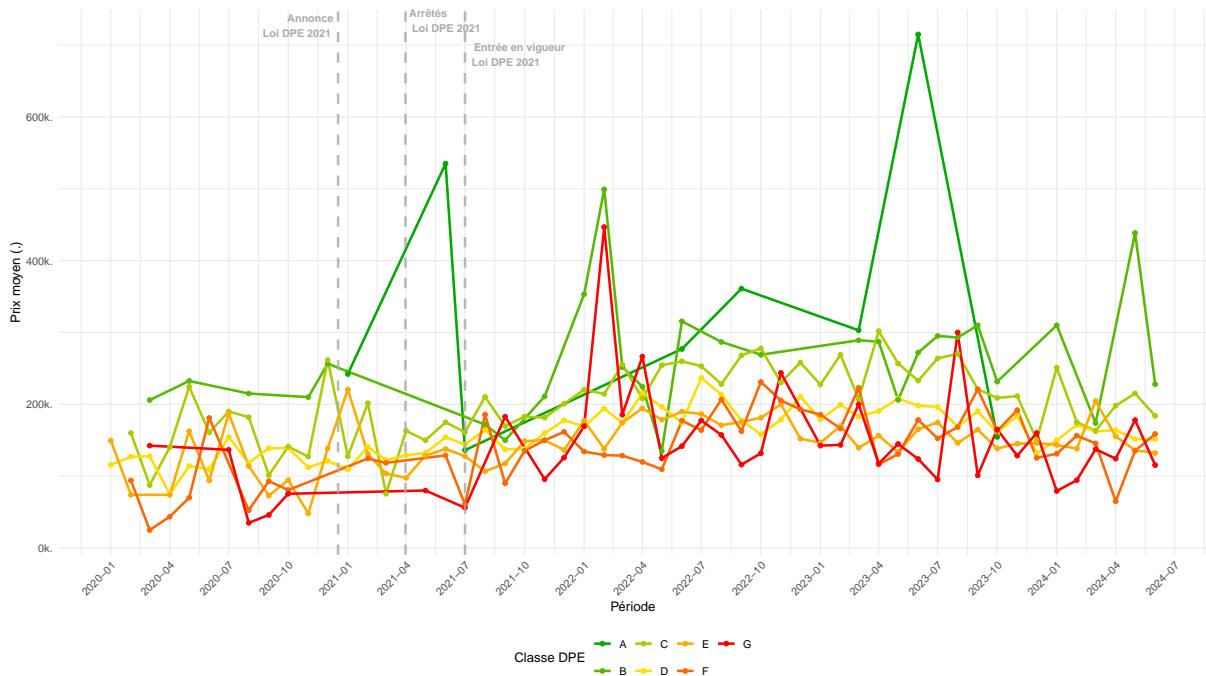


FIGURE 13 – Évolution des prix à Brest par classe DPE

Les figures 14 et 15 montrent une distribution asymétrique et étalée des prix, avec une queue droite longue, ce qui est typique des données immobilières. Les distributions indiquent une concentration des prix autour de la médiane, mais une présence notable de biens haut de gamme, en particulier pour les maisons.

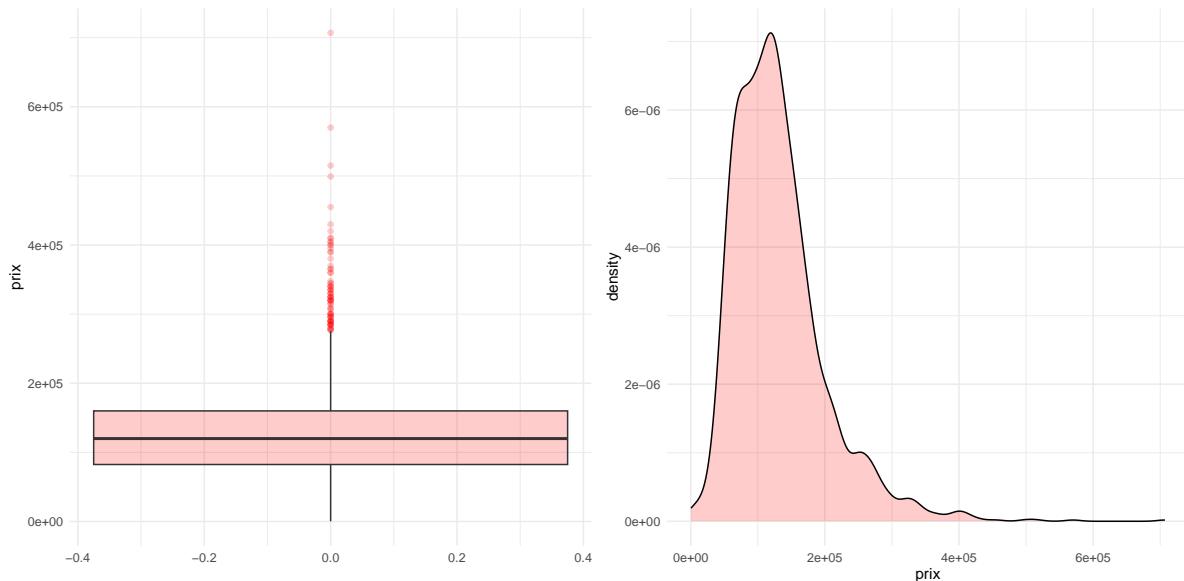


FIGURE 14 – Distribution des prix des appartements à Brest

Min	1er quartile	Médiane	Moyenne	3ème quartile	Max
360	82 500	120 000	131 424	160 061	707 000

TABLE 19 – Statistiques descriptives des prix (€) des appartements

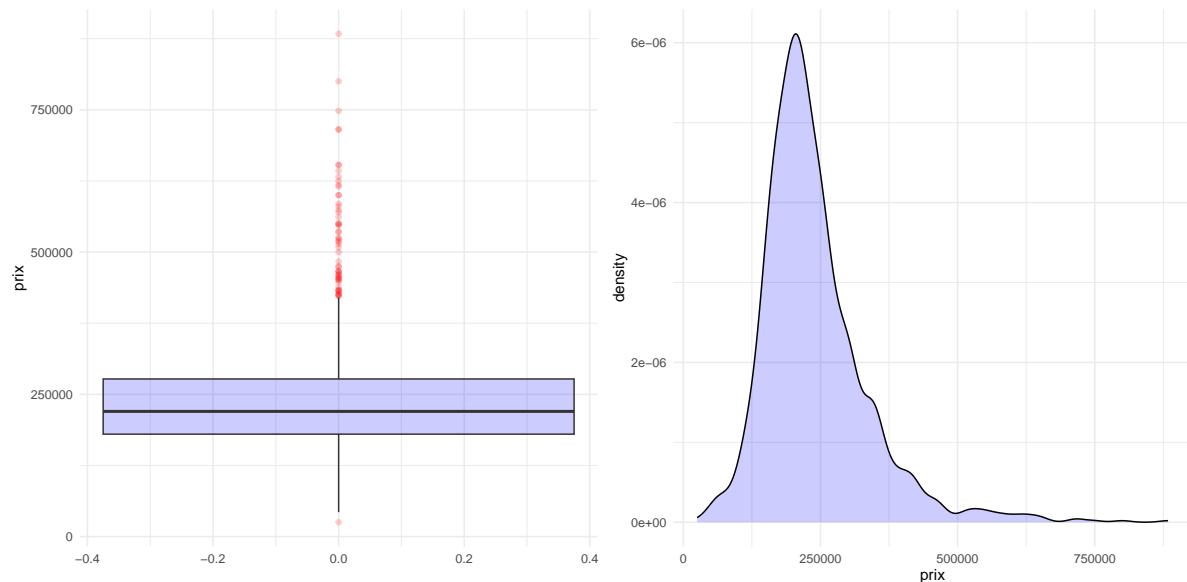


FIGURE 15 – Distribution des prix des maisons à Brest

Min	1er quartile	Médiane	Moyenne	3ème quartile	Max
25 200	180 000	220 000	239 524	277 000	883 500

TABLE 20 – Statistiques descriptives des prix des maisons

L'analyse des statistiques descriptives présentées dans les tableaux 20 et 19 met en évidence des différences notables entre les prix des maisons et ceux des appartements. Pour les maisons, les prix s'échelonnent de 25 200 € à 883 500 €, avec une médiane de 220 000 € et une moyenne légèrement supérieure, traduisant une distribution modérément asymétrique vers les valeurs élevées. Les appartements affichent une gamme de prix plus restreinte, allant de 360€ (*probablement un outlier*) à 707 000 €, avec une médiane de 120 000 € et une moyenne proche (131 424 €), indiquant une distribution moins étirée vers les valeurs extrêmes que celle des maisons. Dans les deux cas, l'écart important entre les valeurs minimales et maximales suggère une forte hétérogénéité des biens, probablement liée à des différences en termes de localisation, de surface et de caractéristiques intrinsèques.

2.3 Étiquettes DPE

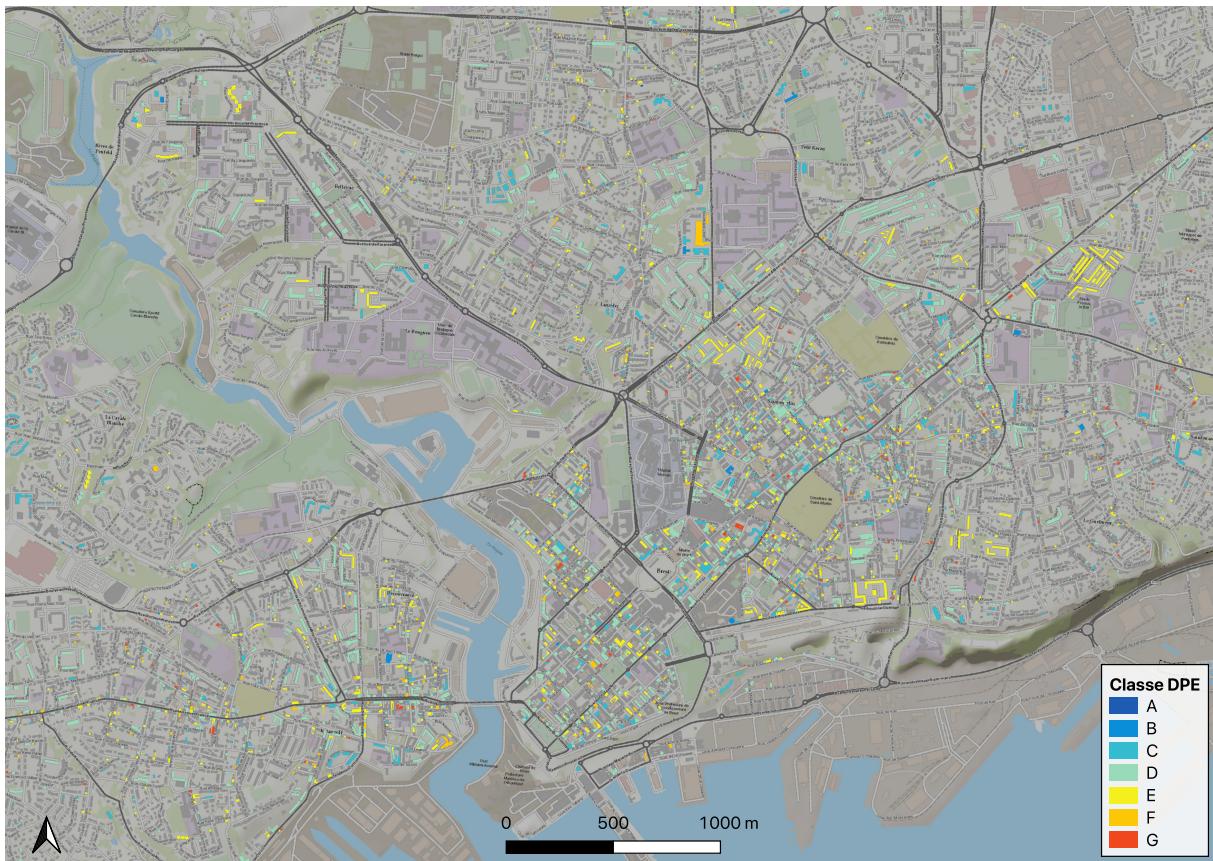


FIGURE 16 – Cartes des étiquettes DPE à Brest : appartements et maisons

Cette carte est une représentation spatiale de la répartition des étiquettes DPE sur le territoire de Brest. On constate une concentration des classes les plus performantes (A à C) dans certains quartiers centraux ou récents, tandis que les logements les moins performants (E à G) sont majoritairement localisés dans des zones périphériques ou anciennes. Cette hétérogénéité spatiale souligne l'importance de prendre en compte l'effet quartier dans l'analyse des prix.

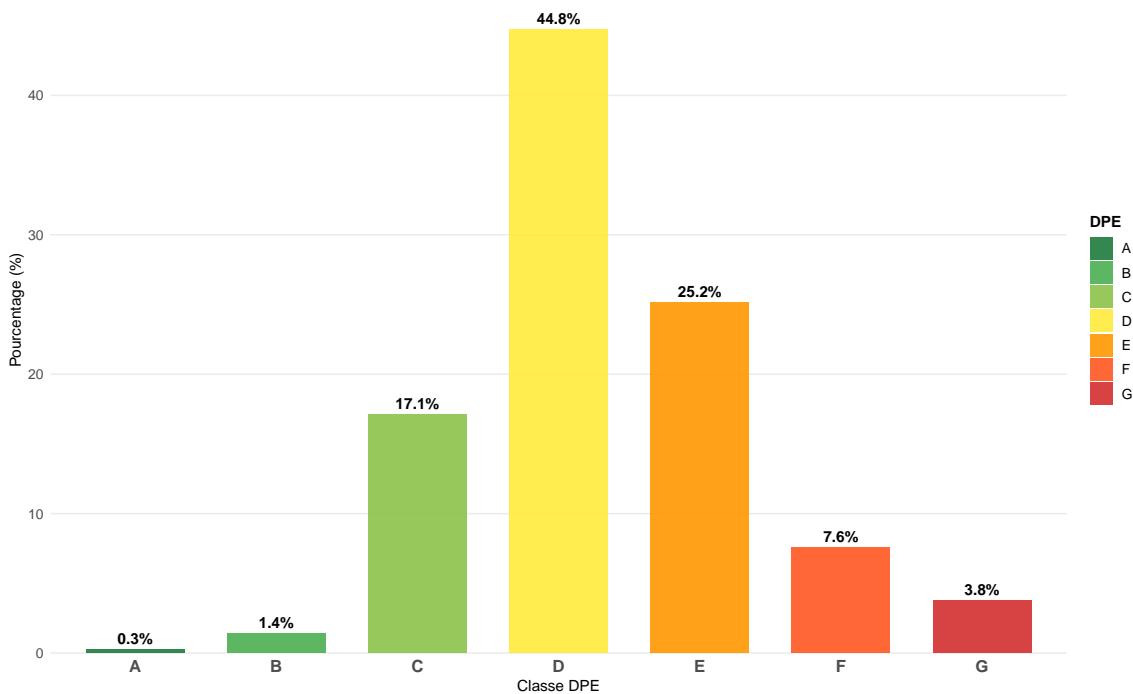


FIGURE 17 – Distribution globale des étiquettes DPE à Brest

La figure 17 présente la distribution agrégée des classes DPE sur l'ensemble du marché immobilier brestois. La classe D domine largement, représentant près de 45% des logements, suivie des classes E et C. À l'inverse, les classes A et B sont extrêmement rares, ce qui traduit la faible présence de bâtiments très récents ou rénovés selon des standards énergétiques élevés. Cette distribution témoigne d'un parc immobilier globalement peu efficient sur le plan énergétique.

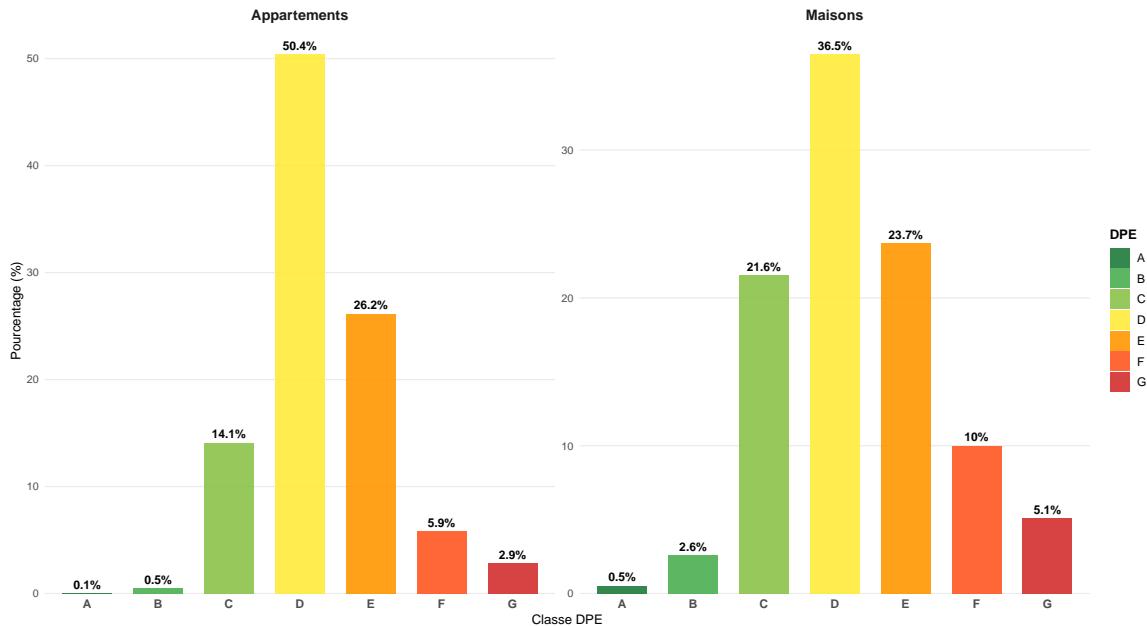


FIGURE 18 – Distribution des étiquettes DPE à Brest : appartements et maisons

La figure 18 compare la distribution des classes DPE selon le type de bien. Les appartements présentent une concentration autour de la classe D (50 %), tandis que les maisons se répartissent plus équitablement entre C, D et E, avec une plus forte présence de l'étiquette C (21,6 %). On observe également une proportion plus importante de maisons en classes F et G, ce qui pourrait refléter une vétusté plus marquée dans ce segment. Cette différence de structure énergétique selon le type de bien est un facteur à considérer dans la modélisation des prix.

2.4 Âge des bâtiments (année)

La distribution des âges montre une concentration autour de 60 à 80 ans pour les deux types de biens, avec un léger pic autour des constructions d'après-guerre. L'âge des bâtiments constitue une variable explicative essentielle, tant pour les performances énergétiques que pour les prix.

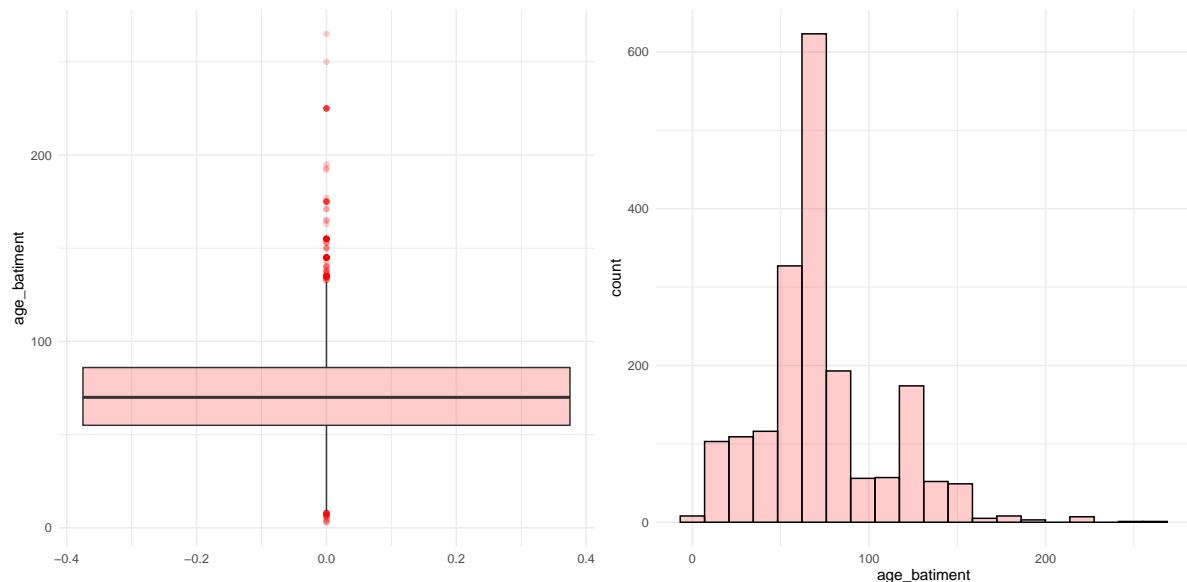


FIGURE 19 – Distribution des âges des appartements à Brest

TABLE 21 – Statistiques descriptives des âges des appartements

Min	1er quartile	Médiane	Moyenne	3ème quartile	Max
3	55	70	74	86	265

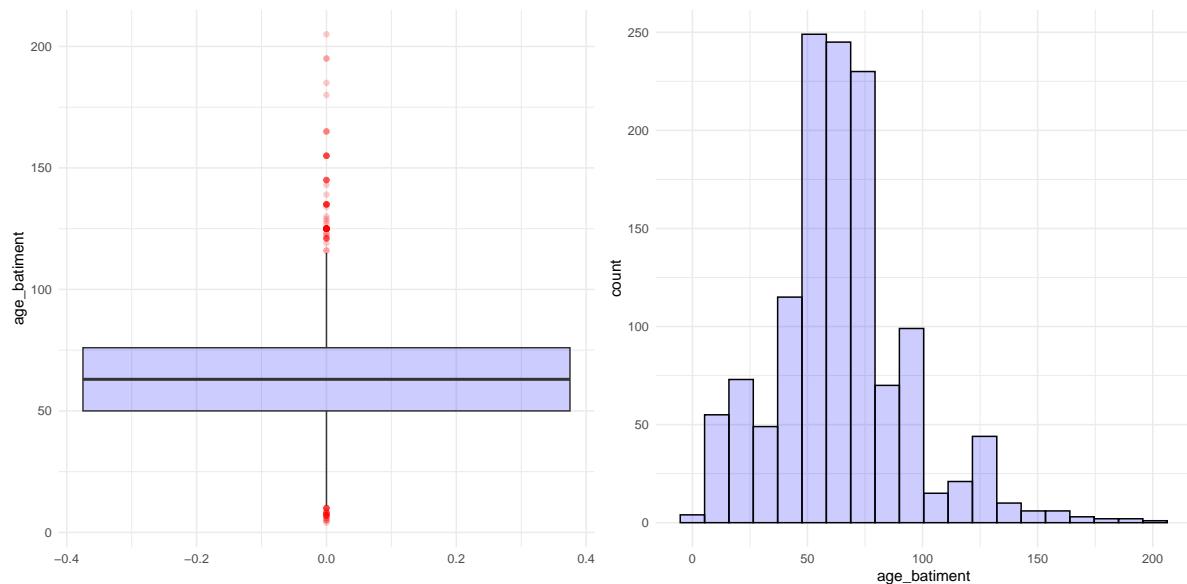


FIGURE 20 – Distribution des âges des maisons à Brest

TABLE 22 – Statistiques descriptives des âges des maisons

Min	1er quartile	Médiane	Moyenne	3ème quartile	Max
4	50	63	65	76	205

Les appartements sont en moyenne légèrement plus anciens que les maisons (74 ans contre 65 ans). Les extrêmes sont importants, certains bâtiments dépassant 200 ans. Cet âge avancé, couplé aux données DPE, pourrait expliquer la forte proportion de "passoires thermiques".

2.5 Surface Habitable (m^2)

Les distributions confirment les observations statistiques : les surfaces sont centrées sur les valeurs médianes, avec une densité maximale autour de 50–60 m^2 pour les appartements et autour de 100–120 m^2 pour les maisons. La forme log-normale de ces distributions est typique du marché immobilier.

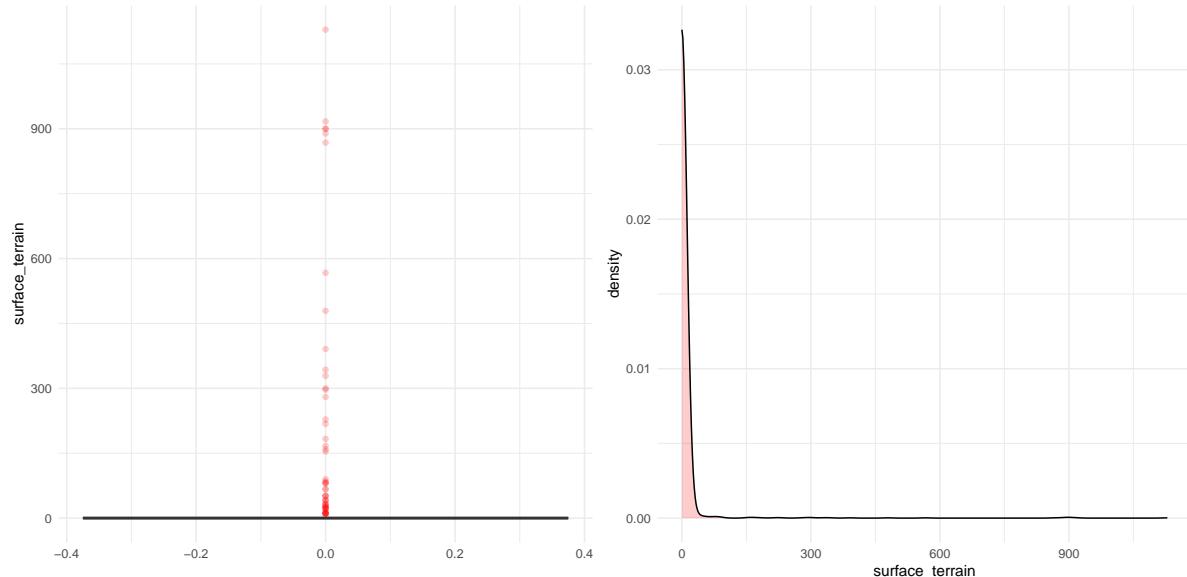


FIGURE 21 – Distribution de la surface habitable des appartements à Brest

TABLE 23 – Statistiques descriptives des surfaces habitables des appartements

Min	1er quartile	Médiane	Moyenne	3ème quartile	Max
10	41	59	64	75	4229

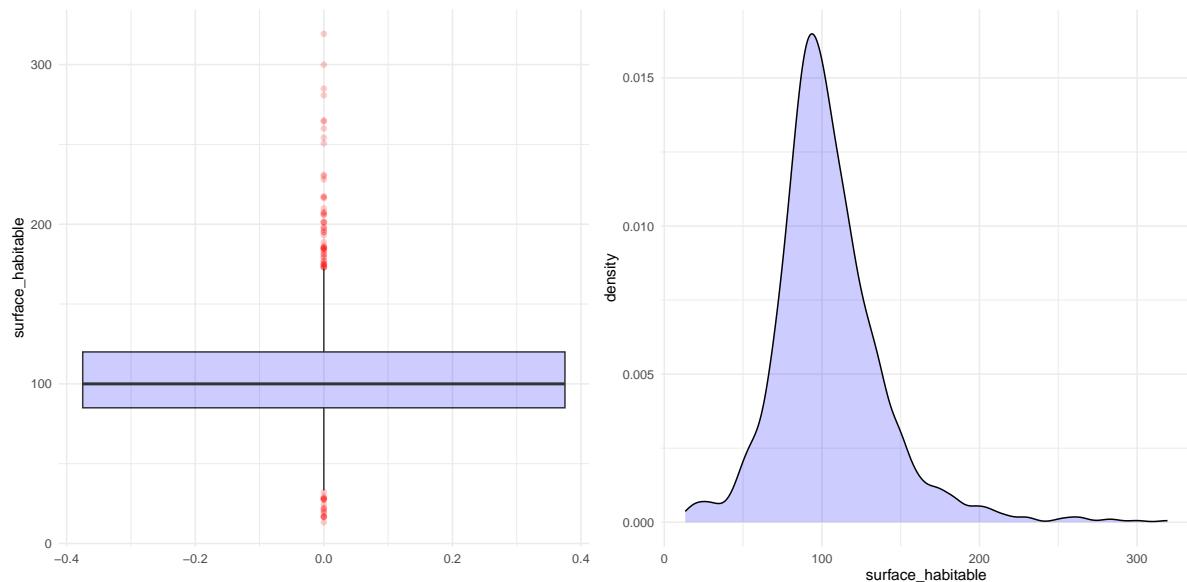


FIGURE 22 – Distribution de la surface habitable des maisons à Brest

TABLE 24 – Statistiques descriptives des surfaces habitables des maisons

Min	1er quartile	Médiane	Moyenne	3ème quartile	Max
13	85	100	105	120	319

La surface habitable médiane est de 59 m² pour les appartements et 100 m² pour les maisons, ce qui confirme nos interprétations graphiques, des résultats conformes aux standards attendus. Cependant, l'écart type est plus marqué pour les appartements, où l'on trouve des biens allant de 10 m² à plus de 4000 m², probablement des erreurs ou cas atypiques (bureaux ou locaux commerciaux).

2.6 Surface de Terrain (m^2)

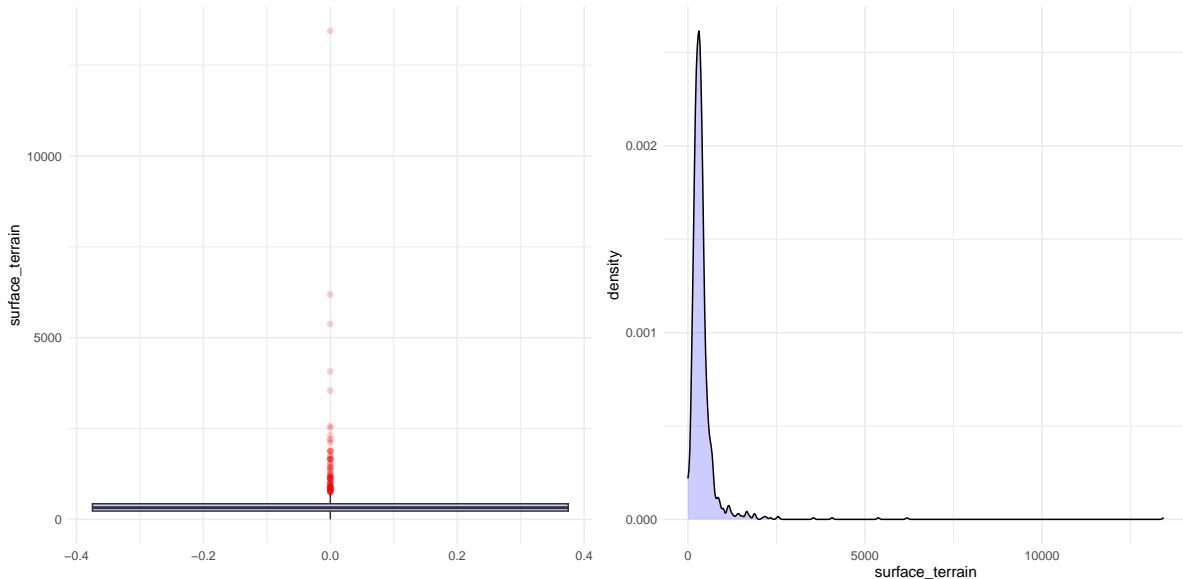


FIGURE 23 – Distribution de la surface de terrain des maisons à Brest

Les maisons disposent en moyenne de 399 m^2 de terrain, avec des cas extrêmes allant jusqu'à plus de 13 000 m^2 . La figure 23 montre une forte asymétrie de la distribution, avec une concentration autour de 300–500 m^2 . La surface de terrain, bien que fortement hétérogène, reste un critère de valorisation important pour les maisons.

TABLE 25 – Statistiques descriptives des surfaces de terrain des maisons

Min	1er quartile	Médiane	Moyenne	3ème quartile	Max
0	228,5	322	399,2	430,5	13 440

2.7 Autres variables

Afin de ne pas alourdir l'analyse les graphiques des variables suivantes sont disponibles en annexes :

- **Nombre d'étages** : Annexes 38 et 39
- **Nombre de pièces** : Annexes 40 et 41
- **Distance (en km) aux entités "vertes"** : Annexes 42 et 43
- **Distance (en km) aux entités "bleus"** : 44 et 45
- **Distance (en km) aux entités "rouges"** : Annexes 46 et 47

Les trois dernières variables sont inspirées d'une étude sur les déterminants des prix immobiliers dans la ville de Dijon [1]. Ces variables consistent à mesurer la distance (*en kilomètres*) entre chaque bâtiment (*représenté sous forme de polygone*) et différents groupes d'entités spatiales, en s'appuyant sur la base de données Marine Land Cover de Copernicus [7], qui recense les types d'occupation du sol à l'échelle européenne.

À l'aide d'un script Python, les différents types de sols ont été regroupés en catégories selon des codes couleurs, de la manière suivante :

Vert : zones à forte composante végétale

- 141 : espaces verts urbains
- 142 : équipements sportifs et de loisirs
- 311, 312, 313 : forêts
- 321, 322, 323, 324 : autres types de végétation
- 331 : plages, dunes, sable

Bleu : milieux aquatiques

- 523 : mer et océan
- 500 et suivants : autres zones humides ou aquatiques

Rouges : infrastructures anthropiques et zones industrielles

- 122 : réseaux routiers et ferroviaires
- 124 : aéroports
- 131 : sites d'extraction de matériaux
- 132 : décharges
- 400 et suivants : zones de marais et milieux humides anthropisés

Les codes utilisés correspondent à ceux de la base Marine Land Cover.

2.8 Matrice des corrélations

Les deux matrices de corrélation [25](#) et [24](#) fournissent une analyse comparative des relations entre les variables structurantes et géographiques associées aux biens immobiliers, distinguant les maisons des appartements.

Pour les maisons, la variable prix montre une forte corrélation positive avec la surface habitable ($\approx 0,64$) et une corrélation modérée avec le nombre de pièces ($\approx 0,46$), ce qui confirme que la taille du logement et sa capacité d'accueil sont des déterminants majeurs de sa valeur. Une corrélation positive est également observée avec la distance aux nuisances ou "entités rouges" ($\approx 0,26$), bien que de moindre intensité. Autrement dit, plus on s'éloigne des nuisances, plus le prix augmente. À l'inverse, la distance aux entités bleus (*océan, mer, lac*) présente une corrélation négative avec le prix ($\approx 0,14$), ce qui suggère que si on s'éloigne de ces dernières, le prix diminue. Les plus fortes corrélations sont celles entre la distance aux entités bleus et vertes ($\approx 0,69$), le nombre de pièces et la surface habitable ($\approx 0,56$) ou encore la distance entre les entités bleus et rouges ($\approx -0,47$).

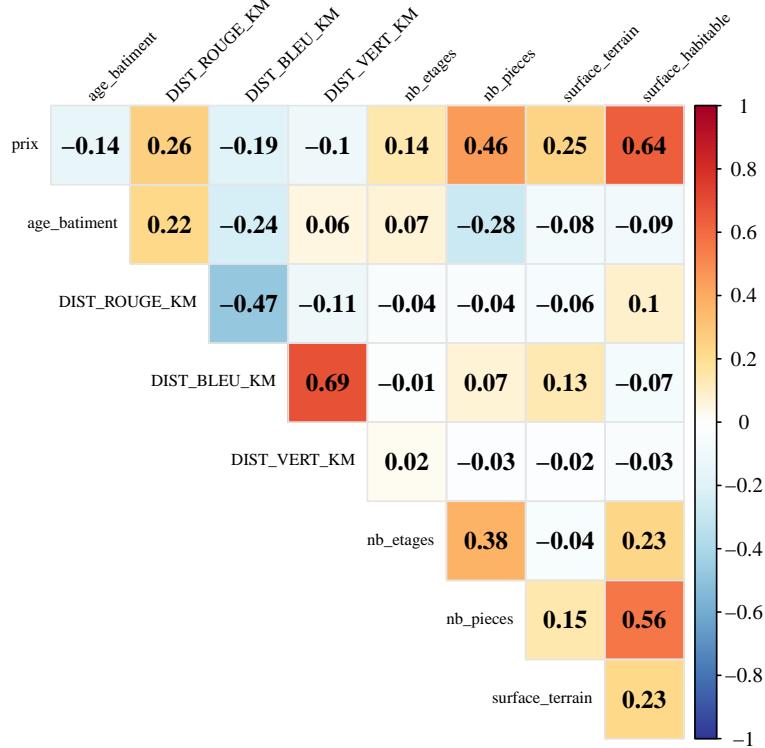


FIGURE 24 – Matrice des corrélations des maisons

Bien que le prix est encore une fois, fortement corrélés aux nombres de pièces ($\approx 0,63$) , ce n'est pas le cas de la surface habitable qui à un impact moindre ($\approx 0,12$) , ce qui est plutôt logique dans le cas d'appartement. On retrouve une nouvelle fois une forte corrélations entre les distances entre entités rouges et bleus ($\approx -0,55$) ainsi que rouge et vertes ($\approx 0,23$).

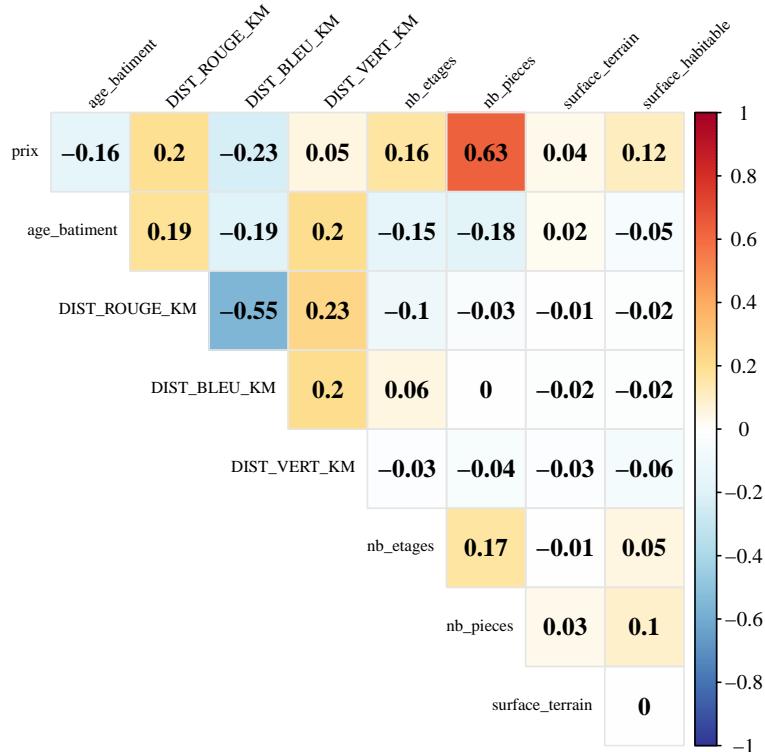


FIGURE 25 – Matrice des corrélations des appartements

En résumé, les maisons présentent des corrélations plus nettes entre les dimensions physiques et le prix, avec une influence modérée de l'âge du bâtiment. Les appartements, quant à eux, montrent une structure de corrélation plus diffuse, centrée presque exclusivement sur le nombre de pièces comme facteur déterminant du prix, et une relative indépendance des autres variables. Ces différences traduisent les logiques distinctes de valorisation entre habitat individuel et habitat collectif.

3 Méthodologie économétrique

L'approche économétrique vise à expliquer le prix des bâtiments (appartements et maisons) à Brest, en tenant compte des interactions potentielles entre bâtiments voisins. L'économétrie spatiale prend en compte l'**hétérogénéité spatiale**, qui reflète la diversité des comportements selon la localisation géographique, ainsi que l'**autocorrélation spatiale**, qui examine le niveau d'interaction entre unités spatiales, ici les bâtiments.

La matrice de pondération (W) est un outil fondamental pour structurer les relations spatiales en définissant les voisins de chaque bâtiment. Plusieurs spécifications de cette matrice peuvent être envisagées :

- La matrice Reine considère comme voisins tous les bâtiments partageant un bord ou un point commun.
- La matrice Tour restreint le voisinage aux bâtiments partageant uniquement un bord commun.
- La matrice des k plus proches voisins repose sur la proximité géographique, sans nécessité de contiguïté physique.

Avant d'estimer les modèles, il est essentiel de tester l'autocorrélation spatiale à l'aide de l'indice de Moran (I), qui teste l'hypothèse nulle d'absence d'autocorrélation. Plusieurs modèles peuvent être envisagés, prenant en compte des effets endogènes, exogènes et non observables :

- Le **modèle d'autorégression spatiale (SAR)** capte les effets spatiaux endogènes, où le prix d'un bâtiment est influencé par le prix des bâtiments voisins. Ce modèle est formalisé par l'équation suivante :

$$y = \rho W y + X\beta + \varepsilon$$

où ρ représente l'intensité de cette influence, y le prix (ou log-prix) du bâtiment, W la matrice de pondération spatiale, X les variables explicatives (ex. surface, année de construction), et ε le terme d'erreur.

- Le **modèle d'erreur spatiale (SEM)** prend en compte l'impact des résidus d'un bâtiment sur ceux de ses voisins. Il est formalisé ainsi :

$$y = X\beta + u, \quad u = \lambda W u + \varepsilon$$

où λ mesure l'autocorrélation spatiale des erreurs.

- Le **modèle de Durbin spatial (SDM)** inclut à la fois les variables explicatives et la variable dépendante spatialement décalée, permettant de capter les effets directs et indirects des variables explicatives. Il est formulé comme suit :

$$y = \rho W y + X\beta + W X\theta + \varepsilon$$

où θ représente les interactions exogènes entre voisins.

- Le modèle **SLX (Spatial Lag of X)** est utilisé pour identifier la présence d'effets spatiaux exogènes, où le prix d'un bâtiment est influencé par les caractéristiques observables des bâtiments voisins. Sa forme est donnée par :

$$y = X\beta + W X\theta + \varepsilon$$

La sélection du modèle le plus approprié peut s'appuyer sur les tests de Lagrange ainsi que sur la méthodologie d'Elhorst [10], qui propose une approche mixte.

4 Modélisation spatiale

Compte tenu de la nature des données utilisées, à savoir des polygones représentant les bâtiments (*maisons et appartements*) situés dans la ville de Brest, l'utilisation de matrices de contiguïté de type Reine ou Tour ne semble pas appropriée dans le cadre de cette analyse (*voir section 3*). Ces matrices reposent sur la notion de contiguïté topologique, selon laquelle deux entités sont considérées comme voisines si elles partagent une frontière, un sommet ou un côté. Or, contrairement à des unités spatiales classiques comme les pays, régions ou départements, les polygones représentant les bâtiments ne sont généralement pas adjacents. Ils sont souvent séparés par des espaces publics tels que des routes, des trottoirs ou des jardins. Par conséquent, la notion de voisin fondée sur un contact géométrique direct perd ici sa pertinence.

Dans ce contexte, il est plus judicieux de recourir à des matrices basées sur les plus proches voisins, en particulier les matrices à 1 et 3 plus proches voisins (qu'on appellera, respectivement PPV1 et PPV3). Celles-ci définissent les relations spatiales à partir de la distance euclidienne entre les centroïdes des polygones (*section 3*). Cette approche permet de mieux capturer les interactions spatiales potentielles entre entités proches, ce qui est plus cohérent avec la réalité urbaine observée. Elle assure également une structure de dépendance spatiale mieux adaptée aux phénomènes étudiés, notamment lorsque l'effet de proximité prime sur la simple contiguïté topologique.

4.1 Mesure de l'autocorrélation spatiale globale

Les tableaux 26 et 27 présentent les résultats des tests de Moran réalisés pour évaluer l'autocorrélation spatiale des prix des logements, selon deux types de biens : appartements et maisons. Deux matrices de pondération spatiale ont été considérées pour chaque cas : celle du plus proche

voisin (PPV1) et celle des trois plus proches voisins (PPV3).

4.1.1 Appartements

TABLE 26 – Tests de Moran pour différentes matrices de pondération spatiale

Matrice	Indice de Moran		
	I Moran	Statistic z	p-value
1 plus proche voisin	0,35	12,17	< 2,2e-16
3 plus proches voisins	0,30	17,08	< 2,2e-16

Dans le cas des appartements (*tableau 26*), les résultats indiquent une dépendance spatiale significative. Pour la matrice basée sur le plus proche voisin, l'indice de Moran s'élève à 0,35, avec une statistique z de 12,17 et une p-valeur inférieure à 2,2e-16. Ces résultats révèlent une structure spatiale fortement non aléatoire, où les observations similaires tendent à se regrouper géographiquement. Une tendance comparable est observée avec la matrice des trois plus proches voisins, bien que l'intensité de l'autocorrélation diminue légèrement ($I = 0,30$), tandis que la statistique z augmente (17,08), signe d'une robustesse accrue de l'effet spatial à cette échelle.

4.1.2 Maisons

TABLE 27 – Indice de Moran selon différentes matrices de pondération

Matrice	Indice de Moran		
	I Moran	Statistic z	p-value
1 plus proche voisin	0,26	7,54	2,227e-14
3 plus proches voisins	0,21	10,23	2,2e-16

Pour les maisons (*tableau 27*), on observe également une autocorrélation spatiale positive significative, quoique légèrement moins marquée que pour les appartements. L'indice de Moran est de 0,26 avec le plus proche voisin ($z = 7,54, p < 2,23e - 14$) et de 0,21 avec les trois plus proches voisins ($z = 10,23, p < 2,2e - 16$). Ces résultats confirment l'existence d'une structure spatiale dans la distribution des prix, bien que d'intensité plus modérée.

Dans l'ensemble, l'ensemble des tests sont significatifs au seuil de 1%, ce qui permet de rejeter l'hypothèse d'indépendance spatiale. Les indices de Moran positifs dans tous les cas suggèrent une tendance au regroupement géographique des valeurs similaires. La comparaison entre les deux types de biens met en évidence une autocorrélation plus élevée pour les appartements, ce qui pourrait s'expliquer par une concentration plus forte en milieu urbain dense, où les effets de voisinage sont plus marqués. Enfin, la légère baisse de l'indice de Moran lorsqu'on passe de 1 à 3 voisins reflète un affaiblissement progressif de l'effet spatial à mesure que l'on élargit le périmètre de voisinage, ce qui est cohérent avec les dynamiques généralement observées dans les

modèles spatiaux.

Les diagrammes de Moran sont disponibles en annexes [48](#), [49](#) et [50](#) pour les maisons et [51](#), [52](#) et [53](#) pour les appartements.

4.2 Sélection des modèles

Cette section [4.2](#) examine la sélection des modèles économétriques spatiaux les plus appropriés pour analyser les déterminants du prix des biens (*appartements et maisons*) à Brest. La sélection des modèles repose souvent sur la détection et la quantification de l'autocorrélation spatiale, c'est-à-dire la mesure dans laquelle les observations spatiales sont influencées par leurs voisines. Plusieurs modèles sont couramment mobilisés dans ce cadre, notamment les modèles SAR, SEM, SDM ET SLX. Chacun de ces modèles permet de capturer des formes spécifiques de dépendance spatiale : dans la variable dépendante pour le SAR, dans les termes d'erreur pour le SEM, dans les variables explicatives ainsi que dans la variable dépendante pour le SDM, et uniquement dans les variables explicatives pour le SLX.

4.2.1 Appartements

TABLE 28 – Tests de Moran sur les résidus de la régression OLS

Matrices de poids	Test de Moran sur les résidus				
	I Moran	Expected Value	Variance	z-statistic	p-value
1 plus proche voisin	0.235	-0.003	0.0009	8.14	3.875e-16
3 plus proches voisins	0.197	-0.003	0.0003	11.49	< 2.2e-16

Le tableau [28](#) présente les résultats des tests de Moran appliqués aux résidus du modèle OLS pour deux matrices de pondération spatiale : le plus proche voisin et les trois plus proches voisins. Dans les deux cas, les résultats mettent en évidence une autocorrélation spatiale résiduelle forte et significative, ce qui suggère que le modèle initial ne parvient pas à capturer convenablement la structure spatiale sous-jacente.

Avec la matrice du plus proche voisin, l'indice de Moran observé est de 0,235, nettement supérieur à sa valeur attendue sous l'hypothèse nulle d'indépendance spatiale (-0,003), avec une statistique z de 8,14 et une p-valeur de 3,875e-16, largement inférieure au seuil de 1%. De manière similaire, la matrice des trois plus proches voisins aboutit à un indice de Moran de 0,197, une statistique z encore plus élevée (11,49) et une p-valeur inférieure à 2,2e-16. Ces résultats indiquent que les résidus présentent un schéma de regroupement spatial non aléatoire.

TABLE 29 – Tests de dépendance spatiale utilisant différentes matrices de pondération

Models	Weight matrices : p-value	
	NN1	NN3
LMerr	8.882e-16	< 2.2e-16
LMLag	3.331e-16	< 2.2e-16
RLMerr	0.01481	1.606e-05
RLMLag	0.005219	0.0005013

Le tableau 29 présente les résultats des tests de dépendance spatiale effectués à l'aide de deux matrices de pondération différentes : PPV1 et PPV3. Les tests de type Lagrange Multiplier (LMerr et LMLag) permettent de détecter la présence d'autocorrélation spatiale, soit dans les erreurs (LMerr), soit dans la variable dépendante (LMLag). Les versions robustes de ces tests (RLMerr et RLMLag) sont utilisées pour tenir compte de la possibilité d'une autocorrélation de nature différente.

L'analyse des p-values indique que, pour les deux matrices de pondération, tous les tests classiques (LMerr et LMLag) sont hautement significatifs, avec des p-values proches de zéro (< 2.2e-16 dans certains cas), indiquant une dépendance spatiale forte. Cependant, ces tests peuvent être biaisés en présence d'un type alternatif de dépendance. Il est donc crucial de se référer aux versions robustes pour orienter le choix du modèle le plus pertinent.

Dans le cas de la matrice PPV1, le test RLMLag présente une p-value de 0.005219, la plus faible parmi les tests robustes pour cette matrice. Ce résultat indique une dépendance spatiale significative dans la variable dépendante, même après correction pour une éventuelle autocorrélation dans les erreurs. Le test RLMerr, quant à lui, présente une p-value plus élevée (0.01481), bien qu'encore significative au seuil de 5%. Par conséquent, le modèle SAR, associé à une dépendance dans la variable dépendante, est à privilégier pour cette matrice étant donné que la p-value du test RLMLag est la plus faible.

Concernant la matrice PPV3, les deux tests robustes sont significatifs, mais le test RLMerr affiche la p-value la plus faible (1.606e-05), contre 0.0005013 pour RLMLag. Ce résultat suggère cette fois-ci une dépendance spatiale plus marquée dans les erreurs, ce qui justifie l'utilisation d'un modèle SEM dans ce cas.

En résumé, étant donné que les tests robustes sont tous significatifs, nous retenons, pour chaque matrice, le modèle correspondant à la p-value la plus faible : **un modèle SAR pour PPV1, et un modèle SEM pour PPV3**. Cette approche permet d'adapter le choix du modèle spatial à la structure de dépendance détectée de manière rigoureuse.

TABLE 30 – Tests de rapport de vraisemblance (LR) et critères AIC pour la sélection de modèles spatiaux pour les appartements

Matrice	Comparaison	p-value (LR)	AIC Modèle 1	AIC Modèle 2
PPV1	SAR vs SDM	0.5306	2232.58 (SAR)	2247.63 (SDM)
PPV3	SEM vs SDM	0.0303*	2231.98 (SEM)	2234.52 (SDM)

Le tableau 33 présente les résultats des tests de rapport de vraisemblance (Likelihood Ratio, LR) ainsi que les valeurs du critère d'information d'Akaike (AIC), pour comparer différents modèles spatiaux en fonction de la matrice de pondération utilisée. L'objectif de cette analyse est de déterminer si un modèle plus complexe, en l'occurrence le modèle SDM, s'impose significativement face à un modèle plus restreint, tel que le SAR ou le SEM. Le test LR repose sur la comparaison de la vraisemblance des deux modèles imbriqués, tandis que l'AIC permet de juger de la qualité d'ajustement tout en pénalisant la complexité du modèle.

Pour la matrice PPV1, la comparaison oppose le modèle SAR au modèle SDM. Le test LR donne une p-value de 0.5306, ce qui est bien au-dessus des seuils usuels de significativité. Cela indique que le modèle SDM ne fournit pas un ajustement significativement meilleur que le modèle SAR, malgré sa complexité accrue. Cette conclusion est confirmée par la comparaison des valeurs d'AIC : le modèle SAR présente une AIC plus faible (2232.58 contre 2247.63), ce qui plaide en faveur de ce dernier. Ces résultats confortent le choix du modèle SAR pour la matrice PPV1.

En revanche, pour la matrice PPV3, la comparaison est effectuée entre le modèle SEM et le modèle SDM. Le test de rapport de vraisemblance est cette fois significatif au seuil de 5%, avec une p-value de 0.0303, indiquant que le modèle SDM offre un ajustement statistiquement meilleur que le SEM. Toutefois, l'écart entre les valeurs d'AIC est relativement faible : le modèle SEM obtient une AIC légèrement inférieure (2231.98 contre 2234.52). Bien que le critère AIC privilégie le modèle SEM, la significativité du test LR justifie l'adoption du modèle SDM, puisqu'il capture une complexité supplémentaire pertinente dans les données. Il convient néanmoins de souligner que le modèle SEM reste compétitif en termes de parcimonie.

En conclusion, **pour la matrice PPV1, le modèle SAR est retenu** comme le plus adapté, tandis que **pour la matrice PPV3, le modèle SDM s'impose**, bien que le modèle SEM offre un compromis intéressant.

4.2.2 Maisons

TABLE 31 – Tests de Moran sur les résidus

Matrices de poids	Test de Moran sur les résidus				
	I Moran	Expected Value	Variance	z-statistic	p-value
1 plus proche voisin	0.13	-0.003	0.001	3.99	6.566e-05
3 plus proches voisins	0.12	-0.003	0.0004	6.03	1.593e-09

Le tableau 31 présente les résultats obtenus les deux matrices de pondérations PPV1 et PPV3. Dans les deux cas, les résultats révèlent une autocorrélation spatiale positive significative des résidus, indiquant que les erreurs du modèle tendent à se regrouper spatialement au lieu d'être distribuées de manière aléatoire.

Avec la matrice PPV1, l'indice de Moran observé est de 0,13, supérieur à la valeur attendue sous l'hypothèse nulle d'indépendance (-0,003), avec une statistique z de 3,99 et une p-valeur de 6,566e-05. Ces résultats sont significatifs au seuil de 1%, ce qui permet de rejeter l'hypothèse nulle d'absence d'autocorrélation spatiale. Le même constat s'applique à la matrice PPV3, où l'indice de Moran reste élevé (0,12), accompagné d'une statistique z encore plus forte (6,03) et d'une p-valeur très faible (1,593e-09).

TABLE 32 – Tests de dépendance spatiale utilisant différentes matrices de pondération

Models	Weight matrices : p-value	
	PPV1	PPV3
LMerr	0.0001053	5.529e-09
LMLag	1.127e-06	9.347e-09
RLMerr	0.9369	0.02832
RLMLag	0.003244	0.05164

Le tableau 32 présente les résultats des tests de dépendance spatiale effectués à l'aide de deux matrices de pondération différentes : PPV1 et PPV3.

Pour la matrice PPV1, les deux tests classiques (LMerr et LMLag) sont significatifs, avec des p-values très faibles (respectivement 0.0001053 et 1.127e-06), indiquant une dépendance spatiale significative. Toutefois, seul le test RLMLag reste significatif dans sa version robuste ($p = 0.003244$), tandis que RLMerr n'est pas significatif ($p = 0.9369$). Ce résultat suggère que la dépendance détectée est essentiellement liée à la variable dépendante, ce qui justifie l'adoption d'un modèle SAR pour cette configuration.

Pour la matrice PPV3, tous les tests classiques sont également hautement significatifs (p-values de l'ordre de 10^{-9}), et les deux tests robustes sont proches du seuil de significativité. Le test RLMerr est significatif au seuil de 5% ($p = 0.02832$), tandis que RLMLag atteint la significativité au seuil de 10% ($p = 0.05164$). Toutefois, la p-value la plus faible revient au test RLMerr, ce qui indique que la dépendance spatiale est davantage présente dans les erreurs pour cette matrice. Le modèle le plus approprié dans ce cas est donc un modèle SEM.

En conclusion, comme pour les précédentes analyses, tous les tests robustes étant significatifs (ou proches du seuil), nous retenons, pour chaque matrice, le modèle correspondant à la p-value la plus faible : **un modèle SAR pour PPV1 et un modèle SEM pour PPV3**. Cette approche garantit une spécification rigoureuse des modèles spatiaux, adaptée à la nature de la dépendance identifiée dans les données.

TABLE 33 – Tests de rapport de vraisemblance (LR) et critères AIC pour la sélection de modèles spatiaux

Matrice	Comparaison	p-value (LR)	AIC Modèle 1	AIC Modèle 2
PPV1	SAR vs SDM	0.0021 ***	299.28 (SAR)	293.33 (SDM)
PPV3	SEM vs SDM	0.0444 **	289.64 (SEM)	293.53 (SDM)

Le tableau 33 présente les résultats des tests de rapport de vraisemblance et AIC pour comparer des modèles spatiaux, en fonction de deux matrices de pondération pondérées par distances : PPV1 et PPV3.

Pour la matrice PPV1, le test de rapport de vraisemblance comparant le modèle SAR au modèle SDM donne une p-value de 0.0021, hautement significative au seuil de 1%. Cela indique que le modèle SDM fournit un ajustement significativement supérieur à celui du SAR. Cette conclusion est renforcée par les valeurs d'AIC : le modèle SDM présente une AIC plus faible (293.33) que le modèle SAR (299.28), ce qui confirme son avantage en termes de qualité d'ajustement. Ces résultats suggèrent que le modèle SDM est à privilégier pour la matrice PPV1.

Concernant la matrice PPV3, la comparaison porte cette fois sur les modèles SEM et SDM. Le test LR indique une p-value de 0.0444, significative au seuil de 5%, suggérant que le modèle SDM offre un meilleur ajustement que le modèle SEM. Toutefois, à l'inverse de PPV1, la comparaison des valeurs d'AIC montre un avantage pour le modèle SEM (289.64 contre 293.53 pour le SDM). Ce contraste souligne une tension classique entre les deux critères : si le test LR privilégie la richesse explicative du SDM, le critère AIC penche pour la parcimonie du modèle SEM. Néanmoins, la significativité du test justifie, dans une logique statistique rigoureuse, la préférence pour le modèle SDM dans ce cas également.

En résumé, pour les deux matrices, les tests de rapport de vraisemblance concluent à la **supériorité du modèle SDM**, qui permet de mieux capturer la complexité de la dépendance spatiale présente dans les données.

4.3 Interprétation des modèles

4.3.1 Appartements

Modèle SAR : matrice PPV1

TABLE 34 – Résultats du modèle SAR estimé avec la matrice de pondération PPV1 pour les maisons

Variable	Estimation	Erreur standard	z-value	p-value
(Intercept)	10.0487	0.1883	53.3595	< 2.2e-16
dpeA	-0.3727	0.5387	-0.6919	0.4890
dpeB	0.2819	0.1381	2.0417	0.0412**
dpeC	0.0558	0.0304	1.8322	0.0669*
dpeE	-0.0347	0.0243	-1.4274	0.1535
dpeF	-0.1751	0.0439	-3.9881	6.66e-05***
dpeG	-0.1736	0.0612	-2.8377	0.0045***
age_batiment	-0.0019	0.0003	-5.9284	3.06e-09***
surface_habitable	0.0002	0.0001	2.3345	0.0196**
surface_terrain	0.0003	0.0002	1.5042	0.1325
nb_etages	0.0127	0.0056	2.2860	0.0223**
nb_pieces	0.2466	0.0086	28.5142	< 2.2e-16***
DIST_VERT_KM	0.0836	0.0203	4.1100	3.96e-05***
DIST_BLEU_KM	-0.1231	0.0172	-7.1670	7.67e-13***
DIST_ROUGE_KM	0.0487	0.0159	3.0586	0.0022***
Rho	0.0803 (Erreur standard = 0.0154; z = 5.23; p = 1.72e-07)			
Test LR (vs OLS)	Valeur du test = 29.497; p-value = 5.60e-08			

Les résultats du modèle SAR estimé avec la matrice PPV1 (tableau 34) indiquent la présence d'une dépendance spatiale significative dans les prix immobiliers. Le paramètre d'autocorrélation spatiale est positif et hautement significatif ($\rho = 0,0803, p = 1,72 \times 10^{-7}$), ce qui suggère que les propriétés situées à proximité tendent à présenter des valeurs similaires.

L'analyse des coefficients associés aux variables explicatives montre que certaines catégories du diagnostic de performance énergétique (DPE) influencent significativement les prix. Comparativement à la catégorie de référence (D), les biens classés en **B** présentent un effet positif et statistiquement significatif (0,2819, $p = 0,0412$), tandis que la catégorie **C** affiche un effet positif mais marginalement significatif (0,0558, $p = 0,0669$). En revanche, les classifications énergétiques **F** ($-0,1751, p = 6,66 \times 10^{-5}$) et **G** ($-0,1736, p = 0,0045$) sont associées à des prix inférieurs, avec des effets négatifs statistiquement robustes. Les catégories **A** et **E** ne présentent pas d'effet significatif.

Les caractéristiques physiques du bien jouent également un rôle important. L'âge du bâtiment exerce un effet négatif marqué ($-0,0019, p = 3,06 \times 10^{-9}$), indiquant que les constructions plus anciennes tendent à être moins valorisées. À l'inverse, la surface habitable (0,0002, $p = 0,0196$), le nombre d'étages (0,0127, $p = 0,0223$) et le nombre de pièces (0,2466, $p < 2,2 \times 10^{-16}$) sont positivement et significativement corrélés au prix. La surface du terrain ne montre pas d'effet statistiquement significatif.

Les variables de distance aux points d'intérêt présentent également des relations significatives avec le prix : la distance aux points « VERT » ($0,0836$, $p = 3,96 \times 10^{-5}$) et « ROUGE » ($0,0487$, $p = 0,0022$) est positivement associée aux prix, tandis que la distance au point « BLEU » ($-0,1231$, $p = 7,67 \times 10^{-13}$) présente une relation négative.

Le test de rapport de vraisemblance comparant ce modèle SAR à une estimation OLS est hautement significatif (statistique = $29,497$, $p = 5,60 \times 10^{-8}$), ce qui confirme l'intérêt d'intégrer la dépendance spatiale dans la modélisation.

Il convient toutefois de rappeler que, dans un modèle SAR, les coefficients estimés ne peuvent pas être directement interprétés comme des effets marginaux unitaires. L'autocorrélation spatiale implique que toute variation d'une variable explicative se répercute non seulement sur l'unité spatiale considérée mais également sur ses voisines via le terme ρWy . Pour évaluer précisément l'impact réel de chaque variable, il est nécessaire de calculer les effets directs, indirects et totaux, qui prennent en compte cette propagation spatiale (*voir tableau 35*).

TABLE 35 – Effets directs et indirects du modèle SAR (impacts), matrice PPV1

Variable	Direct	Indirect	Total
dpeA	-0.3742	-0.0310	-0.4052
dpeB	0.2830	0.0235	0.3065
dpeC	0.0560	0.0046	0.0607
dpeE	-0.0349	-0.0029	-0.0378
dpeF	-0.1758	-0.0146	-0.1904
dpeG	-0.1743	-0.0145	-0.1888
age_batiment	-0.0019	-0.0002	-0.0020
surface_habitable	0.0002	0.00002	0.0003
surface_terrain	0.0003	0.00002	0.0003
nb_etages	0.0128	0.0011	0.0139
nb_pieces	0.2476	0.0205	0.2682
DIST_VERT_KM	0.0839	0.0070	0.0909
DIST_BLEU_KM	-0.1236	-0.0103	-0.1339
DIST_ROUGE_KM	0.0489	0.0041	0.0530

L'estimation des impacts directs, indirects et totaux du modèle SAR avec matrice de pondération PPV1 (tableau 35) met en évidence la manière dont les caractéristiques énergétiques, physiques et géographiques influencent les prix immobiliers, tout en intégrant les interdépendances spatiales. Les performances énergétiques, mesurées par les catégories du DPE, révèlent des effets différenciés : les biens classés en **B** et **C** présentent des effets directs positifs (respectivement 0,2830 et 0,0560) et des effets indirects également positifs (0,0235 et 0,0046), conduisant à des effets totaux positifs (0,3065 et 0,0607). Cela traduit une valorisation des meilleures performances énergétiques, tant localement qu'à travers un effet de diffusion modeste

sur les prix des propriétés voisines. À l'inverse, les catégories F et G affichent des effets directs négatifs (-0,1758 et -0,1743) et des effets indirects eux aussi négatifs (-0,0146 et -0,0145), ce qui se traduit par des effets totaux négatifs (-0,1904 et -0,1888), indiquant que les mauvaises performances énergétiques déprécient les biens, avec un léger effet défavorable sur le voisinage. Ces interprétations se font par rapport à la catégorie de référence, ici l'étiquette D.

Les caractéristiques physiques du bien confirment les tendances attendues. L'âge du bâtiment exerce un effet direct négatif (-0,0019) et un effet indirect très faible (-0,0002), conduisant à un effet total négatif (-0,0020), ce qui corrobore l'idée que les constructions plus anciennes tendent à être moins valorisées. La surface habitable, le nombre d'étages et le nombre de pièces sont associés à des effets directs et indirects positifs, se traduisant par des effets totaux positifs (0,0003, 0,0139 et 0,2682), confirmant que des biens plus spacieux ou offrant plus de pièces et d'étages sont globalement mieux valorisés, avec un effet de diffusion positif mais limité. La surface du terrain présente un effet total positif de très faible amplitude (0,0003).

Enfin, les variables de distance aux points d'intérêt mettent en évidence des relations contrastées. La distance aux points « VERT » et « ROUGE » est associée à des effets directs et indirects positifs, pour des effets totaux respectifs de 0,0909 et 0,0530, suggérant qu'un éloignement accru par rapport à ces éléments est légèrement valorisé sur le marché immobilier, avec un effet de propagation positif. À l'inverse, la distance aux points « BLEU » présente un effet direct négatif (-0,1236) et un effet indirect également négatif (-0,0103), pour un effet total de -0,1339, ce qui indique qu'un éloignement accru de ces éléments est associé à une baisse du prix des biens, avec une répercussion défavorable sur les propriétés voisines.

Modèle SDM : matrice PPV3

L'estimation du modèle SDM avec une matrice PPV3 (tableau 37) permet d'évaluer l'influence conjointe des caractéristiques intrinsèques des biens immobiliers et de celles de leur voisinage immédiat sur les prix, tout en intégrant les interdépendances spatiales. Le paramètre d'autocorrélation spatiale ρ , estimé à 0,1454 et hautement significatif ($p = 2,45 \times 10^{-7}$), confirme l'existence d'une dépendance spatiale positive : les prix des biens immobiliers tendent à être corrélés à ceux des propriétés voisines.

Les effets directs montrent que certaines catégories du diagnostic de performance énergétique sont significativement associées aux prix. Par rapport à la catégorie de référence, les biens classés en B ($0,2867, p = 0,0374$) et en C ($0,0639, p = 0,0376$) présentent une valorisation positive, tandis que les catégories F ($-0,1753, p = 5,84 \times 10^{-5}$) et G ($-0,1756, p = 0,0039$) sont associées à une décote marquée. Les variables DPE A et E ne présentent pas d'effet significatif. Parmi les autres caractéristiques physiques, l'âge du bâtiment exerce un effet négatif significatif ($-0,0016, p = 1,25 \times 10^{-6}$), et la surface habitable ($0,0002, p = 0,0237$) ainsi que le nombre de pièces ($0,2450, p < 2,2 \times 10^{-16}$) sont positivement et significativement corrélés aux prix. La surface du terrain et le nombre d'étages n'ont pas d'effet direct significatif. Concernant les variables de localisation, seule la distance aux points « BLEU » présente une relation négative avec le prix ($-0,4470, p = 0,0740$), significative au seuil de 10%, suggérant qu'un éloignement accru à ces points est associé à une légère diminution de valeur.

TABLE 36 – Résultats du modèle SDM estimé avec la matrice de pondération PPV3

Variable	Estimation	Erreur standard	z-value	p-value
Effets directs				
(Intercept)	9.3725	0.3219	29.1148	< 2.2e-16***
dpeA	-0.5887	0.5556	-1.0595	0.2894
dpeB	0.2867	0.1377	2.0818	0.0374**
dpeC	0.0639	0.0307	2.0795	0.0376**
dpeE	-0.0314	0.0242	-1.2949	0.1953
dpeF	-0.1753	0.0436	-4.0192	5.84e-05***
dpeG	-0.1756	0.0608	-2.8866	0.0039***
age_batiment	-0.0016	0.0003	-4.8473	1.25e-06***
surface_habitable	0.0002	0.0001	2.2626	0.0237**
surface_terrain	0.0002	0.0002	1.3801	0.1676
nb_etages	0.0080	0.0066	1.2192	0.2228
nb_pieces	0.2450	0.0087	28.1053	< 2.2e-16***
DIST_VERT_KM	-0.1240	0.2542	-0.4879	0.6256
DIST_BLEU_KM	-0.4470	0.2502	-1.7868	0.0740*
DIST_ROUGE_KM	0.1422	0.2379	0.5977	0.5500
Effets spatiaux (lag)				
lag.dpeA	-0.7685	1.0084	-0.7621	0.4460
lag.dpeB	0.0673	0.2649	0.2541	0.7994
lag.dpeC	-0.0674	0.0512	-1.3156	0.1883
lag.dpeE	-0.0517	0.0411	-1.2565	0.2089
lag.dpeF	-0.0545	0.0712	-0.7643	0.4447
lag.dpeG	0.0914	0.1082	0.8445	0.3984
lag.age_batiment	-0.0007	0.0005	-1.4406	0.1497
lag.surface_habitable	0.0002	0.0002	1.0038	0.3155
lag.surface_terrain	0.0006	0.0003	1.9408	0.0523*
lag.nb_etages	0.0085	0.0088	0.9636	0.3352
lag.nb_pieces	-0.0216	0.0157	-1.3735	0.1696
lag.DIST_VERT_KM	0.2206	0.2562	0.8610	0.3893
lag.DIST_BLEU_KM	0.3264	0.2515	1.2976	0.1944
lag.DIST_ROUGE_KM	-0.0985	0.2386	-0.4129	0.6797
Paramètres spatiaux				
Rho	0.1454 (Erreur standard = 0.0282 ; z = 5.1617 ; p = 2.45e-07)			
Test LR (vs OLS)	Valeur du test = 27.757 ; p-value = 1.375e-07			
AIC	2234.5 (AIC pour lm : 2260.3)			

Les effets spatiaux décalés ($WX\theta$), qui capturent l'influence des caractéristiques des unités voisines, indiquent qu'une surface de terrain plus importante chez les voisins tend à exercer un effet positif marginal sur le prix du bien considéré (0,0006, $p = 0,0523$). Les autres variables spatiales décalées, qu'il s'agisse des DPE, des caractéristiques physiques ou des distances, ne présentent pas d'effet significatif, ce qui suggère que l'influence des caractéristiques des voisins est, dans ce modèle, faible ou déjà largement captée par le terme d'autocorrélation spatiale.

Le test de rapport de vraisemblance comparé à une estimation OLS est hautement significatif (statistique = 27,757; $p = 1,375 \times 10^{-7}$), et l'AIC est plus faible pour le modèle SDM (2234,5) que pour le modèle OLS (2260,3), ce qui atteste de la supériorité statistique et de l'adéquation accrue du modèle SDM.

TABLE 37 – Effets directs et indirects du modèle SDM (impacts)

Variable	Direct	Indirect	Total
dpeA	-0.6212	-0.9669	-1.5882
dpeB	0.2909	0.1234	0.4143
dpeC	0.0617	-0.0657	-0.0040
dpeE	-0.0335	-0.0637	-0.0972
dpeF	-0.1783	-0.0905	-0.2688
dpeG	-0.1731	0.0746	-0.0985
age_batiment	-0.0017	-0.0010	-0.0027
surface_habitable	0.0002	0.0002	0.0005
surface_terrain	0.0003	0.0007	0.0010
nb_etages	0.0084	0.0109	0.0193
nb_pieces	0.2455	0.0159	0.2614
DIST_VERT_KM	-0.1163	0.2293	0.1130
DIST_BLEU_KM	-0.4370	0.2959	-0.1412
DIST_ROUGE_KM	0.1392	-0.0881	0.0511

Les mesures d'impact issues du modèle SDM PPV3 (voir tableau 39) permettent de distinguer les effets directs, mesurant l'influence d'une variable explicative sur le prix de l'immobilier dans la même unité spatiale, des effets indirects, ou de débordement, qui capturent l'influence exercée sur les prix des unités voisines, ainsi que des effets totaux, qui correspondent à la somme des deux précédents. Pour rappel, l'étiquette DPE "D" sert de référence, de sorte que les coefficients mesurent l'effet relatif des autres catégories par rapport à celle-ci.

Les résultats mettent en évidence des relations contrastées entre performance énergétique et prix immobilier. Les logements classés **B** affichent un effet direct positif (0,2909) et un effet indirect positif (0,1234), conduisant à un effet total de 0,4143, indiquant une valorisation notable par rapport aux biens de classe **D**, tant localement qu'à travers un effet spatial positif. La catégorie **C** présente un léger effet direct positif (0,0617) mais un effet indirect négatif (-0,0657),

entraînant un effet total quasi nul (-0,0040), ce qui suggère que l'avantage local est presque entièrement annulé par un effet défavorable lié aux voisins. Les catégories **E** (-0,0972 en effet total), **F** (-0,2688) et **G** (-0,0985) sont globalement associées à une décote par rapport à **D**, en accord avec l'idée qu'une performance énergétique inférieure est sanctionnée par le marché. Le cas de **G** est particulier, car l'effet indirect positif (0,0746) compense partiellement l'effet direct négatif (-0,1731), possiblement en raison de caractéristiques spatiales spécifiques. Enfin, la catégorie **A** présente un effet total nettement négatif (-1,5882) par rapport à **D**, ce qui est contre-intuitif compte tenu de sa performance énergétique théorique supérieure. Ce résultat pourrait s'expliquer par la rareté des biens de cette catégorie dans notre échantillon.

S'agissant des caractéristiques physiques des biens, l'âge du bâtiment est associé à un effet total négatif (-0,0027), confirmant qu'un bâti plus ancien tend à être moins valorisé. La surface habitable (0,0005) et la surface du terrain (0,0010) présentent toutes deux des effets totaux positifs, traduisant une valorisation croissante avec la taille, y compris via l'effet positif des surfaces plus importantes dans les propriétés voisines. Le nombre d'étages (0,0193) et le nombre de pièces (0,2614) contribuent également à une augmentation des prix, le premier via des effets directs et indirects similaires, le second principalement par un effet direct marqué.

En ce qui concerne les distances aux points d'intérêt, l'éloignement des espaces verts (DIST VERT KM) a un effet direct négatif (-0,1163), indiquant que la proximité est valorisée, mais un effet indirect positif important (0,2293), aboutissant à un effet total positif (0,1130) qui suggère une relation spatiale complexe. L'éloignement des points d'intérêt « bleus » (DIST BLEU KM) entraîne un effet direct négatif prononcé (-0,4370) et un effet indirect positif (0,2959), pour un effet total négatif (-0,1412), montrant que la valorisation liée à la proximité prime sur les bénéfices éventuels liés à la situation des voisins. Enfin, l'éloignement des points « rouges » (DIST ROUGE KM) est associé à un effet direct positif (0,1392), possiblement lié à une perception négative de ces points, mais à un effet indirect négatif (-0,0881), de sorte que l'effet total (0,0511) reste positif mais atténué.

4.3.2 Maisons

Modèle SDM : matrice PPV1

TABLE 38 – Résultats du modèle SDM estimé avec la matrice de pondération PPV1 pour les maisons

Variable	Estimation	Erreur standard	z-value	p-value
Effets directs				
(Intercept)	10.8770	0.2605	41.7520	< 2.2e-16***
dpeA	0.1738	0.1018	1.7073	0.0878*
dpeB	0.1526	0.0507	3.0075	0.0026***
dpeC	0.0695	0.0203	3.4262	0.0006***
dpeE	-0.0752	0.0196	-3.8303	0.0001***
dpeF	-0.1528	0.0267	-5.7308	< 1e-08***
dpeG	-0.2037	0.0360	-5.6575	< 1e-08***
age_batiment	-0.0016	0.0003	-4.8097	< 1e-06***
surface_habitable	0.0044	0.0003	15.9062	< 2.2e-16***
surface_terrain	0.0001	0.00002	6.0984	< 1e-09***
nb_etages	-0.0038	0.0163	-0.2358	0.8136
nb_pieces	0.0461	0.0077	5.9605	< 1e-08***
DIST_VERT_KM	-0.3237	0.1449	-2.2340	0.0255**
DIST_BLEU_KM	-0.0147	0.1471	-0.0998	0.9205
DIST_ROUGE_KM	0.6681	0.1482	4.5074	< 1e-05***
Effets spatiaux (lag)				
lag.dpeA	-0.0292	0.0960	-0.3043	0.7609
lag.dpeB	0.0493	0.0478	1.0306	0.3027
lag.dpeC	0.0101	0.0207	0.4851	0.6276
lag.dpeE	-0.0025	0.0195	-0.1300	0.8966
lag.dpeF	0.0083	0.0266	0.3128	0.7544
lag.dpeG	0.0087	0.0377	0.2299	0.8182
lag.age_batiment	0.00007	0.0003	0.2112	0.8327
lag.surface_habitable	0.0002	0.0003	0.8204	0.4120
lag.surface_terrain	0.000003	0.00002	0.1702	0.8648
lag.nb_etages	0.0109	0.0162	0.6721	0.5015
lag.nb_pieces	-0.0040	0.0077	-0.5173	0.6050
lag.DIST_VERT_KM	0.3637	0.1445	2.5178	0.0118**
lag.DIST_BLEU_KM	-0.0422	0.1478	-0.2857	0.7751
lag.DIST_ROUGE_KM	-0.6215	0.1481	-4.1973	< 1e-05***
Paramètres spatiaux				
Rho	0.0651 (Erreur standard = 0.0216 ; z = 3.0128 ; p = 0.0026)			
Test LR (vs OLS)	Valeur du test = 8.5357 ; p-value = 0.0035			
AIC	293.33 (AIC pour lm : 299.87)			
Log-likelihood	-115.6652			

Le modèle SDM estimé avec la matrice de pondération PPV1 met en évidence plusieurs rela-

tions statistiquement significatives entre les caractéristiques des logements et leur prix, exprimées en logarithme, tout en intégrant la dépendance spatiale via le terme autorégressif ρ . Encore une fois, l'étiquette DPE "D" sert de référence, de sorte que les coefficients associés aux autres classes mesurent leur effet relatif par rapport à cette référence. Les coefficients estimés sont ici les paramètres bruts du modèle, et ne peuvent donc pas être interprétés directement comme des effets totaux sans passage par le calcul des impacts ; néanmoins, leur signe et leur significativité renseignent sur la direction et la robustesse des associations.

Les résultats du tableau 38 montrent que, toutes choses égales par ailleurs, les logements classés A, B et C présentent des coefficients positifs par rapport à la classe D, traduisant une association favorable avec le prix, mais seule la catégorie A n'est significative qu'au seuil de 10%, tandis que B et C le sont à 1%. À l'inverse, les catégories E, F et G sont associées à des coefficients négatifs et fortement significatifs, confirmant qu'une performance énergétique inférieure est corrélée à une moindre valeur marchande. Parmi les variables continues, l'âge du bâtiment exerce un effet négatif significatif, indiquant qu'un accroissement de l'ancienneté est lié à une baisse du prix, tandis que la surface habitable et la surface du terrain présentent des coefficients positifs et significatifs, ce qui traduit une valorisation des biens plus spacieux. Le nombre de pièces est également positivement associé au prix, alors que le nombre d'étages ne présente pas de relation statistiquement significative dans ce modèle.

Concernant les variables de localisation, une plus grande distance aux espaces verts (DIST VERT KM) est associée à un coefficient direct négatif et significatif, suggérant qu'une proximité accrue à ces aménités est valorisée. La distance aux points d'intérêt « bleus » (DIST BLEU KM) n'est pas significative, tandis que la distance aux points « rouges » (DIST ROUGE KM) présente un coefficient positif significatif, compatible avec l'idée que ces points constituent des sources de nuisance dont on valorise l'éloignement.

Du côté des effets spatiaux (*variables en lag*), peu de coefficients sont significatifs, ce qui indique que, dans ce modèle, la plupart des caractéristiques des logements voisins n'exercent pas d'effet direct substantiel sur le prix local lorsqu'elles sont introduites conjointement aux caractéristiques propres du bien. Deux exceptions notables apparaissent : la distance moyenne aux espaces verts dans le voisinage (lag.DIST VERT KM), qui présente un coefficient positif significatif, et la distance aux points rouges dans le voisinage (lag.DIST ROUGE KM), dont le coefficient négatif et significatif suggère que des voisins plus proches de ces points connaissent une moindre valorisation, ce qui peut se répercuter sur le prix local.

Le coefficient de dépendance spatiale ρ est positif (0,0651) et significatif, confirmant la présence d'une autocorrélation spatiale des prix immobiliers, bien que d'ampleur modérée. Le test de rapport de vraisemblance (LR) indique que le SDM fournit un ajustement significativement meilleur qu'un modèle OLS sans structure spatiale, tandis que l'AIC inférieur à celui du modèle linéaire confirme cette amélioration. Dans l'ensemble, ces résultats indiquent que la performance énergétique, les caractéristiques physiques et certaines proximités environnementales jouent un rôle important dans la formation des prix, tandis que les interactions spatiales se manifestent

surtout à travers le processus autorégressif global et quelques effets spécifiques liés aux distances aux aménités.

TABLE 39 – Effets directs et indirects du modèle SDM (impacts), matrice PPV3

Variable	Direct	Indirect	Total
dpeA	0.1731	-0.0184	0.1547
dpeB	0.1550	0.0609	0.2159
dpeC	0.0701	0.0150	0.0851
dpeE	-0.0755	-0.0076	-0.0831
dpeF	-0.1529	-0.0017	-0.1545
dpeG	-0.2039	-0.0047	-0.2086
age_batiment	-0.0016	-0.00003	-0.0016
surface_habitable	0.0044	0.0005	0.0049
surface_terrain	0.0001	0.00001	0.0001
nb_etages	-0.0034	0.0109	0.0075
nb_pieces	0.0460	-0.0010	0.0450
DIST_VERT_KM	-0.3098	0.3526	0.0428
DIST_BLEU_KM	-0.0164	-0.0444	-0.0609
DIST_ROUGE_KM	0.6446	-0.5948	0.0498

L’analyse des mesures d’impact issues du modèle SDM avec la matrice de pondération PPV1 permet de distinguer les effets directs, représentant l’influence moyenne d’une variable sur le prix des logements au sein de la même unité spatiale, des effets indirects, ou de débordement, qui reflètent l’impact moyen exercé sur les prix des logements voisins, ainsi que des effets totaux, somme des deux précédents. L’étiquette DPE de référence est toujours l’étiquette D.

Les résultats montrent que les logements classés A, B et C présentent des effets directs positifs par rapport à la classe D, avec des effets totaux respectifs de 0,1547, 0,2159 et 0,0851, traduisant une valorisation immobilière associée à une meilleure performance énergétique. L’effet indirect est faiblement négatif pour la catégorie A, modérément positif pour la catégorie B et légèrement positif pour la catégorie C, indiquant que la présence de voisins plus performants énergétiquement, notamment en classe B, est corrélée à une hausse des prix locaux. À l’inverse, les catégories E, F et G présentent des effets totaux négatifs (respectivement -0,0831, -0,1545 et -0,2086), principalement portés par des effets directs défavorables, ce qui confirme qu’une performance énergétique inférieure se traduit par une décote sur le marché, sans compensation notable par les effets indirects.

Parmi les variables continues, l’âge du bâtiment exerce un effet total négatif faible (-0,0016), cohérent avec une dépréciation progressive des biens plus anciens. La surface habitable et la surface du terrain présentent toutes deux des effets totaux positifs (0,0049 et 0,0001), traduisant une valorisation croissante avec la taille, renforcée dans le cas de la surface habitable par un léger effet positif lié aux logements voisins plus spacieux. Le nombre d’étages a un effet direct

légèrement négatif mais un effet indirect positif, aboutissant à un effet total de 0,0075, suggérant que si la verticalité d'un logement n'accroît pas sa valeur propre, elle peut être associée à un environnement bâti valorisant. Le nombre de pièces a un effet total positif (0,0450), porté presque exclusivement par l'effet direct, indiquant une préférence marquée pour des logements plus compartimentés.

Les variables de localisation révèlent des dynamiques spatiales plus complexes. L'éloignement aux espaces verts (DIST VERT KM) est associé à un effet direct négatif important (-0,3098), confirmant que la proximité à ces aménités est valorisée localement, mais l'effet indirect positif substantiel (0,3526) compense cette tendance dans le calcul de l'effet total, qui devient légèrement positif (0,0428), traduisant un possible effet de substitution ou de composition spatiale. La distance aux points d'intérêt « bleus » (DIST BLEU KM) présente un effet total négatif (-0,0609), dominé par l'effet indirect négatif, suggérant que la proximité à ces éléments est favorable, tant localement que dans le voisinage. Enfin, la distance aux points « rouges » (DIST ROUGE KM) affiche un effet direct positif marqué (0,6446), indiquant que l'éloignement de ces sources de nuisance est apprécié, mais cet effet est en partie annulé par un effet indirect négatif (-0,5948), de sorte que l'effet total (0,0498) demeure légèrement positif.

Dans l'ensemble, ces résultats confirment que la performance énergétique et les caractéristiques physiques des logements constituent des déterminants majeurs des prix immobiliers, tandis que certaines aménités ou nuisances environnementales induisent des effets spatiaux différenciés, dont la prise en compte est essentielle pour une analyse fine des mécanismes de valorisation.

Modèle SDM : matrice PPV3

TABLE 40 – Résultats du modèle SDM (PPV3)

Variable	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	9.6973	0.4001	24.2362	< 2.2e-16
dpeA	0.1339	0.1043	1.2842	0.1991
dpeB	0.1402	0.0520	2.6939	0.0071
dpeC	0.0672	0.0203	3.3049	0.0010
dpeE	-0.0784	0.0195	-4.0175	5.88e-05
dpeF	-0.1556	0.0265	-5.8736	4.26e-09
dpeG	-0.1984	0.0359	-5.5267	3.26e-08
age_batiment	-0.0016	0.0003	-4.7200	2.36e-06
surface_habitable	0.0045	0.0003	16.2406	< 2.2e-16
surface_terrain	0.0001	0.00002	5.8251	5.71e-09
nb_etages	-0.0041	0.0162	-0.2510	0.8018
nb_pieces	0.0445	0.0077	5.7518	8.83e-09
DIST_VERT_KM	-0.0027	0.1463	-0.0188	0.9850
DIST_BLEU_KM	-0.0501	0.1352	-0.3707	0.7108
DIST_ROUGE_KM	0.4639	0.1409	3.2915	0.0010
lag.dpeA	-0.3556	0.1811	-1.9633	0.0496
lag.dpeB	0.1328	0.0763	1.7401	0.0818
lag.dpeC	-0.0081	0.0358	-0.2266	0.8207
lag.dpeE	0.0082	0.0346	0.2374	0.8123
lag.dpeF	0.0463	0.0440	1.0529	0.2924
lag.dpeG	-0.0637	0.0624	-1.0197	0.3079
lag.age_batiment	0.0001	0.0005	0.1760	0.8603
lag.surface_habitable	-0.0002	0.0005	-0.4605	0.6452
lag.surface_terrain	-0.0001	0.00004	-1.4443	0.1487
lag.nb_etages	-0.0074	0.0296	-0.2496	0.8029
lag.nb_pieces	-0.0041	0.0134	-0.3078	0.7582
lag.DIST_VERT_KM	0.0448	0.1470	0.3047	0.7606
lag.DIST_BLEU_KM	-0.0046	0.1358	-0.0342	0.9727
lag.DIST_ROUGE_KM	-0.4217	0.1411	-2.9898	0.0028
Rho :		0.16967 (ASE : 0.033403)		
Test du rapport de vraisemblance :		22.784 (p-value : 1.8127e-06)		
Statistique de Wald :		25.802 (p-value : 3.7835e-07)		
Log-vraisemblance :		-115.763		
AIC :		293.53 (AIC LM : 314.31)		
Test LM d'autocorrélation résiduelle :		2.3904 (p-value : 0.12208)		

L'estimation du modèle Spatial Durbin avec la matrice de pondération PPV3 met en évidence l'existence d'une dépendance spatiale significative des prix immobiliers, comme l'indique le coefficient autorégressif ρ , positif (0,1697) et significatif, ainsi que les tests du rapport de vraisemblance et de Wald qui confirment la supériorité de ce modèle par rapport à une spécification

sans structure spatiale. Les coefficients présentés correspondent aux estimations brutes du SDM et ne peuvent pas être directement interprétés comme des effets marginaux totaux, mais leur signe et leur significativité renseignent sur la direction et la robustesse des associations.

En prenant la classe D comme référence pour le diagnostic de performance énergétique, les coefficients associés aux classes B et C sont positifs et significatifs, suggérant une valorisation relative de ces niveaux de performance énergétique par rapport à la catégorie de référence. La classe A présente également un coefficient positif mais non significatif, tandis que les classes E, F et G affichent des coefficients négatifs et fortement significatifs, ce qui traduit une dépréciation associée à une moindre performance énergétique. Parmi les caractéristiques physiques, l'âge du bâtiment présente un coefficient négatif significatif, confirmant qu'une plus grande ancienneté est corrélée à une baisse du prix. La surface habitable et la surface du terrain sont toutes deux associées à des coefficients positifs et significatifs, traduisant une valorisation des biens plus grands. Le nombre de pièces présente également un effet positif significatif, alors que le nombre d'étages ne montre pas de relation statistiquement significative avec le prix.

Les variables de localisation indiquent qu'une plus grande distance aux points d'intérêt « rouges » est associée à un coefficient positif significatif, ce qui suggère que l'éloignement de ces sources de nuisance est apprécié par le marché. En revanche, les distances aux espaces verts et aux points « bleus » ne sont pas significativement associées au prix dans cette spécification.

Concernant les effets spatiaux représentés par les variables en lag, peu de relations significatives sont observées. La principale exception concerne lag.dpeA, qui présente un coefficient négatif significatif au seuil de 5%, indiquant qu'avoir des voisins classés en A est associé à un ajustement à la baisse du prix local, toutes choses égales par ailleurs. De manière similaire, lag.DIST ROUGE KM est négatif et significatif, ce qui suggère que la proximité des voisins à ces points d'intérêt « rouges » est corrélée à une baisse du prix local, peut-être par un effet de contamination des nuisances dans l'espace. Les autres variables en lag, qu'il s'agisse des DPE, des caractéristiques physiques ou des distances environnementales, ne présentent pas de relation statistiquement significative dans cette spécification.

Globalement, ce modèle met en évidence l'importance de la performance énergétique et des caractéristiques intrinsèques du logement dans la formation des prix, avec un rôle plus limité des interactions spatiales spécifiques entre caractéristiques voisines, même si l'effet autorégressif global reste significatif et non négligeable.

TABLE 41 – Effets directs et indirects du modèle SDM (impacts), matrice PPV3

Variable	Direct	Indirect	Total
dpeA	0.1191	-0.3862	-0.2670
dpeB	0.1471	0.1816	0.3287
dpeC	0.0674	0.0038	0.0712
dpeE	-0.0786	-0.0059	-0.0845
dpeF	-0.1547	0.0231	-0.1315
dpeG	-0.2027	-0.1129	-0.3156
age_batiment	-0.0016	-0.0002	-0.0018
surface_habitable	0.0045	0.0006	0.0051
surface_terrain	0.00009	-0.00004	0.00005
nb_etages	-0.0044	-0.0094	-0.0138
nb_pieces	0.0446	0.0040	0.0486
DIST_VERT_KM	-0.0008	0.0514	0.0506
DIST_BLEU_KM	-0.0507	-0.0152	-0.0659
DIST_ROUGE_KM	0.4486	-0.3979	0.0508

L'estimation des effets directs, indirects et totaux du modèle Spatial Durbin avec matrice PPV3 permet de quantifier l'impact marginal de chaque variable explicative sur le prix, en distinguant l'effet local sur la commune considérée et l'effet de diffusion vers les communes voisines. En prenant la classe D comme référence pour le diagnostic de performance énergétique, les résultats montrent que les classes B et C sont associées à des effets totaux positifs sur le prix, traduisant une valorisation par rapport à D, avec un impact direct dominant pour C et un effet combiné direct et indirect pour B. La classe A présente en revanche un effet total négatif, dû à un impact indirect fortement négatif qui contrebalance un effet direct positif, ce qui suggère que la présence de logements voisins très performants énergétiquement pourrait exercer une pression à la baisse sur les prix locaux. Les classes E, F et G affichent des effets totaux négatifs, plus marqués pour G, reflétant une pénalisation des logements moins performants sur le marché immobilier, bien que pour F l'impact indirect soit légèrement positif.

Parmi les caractéristiques physiques, l'âge du bâtiment présente un effet total négatif, confirmant qu'une plus grande ancienneté entraîne une dépréciation. La surface habitable a un effet total positif, de même que le nombre de pièces, traduisant la valorisation des logements plus spacieux. La surface du terrain a un effet total très faible et peu significatif sur le plan économique. Le nombre d'étages montre un effet total légèrement négatif, suggérant que la verticalité du bâtiment ne constitue pas un facteur de valorisation dans cet échantillon.

S'agissant des variables de localisation, la distance aux espaces verts présente un effet total légèrement positif, ce qui peut indiquer qu'un éloignement modéré de ces espaces n'est pas pénalisant, voire marginalement valorisé dans certains contextes. En revanche, la distance aux points d'intérêt « bleus » est associée à un effet total négatif, traduisant une dépréciation liée à l'éloignement de ces aménités. Enfin, la distance aux points « rouges » présente un effet total faible et

positif, résultant d'un fort effet direct positif atténué par un effet indirect négatif, ce qui pourrait refléter des préférences différencierées entre localisation propre et influence des localisations voisines.

Dans l'ensemble, les résultats confirment le rôle central de la performance énergétique, de la taille et du nombre de pièces dans la formation des prix, tout en montrant que certains effets de voisinage peuvent atténuer ou inverser l'impact attendu de certaines variables, en particulier pour les classes énergétiques extrêmes comme A et G.

5 Modélisation machine-learning

Cette partie 5 propose une méthodologie rigoureuse d'analyse prédictive des prix immobiliers par apprentissage supervisé. La variable cible est transformée en logarithme afin de stabiliser la variance des résidus, de normaliser une distribution généralement log-normale et de linéariser les relations multiplicatives entre variables explicatives et prix. Le modèle prédit ainsi le logarithme du prix, nécessitant une exponentiation pour retrouver les valeurs initiales, ce qui introduit un biais de retransformation à considérer dans l'interprétation. Les étiquettes DPE sont encodés de façon ordinaire, de G (1) à A (7), afin de refléter la continuité de l'efficacité énergétique et son effet graduel sur la valeur. La gestion des valeurs aberrantes repose sur une winsorisation appliquée aux 5e et 95e percentiles calculés uniquement sur l'échantillon d'entraînement, puis appliqués au test pour éviter tout data leakage [2].

Deux approches de modélisation sont comparées. La régression linéaire multiple suppose une relation additive entre variables standardisées et logarithme du prix ; ses coefficients, évalués par tests de Student, permettent une interprétation statistique, mais l'approche reste contrainte par ses hypothèses de linéarité, d'homoscédasticité et de normalité des résidus. À l'inverse, la méthode Random Forest, non paramétrique et fondée sur le bagging et la sélection aléatoire de variables, capture les relations non linéaires et interactions complexes. L'agrégation des prédictions d'arbres multiples confère une robustesse adaptée à la complexité du marché immobilier.

L'optimisation du Random Forest s'appuie sur une recherche aléatoire avec validation croisée à trois plis, plus efficiente qu'une recherche exhaustive. Les hyperparamètres ciblés régulent la complexité du modèle : nombre d'arbres (`n_estimators`), profondeur (`max depth`), taille minimale des divisions (`min samples split` et `min samples leaf`) et nombre de variables testées (`max features`). L'optimisation repose sur la minimisation de l'erreur quadratique moyenne négative, cohérente avec la métrique d'évaluation finale.

Enfin, l'interprétabilité du modèle repose sur un ensemble de techniques complémentaires. Les Partial Dependence Plots (PDP) mettent en évidence les effets marginaux moyens des variables, et leurs versions bidimensionnelles révèlent des interactions, par exemple entre surface habitable et terrain. L'importance par permutation, plus robuste que l'importance native des arbres, mesure quant à elle la contribution effective de chaque variable à la performance prédictive.

Au-delà de ces analyses globales, la véritable plus-value de cette approche économétrique réside dans l'interprétation locale, c'est-à-dire la capacité à expliquer les prédictions individuelles du modèle, notamment avec les Shapley Values. Cette lecture observation par observation permet de comprendre quels facteurs spécifiques déterminent la valeur prédictive pour un bien donné, offrant ainsi un outil opérationnel d'évaluation et de décision en contexte immobilier.

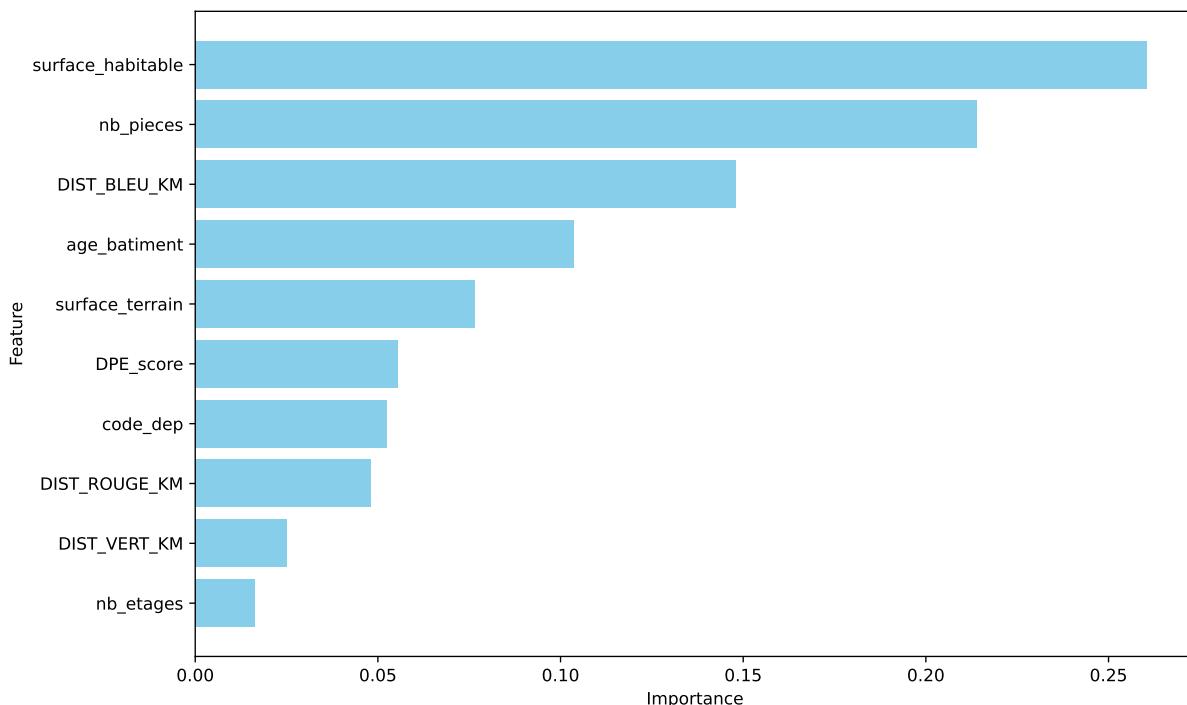


FIGURE 26 – Importance des features dans le modèle Random Forest

Le graphique issu du modèle Random Forest (*Figure 26*) met en évidence l'importance relative des différentes variables explicatives dans la prédiction du prix. Il apparaît que la surface habitable est de loin la variable la plus déterminante, suivie de près par le nombre de pièces, ce qui souligne le rôle central des caractéristiques dimensionnelles du logement dans la formation des valeurs. La distance aux points d'intérêt « bleus » occupe la troisième position, suggérant que la proximité à ce type d'aménités influence notablement les prix. L'âge du bâtiment arrive ensuite, traduisant l'effet non négligeable de la vétusté sur la valorisation immobilière. La surface du terrain joue également un rôle significatif, bien que moindre que celui des variables précédentes.

En comparaison, l'étiquette DPE et le départements ont un poids relativement modéré dans le modèle, ce qui indique que, dans ce contexte, ils contribuent à la prédiction mais restent moins influents que la taille, la configuration ou la localisation fine par rapport aux aménités. La distance aux points d'intérêt « rouges » et la distance aux espaces verts ont un impact encore plus faible, reflétant une importance secondaire dans la hiérarchisation des facteurs explicatifs. Enfin, le nombre d'étages se révèle être la variable la moins déterminante, ce qui suggère qu'il n'apporte que peu d'information supplémentaire dans la prédiction du prix par rapport aux autres caractéristiques.

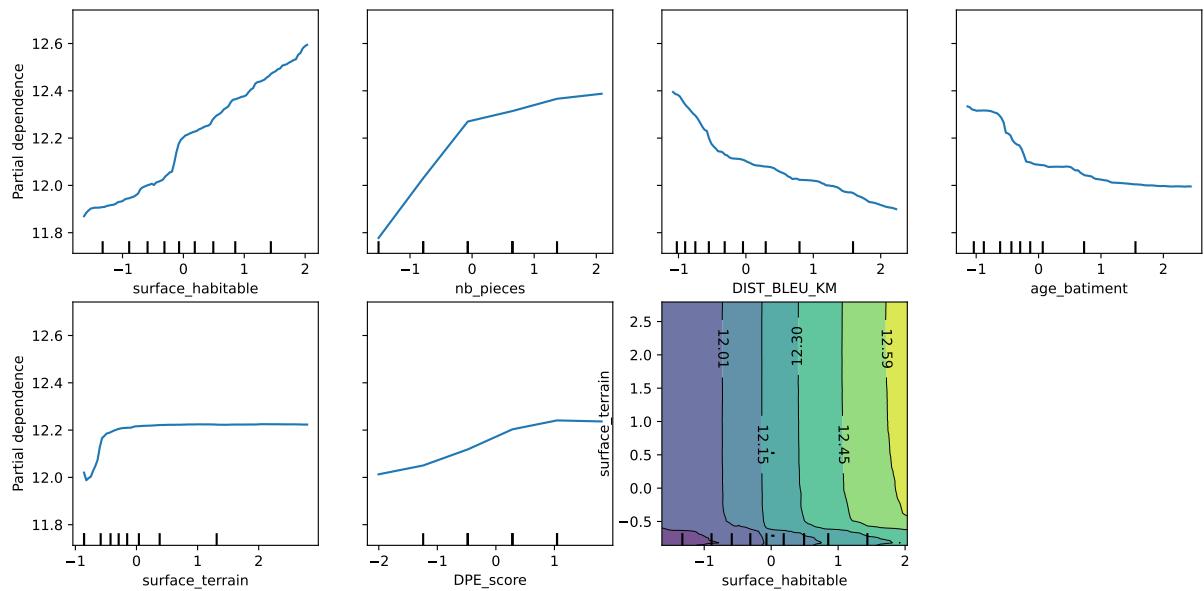


FIGURE 27 – Partial Dependence Plots (PDP)

Les PDP (Figure 27) permettent d'isoler et de visualiser l'effet marginal de chaque variable sur le prix des logements, toutes choses égales par ailleurs. La surface habitable présente une relation croissante quasi linéaire avec le prix, indiquant qu'une augmentation de la superficie se traduit par une hausse continue et marquée de la valeur du bien. Le nombre de pièces montre également une tendance positive, mais avec un effet plus fort dans les faibles valeurs, suggérant un gain de prix plus important lors du passage de petites à moyennes surfaces habitables, puis un effet qui se stabilise.

La distance aux entités bleues affiche une relation décroissante : plus un logement est éloigné de la mer, de l'océan ou des lacs, plus son prix tend à diminuer, ce qui reflète la forte valorisation de la proximité avec ces aménités. L'âge du bâtiment a également un effet négatif, plus marqué pour les biens les plus récents et qui tend à se stabiliser pour les bâtiments plus anciens, traduisant une décote progressive avec la vétusté.

La surface du terrain présente une relation globalement croissante avec le prix, mais avec un effet marginal décroissant : l'augmentation initiale de la surface du terrain entraîne une hausse notable de la valeur, puis l'impact supplémentaire se réduit. L'étiquette DPE, codé de 1 (G) à 7 (A), a un effet positif sur le prix, ce qui signifie qu'une meilleure performance énergétique est associée à une valorisation plus élevée du logement. Enfin, le graphique bidimensionnel combinant surface habitable et surface du terrain montre un effet complémentaire : les prix les plus élevés s'observent lorsque les deux dimensions sont importantes, confirmant que la combinaison d'un logement spacieux et d'un grand terrain constitue un facteur de forte valorisation.

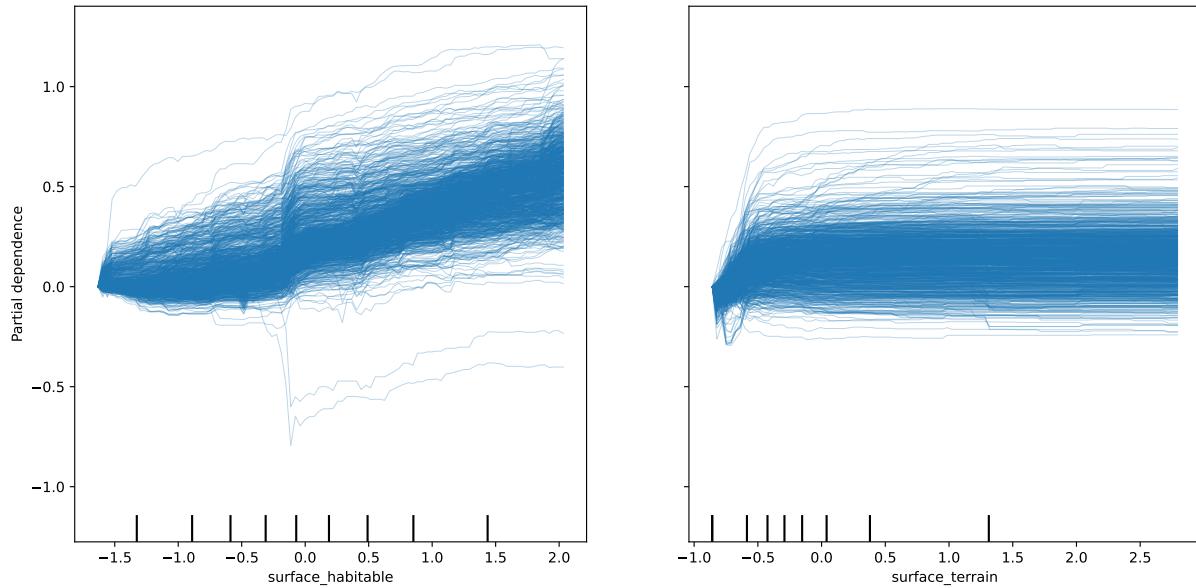


FIGURE 28 – Partial Dependence Plots (PDP) pour la surface habitable et la surface de terrain

Pour la surface habitable, on observe une tendance globale croissante : l'augmentation de la surface est associée à une hausse du prix prédict, ce qui confirme les tendances précédentes. Cependant, l'éventail des courbes individuelles met en évidence une variabilité importante selon les logements : l'effet positif moyen s'accompagne d'un écart type non négligeable, en particulier pour les surfaces élevées, ce qui traduit une influence du contexte ou d'interactions avec d'autres variables (par exemple, la localisation ou l'état du bien).

Pour la surface du terrain, l'effet moyen est également positif mais présente une saturation : le prix augmente surtout pour les petites et moyennes surfaces, puis l'effet marginal tend à se stabiliser. Ici encore, la dispersion des courbes individuelles est visible, bien que moins marquée que pour la surface habitable, indiquant un effet plus homogène entre logements.

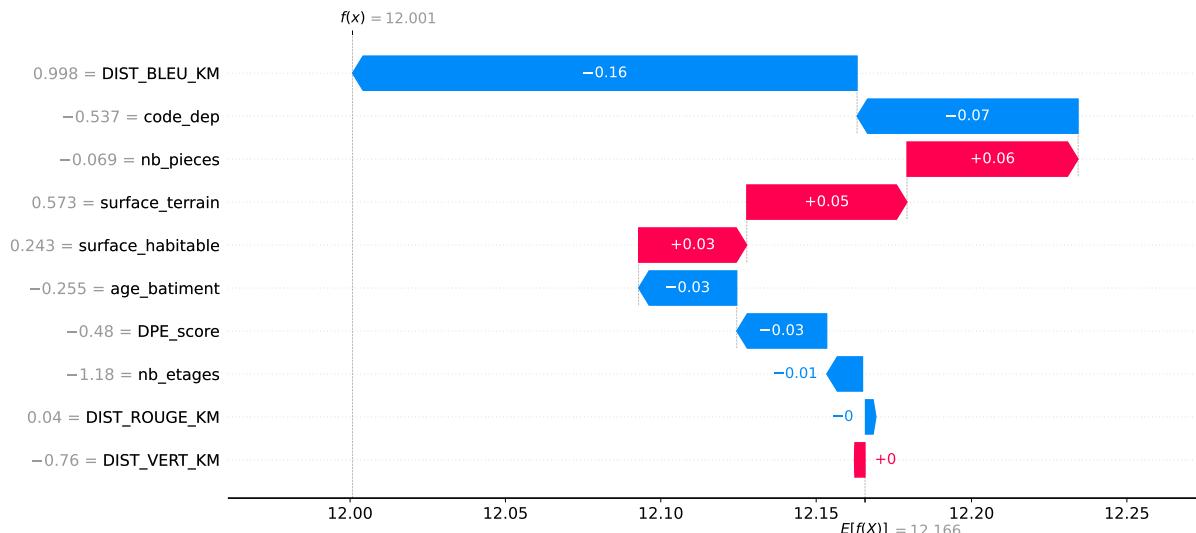


FIGURE 29 – Shapley Values pour l'individu 50

La figure 29 présente les contributions locales des variables explicatives à la prédiction du logarithme du prix pour l'individu 50, telles que calculées par les valeurs de Shapley. La prédiction du modèle pour cet individu est de 12.001, soit légèrement inférieure à la valeur moyenne attendue $E[f(X)] = 12.166$. Autrement dit, l'ensemble des caractéristiques de ce logement tend globalement à réduire sa valeur par rapport à la moyenne de l'échantillon.

La variable qui exerce l'effet négatif le plus marqué est la distance aux entités bleues (DIST BLEU KM), qui contribue à diminuer la prédiction de -0.16. Cela reflète le fait que ce logement, situé à environ 1 km de ces aménités, est relativement éloigné des zones attractives de type littoral ou lacustre, ce qui réduit sa valeur. La variable départementale (code dep) a également un effet négatif important (-0.07), suggérant que l'appartenance à ce territoire spécifique est associée à des prix inférieurs à la tendance générale.

À l'inverse, certaines caractéristiques contribuent positivement. C'est le cas du nombre de pièces (+0.06), de la surface du terrain (+0.05) et dans une moindre mesure de la surface habitable (+0.03). Ces éléments traduisent un profil de logement relativement spacieux, ce qui accroît la valeur prédictive.

D'autres variables, comme l'âge du bâtiment (-0.03), le score DPE (-0.03) et le nombre d'étages (-0.01), exercent de faibles effets négatifs, cohérents avec une dépréciation liée à la vétusté et à une performance énergétique médiocre. Enfin, les distances aux entités rouges et vertes n'ont quasiment aucun impact dans ce cas particulier.

Au total, cette décomposition permet de comprendre que, pour l'individu 50, les atouts liés à la taille du logement sont insuffisants pour compenser les effets négatifs de la localisation et de la performance énergétique. Ce sont donc surtout la distance à l'eau et le contexte départemental qui expliquent que la prédiction finale du modèle se situe légèrement en dessous de la moyenne.

6 Conclusion

Cette recherche a étudié les déterminants des prix immobiliers à Brest, en mettant l'accent sur l'impact du Diagnostic de Performance Énergétique (DPE) dans un contexte marqué par la loi Climat et Résilience. L'articulation entre économétrie spatiale et Machine Learning a permis d'apporter un éclairage original sur la manière dont les caractéristiques intrinsèques, environnementales et énergétiques structurent la valeur immobilière.

L'analyse a d'abord mis en évidence une autocorrélation spatiale significative des prix, confirmant que les biens immobiliers ne s'évaluent pas isolément mais en interaction avec leur voisinage. L'usage de modèles spatiaux adaptés a permis de décomposer ces interactions en effets directs et indirects, révélant la propagation de l'influence des variables explicatives à travers l'espace. Le DPE apparaît alors comme un critère structurant : les logements performants bénéficient d'une

prime de valorisation, tandis que les passoires thermiques subissent une décote nette. À côté de cette dimension énergétique, la surface habitable, le nombre de pièces, l'âge du bâtiment et la proximité au littoral se sont affirmés comme des déterminants centraux et robustes du prix.

L'apport du Machine Learning a complété ces résultats de deux manières. D'une part, il a confirmé la prépondérance des caractéristiques physiques et géographiques dans la formation des prix. D'autre part, il a enrichi l'interprétation par une analyse plus locale : au-delà des tendances globales, les méthodes mobilisées ont permis de comprendre, bien par bien, comment chaque facteur contribue à la valeur. Cette lecture individualisée, qui révèle la diversité des situations, constitue un atout majeur pour une compréhension opérationnelle des dynamiques de marché.

Au-delà de leurs apports scientifiques, ces résultats ont une portée concrète pour le Crédit Mutuel Arkéa. Ils fournissent des outils pour anticiper les évolutions du marché, adapter les politiques de financement et mieux intégrer le risque lié aux biens énergivores dans l'évaluation des actifs. En ce sens, ils permettent d'aligner la stratégie de financement sur les enjeux économiques, environnementaux et sociaux, tout en accompagnant les ménages dans la transition énergétique.

Enfin, des perspectives de recherche s'ouvrent autour de l'élargissement du périmètre d'analyse à d'autres territoires, de l'intégration de séries temporelles plus longues ou encore de l'enrichissement des données par les travaux de rénovation énergétique. De tels prolongements permettraient de mesurer plus finement la dynamique de valorisation liée au DPE et de renforcer la compréhension des interactions entre énergie, habitat et marché immobilier.

Références

- [1] Catherine BAUMONT, Masha MASLIANSKAIA-PAUTREL et Pierre VOYÉ. *Hedonic estimation of the green value of residential housing*. Rapp. tech. 2019.
- [2] Mohamed Aly BOUKE, Saleh Ali ZAID et Azizol ABDULLAH. "Implications of data leakage in machine learning preprocessing : a multi-domain investigation". In : (2024).
- [3] CHAMBRE DES NOTAIRES DE BRETAGNE. *Baromètre de l'immobilier - Janvier 2025*. Notaire et Breton. Jan. 2025. URL : <https://www.notaireetbreton.bzh/actualites/barometre-de-l-immobilier-janvier-2025> (visité le 14/08/2025).
- [4] COMMISSARIAT À LA COHÉSION DES TERRITOIRES. *Rapport de Sûreté 2024*. Consulté le 19/08/2025. Commissariat à la Cohésion des Territoires (CCR), juill. 2024.
- [5] COMMISSARIAT GÉNÉRAL AU DÉVELOPPEMENT DURABLE. *Chiffres clés des risques naturels – 19. Mouvements de terrain hors retrait-gonflement des argiles*. Ministère de la Transition Écologique et de la Cohésion des Territoires. 2023. URL : <https://www.statistiques.developpement-durable.gouv.fr/edition-numerique/chiffres-cles-risques-naturels/19-mouvements-de-terrain-hors-retrait> (visité le 19/08/2025).
- [6] COMMISSARIAT GÉNÉRAL AU DÉVELOPPEMENT DURABLE. *Chiffres clés des risques naturels – Sinistralité liée aux inondations*. Consulté le 19/08/2025. Ministère de la Transition Écologique et de la Cohésion des Territoires, 2023.
- [7] COPERNICUS LAND MONITORING SERVICE. *Corine Land Cover (CLC)*. <https://land.copernicus.eu/en/products/corine-land-cover>. Accessed : 2025-08-04. 2023.
- [8] DIRECTION INTERMINISTÉRIELLE DU NUMÉRIQUE (DINUM). *Base Adresse Nationale*. 2024. URL : <https://api.gouv.fr/les-api/base-adresse-nationale> (visité le 19/08/2025).
- [9] DRIAS — LES FUTURS DU CLIMAT. *Accompagnement*. 2023. URL : <https://www.drias-climat.fr/acc陪agnement/sections/181> (visité le 19/08/2025).
- [10] J Paul ELHORST. "Applied spatial econometrics : raising the bar". In : *Spatial economic analysis* 5.1 (2010), p. 9-28.
- [11] Ehsan FOROUTAN et al. "Assessing heat vulnerability in Philadelphia using geographically weighted principal component analysis (GWPCA) : A geospatial big data-driven approach". In : *International Journal of Applied Earth Observation and Geoinformation* 127 (2024), p. 103653.
- [12] FRANCE 3 PARIS ÎLE-DE-FRANCE. *Interdiction de louer des logements classés F et G : une "véritable catastrophe" pour les propriétaires et locataires*. Article sur les conséquences de l'interdiction de location des passoires thermiques en Île-de-France. Franceinfo. 20 juin 2025. URL : <https://france3-regions.franceinfo.fr/paris-ile-de-france/paris/interdiction-de-louer-des-logements-classes-f-et-g-une-veritable-catastrophe-pour-les-proprietaires-et-locataires-3085171.html> (visité le 14/08/2025).
- [13] FRANCE ASSUREURS. *L'assurance des événements naturels en 2022*. Consulté le 19/08/2025. France Assureurs, 2023.

- [14] Revue Politique Internationale LESSOR. “Finance et risque climatique : la “tragédie des horizons””. In : *PolitiqueInternationale, Revue n°162, dossier spécial “Investissement responsable – Lessor”* (2023). mise en ligne le 4mai2023.
- [15] MÉTÉO-FRANCE. *Canicule, vague ou pic de chaleur*. 2023. URL : <https://meteofrance.com/actualites-et-dossiers/comprendre-la-meteo/canicule-vague-ou-pic-de-chaleur> (visité le 19/08/2025).
- [16] Sachi PERERA et al. “Deriving Drought Vulnerability Index using Geographically Weighted Principal Component Analysis (GWPCA) and K-Means Clustering for Nile Basin”. In : *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2022, p. 3187-3190.
- [17] Subham ROY et al. “GWPCA-based spatial analysis of urban vitality : a comparative assessment of three high-altitude Himalayan towns in India”. In : *Journal of Spatial Science* 69.2 (2024), p. 593-620.
- [18] SELoger. *Data DPE : Tout savoir sur les données de performance énergétique*. Données et analyses sur les Diagnostics de Performance Énergétique (DPE) en France. SeLoger. 2023. URL : <https://www.seloger.com/data-dpe/> (visité le 14/08/2025).

Table des figures

1	Organigramme fonctionnel - Focus sur les Risques	6
2	Capture d'écran de l'outil PRISM simulation v.1.1.2	11
3	Comparaison entre les différentes méthodes pour l'aléa inondation	23
4	Capture d'écran de l'application en ligne	24
5	Capture d'écran de l'application en ligne	25
6	Capture d'écran de l'application en ligne	26
7	Calendrier mensuel du stage (Mars à Aout 2025)	30
8	Évolution des ventes immobilières à Brest par type de bien	34
9	Évolution des ventes immobilières à Brest par classe DPE et type de bien	35
10	Cartographie des prix des appartements à Brest	36
11	Cartographie des prix des maisons à Brest	36
12	Évolution des prix à Brest par type de biens	37
13	Évolution des prix à Brest par classe DPE	38
14	Distribution des prix des appartements à Brest	38
15	Distribution des prix des maisons à Brest	39
16	Cartes des étiquettes DPE à Brest : appartements et maisons	40
17	Distribution globale des étiquettes DPE à Brest	41
18	Distribution des étiquettes DPE à Brest : appartements et maisons	42
19	Distribution des âges des appartements à Brest	43
20	Distribution des âges des maisons à Brest	43
21	Distribution de la surface habitable des appartements à Brest	44
22	Distribution de la surface habitable des maisons à Brest	45
23	Distribution de la surface de terrain des maisons à Brest	46
24	Matrice des corrélations des maisons	48
25	Matrice des corrélations des appartements	48
26	Importance des features dans le modèle Random Forest	71
27	Partial Dependence Plots (PDP)	72
28	Partial Dependence Plots (PDP) pour la surface habitable et la surface de terrain	73
29	Shapley Values pour l'individu 50	73
30	Pondération au code postal d'un indicateur DRIAS	83
31	Comparaison entre PRISM, les CATNAT et l'ACP spatialisée (GWPCA) sur l'aléa RGA	83
32	Schéma explicatif des différentes dynamiques de la modélisation de la NGFS	84
33	Schéma explicatif des différentes étapes de modélisation de la NGFS	85
34	Impact des étiquettes DPE sur le prix des maisons en Bretagne, de 2020 à 2024	86
35	Impact des étiquettes DPE sur le prix des appartements en Bretagne, de 2020 à 2024	86
36	Impact des différentes caractéristiques sur le prix de vente des maisons de 2020 à 2024	87
37	Impact des différentes caractéristiques sur le prix de vente des appartements de 2020 à 2024	87
38	Distribution du nombre d'étages des appartements à Brest	88
39	Distribution du nombre d'étages des maisons à Brest	88

40	Distribution du nombre de pièces des maisons à Brest	89
41	Distribution du nombre de pièces des maisons à Brest	89
42	Distance (km) aux entités vertes des appartements à Brest	90
43	Distance (km) aux entités vertes des maisons à Brest	90
44	Distance (km) aux entités bleu des maisons à Brest	91
45	Distance (km) aux entités bleus des maisons à Brest	91
46	Distance (km) aux entités rouges des appartements à Brest	92
47	Distance (km) aux entités rouges des maisons à Brest	92
48	Indice de Moran pour la matrice Reine sur les maisons	93
49	Indice de Moran pour la matrice PPV1 sur les maisons	94
50	Indice de Moran pour la matrice PPV3 sur les maisons	95
51	Indice de Moran pour la matrice Reine sur les appartements	96
52	Indice de Moran pour la matrice PPV1 sur les appartements	97
53	Indice de Moran pour la matrice PPV3 sur les appartements	98

Liste des tableaux

1	Scoring de l'aléa Inondation	13
2	Scoring de l'aléa Retrait Gonflement des Argiles	14
3	Scoring de l'aléa Tempête-Grêle-Neige	15
4	Scoring de l'aléa Feux de Forêts	15
5	Scoring de l'aléa Glissements de terrain	16
6	Scoring de l'aléa Sécheresse	16
7	Scoring de l'aléa Vague de chaleur	17
8	Scoring de l'aléa Vague de froid	17
9	Scoring de l'aléa Vague de gel	18
10	Scoring de l'aléa Variation des températures	18
11	Scoring de l'aléa Variation des précipitations	19
12	Scoring de l'aléa Stress hydrique	19
13	Scoring de l'aléa Érosion du trait de côte	20
14	Scoring de l'aléa Évolution des régimes des vents	20
15	Scoring de l'aléa Hauteur du niveau de la mer	21
16	Ratios de sinistres pour les inondations (en pourcentage de la valeur du bien)	28
17	Ratios de sinistres pour les inondations par niveau de risque PRISM (en pourcentage de la valeur du bien)	28
18	Analyse des ratios de sinistres et charges par niveau de risque PRISM pour les inondations	29
19	Statistiques descriptives des prix (€) des appartements	38
20	Statistiques descriptives des prix des maisons	39
21	Statistiques descriptives des âges des appartements	43
22	Statistiques descriptives des âges des maisons	43
23	Statistiques descriptives des surfaces habitables des appartements	44
24	Statistiques descriptives des surfaces habitables des maisons	45
25	Statistiques descriptives des surfaces de terrain des maisons	46
26	Tests de Moran pour différentes matrices de pondération spatiale	51
27	Indice de Moran selon différentes matrices de pondération	51
28	Tests de Moran sur les résidus de la régression OLS	52
29	Tests de dépendance spatiale utilisant différentes matrices de pondération	53
30	Tests de rapport de vraisemblance (LR) et critères AIC pour la sélection de modèles spatiaux pour les appartements	54
31	Tests de Moran sur les résidus	54
32	Tests de dépendance spatiale utilisant différentes matrices de pondération	55
33	Tests de rapport de vraisemblance (LR) et critères AIC pour la sélection de modèles spatiaux	56
34	Résultats du modèle SAR estimé avec la matrice de pondération PPV1 pour les maisons	57
35	Effets directs et indirects du modèle SAR (impacts), matrice PPV1	58
36	Résultats du modèle SDM estimé avec la matrice de pondération PPV3	60

37	Effets directs et inidrects du modèle SDM (impacts)	61
38	Résultats du modèle SDM estimé avec la matrice de pondération PPV1 pour les maisons	63
39	Effets directs et indirects du modèle SDM (impacts), matrice PPV3	65
40	Résultats du modèle SDM (PPV3)	67
41	Effets directs et indirects du modèle SDM (impacts), matrice PPV3	69

7 Annexes

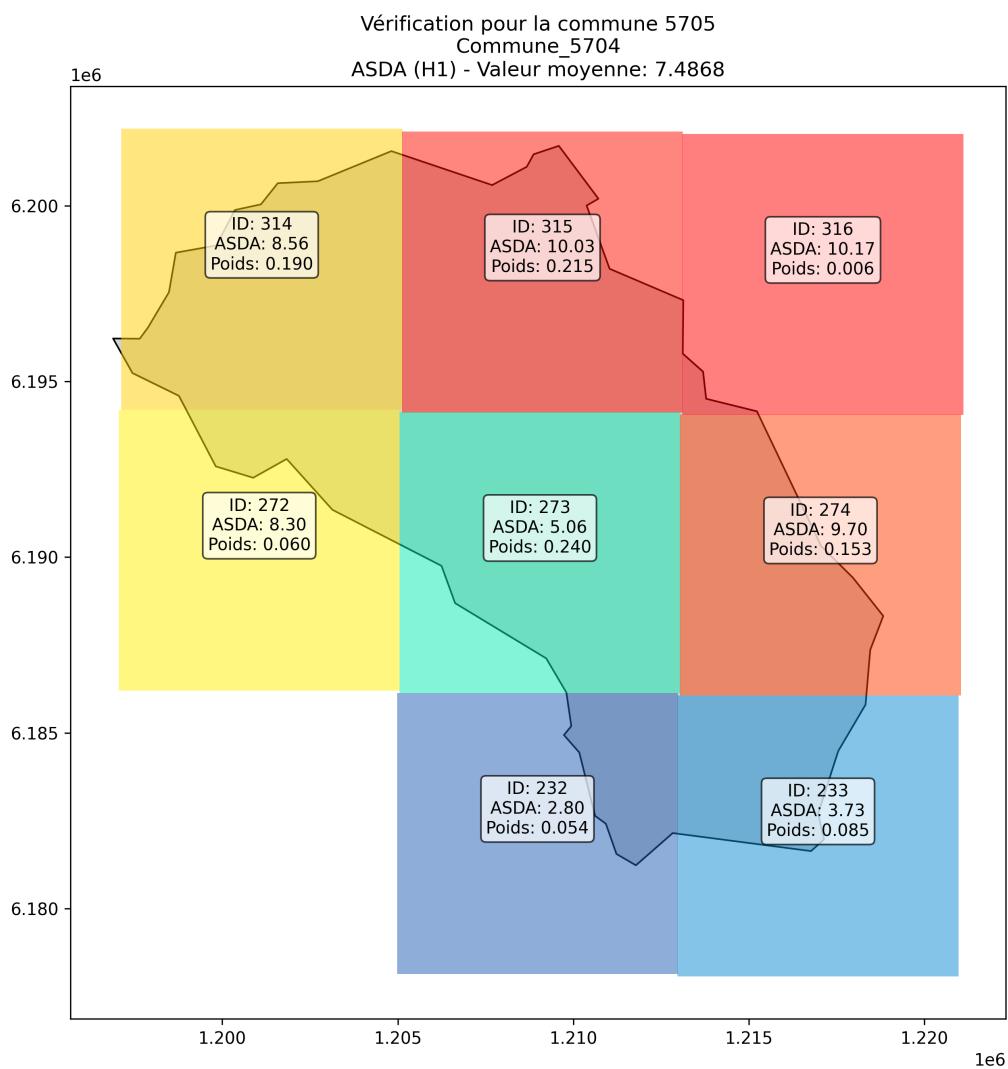


FIGURE 30 – Pondération au code postal d'un indicateur DRIAS

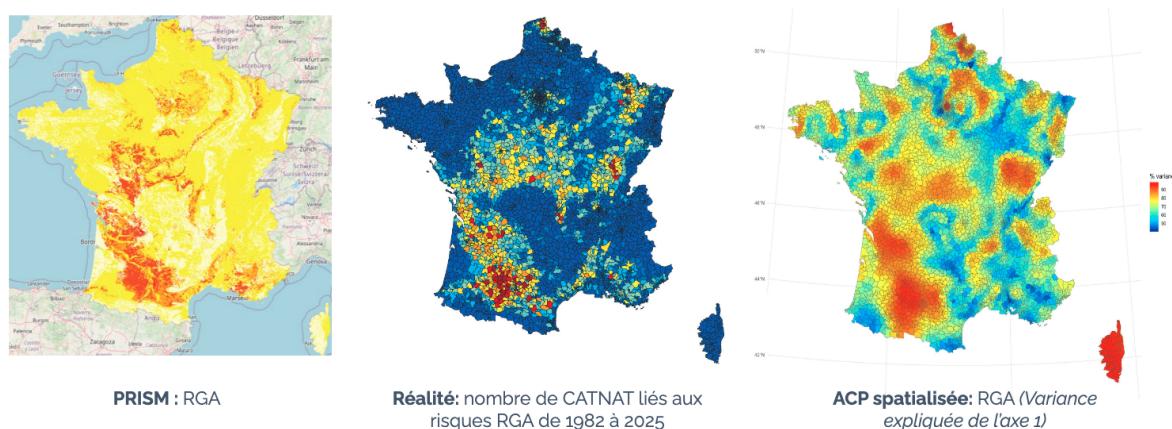
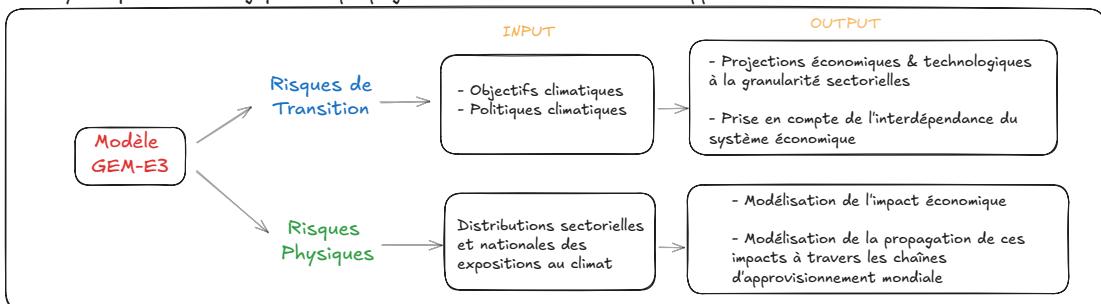
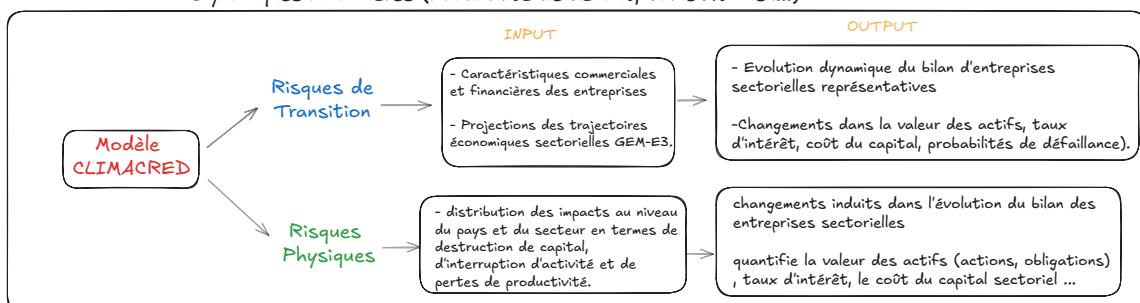


FIGURE 31 – Comparaison entre PRISM, les CATNAT et l'ACP spatialisée (GWPCA) sur l'aléa RGA

Dynamiques technologiques et propagation à travers les chaînes d'approvisionnement mondiale



Dynamiques Financières (Probabilité de défaut, Value At Risk...)



Dynamique Macro-Financière par continent

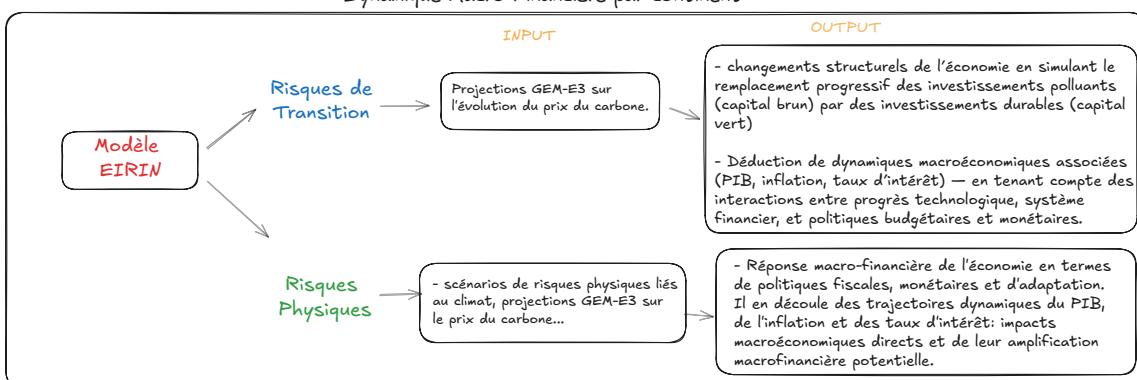


FIGURE 32 – Schéma explicatif des différentes dynamiques de la modélisation de la NGFS

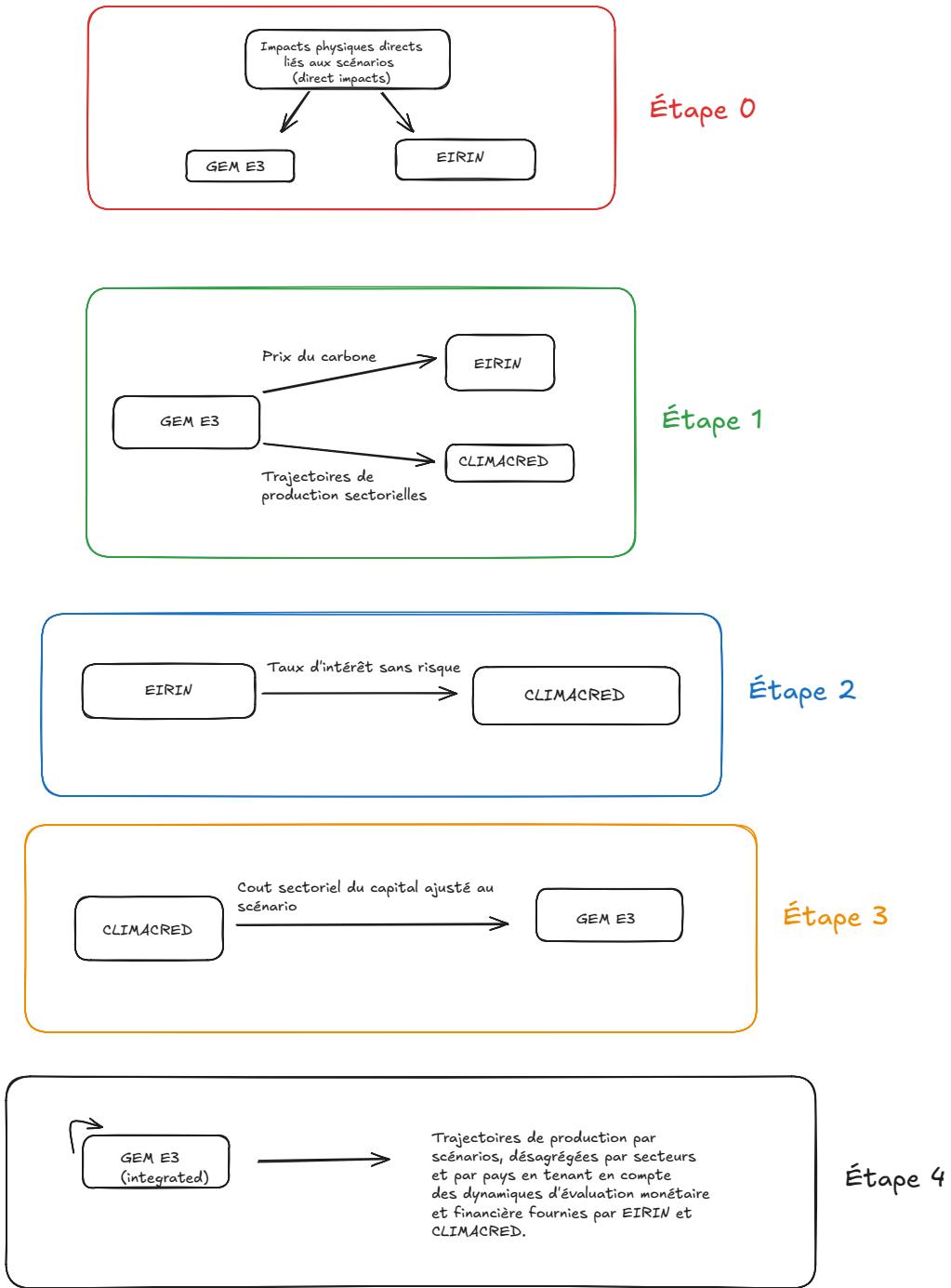


FIGURE 33 – Schéma explicatif des différentes étapes de modélisation de la NGFS

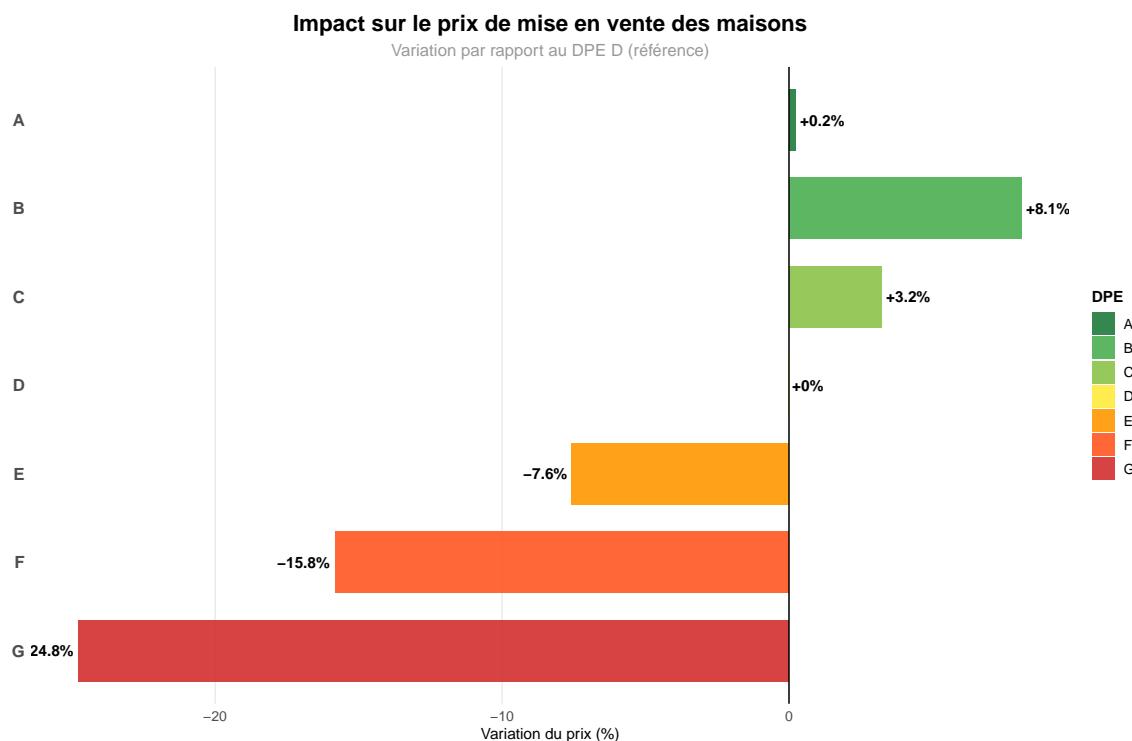


FIGURE 34 – Impact des étiquettes DPE sur le prix des maisons en Bretagne, de 2020 à 2024.
Note : le raisonnement se fait par rapport à l'étiquette D, utilisée comme référence

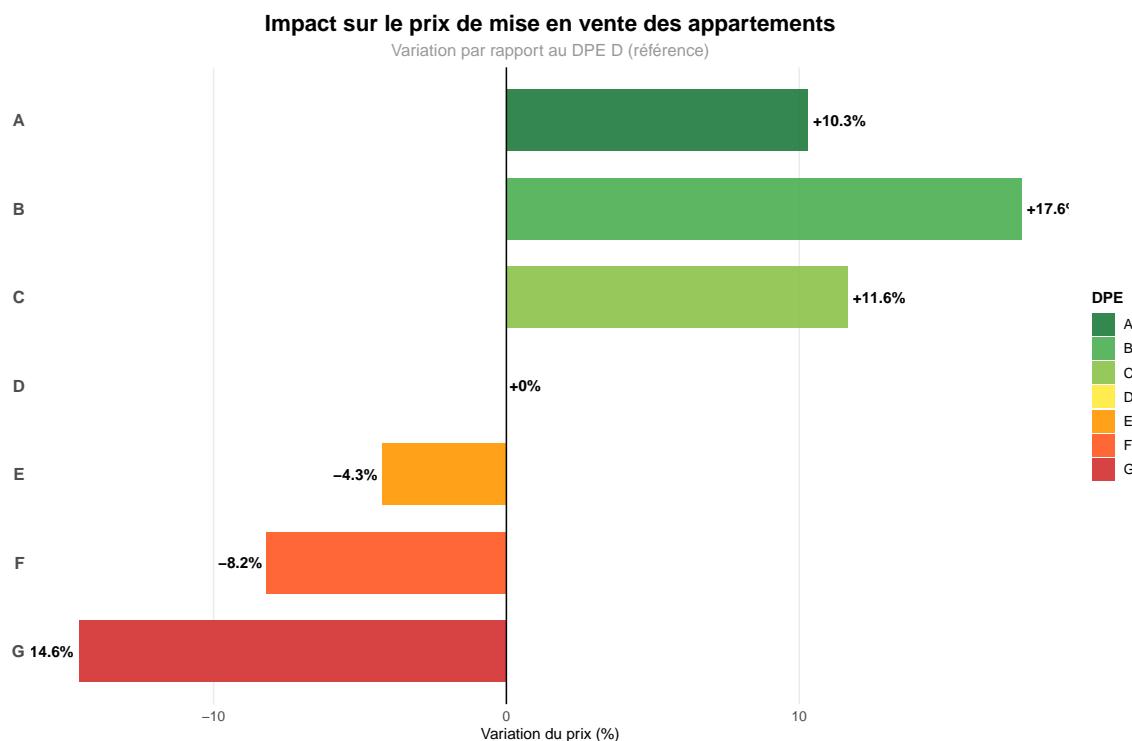


FIGURE 35 – Impact des étiquettes DPE sur le prix des appartements en Bretagne, de 2020 à 2024.
Note : le raisonnement se fait par rapport à l'étiquette D, utilisée comme référence

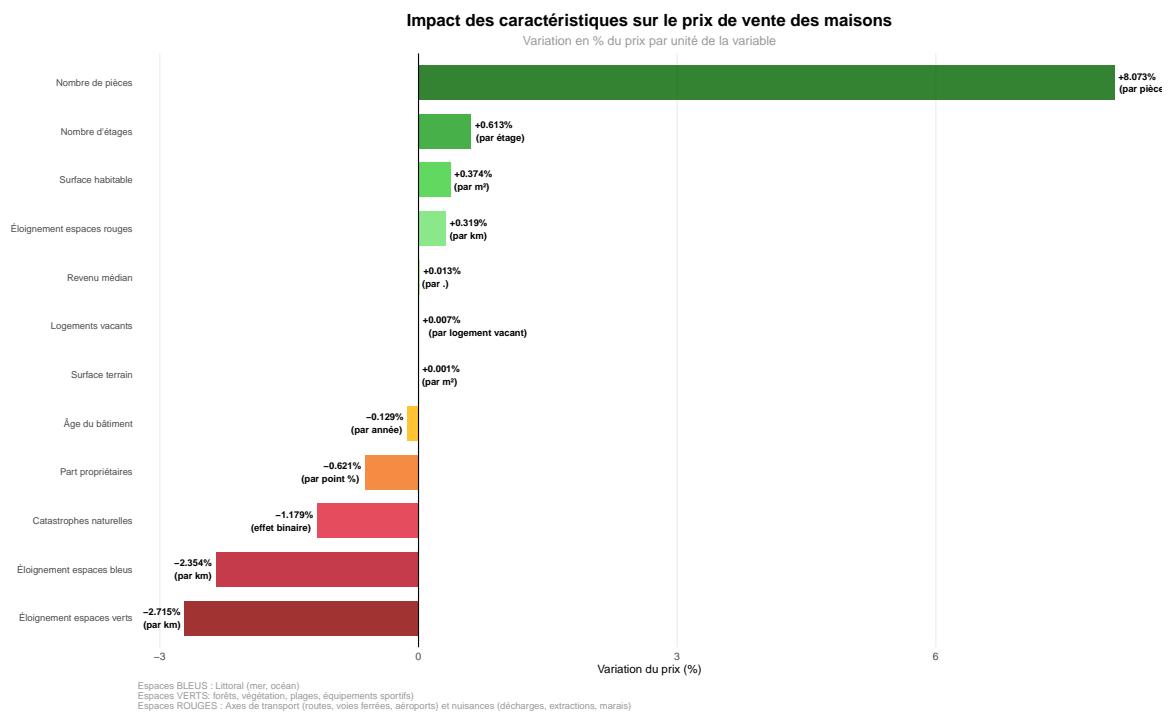


FIGURE 36 – Impact des différentes caractéristiques sur le prix de vente des maisons de 2020 à 2024

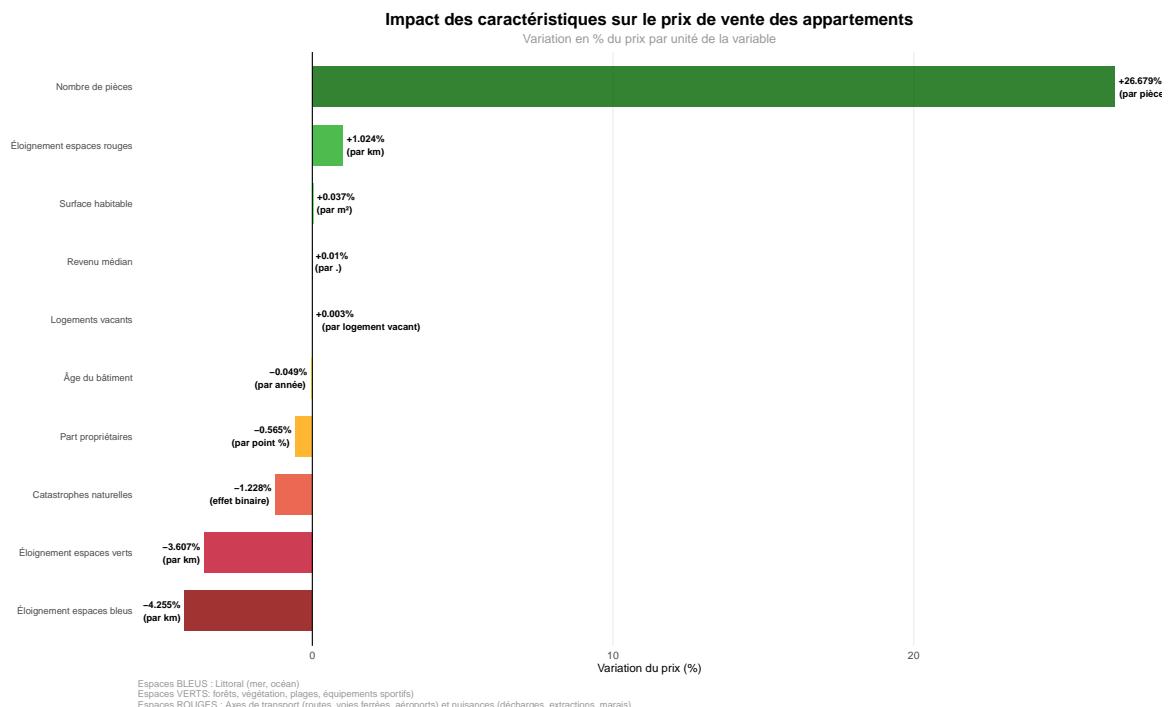


FIGURE 37 – Impact des différentes caractéristiques sur le prix de vente des appartements de 2020 à 2024

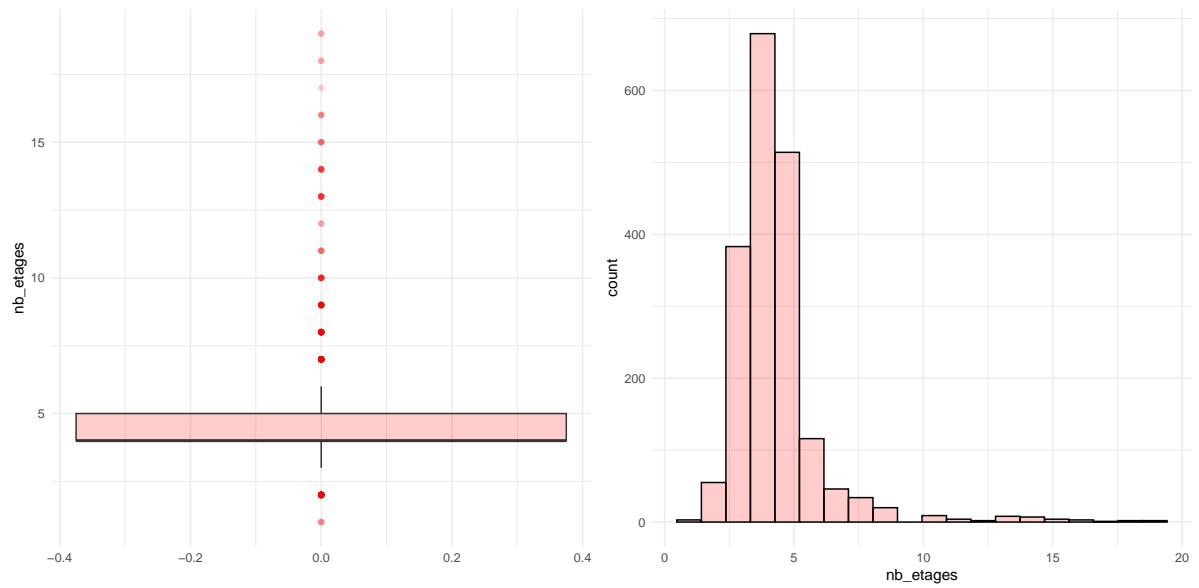


FIGURE 38 – Distribution du nombre d'étages des appartements à Brest

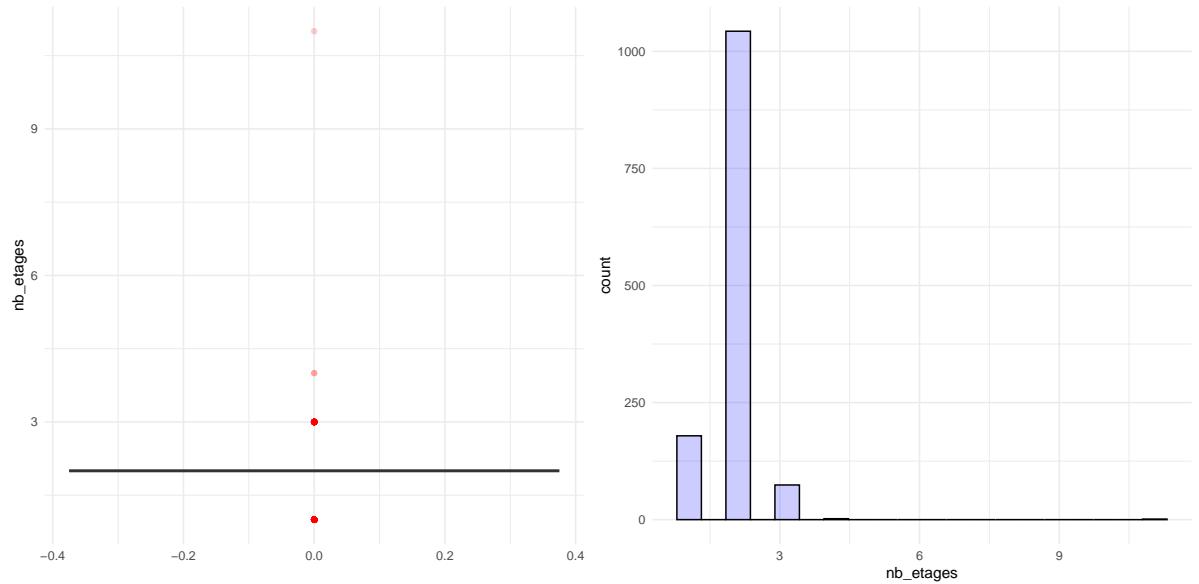


FIGURE 39 – Distribution du nombre d'étages des maisons à Brest

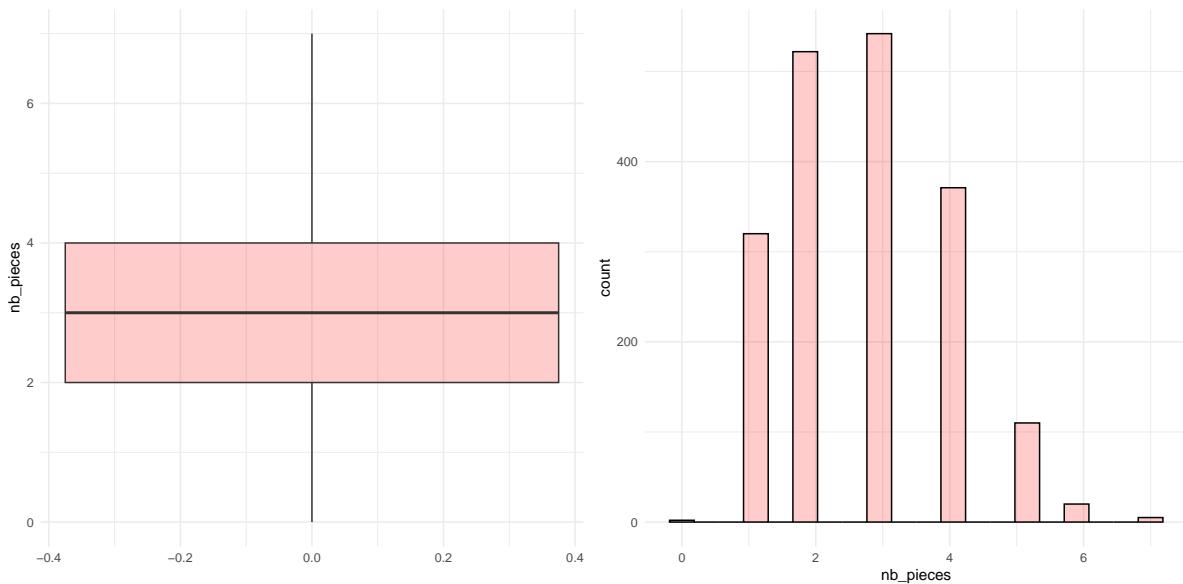


FIGURE 40 – Distribution du nombre de pièces des maisons à Brest

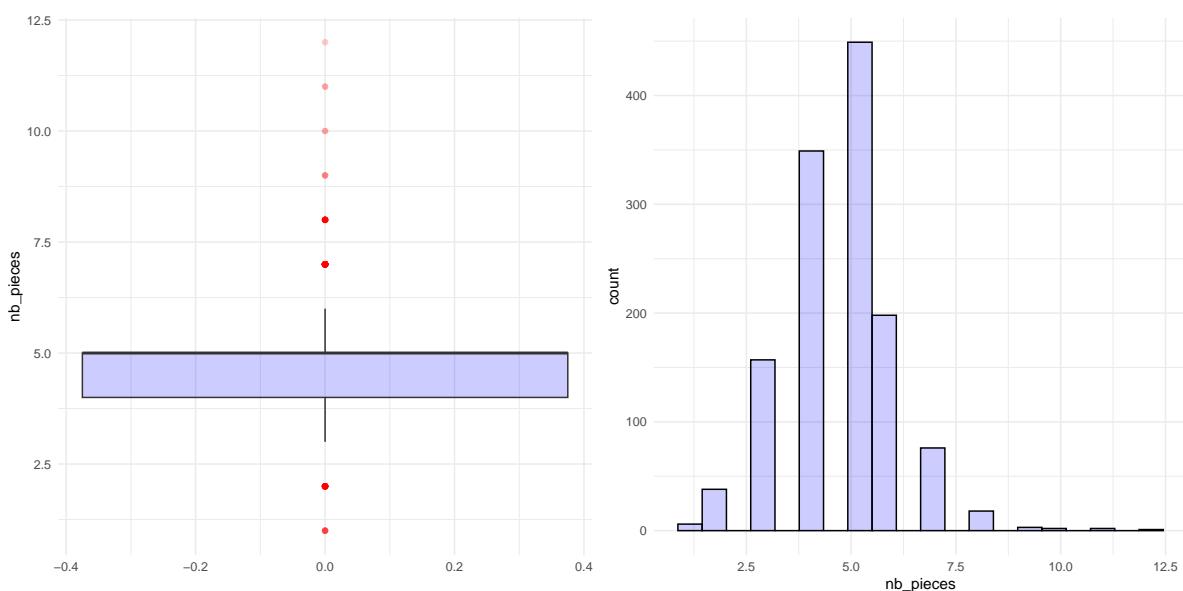


FIGURE 41 – Distribution du nombre de pièces des maisons à Brest

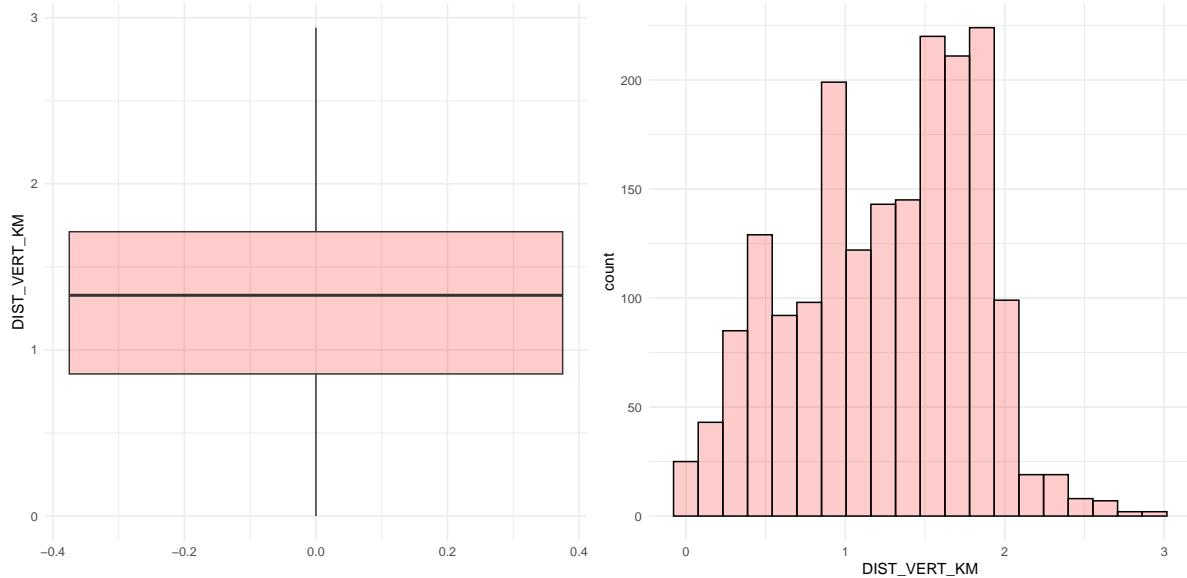


FIGURE 42 – Distance (km) aux entités vertes des appartements à Brest

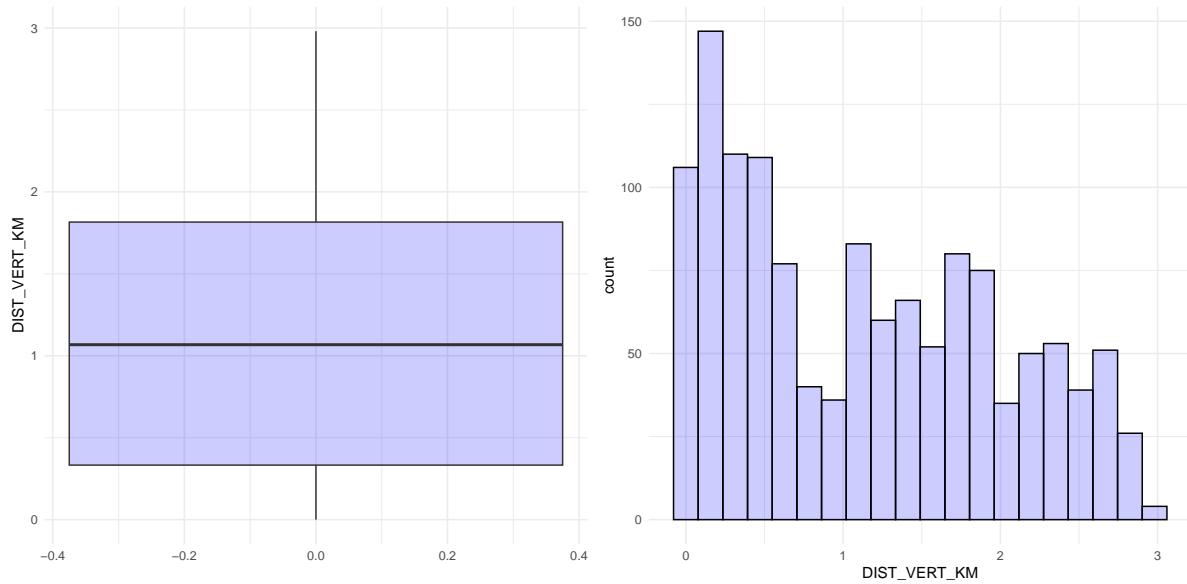


FIGURE 43 – Distance (km) aux entités vertes des maisons à Brest

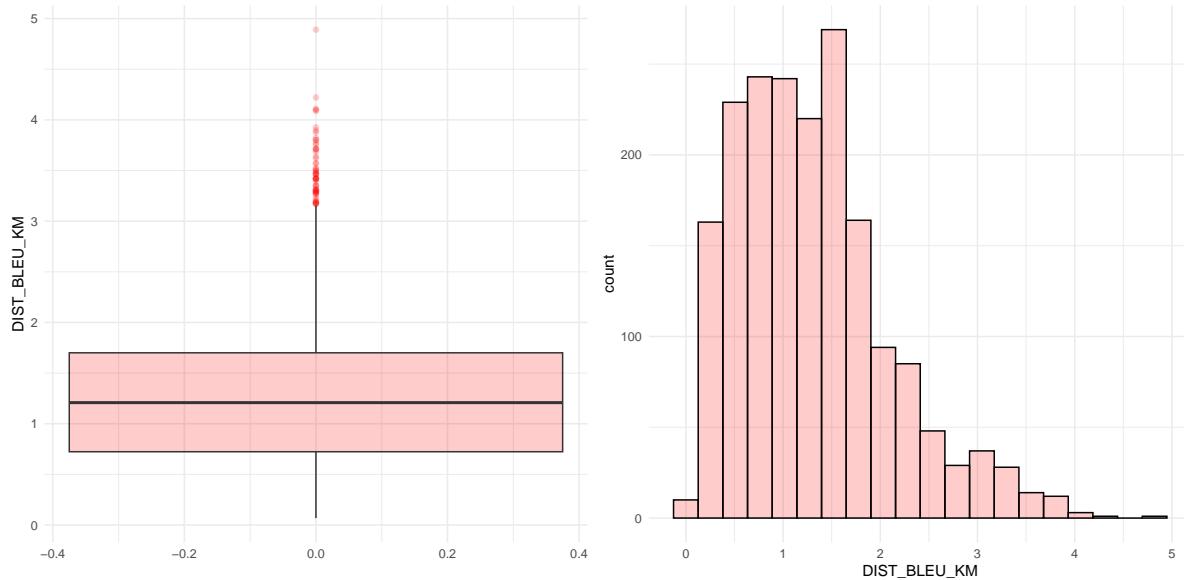


FIGURE 44 – Distance (km) aux entités bleu des maisons à Brest

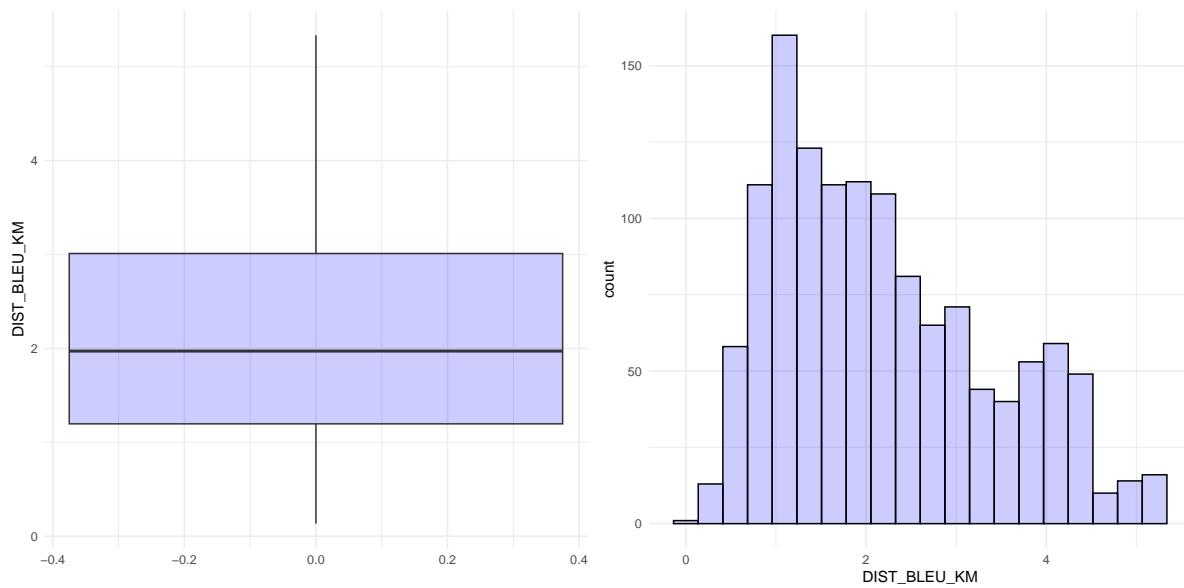


FIGURE 45 – Distance (km) aux entités bleus des maisons à Brest

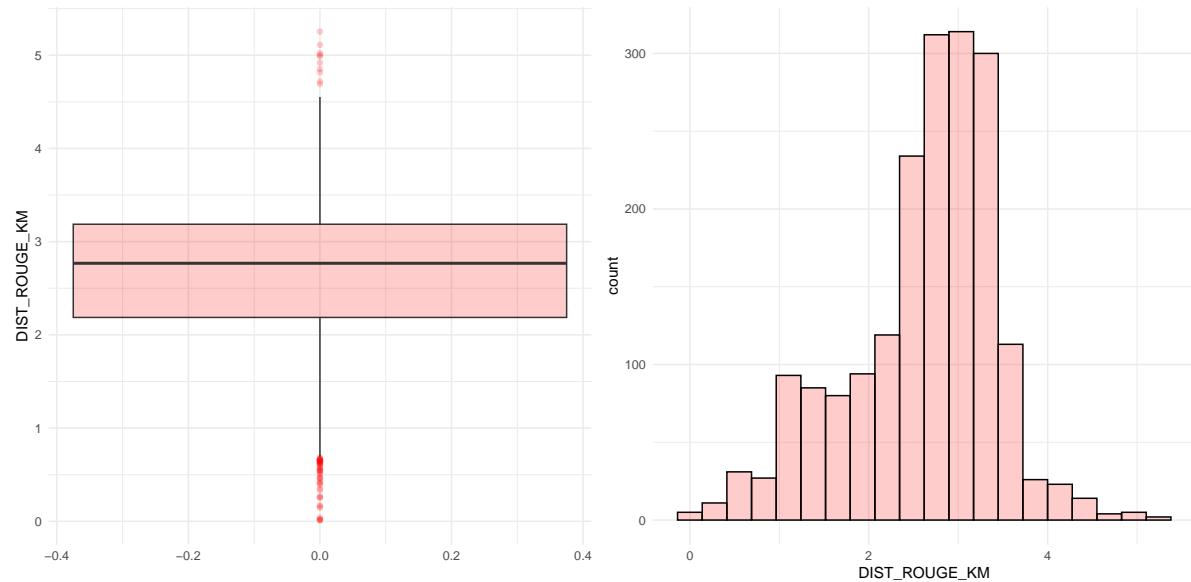


FIGURE 46 – Distance (km) aux entités rouges des appartements à Brest

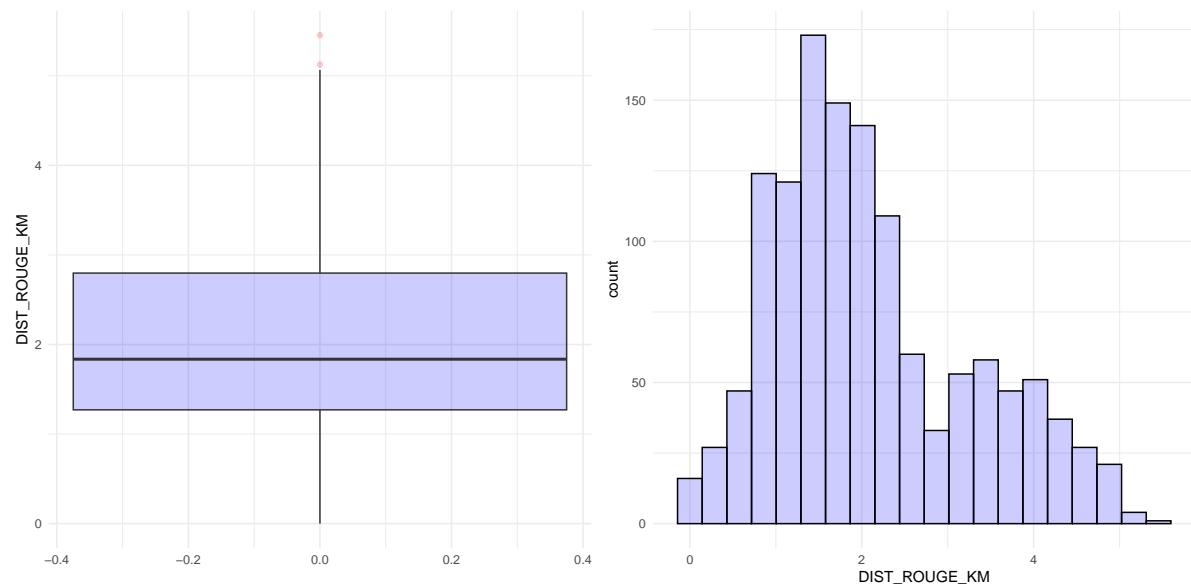
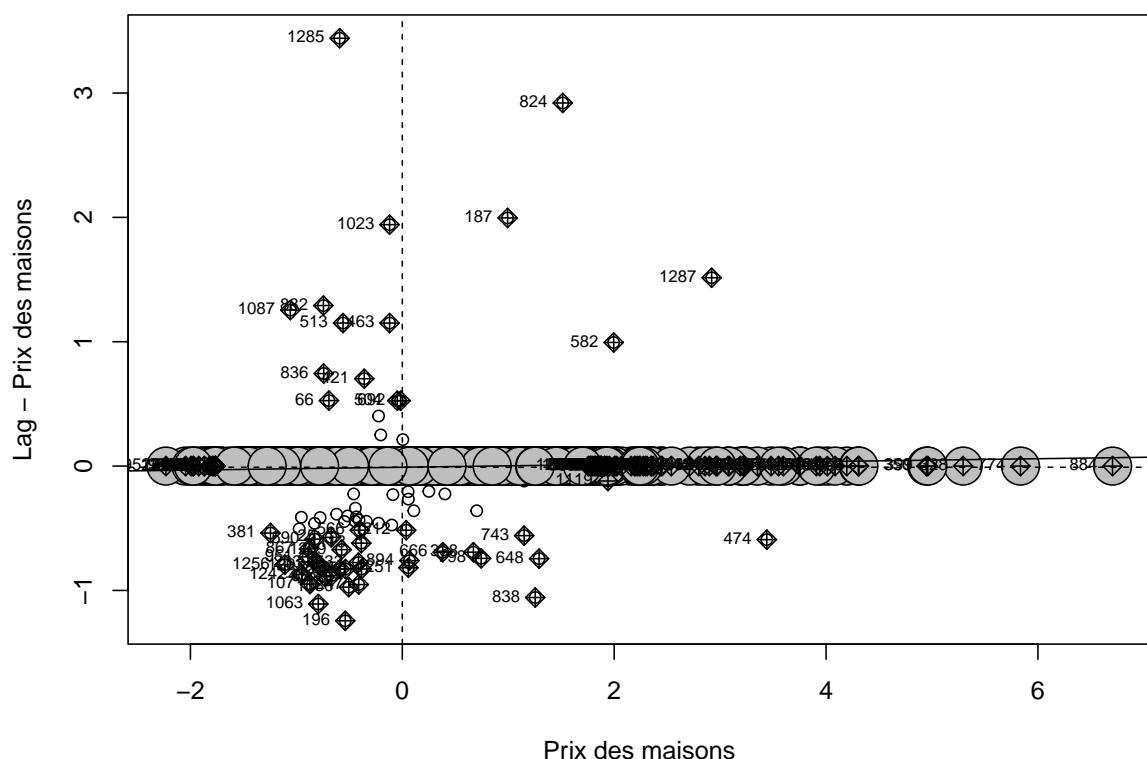


FIGURE 47 – Distance (km) aux entités rouges des maisons à Brest

Matrice type Reine



Matrice type PPV1

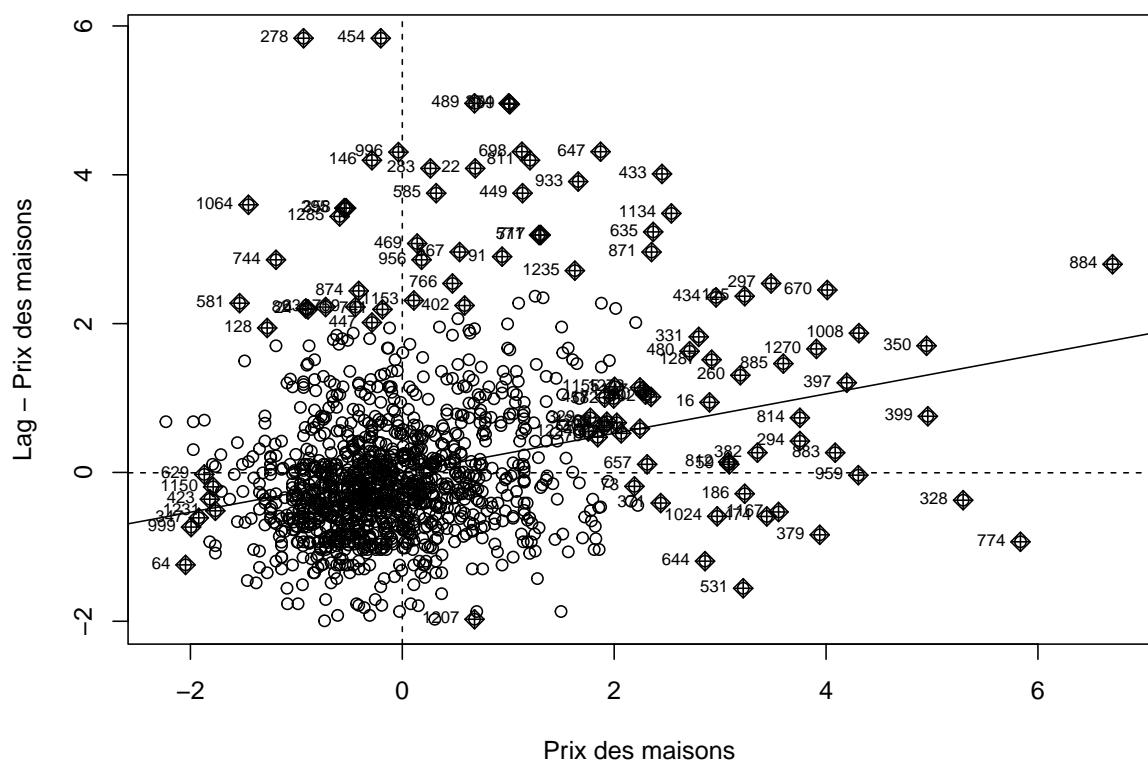


FIGURE 49 – Indice de Moran pour la matrice PPV1 sur les maisons

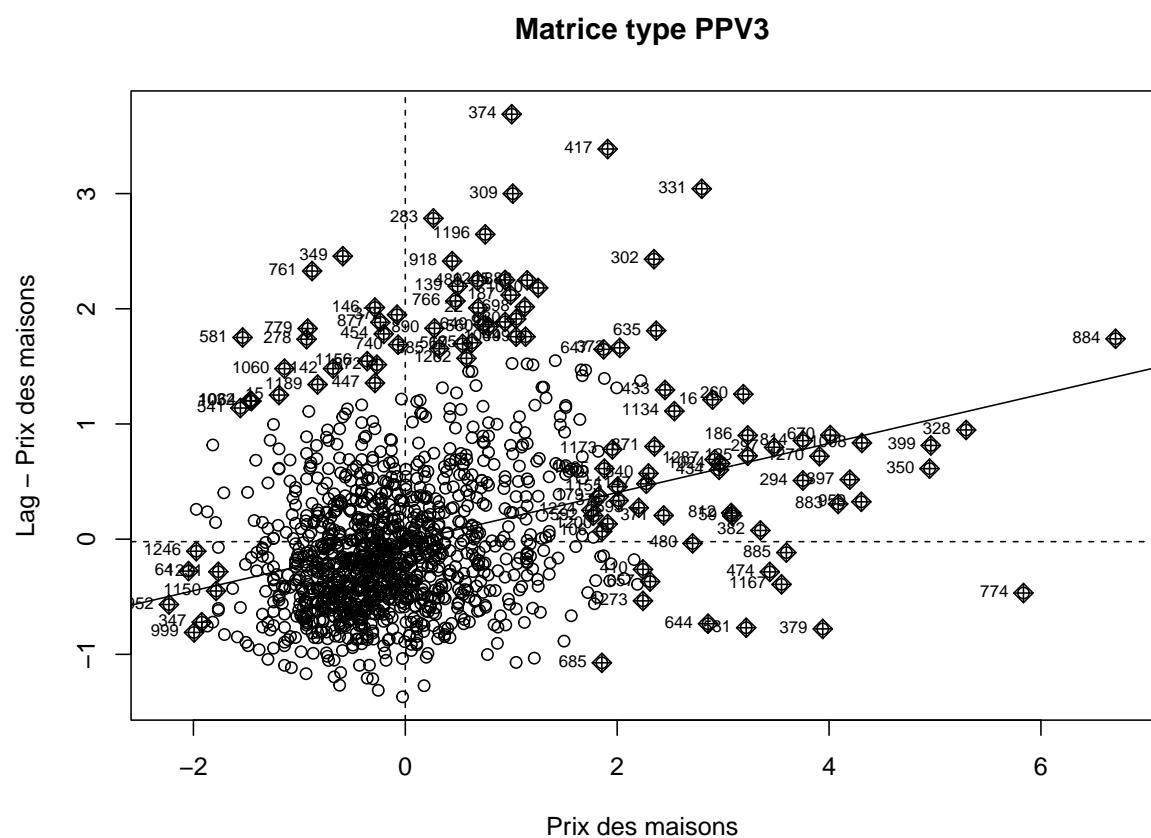


FIGURE 50 – Indice de Moran pour la matrice PPV3 sur les maisons

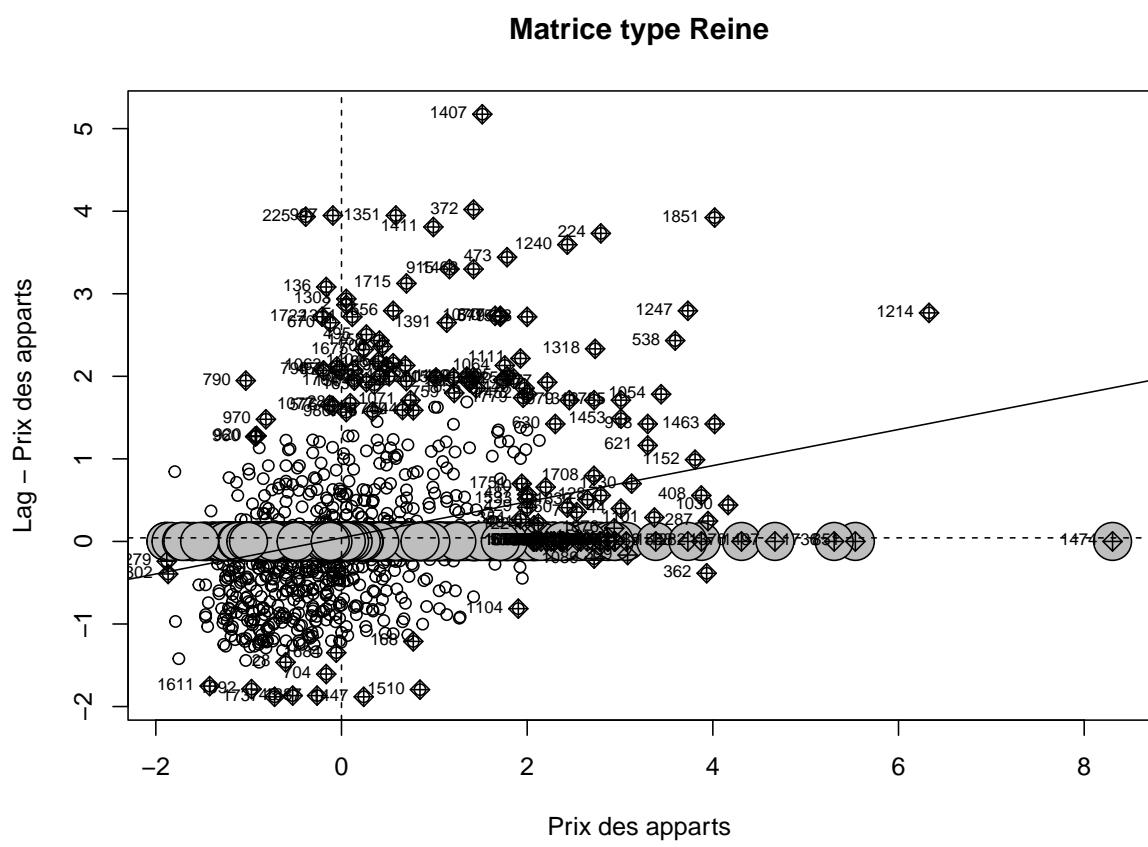
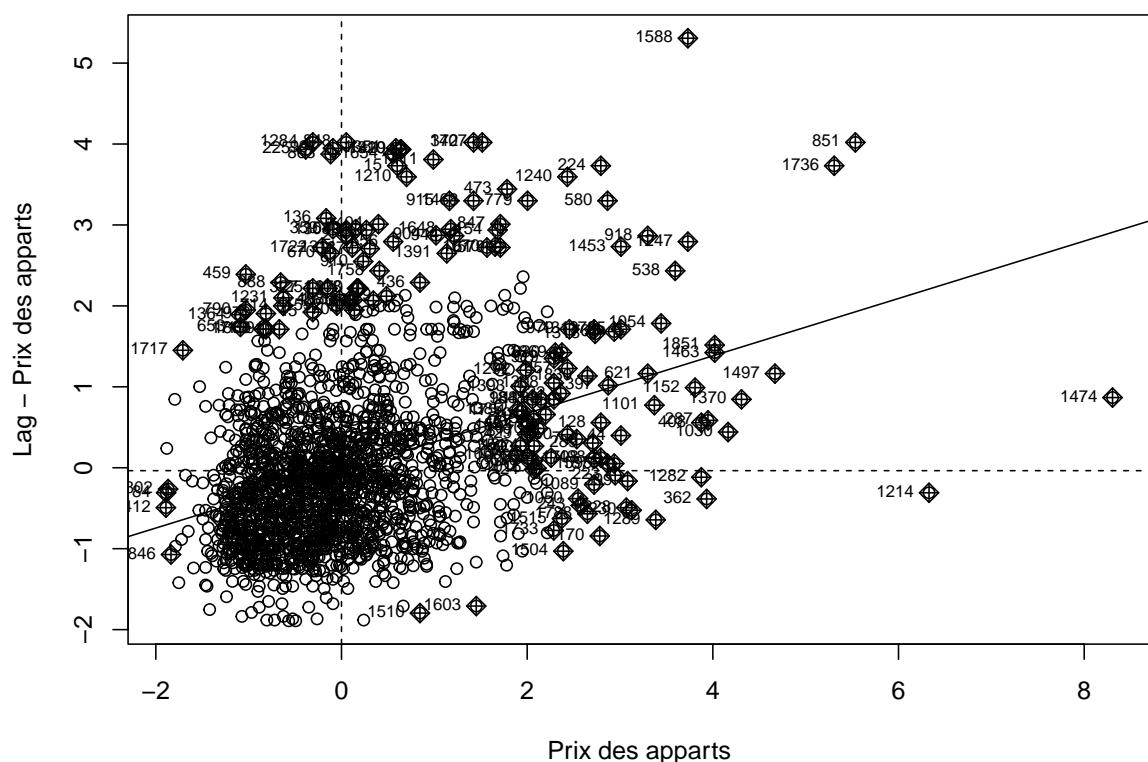


FIGURE 51 – Indice de Moran pour la matrice Reine sur les appartements

Matrice type PPV1



Matrice type PPV3

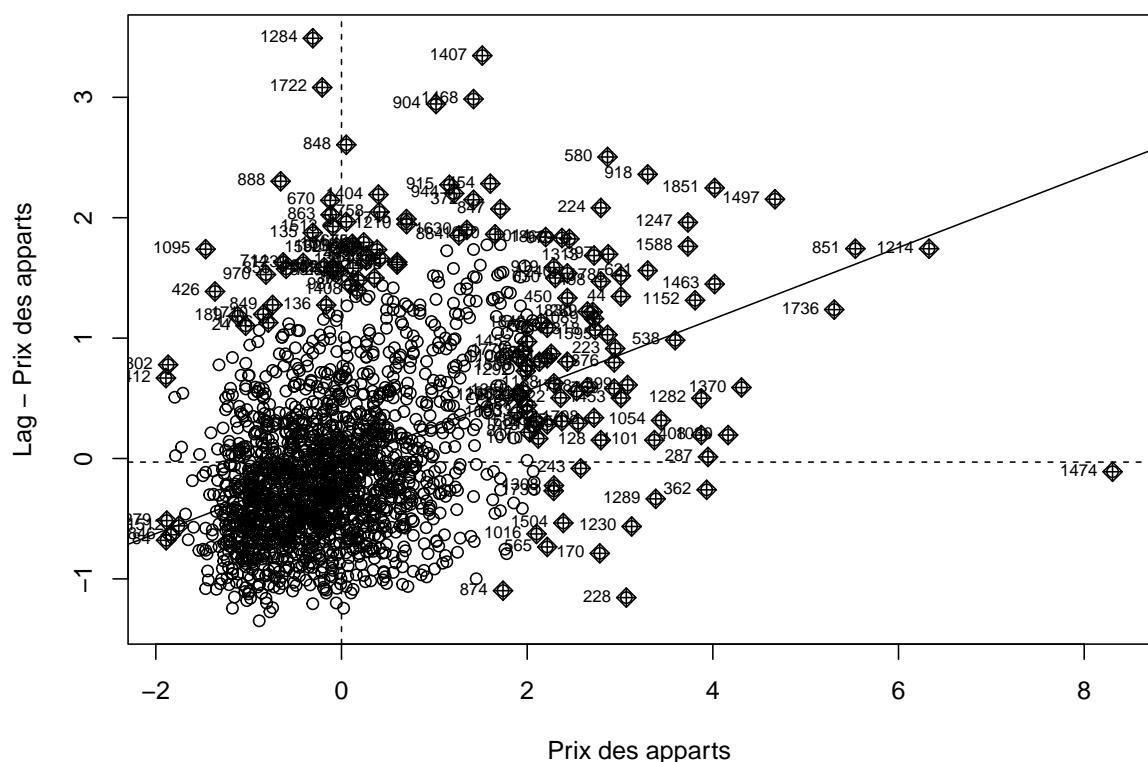


FIGURE 53 – Indice de Moran pour la matrice PPV3 sur les appartements