



# APPORT DE DONNÉES TÉLÉMATIQUES DANS LA MODÉLISATION DU RISQUE GÉOGRAPHIQUE EN ASSURANCE AUTOMOBILE

Junior ASSI

Mardi 17 octobre 2023

## SOMMAIRE

- 📌 **1 - Contexte et enjeux du mémoire**
- 2 - Présentation des données externes : *Smart Road Data*
- 3 - Evaluations techniques de l'apport de la télématique
- 4 - Evaluations opérationnelles de l'apport de la télématique
- 5 - Conclusion et ouverture

- 1 • Contexte et enjeux du mémoire

**addactis**  
THE RISKTECH FOR INSURANCE

En partenariat avec



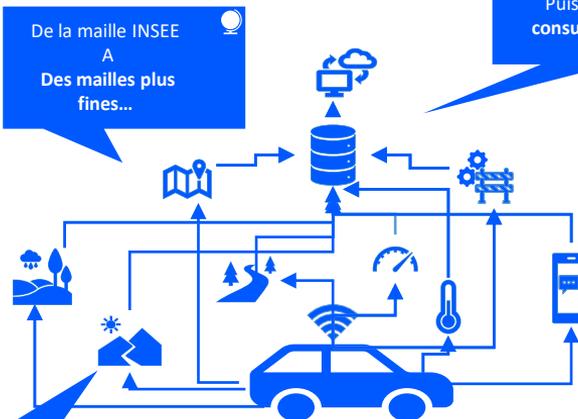
Différents contextes:

- ⇒ Météo
- ⇒ Urbanité
- ⇒ Relief
- ⇒ Courbure de route
- ⇒ Saisonnalité

Impactant différemment la sinistralité...

De la maille INSEE  
A  
Des mailles plus fines...

Des données télématiques  
recueillies par les experts  
de MICHELIN  
Puis analysées par les  
consultants D'ADDACTIS



- ⇒ Nouvelles variables tarifaires
- ⇒ Zoniers innovants
- ⇒ Modèles de scoring
- ⇒ Campagnes de prévention

Mieux Comprendre le  
comportement des  
conducteurs au volant  
Grâce à  
Des scores de conduite  
finement calculés

## SOMMAIRE

1 - Contexte et enjeux du mémoire

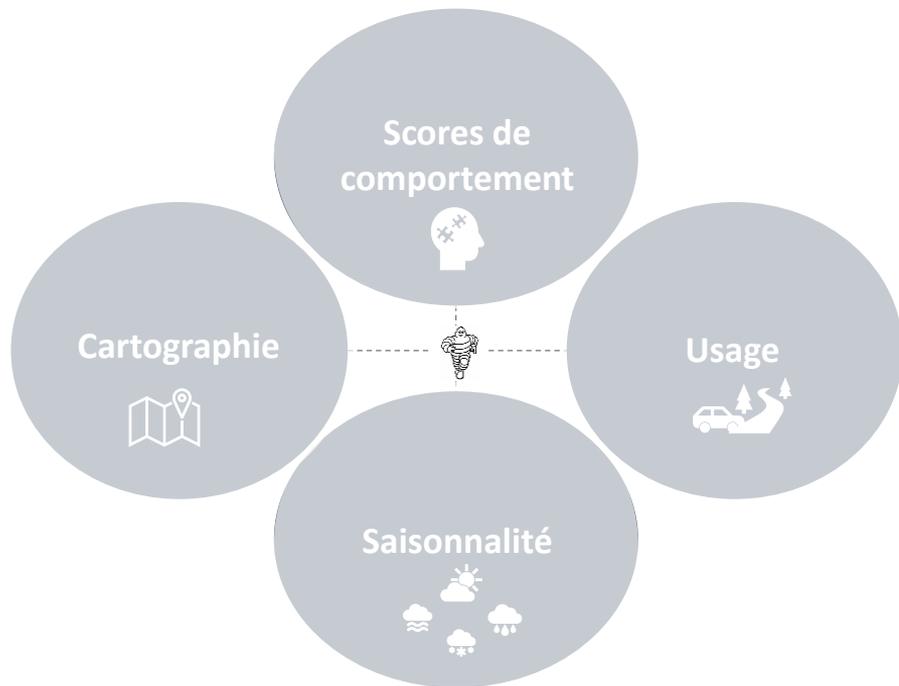
 **2 - Présentation des données externes : Smart Road Data**

3 - Evaluations techniques de l'apport de la télématique

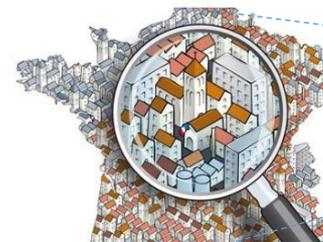
4 - Evaluations opérationnelles de l'apport de la télématique

5 - Conclusion et ouverture

- 2 • Présentation des données externes: Smart Road Data



Les différentes catégories de variables



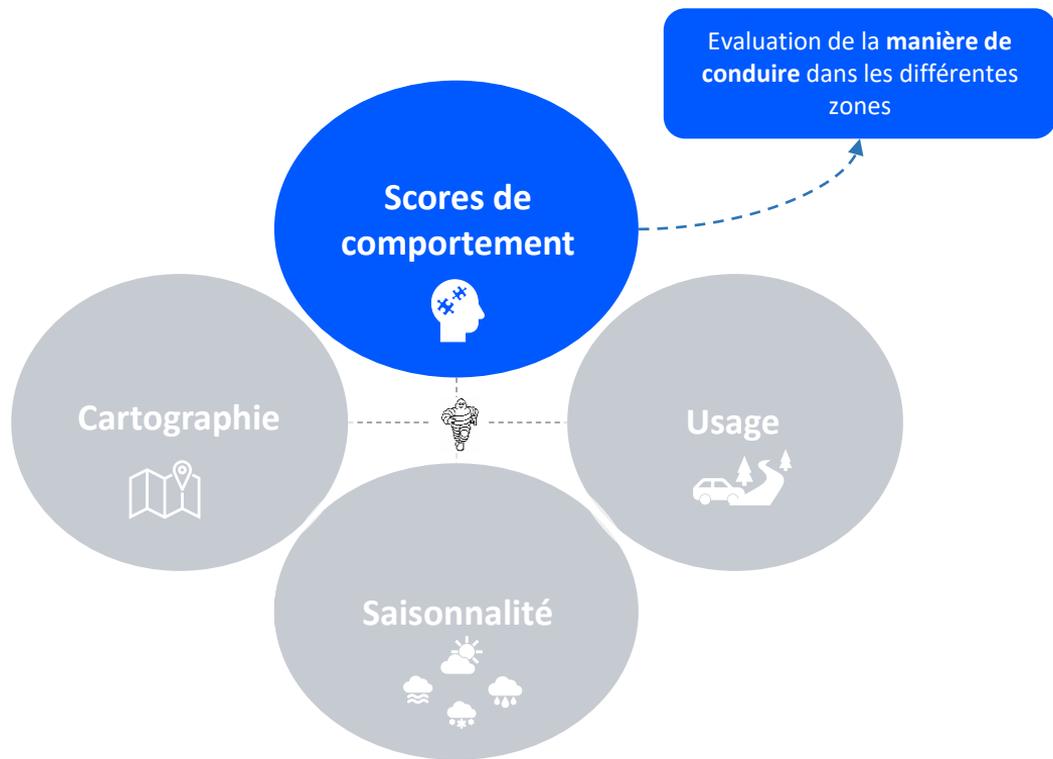
Maille de l'étude:  
**INSEE**



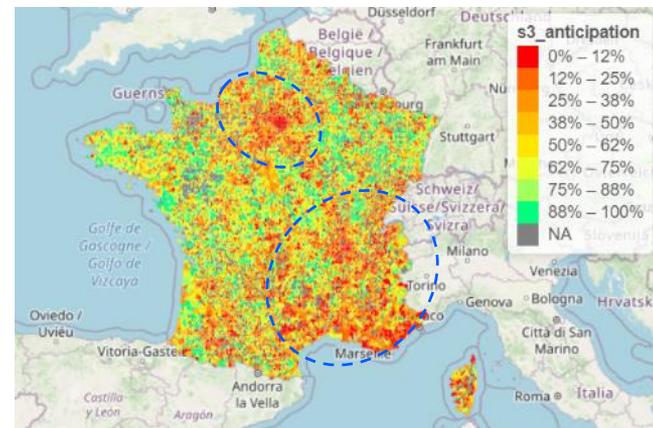
**Massification de la donnée** sur tout le territoire:

- par des méthodes statistiques
- via données brutes (mis à jour Michelin)

- 2 • Présentation des données externes: Smart Road Data



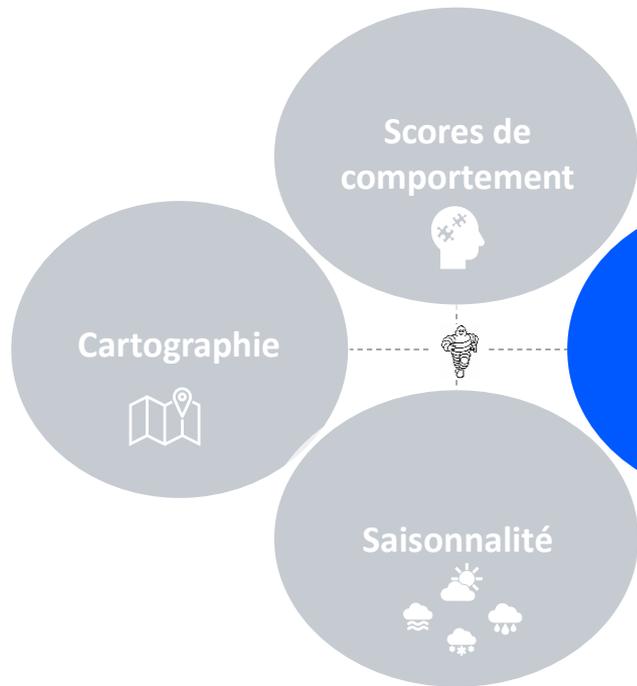
Les différentes catégories de variables



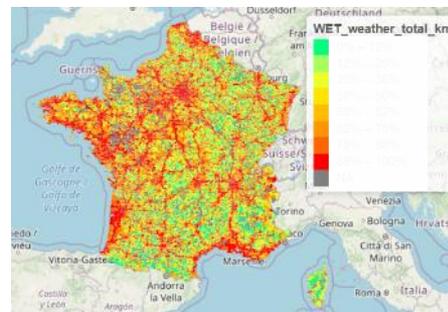
Score général de freinage

Il se dégage quelques **zones où le comportement de freinage** (distance, intensité, fréquence,...) **est assez prononcé** comme l'Île-de-France, le Massif central et Marseille

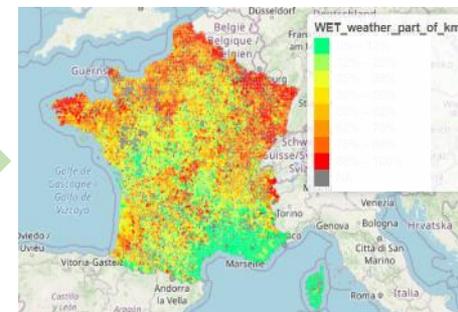
- 2 • Présentation des données externes: Smart Road Data



Les différentes catégories de variables



Total de kilomètres parcourus en temps humide



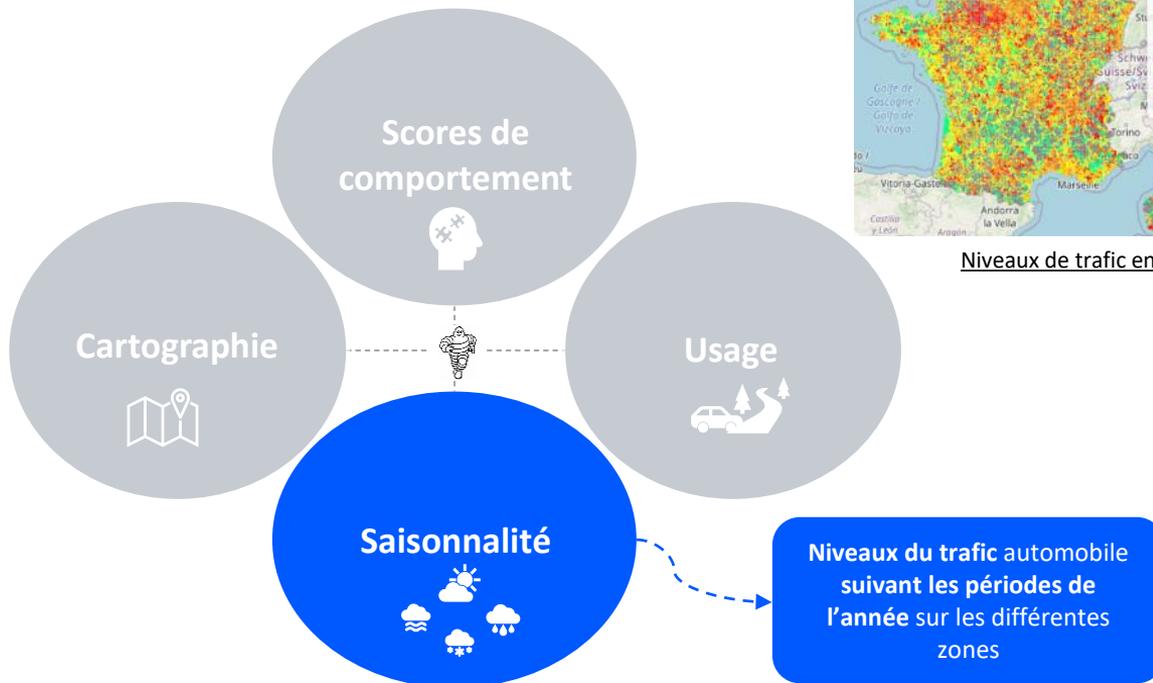
Taux de kilomètres parcourus en temps humide

**Kilomètres parcourus et durées de trajets selon différents contextes : météo, relief, urbanité, type de route...**

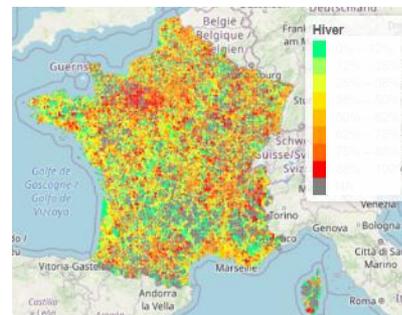
**Sur la carte de gauche** (donnée initiale) l'aspect météorologique de la donnée n'est pas visible

**Sur la carte de droite** (variable transformée) on discerne mieux les zones où il est plus (resp moins) probable de conduire en **temps humide**

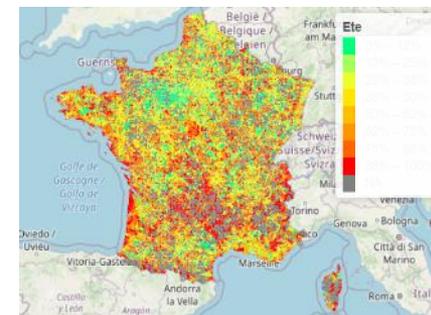
- 2 • Présentation des données externes: Smart Road Data



Les différentes catégories de variables



Niveaux de trafic en hiver

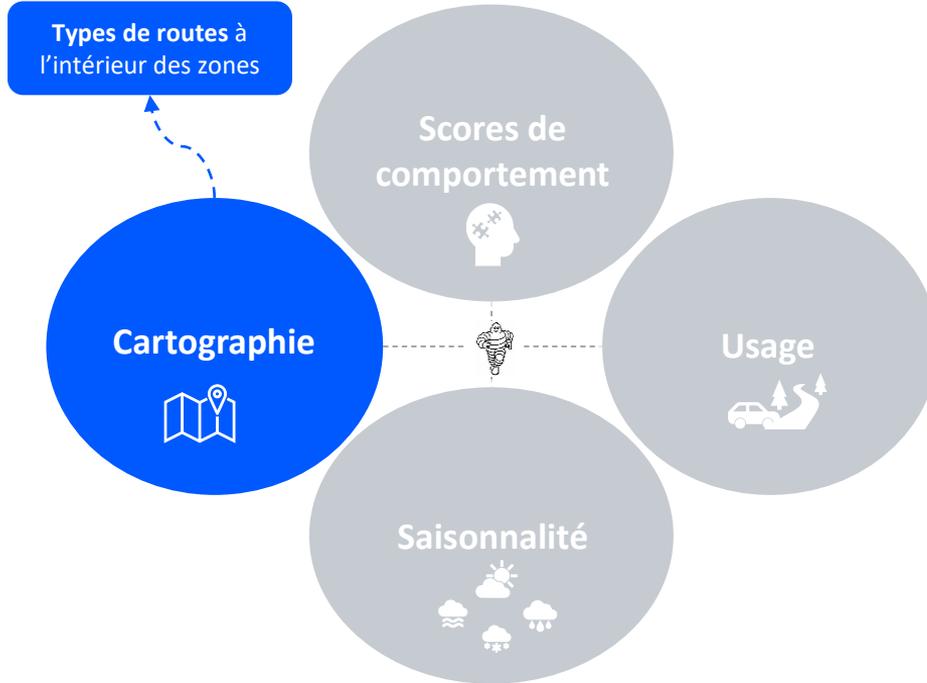


Niveaux de trafic en été

En été  
le trafic diminue sur certaines zones telles que l'Île-de-France et s'accroît sur d'autres zones comme les villes côtières de l'ouest et dans le sud (effet des vacances)

Présentation de mémoire | IA

- 2 • Présentation des données externes: Smart Road Data



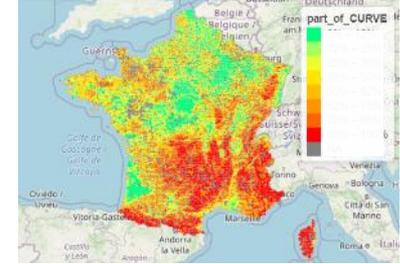
Les différentes catégories de variables



Distance des virages



Distances des lignes droites

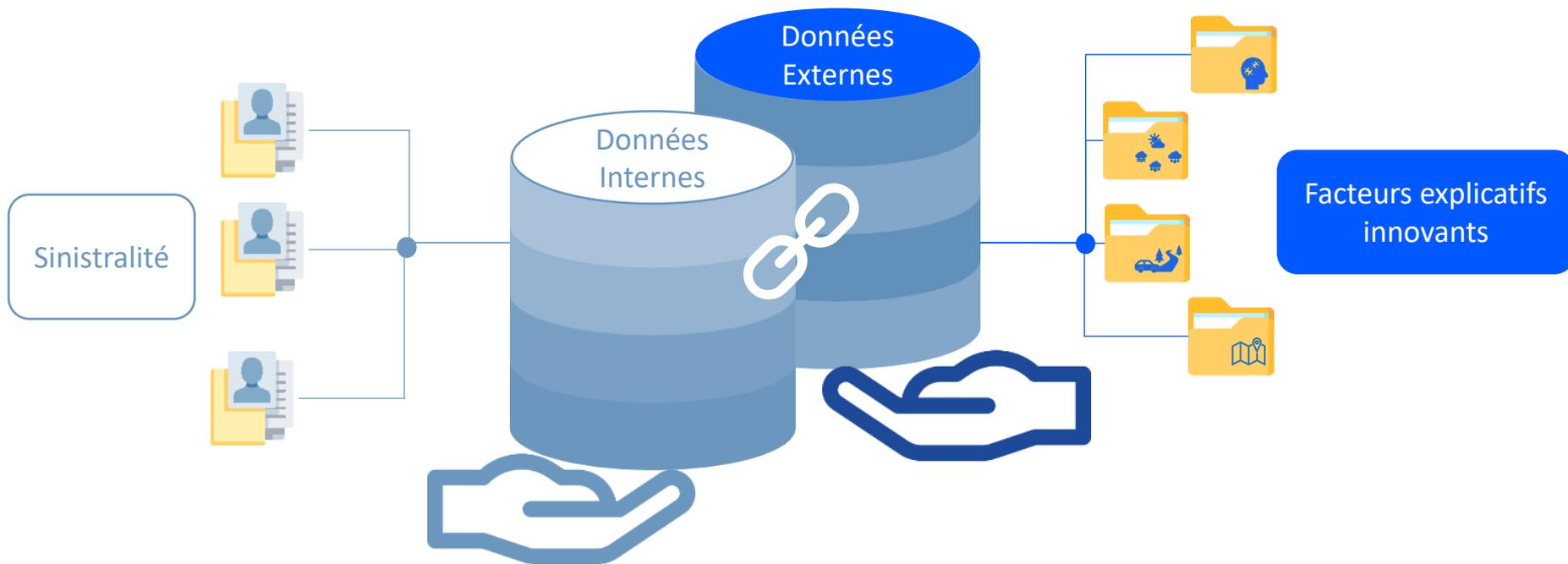


Taux de Virages

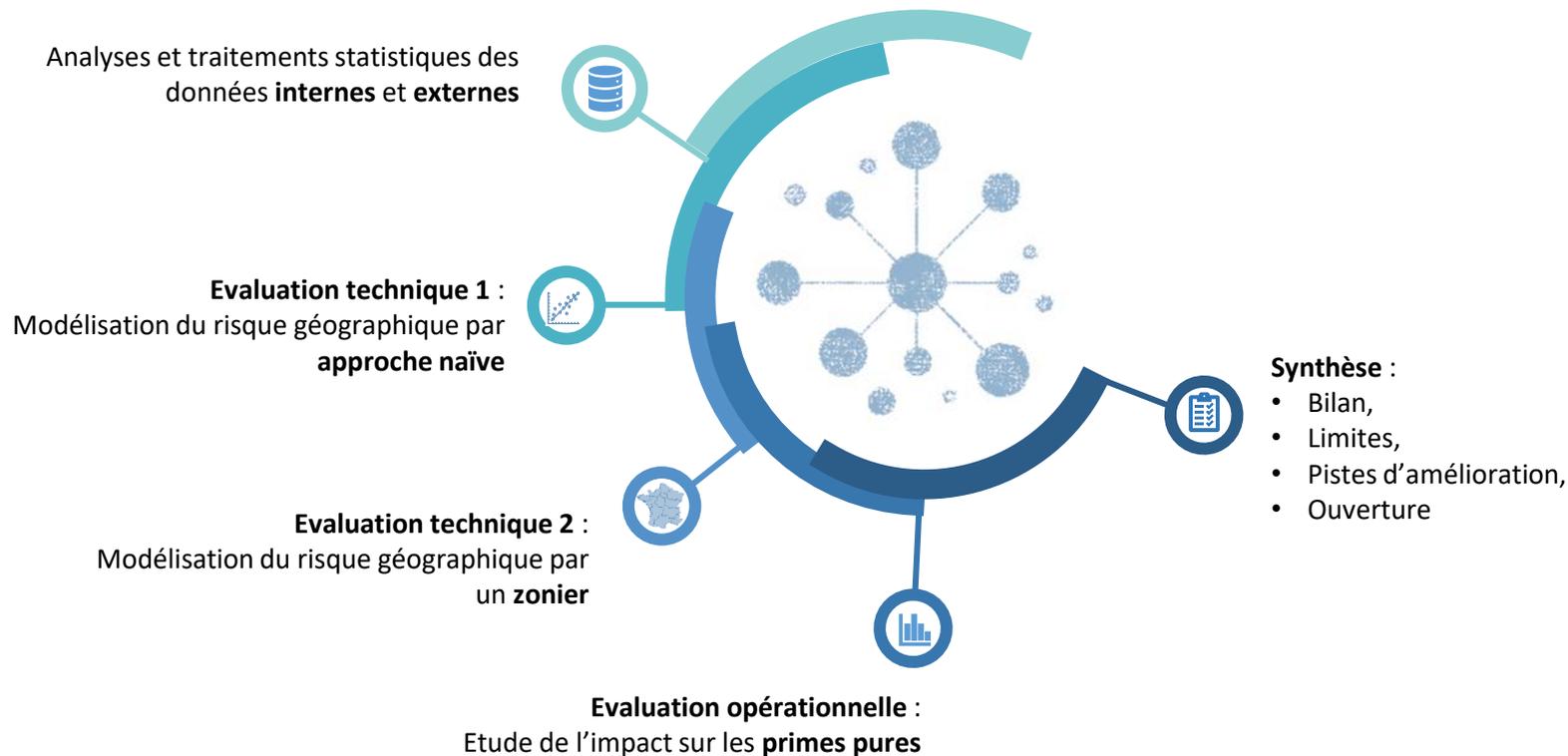
Les cartes de gauche (données initiales) masquent l'information sur les courbures des routes

Sur la carte de droite (variable transformée) on discerne mieux les zones où la probabilité d'emprunter un virage est élevée (resp faible)

- 2 • Présentation des données externes: Smart Road Data



- 2 • Présentation des données externes: Smart Road Data



## SOMMAIRE

1 - Contexte et enjeux du mémoire

2 - Présentation des données externes : Smart Road Data

🔊 **3 - Evaluations techniques de l'apport de la télématique**

4 - Evaluations opérationnelles de l'apport de la télématique

5 - Conclusion et ouverture

- 3 • Evaluations techniques de l'apport de la télématique

## MÉTHODE 1: APPROCHE NAÏVE

### Modèle initial de fréquence de sinistres

$$g(\mathbb{E}(Y|X_1, \dots, X_p)) = \beta_0 + X_1\beta_1 + \dots + X_p\beta_p$$

#### GLM Log Poisson:

Données recueillies à la souscription  
 sans les **variables géographiques**

#### Variables explicatives :

- Age du conducteur
- Age du véhicule
- Classe SRA
- Coefficient novice
- Catégorie socio-professionnelle
- Puissance du véhicule

Intégration d'informations télématiques  
 géo-spatialisées dans le modèle initial de  
 fréquence de sinistres RCM

### Approche naïve

$$g(\mathbb{E}(Y|X_1, \dots, X_p, X_{p+1}, \dots, X_{p+q})) = \beta_0 + X_1\beta_1 + \dots + X_p\beta_p + X_{p+1}\beta_{p+1} + \dots + X_{p+q}\beta_{p+q}$$

Modèle initial

Données télématiques

#### Ajout de 7 variables télématiques :

- Score global de freinage
- Score global d'allure de conduite
- Total de kilomètres parcourus
- Fréquence de conduite
- Taux de kilomètres parcourus en temps humide
- Taux de routes urbaines
- Taux de virages

- 3 • Evaluations techniques de l'apport de la télématique

## MÉTHODE 1: APPROCHE NAÏVE



Analyse de la fréquence de sinistres en fonction du score général de freinage

Plus le score de freinage s'améliore, plus la fréquence de sinistres diminue  
(courbe rouge)

Les prédictions du modèle initial ne suivent pas les tendances des fréquences observées (courbe bleue)



Analyse de la fréquence de sinistres en fonction du total de km parcourus

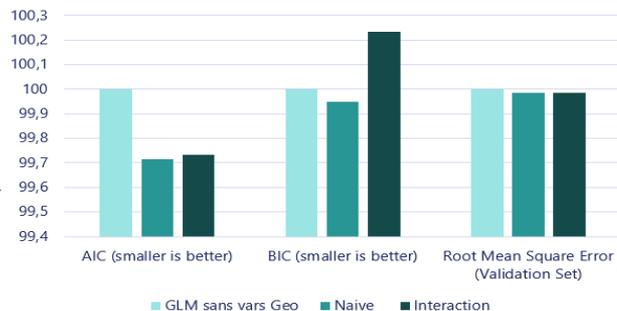
Plus le total de kilomètres parcourus augmente, plus grande est la probabilité d'avoir un sinistre  
(courbe rouge)



Cette approche est complétée par une étude des **interactions entre les variables** internes et externes

- 3 • Evaluations techniques de l'apport de la télématique

## MÉTHODE 1: APPROCHE NAÏVE



Comparaison des indicateurs de qualité (base 100)

Plus ces indicateurs augmentent, meilleurs ils sont

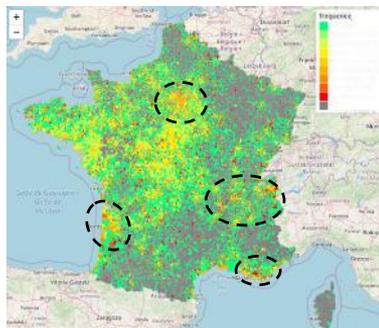


Comparaison des indices de Gini normalisés (base 100)

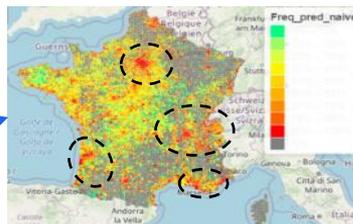
De manière globale,  
**l'intégration des données télématiques dans le modèle initial améliore ses performances statistiques**

- 3 • Evaluations techniques de l'apport de la télématrique

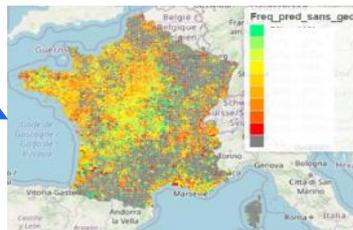
## MÉTHODE 1: APPROCHE NAÏVE



Fréquences de sinistres observées



Fréquences de sinistres estimées par le modèle naïf



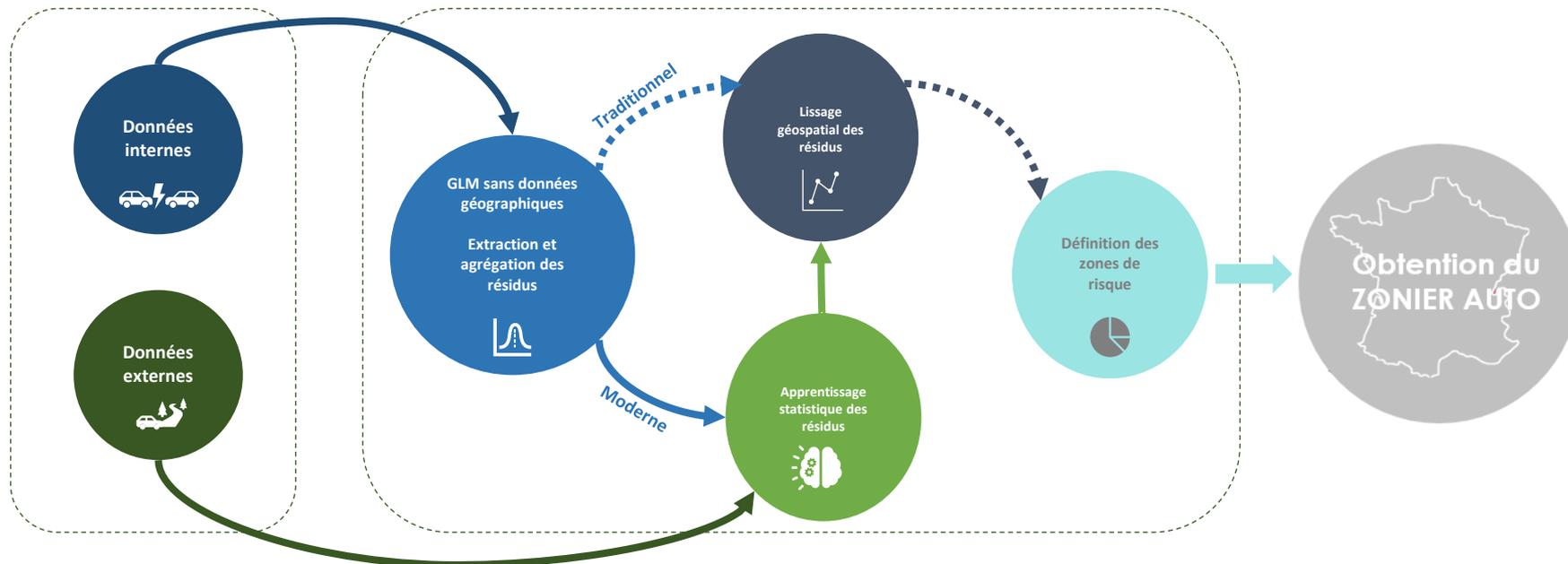
Fréquences de sinistres estimées par le modèle initial

le **modèle naïf** retrouve bien les **points d'impact importants** pour la fréquence de sinistres

Toutefois, Cette approche n'est pas la plus adaptée pour une mise en production...

- 3 • Evaluations techniques de l'apport de la télématique

## MÉTHODE 2: ZONIER



- 3 • Evaluations techniques de l'apport de la télématrique

## MÉTHODE 2: ZONIER

Zonier traditionnel



Lissage par crédibilité

$$\hat{R}_i = c(E_i)R_i + (1 - c(E_i)) \frac{\sum_{j=1}^{n-1} R_j E_j f(d_{ij})}{\sum_{j=1}^{n-1} E_j f(d_{ij})}$$

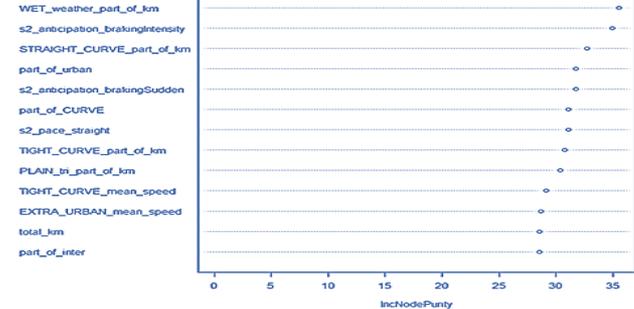
Ici, le risque géographique sur les différentes zones s'explique uniquement par les distances entre celles-ci

VS

Zonier innovant



Random Forest + Lissage par crédibilité



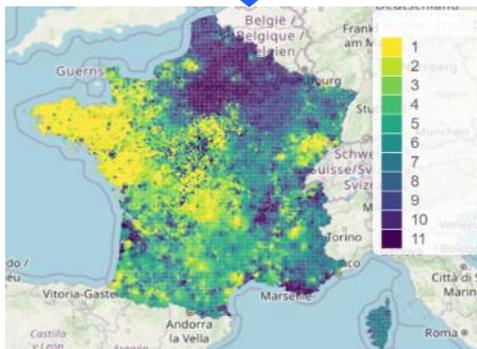
Au-delà des distances, **bien d'autres facteurs peuvent expliquer le risque géographique !**

Et pourtant... >>>

- 3 • Evaluations techniques de l'apport de la télémétrie

## MÉTHODE 2: ZONIER

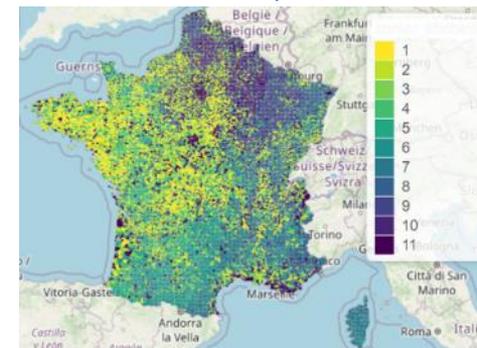
### Zonier traditionnel



Cartographie du zonier traditionnel

VS

### Zonier innovant



Cartographie du zonier innovant

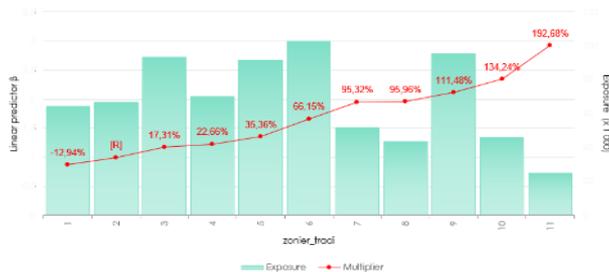
**Le zonier innovant propose des zones plus affinées que le zonier traditionnel**

Toutefois, ces 2 zoniers sont des zoniers techniques qui mériteraient des retouches avant leur mise en production

- 3 • Evaluations techniques de l'apport de la télématique

## MÉTHODE 2: ZONIER

### Zonier traditionnel



Courbe des facteurs multiplicatifs des modalités du zonier traditionnel

L'intégration du **zonier traditionnel** dans le model initial permet d'obtenir le **modèle de référence** de fréquences de sinistres

Spread standard du **zonier traditionnel** = **70,25%**

VS

$$spread = 1 - \frac{\exp(\beta_1)}{\exp(\beta_{11})}$$

+13  
pts

### Zonier innovant



Courbe des facteurs multiplicatifs des modalités du zonier innovant

L'intégration du **zonier innovant** dans le model initial permet d'obtenir le **modèle innovant** de fréquences de sinistres

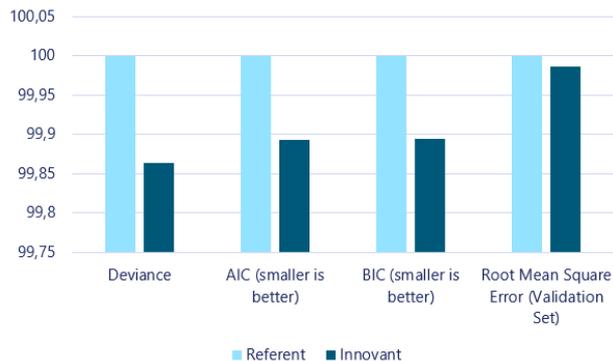
Spread standard du **zonier innovant** = **83,7%**



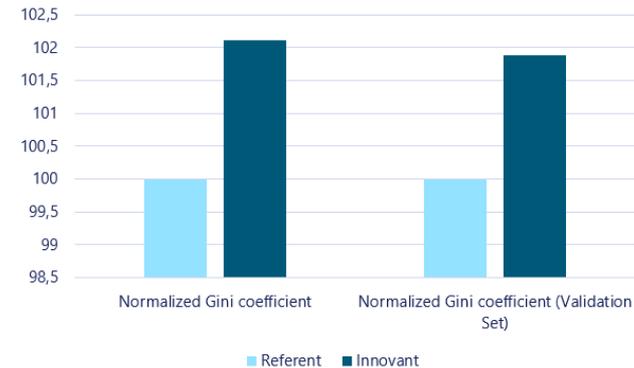
Une formule revue du spread a aussi confirmé ces résultats

- 3 • Evaluations techniques de l'apport de la télématique

## MÉTHODE 2: ZONIER



Comparaison des indicateurs de qualité (base 100)

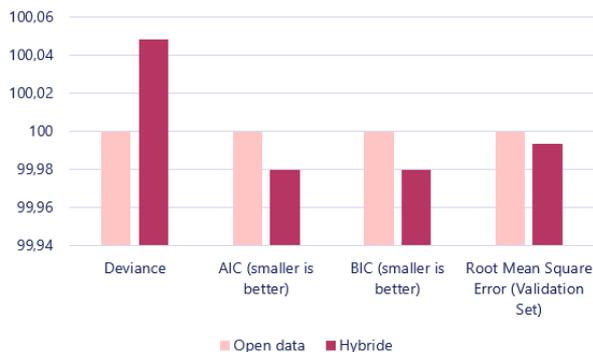


Comparaison des indices de Gini normalisés (base 100)

**Le modèle innovant surperforme le modèle de référence,**  
même si ces écarts de performance sont moindres comparés à ceux obtenus  
dans l'approche naïve

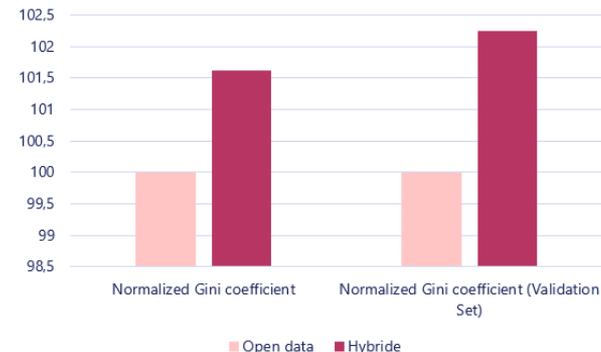
- 3 • Evaluations techniques de l'apport de la télémétrie

## MÉTHODE 2: ZONIER



Comparaison des indicateurs de qualité (base 100)

Plus ces indicateurs augmentent, meilleurs ils sont



Comparaison des indices de Gini normalisés (base 100)

**Modèle Open data** = Modèle contenant un zonier se basant uniquement des Open data.

**Modèle Hybride** = Modèle contenant un zonier se basant sur une combinaison de données télémétriques et des Open data.



La combinaison **Données télémétriques & Open data** est une alliance très prometteuse...

## SOMMAIRE

1 - Contexte et enjeux du mémoire

2 - Présentation des données externes : Smart Road Data

3 - Evaluations techniques de l'apport de la télématique

**4 - Evaluations opérationnelles de l'apport de la télématique**

5 - Conclusion et ouverture

- 4 • Evaluations opérationnelles de l'apport de la télématique

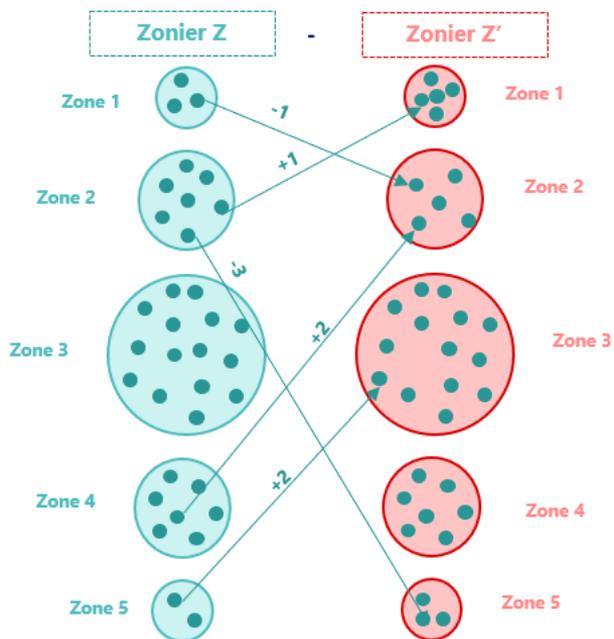
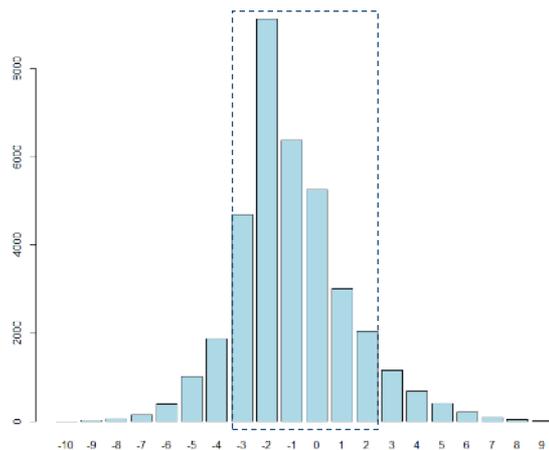


Illustration du *switch* entre 2 zoniers

$$\text{Switch} = \text{zonier}_{\text{traditionnel}} - \text{zonier}_{\text{innovant}}$$

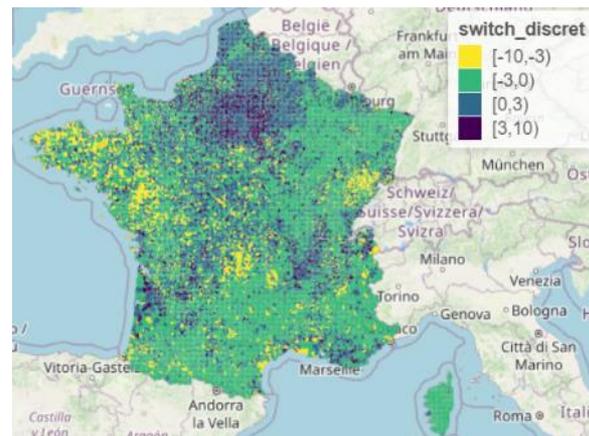
Le *switch* quantifie les **migrations** des communes  
entre deux zoniers.

- 4 • Evaluations opérationnelles de l'apport de la télémétrie



Distribution du switch entre le zonier traditionnel et le zonier innovant

En moyenne, le **sur-classement** maximal du risque est de **3 classes** et le **déclassement** maximal est de **2 classes**

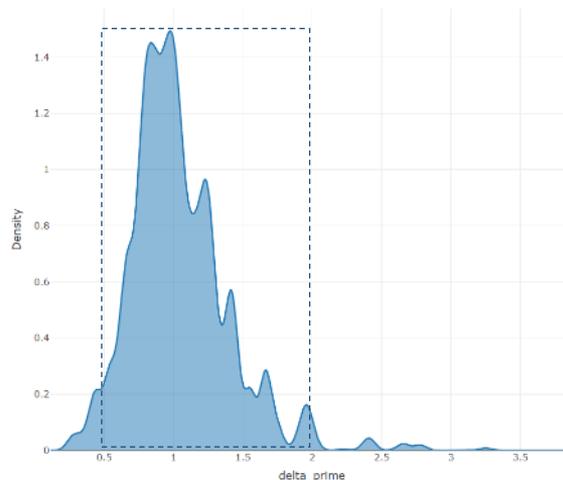


Cartographie du switch entre le zonier traditionnel et le zonier innovant

Les classes de *switch* extrêmes se localisent principalement en **Île-de-France (déclassement extrême)** et en **Bretagne (sur-classement extrême)**

- 4 • Evaluations opérationnelles de l'apport de la télématique

$$\Delta = \frac{Prime_{inno}}{Prime_{ref}}$$



Distribution du delta entre les primes innovantes et les primes de référence

En moyenne, la **minoration** (resp. **majoration**) maximale suggérée par la structure tarifaire **innovante** revient à **diviser** (resp. **multiplier**) le **tarif de référence par 2**

- 4 • Evaluations opérationnelles de l'apport de la télématique



La structure tarifaire **innovante** permet de **corriger** par endroits la structure tarifaire de **référence**

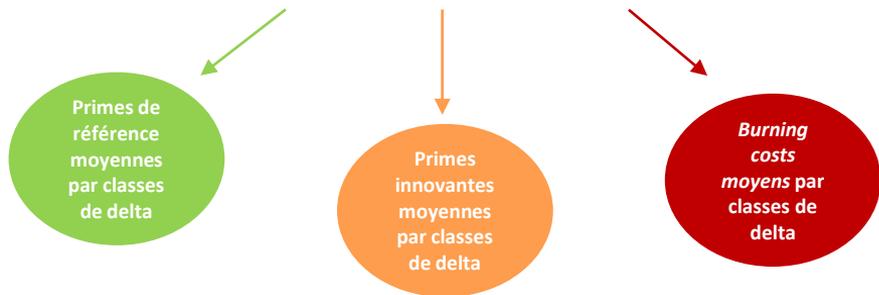
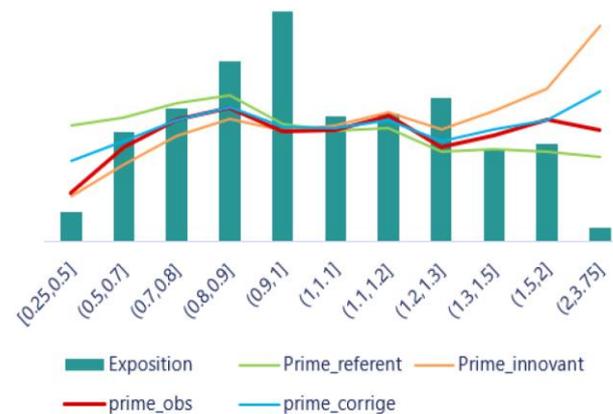


Primes de référence moyennes par classes de delta

Primes innovantes moyennes par classes de delta

*Burning costs* moyens par classes de delta

- 4 • Evaluations opérationnelles de l'apport de la télématique



## SOMMAIRE

1 - Contexte et enjeux du mémoire

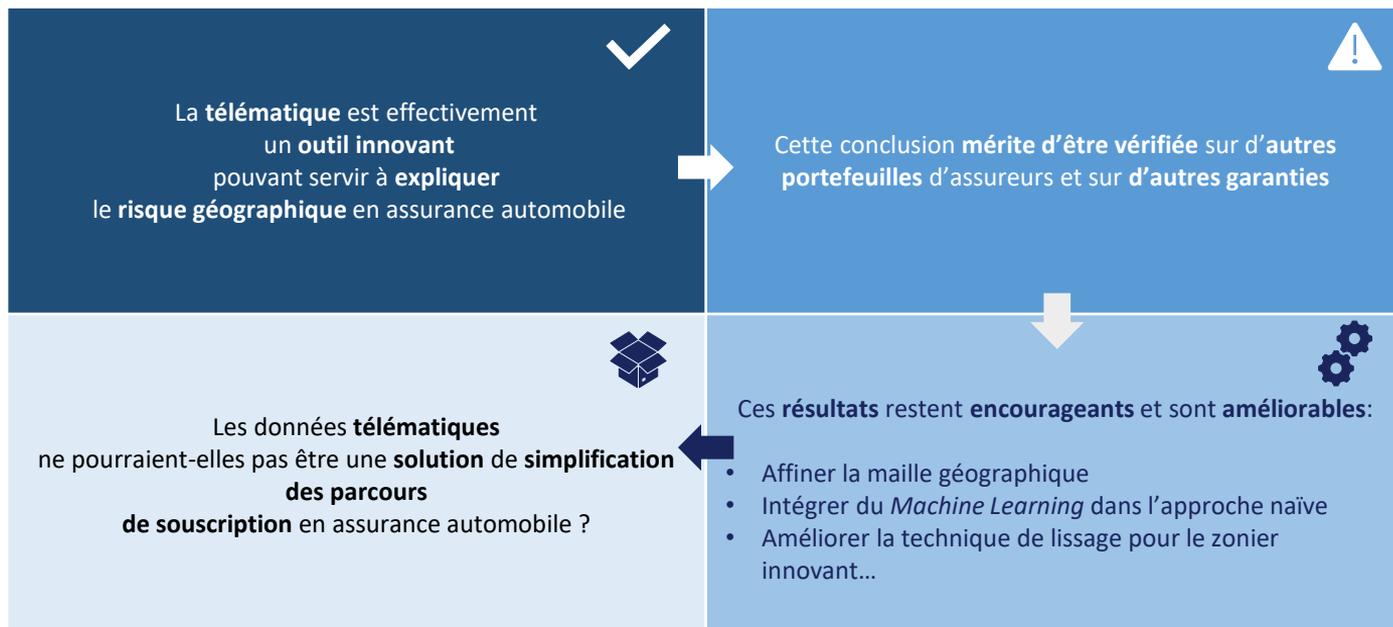
2 - Présentation des données externes : Smart Road Data

3 - Evaluations techniques de l'apport de la télématique

4 - Evaluations opérationnelles de l'apport de la télématique

 **5 - Conclusion et ouverture**

- 5 • Conclusion et ouverture



# Merci de votre attention