

# Evaluation de la qualité des données de l'inventaire BASEMIS® : consommations d'énergie, des émissions de polluants atmosphériques et gaz à effet de serre en Pays de la Loire



Réalisé par : Axel FOURNEYRON

Encadrement professionnel : Corentin LEMAIRE et Sébastien CIBICK

Encadrement pédagogique : Hélène MATHIAN

Jeudi 30 juillet 2020 – Mémoire du Master Géographie Numérique en 2ème année

## Remerciements

Ce travail a été rendu possible grâce à la collaboration, le soutien et la présence de nombreuses personnes. Je souhaite donc leur témoigner ma reconnaissance pour m'avoir permis d'aboutir à ce travail.

Je tiens tout d'abord à remercier mes encadrants au sein de l'association Air Pays de la Loire, Corentin Lemaire, Sébastien Cibick et Céline Puente-Lelièvre pour le temps qu'ils m'ont consacré, que ce soit pour m'apporter des conseils, me faire part de leurs points de vue, ou encore m'aider à la réalisation de ce mémoire. Je remercie aussi Marion, Nathalie, Thierry et Charlotte, qui, malgré les conditions atypiques du télétravail et du chômage partiel, m'ont fourni une ambiance et un cadre de travail idéal mais aussi une fantastique découverte de la ville et ces événements !

Je souhaite également exprimer ma reconnaissance à Hélène Mathian, ma tutrice de stage, qui a su m'accompagner et m'orienter dans la réalisation de ce travail. Par ailleurs j'exprime ma gratitude à l'équipe pédagogique au sein du master GéoNum, pour leur bienveillance, leur altruisme et leurs expertises qui nous ont été transmises ces dernières années.

J'aimerais enfin adresser mes remerciements à mes camarades GéoNum, mes fidèles amis, avec qui j'ai pu dialoguer, me confier et rebondir dans cette situation particulière. Durant ces deux dernières années, j'ai eu un immense plaisir de travailler, découvrir, partager et de m'épanouir à leur côté. C'est ce qui aujourd'hui m'a permis de m'impliquer pleinement dans ma mission et d'y apporter les outils les plus pertinents.

## Synthèse

Air Pays de la Loire, association agréée par l'État pour la surveillance de la qualité de l'air, a mis en place l'inventaire BASEMIS®. Celui-ci fait l'état des lieux des consommations d'énergie et des émissions de polluants et de gaz à effet de serre à l'échelle communale pour la région Pays de la Loire, par secteur d'activité (industriel, agricole, résidentiel, etc.), par usage (combustion, transport, etc.), par type d'énergie (produits pétroliers, électricité, etc.) et activités non énergétiques (élevage, utilisation de peinture dans l'industrie, utilisation de composés fluorés...).

L'objectif d'Air Pays de la Loire est de publier l'inventaire BASEMIS® le plus pertinent et qualitatif compte tenu des données disponibles. C'est à travers cette volonté de recherche qualitative que la mise en place d'un outil automatique permettant d'évaluer la qualité des données de l'inventaire s'est révélée nécessaire. Cet inventaire a la particularité d'être réalisé à travers différents processus en fonction des secteurs d'activité, des usages, des types d'énergies...

Le projet s'est structuré en plusieurs étapes :

1. Une prise en main de la base de données,
2. La construction d'une méthodologie d'analyse,
3. La détection d'anomalies,
4. L'évaluation de la qualité des données,
5. Le développement des outils de visualisation,
6. Les formations sur la prise en main de l'outil.

À l'issue de ce projet, un outil opérationnel et évolutif a été mis à disposition d'Air Pays de la Loire. Cet outil permet de mieux évaluer la pertinence des données utilisées et de mieux identifier les corrections et les points de vigilance nécessaires à la diffusion de l'inventaire.

Cet outil a permis un premier diagnostic de la qualité des données dans trois secteurs importants de BASEMIS® : l'agriculture, le secteur résidentiel et les transports routiers.

Cet outil prend la forme d'une application en permettant de :

- Donner des informations sur les analyses et procédés réalisés sur le secteur
- Evaluer la qualité de précision thématique
- Evaluer la qualité d'évolution temporelle
- Comparer la qualité entre les versions de l'inventaire

L'évaluation de la qualité se fait à l'aide de trois aspects de qualité des données : le nombre d'erreurs détectées, le taux d'erreurs présent, le poids des erreurs.

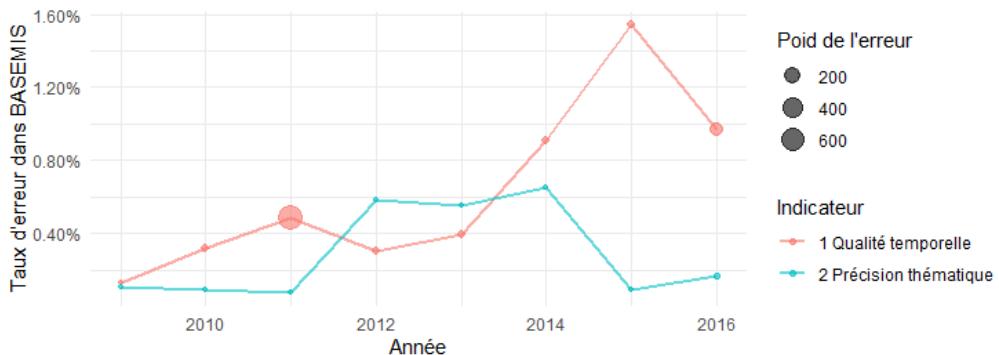
L'application peut afficher des data visualisations interactives pour représenter ces trois indices. Elle permet aussi de lancer la génération d'un rapport automatisé sur un secteur spécifique pour une année choisie de l'inventaire.

Le rapport automatisé contient l'ensemble des analyses réalisées pour détecter les erreurs. Il se décompose en 4 parties :

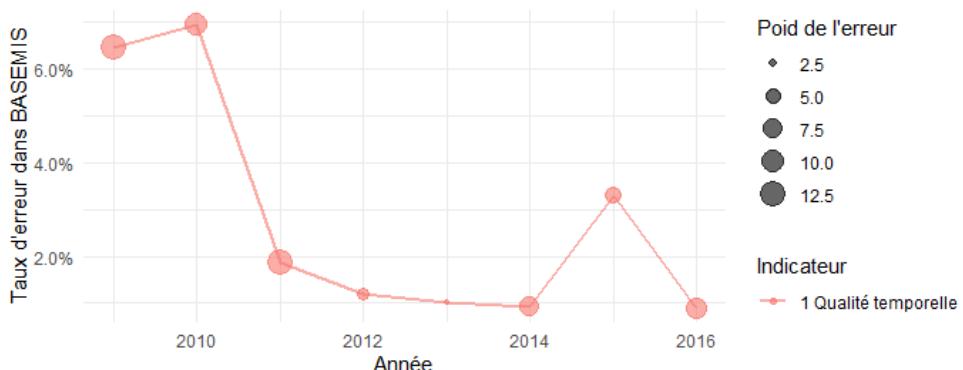
- Informations générales sur le secteur, les données utilisées et les méthodes d'analyse
- Data visualisation des comportements des données (histogramme et boxplot)
- Data visualisation des données avec les données contextuelles (cartographies)
- Listing des données détectées comme anormales
- Affichage des indicateurs de qualité des données

C'est grâce à ces deux outils (application et rapport) que nous avons pu obtenir et interpréter des résultats sur la qualité des données du secteurs résidentiel, routier et agricole.

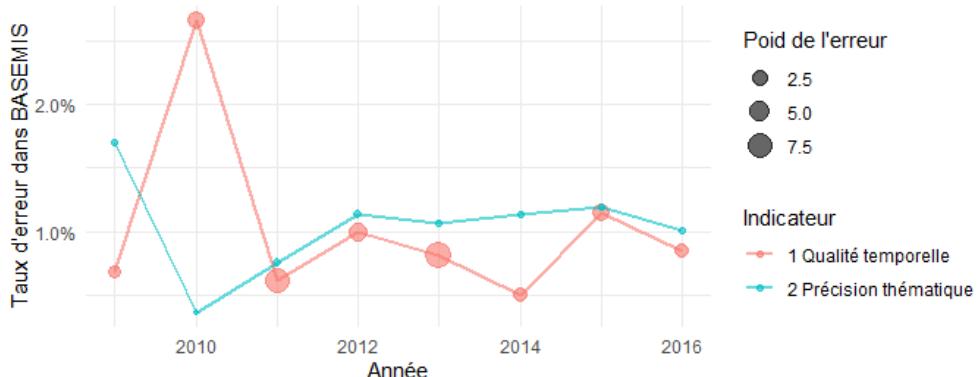
**Sur le secteur résidentiel :** des analyses pertinentes, qui permettent d'évaluer avec précision la qualité des données. La précision thématique présente moins de 2% d'erreurs. Le poids de ces erreurs est faible. La qualité d'évolution temporelle présente moins de 3% d'erreurs, mais certaines erreurs ont un poids assez important (500 fois supérieurs à notre limite).



**Sur le secteur du transport routier :** développement d'une méthode d'analyse qui permet d'évaluer seulement la qualité de l'évolution temporelle des données primaires. L'échelle spatiale du secteur routier permet d'avoir un échantillonnage important de données ce qui consolide la tendance observée. Avec moins de 4% d'anomalies détectées, le résultat est satisfaisant. Il existe de multiples causes qui peuvent expliquer ces anomalies : modification des plans de circulation, travaux de voiries...



**Sur le secteur agricole :** une méthodologie d'analyse mise en œuvre qui nécessite des vigilances voire des améliorations sur le niveau de détail pour certaines catégories. Elle permet cependant d'évaluer la qualité des données. Sur l'analyse de précision thématique, moins de 3% d'erreurs sont détectées avec très peu de variations au fil du temps. La qualité de l'évolution temporelle ne montre pas plus de 5% d'anomalies dans la base. Elles peuvent être liées à des changements pratiques agricoles, des variations de cheptels....



Dans l'ensemble, les résultats montrent que l'inventaire BASEMIS® V5 confirment par la faible présence d'erreur la qualité de l'inventaire. Les outils développés permettent de faciliter le ciblage d'erreurs notamment celle ayant un poids considérable dans l'inventaire. Cependant les outils ne permettent pas de déterminer la nature des erreurs. Les erreurs détectées sont multifactorielles et nécessite le regard d'expert.

Une deuxième partie de ce projet s'est axée sur la transmission des compétences pour permettre à l'équipe BASEMIS® d'utiliser et de faire évoluer l'outil, par la mise en place de formation en interne et mise à disposition de documents d'utilisations.

Ce mémoire, permet d'énumérer et de recenser l'ensemble des recherches faites sur la qualité des données dans ce projet. Il permet aussi de préconiser à Air Pays de la Loire de nouvelles analyses et méthodes qui permettraient de s'affranchir des problématiques rencontrées :

- Analyses en deep learning sur les données primaires des partenaires,
- Détection et correction automatiques des erreurs,
- Identification de la provenance des erreurs.

# Sommaire

<b>1</b>	<b>Air Pays de la Loire .....</b>	<b>7</b>
1.1	Structure et missions .....	7
1.2	Inventaire BASEMIS® .....	9
1.3	Objectifs du stage .....	12
<b>2</b>	<b>Base de données BASEMIS® .....</b>	<b>13</b>
2.1	Architectures de stockages des données.....	13
2.2	Modèle conceptuel des données sollicitées .....	14
<b>3</b>	<b>Qualité des données .....</b>	<b>16</b>
3.1	Critères d'évaluation de la qualité de données .....	17
3.2	Méthode d'analyse qualité.....	18
3.3	Conception de l'indicateur de précision thématique.....	21
3.4	Conception de l'indicateur qualité temporelle .....	22
3.5	Détection d'anomalie .....	23
3.6	Indicateur de qualité des données .....	26
3.7	Outils de reporting .....	27
<b>4</b>	<b>Analyses thématiques .....</b>	<b>33</b>
4.1	Les différents secteurs de l'inventaires BASEMIS® .....	33
4.2	Secteur Résidentiel .....	34
4.3	Secteur Routier.....	38
4.4	Secteur Agricole .....	41
<b>5</b>	<b>Gestion de projet .....</b>	<b>45</b>
5.1	Planification des objectifs .....	45
5.2	Méthodes de travail et organisation .....	47
5.3	Formation et documentation des outils mis en place .....	51
<b>6</b>	<b>Conclusions .....</b>	<b>53</b>
6.1	Evaluation des outils apportés .....	53
6.1	Qualité des données des secteurs analysés.....	54
6.2	L'ensemble du projet de validation de données .....	57
6.3	Perspectives d'amélioration et préconisation .....	58
<b>7</b>	<b>Annexes .....</b>	<b>59</b>
7.1	Annexe 1 : Air Pays de la Loire .....	60
7.2	Annexe 2 : Polluants et Gaz à effet de serre.....	62
7.3	Annexe 3 : Cartes ATMO France .....	63
7.4	Annexe 4 : équipe BASEMIS® .....	64
7.5	Annexe 5 : Rapport secteur résidentiel .....	65
7.6	Annexe 6 : Rapport secteur routier .....	68
7.7	Annexe 7 : Rapport secteur agricole .....	70
7.8	Annexe 8 : Application qualité de la donnée .....	77
7.9	Annexe 9 : Documentation architecture des fichiers .....	81
7.10	Annexe 10 : Guide d'utilisation .....	84
7.11	Annexe 11 : Documentation – logigrammes (extrait) .....	88
7.12	Annexe 12 : Liens de formation et tutoriels .....	90
7.13	Annexe 13 : Table des schémas .....	91
<b>8</b>	<b>Références .....</b>	<b>92</b>
<b>9</b>	<b>Glossaire .....</b>	<b>93</b>

# 1 Air Pays de la Loire

## 1.1 Structure et missions

### 1.1.1 Association Agréées de Surveillance de la Qualité de l'Air (AASQA)

Air Pays de la Loire possède deux missions principales qui constituent aujourd'hui son cœur de métier historique :

1. Sa première mission est la **surveillance de la qualité de l'air** sur le territoire régional. Cette surveillance porte sur l'évaluation de la qualité de l'air dans les Pays de la Loire au regard de substances pouvant entraîner des effets nocifs sur la santé et l'environnement.
2. La seconde grande mission d'Air Pays de la Loire consiste à **informer le public et les autorités compétentes** par la publication fréquente et réactive des résultats obtenus sous la forme de communiqués, bulletins, rapports ou encore via son site internet.

Pour réaliser ses missions Air Pays de la Loire se constitue d'une trentaine d'experts composés d'ingénieurs d'étude, métrologues, modélisateurs et communicants, organisés par processus (Annexe 1 : Air Pays de la Loire). L'association est pilotée par un conseil d'administration de 24 membres qui met en œuvre la politique de l'association décidée en assemblée générale. Ces membres, sont organisés en quatre collèges (Schéma 1).

Aujourd'hui, une grande diversité d'actions sont mises en place pour agir sur la qualité de l'air. C'est plus d'une dizaine d'activités qui sont réalisées au sein de l'association (Schéma 2)



Schéma 1 Organisation d'Air Pays de la Loire



Schéma 2 Panel des activités d'Air Pays de la Loire

## 1.1.2 Une organisation nationale

Les **associations agréées pour la surveillance de la qualité de l'air (AASQA)** mettent en œuvre la surveillance à l'échelle régionale. Elles sont coordonnées à l'échelle nationale par la fédération Atmo France afin de mutualiser leurs pratiques et valoriser leurs travaux. Après la fusion des régions en 2016, on compte aujourd'hui 19 AASQA sur le territoire métropolitain et en outre-mer. (Annexe 3 : Cartes ATMO France)

C'est grâce à cette coordination nationale que le territoire Français peut appliquer la loi LAURE. Cette loi permet d'assurer le droit de pouvoir depuis 1996 : « *de respirer un air qui ne nuise pas à sa santé* »

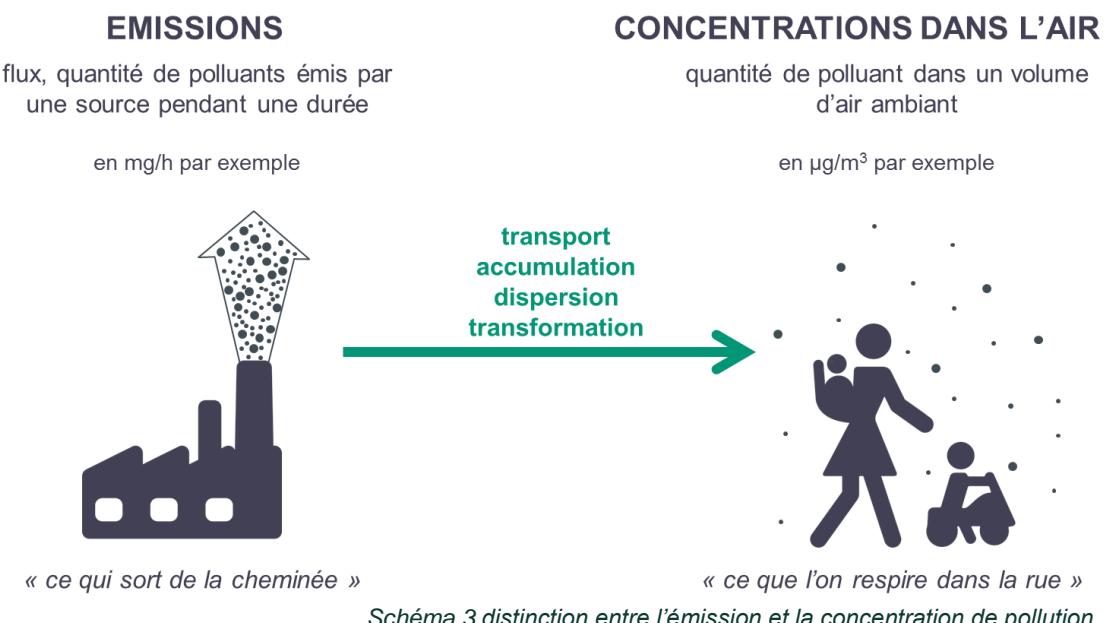
## 1.1.3 La qualité de l'air et enjeu climatique et énergétiques

Afin de répondre à sa mission de **surveillance de la qualité de l'air**, Air Pays de la Loire a mis en place des mesures de concentration de la qualité de l'air. Des stations mesurent les concentrations des polluants dans un volume d'air ambiant. Elles permettent ainsi de caractériser la qualité de l'air respiré par la population (Schéma 3).

Air Pays de la Loire a aussi mis en place un inventaire des consommations d'énergie et d'émissions de polluants et de gaz à effet de serre permettant, quant à eux, de quantifier les polluants émis par une source (industrielle, résidentielle, etc.) pendant une certaine durée (Schéma 3).

Un certain nombre d'activités émettrices de pollution est associé à une consommation d'énergie. L'amélioration de la qualité de l'air et l'adaptation au changement climatique par la transition énergétique sont ainsi étroitement liées.

Les actions qui visent à réduire les émissions de gaz à effet de serre (déplacements motorisés, secteurs résidentiel, tertiaire, industriel et agricole) ont dans l'ensemble un effet positif sur les émissions de polluants atmosphériques locaux.



Cet inventaire, est une des briques des multiples actions de l'association permettant de :

1. **Quantifier**, les sources de pollution du territoire
2. **Modéliser**, les prévisions de pollution atmosphérique
3. **Alerner**, les populations touchées, et les industries concernées
4. **Accompagner**, les membres, collectivités, territoires dans une démarche d'amélioration.

## 1.2 Inventaire BASEMIS®

### 1.2.1 Objectifs de l'inventaire

L'inventaire BASEMIS® permet de détailler les consommations d'énergie, les productions d'énergie renouvelable et les émissions de gaz à effet de serre et de polluants de la région Pays de la Loire. C'est un outil d'expertise et d'aide à la décision pour accompagner les collectivités territoriales et les services de l'État, notamment dans le suivi de différents plans d'actions, tels que les Plans Climat-Air-Energie Territoriaux (PCAET), le Schéma régional Climat-Air-Énergie (SRCAE) ou encore les Plans locaux d'urbanisme (PLU)... Il sert également de données d'entrée dans les outils de modélisation et de prévision de la qualité de l'air.

Des nombreux projets menés par Air Pays de la Loire visent à valoriser l'inventaire BASEMIS® sous différentes formes. On retrouve alors des projets récents tel que la plateforme d'open data (2018), la mutualisation de service dataviz (2019), les communications de dataviz sur les réseaux sociaux (2018)

L'inventaire BASEMIS® est réalisé par une équipe d'inventoriste (Annexe 4 : équipe BASEMIS®)

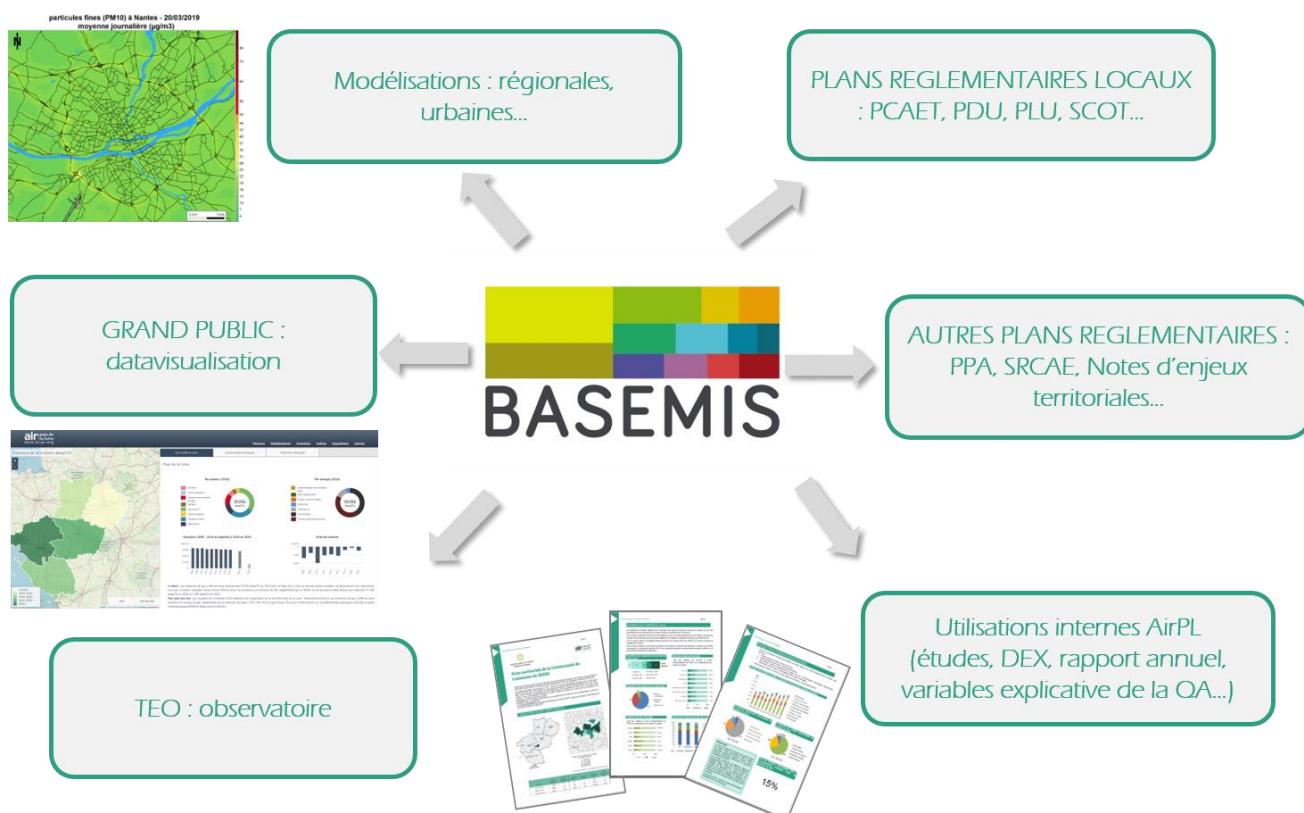


Schéma 4 Les différentes utilisations de l'inventaire BASEMIS®

## 1.2.2 Méthodes de réalisations

La réalisation de l'inventaire BASEMIS® s'effectue suivant les critères de qualité, de vérification et d'amélioration continue développés par la Convention Cadre des Nations Unies sur les Changements Climatiques (CCNUCC)<sup>1</sup>. Ces critères de qualité visent notamment à garantir les caractéristiques suivantes aux inventaires territoriaux :

- **Transparence** : les méthodologies et données d'entrée utilisées pour la construction de l'inventaire doivent être tracées, ainsi que l'incertitude relative aux résultats
- **Exhaustivité** : l'inventaire doit recouvrir l'ensemble des sources d'émissions territoriales et le cas échéant justifier du caractère négligeable des sources non intégrées
- **Cohérence** : la cohérence temporelle des résultats doit être assurée, afin que ceux-ci reflètent les évolutions technologiques et comportementales, et non les évolutions méthodologiques inhérentes à un changement de version d'inventaire
- **Comparabilité** : la cohérence territoriale doit être assurée en appliquant une méthodologie commune et homogène (niveaux infrarégional, régional et national)
- **Précision** : les données locales fines doivent être privilégiées
- **Pérennité** : l'utilisation de données durables et mises à jour régulièrement doit être privilégiée

L'inventaire BASEMIS® est calculé à l'échelle communale, par secteur d'activité, par usage et par type d'énergie. Tous les deux ans depuis 2010, l'inventaire est mis à jour par le calcul des émissions produites pour les deux dernières années. (Schéma 5)

Temporalité	Echelle spatiale		
<ul style="list-style-type: none"> <li>• Année 2008 à 2016</li> <li>• Mise à jour tous les 2 ans</li> <li>• Mise à jour 2018 pour la fin de l'année 2020</li> </ul>	<ul style="list-style-type: none"> <li>• Région Pays de la Loire</li> <li>• Avec une résolution communale</li> </ul>		
Détails d'activités	Usage non énergétique	Usage énergétique	Source d'énergie utilisée
<ul style="list-style-type: none"> <li>• 10 secteurs références (agriculture, routier...)</li> <li>• 60 sous-secteurs (élevage, voiries, ...)</li> </ul>	<ul style="list-style-type: none"> <li>• Répartition des différents usages (études, petit vraquier, ...)</li> </ul>	<ul style="list-style-type: none"> <li>• Répartition des différents usages (chauffage, cuisson,...).</li> </ul>	<ul style="list-style-type: none"> <li>• 9 types d'énergie (gaz, produits pétroliers...),</li> <li>• 35 vecteurs énergétiques (biogaz, fioul...)</li> </ul>
Impact énergétique		Impact climatique	
<ul style="list-style-type: none"> <li>• Consommations énergétiques finales</li> <li>• Consommations énergétiques primaire</li> </ul>		<ul style="list-style-type: none"> <li>• Quantités des émissions de GES</li> <li>• Détails de chaque gaz à effets de serre</li> </ul>	
Impact qualité de l'air			
		<ul style="list-style-type: none"> <li>• Plus de 40 polluants (PM10, NOx, ...)</li> </ul>	

Schéma 5 Descriptions des détails de précision de l'inventaire

<sup>1</sup> Webographie : [https://www.ipcc-nqip.iges.or.jp/public/2019rf/pdf/1\\_Volume1/19R\\_V1\\_Ch06\\_QA\\_QC.pdf](https://www.ipcc-nqip.iges.or.jp/public/2019rf/pdf/1_Volume1/19R_V1_Ch06_QA_QC.pdf)

La réalisation de BASEMIS® consiste en un calcul théorique (Schéma 6) des flux de polluants émis dans l'atmosphère (masse du composé par unité de temps). Il s'agit de faire correspondre à des données dites primaires (statistiques, comptages, enquêtes, besoins énergétiques, etc.), des facteurs d'émissions issus d'expériences métrologiques ou de modélisations.

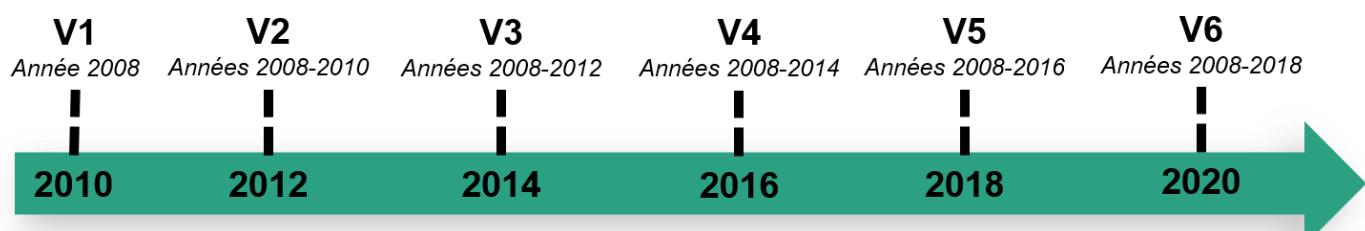
Les émissions sont estimées pour chacune des activités retenues pour l'inventaire au moyen de la formule suivante qui exprime de manière très générale et schématique la méthode utilisée.



### Schéma 6 Méthode de calcul de l'inventaire BASEMIS®

La méthodologie déployée provient du Pôle de coordination national sur les inventaires d'émissions territoriaux<sup>2</sup> (PCIT) qui fixe les lignes directrices pour le calcul des émissions de polluants dans les inventaires territoriaux. Les facteurs d'émissions appliqués sont issus majoritairement du référentiel français d'Organisation et méthodes des inventaires nationaux des émissions atmosphériques en France<sup>3</sup> (OMINEA) élaboré par le Centre interprofessionnel technique d'études de la pollution de l'air (CITEPA).

Les processus et méthodes appliqués sont spécifiques à chaque secteur et sous-secteurs d'activités, le détail de ses processus sont expliqués dans la partie Analyses thématiques pour les secteurs analysés.



### Schéma 7 Chronologies des versions de BASEMIS®

### 1.2.3 L'évolution et l'amélioration de BASEMIS®

L'inventaire BASEMIS® est mis à jour tous les deux ans. Cette mise à jour couvre deux années supplémentaires par rapport à la version précédente. Par ailleurs, cette mise à jour s'accompagne d'évolutions méthodologiques, qui sont prises en compte pour actualiser les résultats des années précédentes. (Schéma 7).

Pour la version 6 en cours de développement, les évolutions méthodologiques principales sont :

- **Intégrations de nouvelles sources** : manipulation de produits pulvérulents dans les ports, prise en compte des engins aéroportuaires, ...
  - **Consolidation de la base de données routière**, évolution du parc roulant routier, intégration des bus des principales agglomérations ...
  - **Evolution des formats** des données énergétiques de l'article 179 de la LTECV

<sup>2</sup> Webographie : <https://www.lcsqa.org/fr/rapport/guide-methodologique-pour-le-laboration-des-inventaires-territoriaux-des-emissions>

<sup>3</sup> Webographie : <https://www.citepa.org/fr/ominea/>

## 1.3 Objectifs du stage

### 1.3.1 Contexte

Dans le cadre de la mise à jour de BASEMIS®, des protocoles de vérification sont mis en œuvre quant à la cohérence des résultats, en termes de logique de base de données. Ces vérifications sont également réalisées sur les nombreuses données primaires reçues et traitées par Air Pays de la Loire.

Air Pays de la Loire a souhaité investiguer la qualité des données de l'inventaire à travers un stage de recherche, analyse et développement. Déterminer la qualité des données permet d'assurer la confiance nécessaire dans l'utilisation des données.

### 1.3.2 Objectifs

Ce stage vise à développer des outils de traitement automatisés pour analyser la qualité et la cohérence des données de l'inventaire BASEMIS®. Ces outils permettront de détecter des erreurs ou incohérences, et d'évaluer de manière générale la qualité de l'inventaire. Ces outils seront testés sur les données disponibles (BASEMIS® V5, 2008-2016) et sur les nouvelles données à intégrer dans la 6eme version (2008-2018).

### 1.3.3 Attentes générales du stage

Air Pays de la Loire souhaite grâce à ce stage avoir une analyse critique de la qualité des données de l'inventaire. Pour cela plusieurs tâches sont à réaliser :

- Faire une analyse critique sur les données de l'inventaire
- Comprendre les enjeux liés à l'inventaire
- Définir les méthodes d'analyses qualités
- Définir des critères d'évaluations de la qualité de la donnée
- Mettre en place des outils de traitements automatisés adaptés à Air Pays de la Loire

### 1.3.4 Attentes sur les outils de traitement automatisés

La mise en place d'un outil de traitements automatisés est un point essentiel du stage. Celui-ci devra prendre en compte plusieurs paramètres tels que :

- La mise en place d'un indicateur spécifique aux thématiques de l'inventaire
- L'analyse et l'intégration des résolutions géographiques et temporelles les plus pertinentes
- La définition des critères d'évaluation des indicateurs
- L'analyse des indicateurs et des critères mis en place à différents niveaux d'information
- La rédaction d'un mode opératoire et de documents récapitulatifs des résultats (rapports)

## 2 Base de données BASEMIS®

### 2.1 Architectures de stockages des données

Le Schéma 8 présente l'architecture de données sollicitées pour la réalisation de l'inventaire d'émissions. Différents serveurs sont mis en place pour modéliser certains comportements (modélisation du trafic routier) afin de générer des données précises pour l'inventaire.

L'inventaire BASEMIS® final est situé dans une table de données '**Résultats**', dans la base de données **BD BASEMIS® V5**. Dans cette base de données, on retrouve d'autres tables ('**Générale**' et '**Contexte**') permettant de réaliser des jointures afin d'attribuer certaines nomenclatures et les classifications nécessaires aux différentes normes Européens. Elles nous permettent aussi d'enrichir les informations liées à une année et aux découpages géographiques.

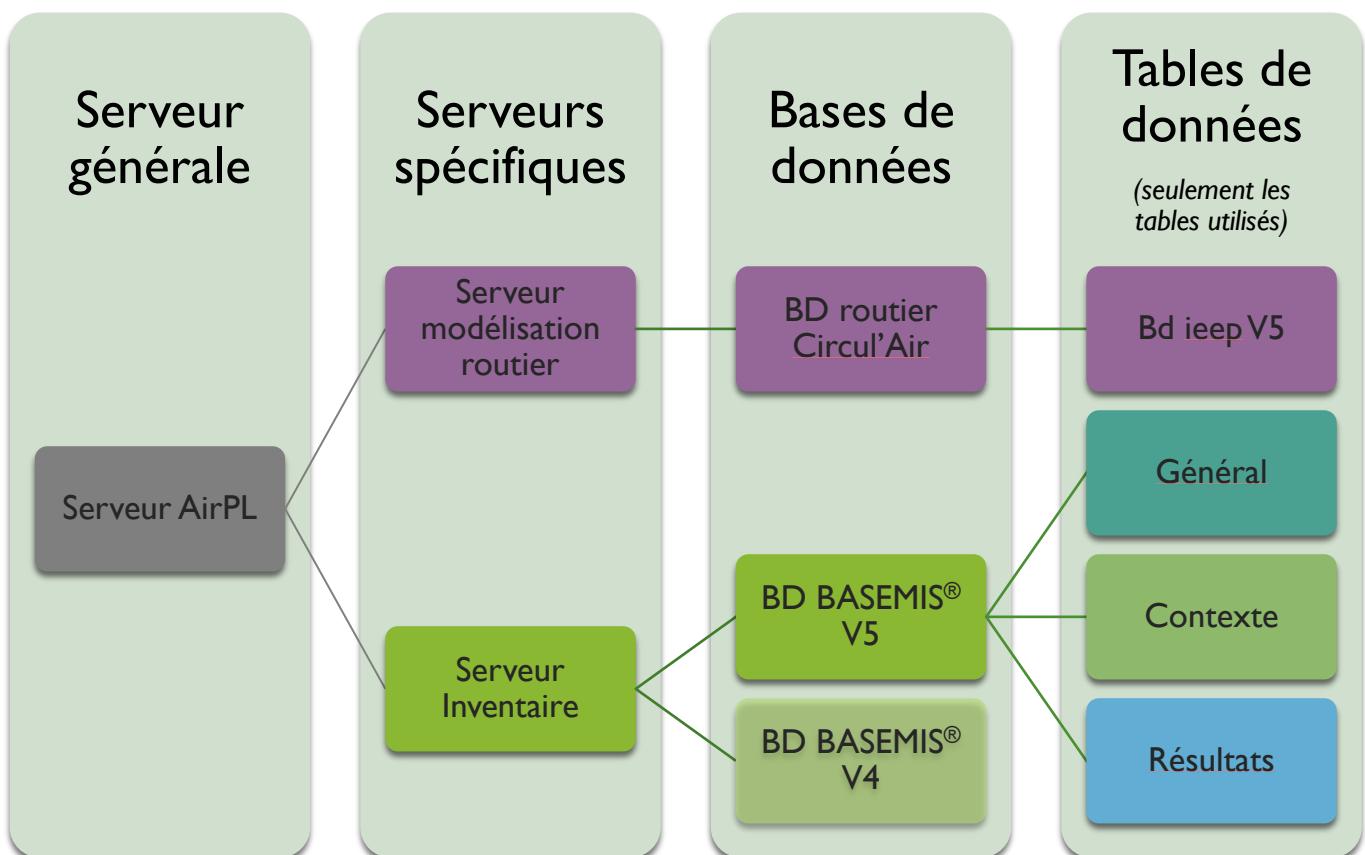


Schéma 8 Représentation de l'architecture des données sollicitées

## 2.2 Modèle conceptuel des données sollicitées

### 2.2.1 Structurations des données sollicitées

Le Schéma 9 ci-dessous présente les données sollicitées dans le cadre du travail réalisé.

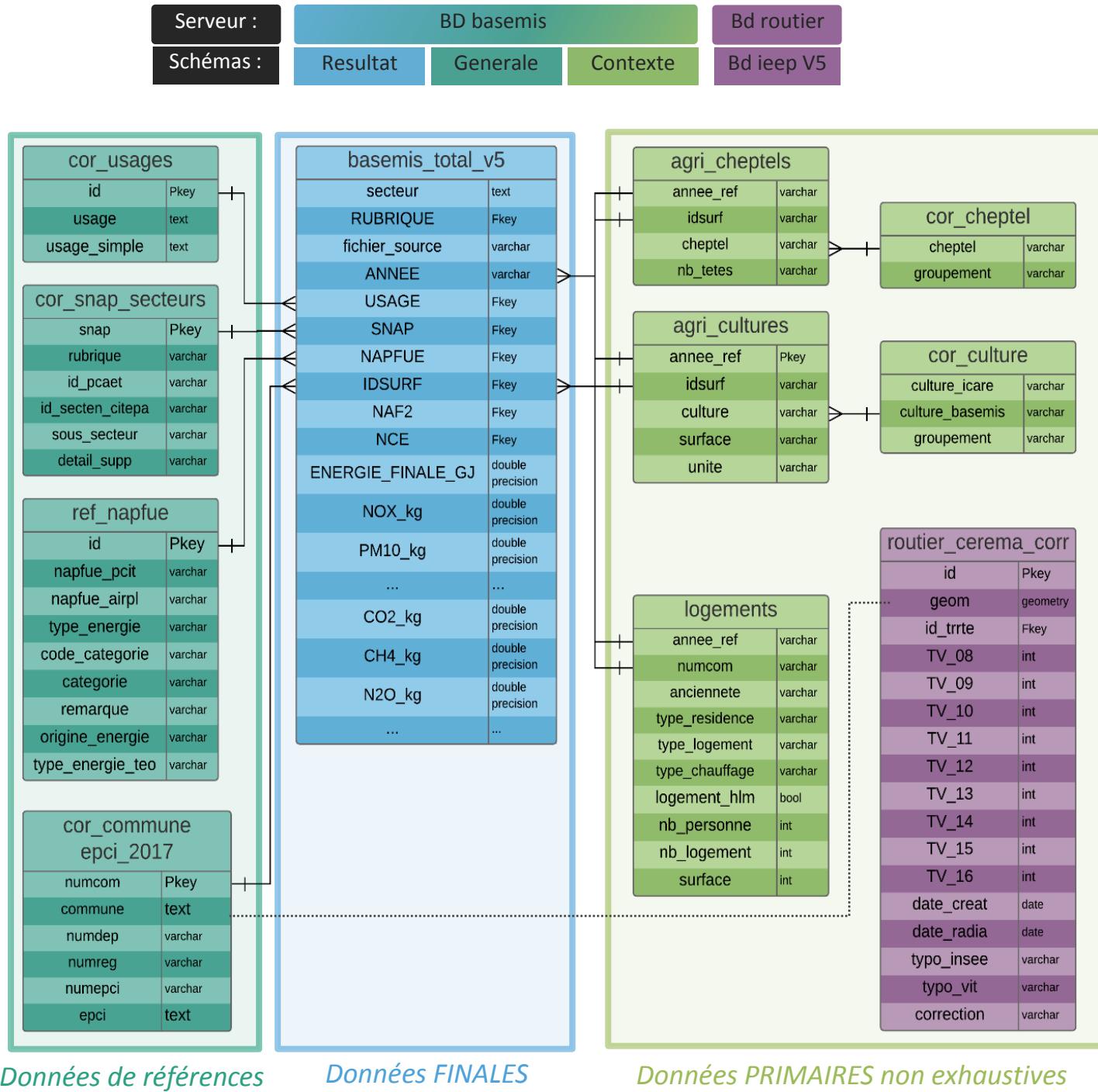


Schéma 9 Modèles conceptuelles des tables de données sollicités

## 2.2.2 Descriptions et explications des données sollicitées

L'objectif principal est d'analyser les **données finales** (Schéma 10 et Schéma 11) de l'inventaire BASEMIS®. L'inventaire est créé à partir des **données primaires** issues des multiples partenaires et publications officiels.

Néanmoins, afin d'analyser ces **données finales**, il est parfois nécessaire de les contextualiser et donc de faire appels aux **données primaires**

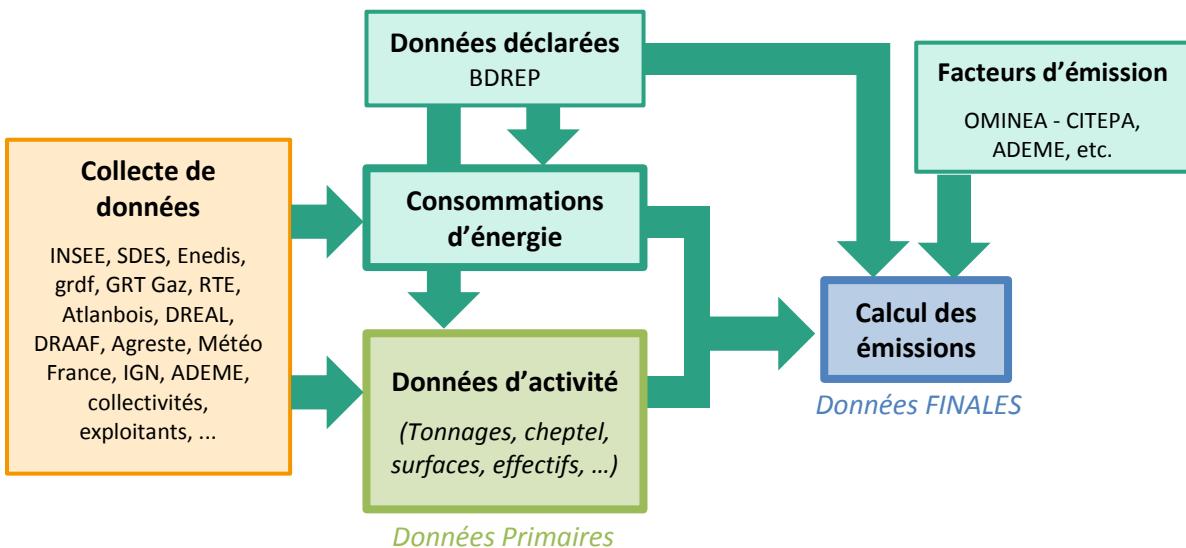


Schéma 10 Méthode de calcul de l'inventaire BASEMIS®

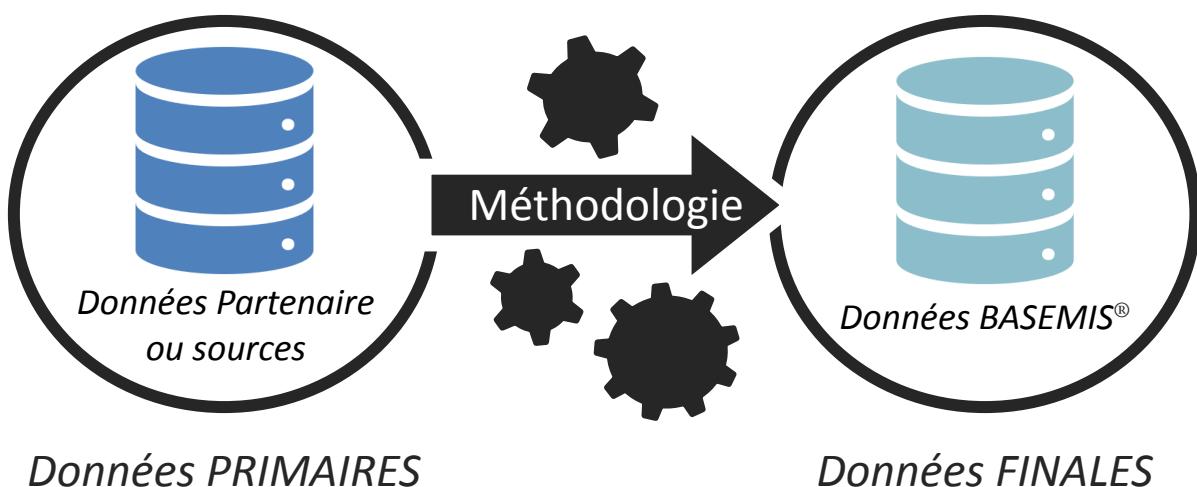


Schéma 11 Coordinations et utilisations des données

### Données Primaires :

Données descriptives : (*ldsurf, usages, rubrique, snap, napfue, ...*)

Données contextuelles : (*nombre de logement, nombre habitant, type de logement, surface logement, nombre de bétail, type de bétail, ...*)

### Données Finales :

Données descriptives : (*ldsurf, usages, rubrique, snap, napfue, ...*)

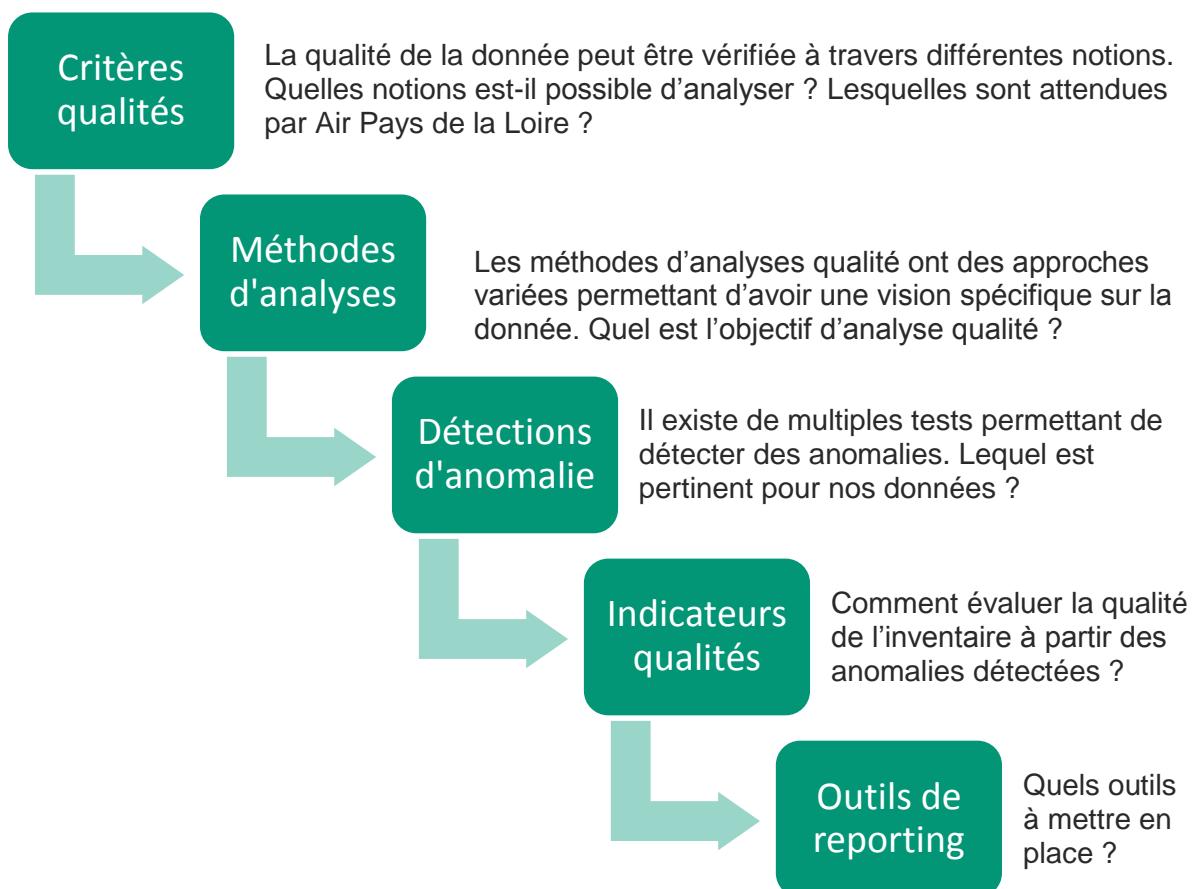
Données d'émissions et de consommations d'énergie : (*co2\_kg, ch4\_kg, n2o\_kg, GJ\_finale, Nox, pm10, ...*)

## 3 Qualité des données

La réalisation de l'inventaire comporte une phase d'évaluation de la qualité de données, de vérification de données et d'amélioration/correction des erreurs détectées. La qualité des résultats dépend fortement de celle des données primaires.

Pour produire et diffuser une donnée fiable et cohérente, il est nécessaire de s'interroger sur les moyens qui permettent de tester la cohérence mais aussi d'évaluer la qualité des données.

La qualité des données peut s'appliquer à travers différents : critères de qualité, méthodes d'analyses, indicateurs d'évaluation et d'outils de représentation. Il est nécessaire dans un premier temps d'identifier les attentes d'Air Pays de la Loire sur l'évaluation de la qualité des données afin de mettre en place la méthodologie la plus pertinente.



C'est à travers l'ensemble de ces questions que nous pouvons déterminer le procédé le plus adapté pour analyser la qualité des données d'Air Pays de la Loire.

Une bibliographie et un état des lieux est apporté sur chacune de ces questions pour aider à déterminer ce procédé d'analyse des données BASEMIS.

### 3.1 Critères d'évaluation de la qualité de données

Différents critères d'évaluation de la qualité de données ont été définis en suivant les préconisations de la norme ISO 19157 et autres références<sup>4</sup>. Le Schéma 12 présente une synthèse de toutes les préconisations. Conformément à cette norme, sept critères d'évaluation peuvent permettre de représenter la qualité des données.

Après avoir présenté ces critères d'évaluations à l'équipe d'ingénieurs s'occupant de l'inventaire BASEMIS®, nous avons pu échanger sur les validations déjà mises en place ainsi que celles attendues lors de mon stage.

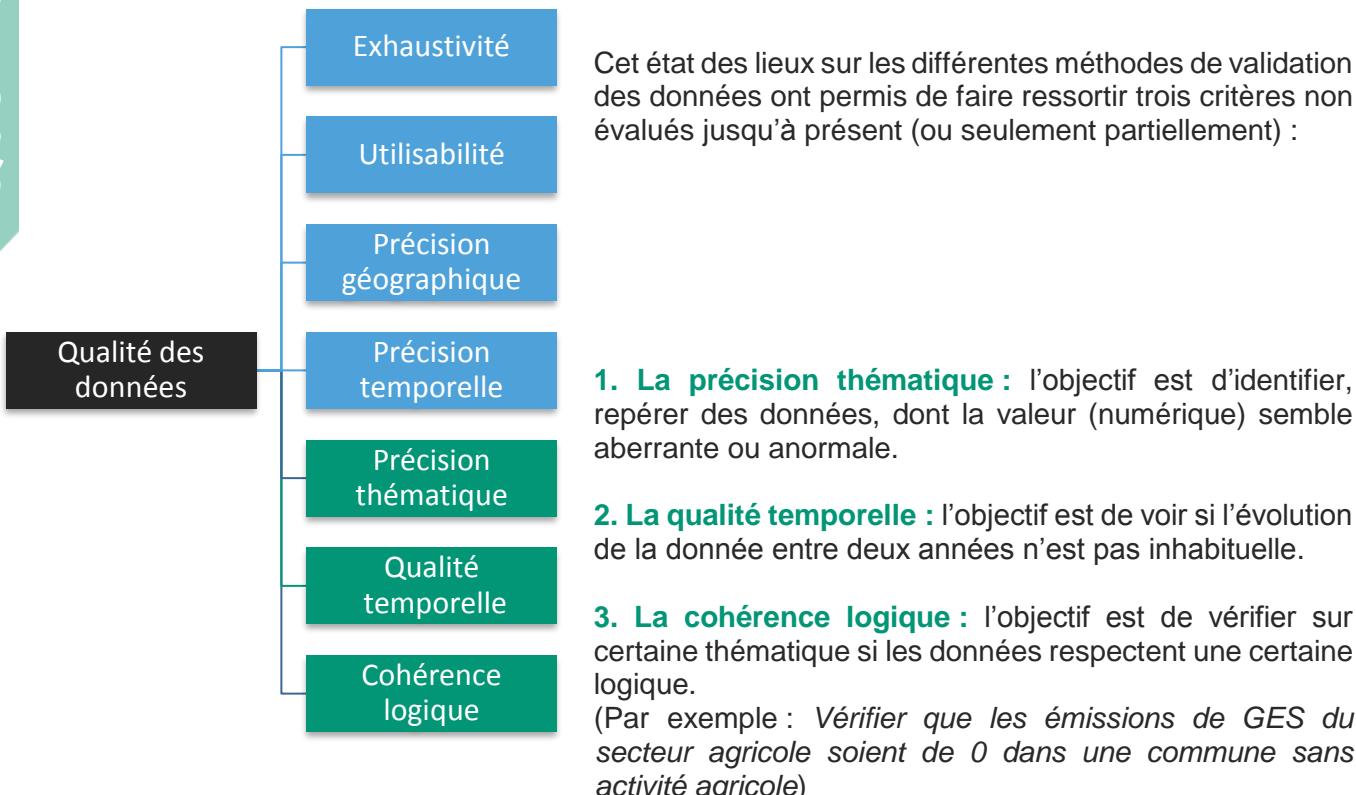


Schéma 12 : Critères d'évaluation de la qualité des données

A partir de ces objectifs généraux, l'enjeux est d'établir une méthode d'analyse permettant d'évaluer ces critères de façon pertinente en prenant en compte toutes les complexités des données.

<sup>4</sup> Document référence en [annexe](#) : « Norme ISO 19157 », « Data Management, qualité des données et compétitivité », « Base de données »

## 3.2 Méthode d'analyse qualité

### 3.2.1 L'état des lieux des améliorations de la qualité des données

Il y a plusieurs méthodes pour aborder la problématique de la qualité de la donnée. On retrouve quatre approches (complémentaires) qui permettent d'apporter une vision sur la qualité de données à différents niveaux. (Schéma 13)

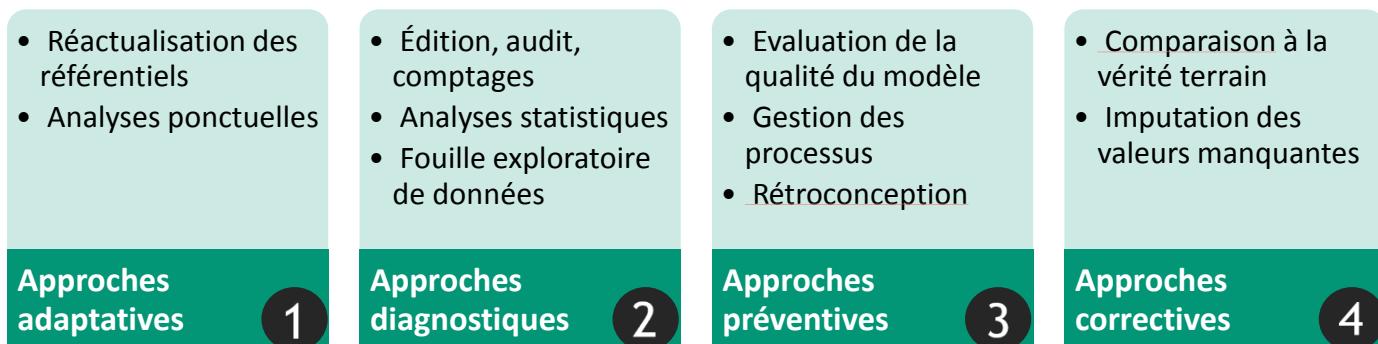


Schéma 13 Inventaires des approches qualités avec les méthodes associées

1

**L'approche adaptative**, n'est pas envisagée. Les guides et référentiels ayant une expertise dans les secteurs et domaines investigués répondent aux normes européennes et nationales. Ce sont des références sur lesquelles s'appuient l'inventaire BASEMIS pour sa construction. On retrouve parmi ces guides : le guide PCIT, les références du CITEPA, les conventions CCNUCC.

2

**L'approche diagnostique**, est partiellement réalisée notamment par l'audit MRV<sup>5</sup> mais également les rapports de cohérence établis à la fin de chaque version de l'inventaire. Elle permet de faciliter la compréhension des données, la provenance des erreurs, de cibler des données pour une approche corrective plus efficace, de faciliter le déroulement d'un audit et d'avoir une certification et document d'appuis pour évaluer la qualité des données.

3

**L'approche préventive**, n'est pas encore réalisée sur l'ensemble des données. Une partie de cette approche a été faite dans le secteur du trafic routier l'an dernier (analyse critique de la base de données du transport routier). Cependant cette approche doit être réalisée lors de l'identification d'une source d'erreur récurrente ? pour répondre à une problématique précise.

4

**L'approche corrective**, est déjà réalisée par les ingénieurs d'études avant la publication des données. Cette tâche nécessite de comprendre les différentes étapes de manipulation des données pour assurer que les corrections apportées aux données reflètent au mieux la situation réelle des émissions de polluants ou de GES sur le territoire.

Afin de mieux appréhender la qualité des données, l'approche diagnostique semble la plus appropriée car elle permet de :

- Visualiser et explorer les données
- Analyser et identifier des comportements
- Détecter et recenser des anomalies
- Evaluer la qualité des données à travers des critères et indicateurs.

<sup>5</sup> Audit MRV (Monitoring, Reporting, Verification) : <http://www.airpl.org/Publications/rapports/29-10-2015-rapport-de-projet-BASEMIS-MRV>

### 3.2.2 Méthodes d'analyses statistiques

Les **analyses de statistiques descriptives** nécessitent d'utiliser la cohérence des données. Il est nécessaire de s'appuyer sur des hypothèses fournies par les inventoriastes manipulant régulièrement les données pour acquérir cette cohérence.

Afin de définir ces hypothèses, une réflexion a été menée pour permettre la comparaison cohérente (en prenant en compte tous les niveaux de détail) des consommations d'énergie ou des émissions de polluants ou GES.

1. Cibler un seul secteur pour voir les consommations énergétiques, les émissions de polluants ou de GES les plus impactant dans le secteur ?

**Exemple secteur agricole (Schéma 32)** : 33% des GES, 95% des polluant NH3 et 45% des PM10 et seulement 5% des consommations d'énergies

2. Comprendre les méthodes de calcul du secteur et sous-secteur. Ces méthodes nous permettent de voir les variables utilisées, calculées ou modifiées.

**Exemple secteur agricole (Schéma 34)** : utilisation du nombre de cheptels dans l'élevage pour définir les émissions de polluants et de GES. Les émissions de NH3 et de PM10 sont donc liées (corrélées) car le facteur d'émission se base sur une variable commune : le nombre de cheptels.

3. Quelles sont les niveaux de détails assignés aux consommations énergétiques et émissions de polluants et de GES.

**Exemple secteur agricole** : il y a deux niveaux de détail (Schéma 14 les différentes activités, puis les rubriques élevages et travaux des cultures) dans l'inventaire BASEMIS qui permettent d'expliquer les émissions des polluants et GES.

4. Quelle hypothèse pouvons-nous émettre sur le secteur ?

**Exemple secteur agricole** : Les émissions des GES et polluants des bovins varient en fonction du nombre de d'animaux pris en compte. Un indicateur de GES par bêtes nous permet de comparer les différentes émissions des communes de façons cohérente. Le choix des GES et des PM10 est pertinent car ce sont les émissions les plus représentatives du secteur.

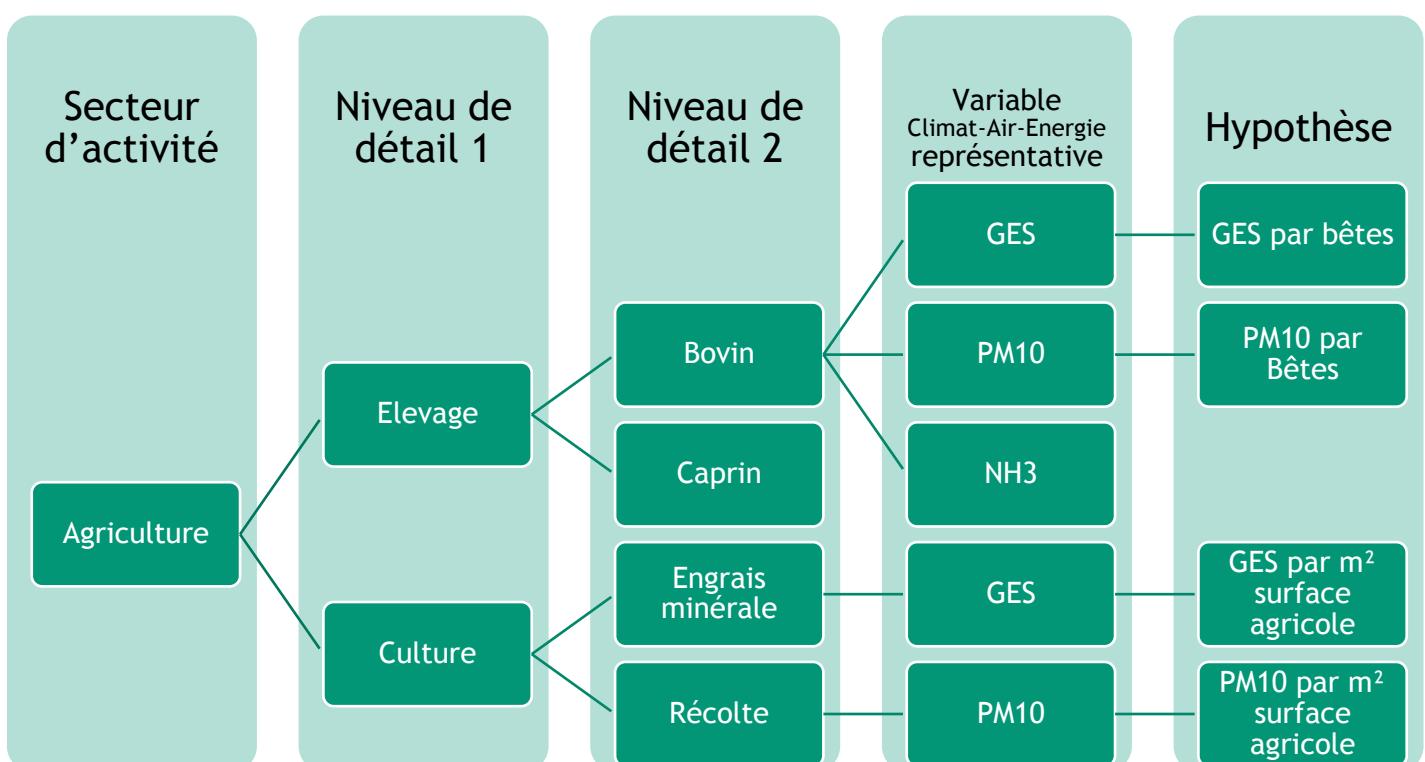


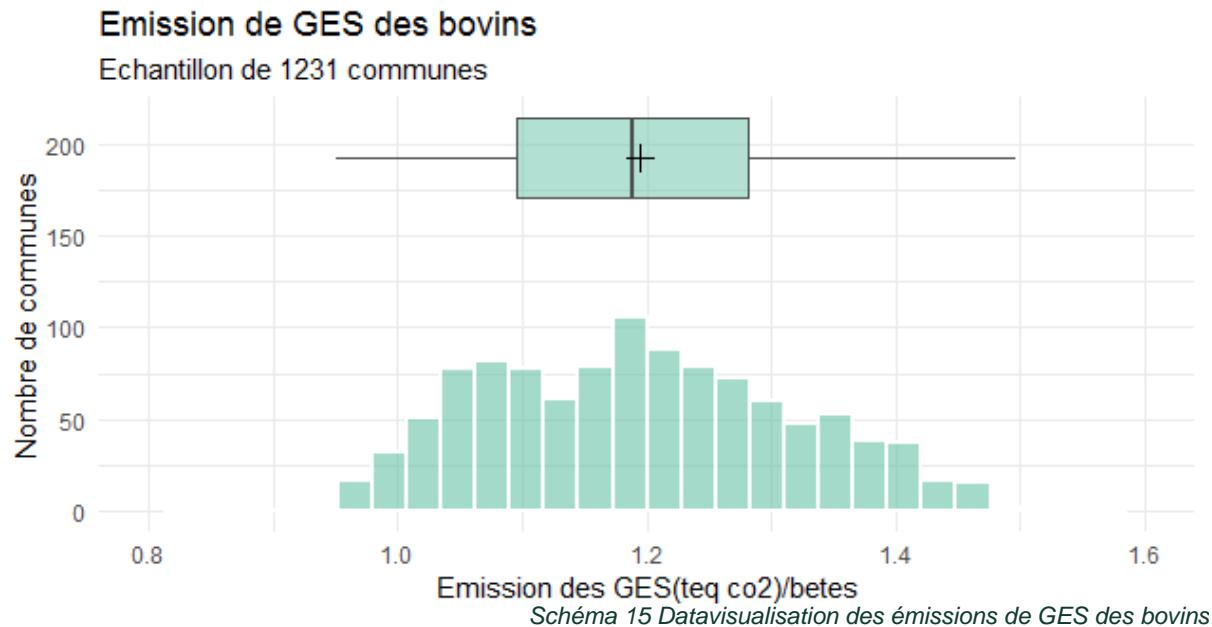
Schéma 14 : exemple de recherche du secteur agricole

Cette démarche permet d'établir un ou des indicateurs cohérents et pertinents représentatifs du secteur. Cet indicateur permet aussi de définir un comportement général des consommations d'énergie et des émissions de polluants et GES.

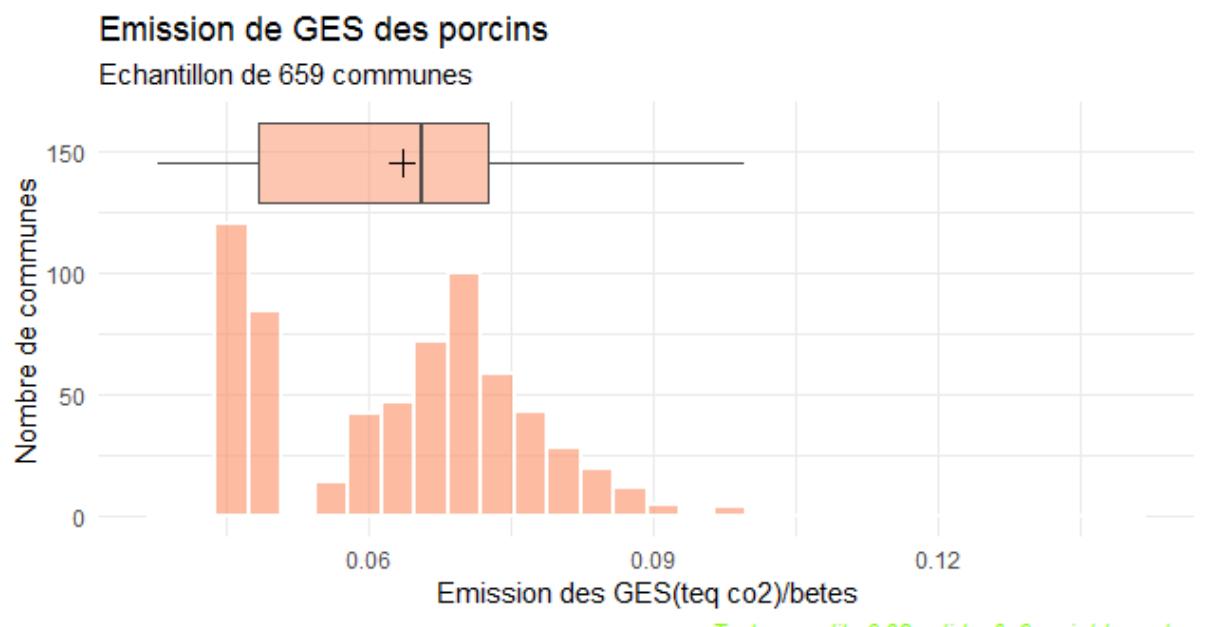
**Le Erreur ! Source du renvoi introuvable.** présente le résultat des émissions de GES des bovins dans le secteur agricole. On peut voir dans ce résultat que chaque bovin émet entre 0.95 et 1.5 teqCO<sub>2</sub>. Nous avons effectué un test de normalité sur l'ensemble de ces données afin de voir si celles-ci se rapprochent d'une loi normale.

L'hypothèse d'analyse est de dire que, les indicateurs mis en place rapprochent les distributions d'une loi normale. Hors une loi normale permet de décrire un comportement général des données.

L'objectif par la suite sera d'analyser les variables en s'éloignant de ce comportement général. Par exemple une commune ayant une émission de 2.5 teqCO<sub>2</sub> /bovin serait une variable éloignée de notre distribution de référence (allant de 0.95 à 1.5 teqCO<sub>2</sub> /bovin pour 99% des données), ce qui implique un questionnement sur cette valeur.



Cependant il faut faire attention à cette distribution. Si celle-ci s'éloigne d'une loi normale cela peut compromettre les analyses. Dans le cas ci-dessous (Schéma 16), la distribution est séparée en deux groupes de valeurs, ce qui peut être dû aux espèces de porcins pris en compte (Truies, porcs, ...). Une analyse sur les données est alors nécessaire pour se rapprocher d'une distribution normale stable. Parfois le niveau de détail de l'inventaire ne permet pas de fixer un indicateur stable.



### 3.3 Conception de l'indicateur de précision thématique

Les indicateurs doivent permettre de représenter aux mieux les enjeux du secteur. Les niveaux de détails qui caractérisent le secteur sont importants. Il faut alors définir un ou des indicateurs, cohérents avec le niveau de détails. C'est dans cet objectif de cohérence, de précision et de représentabilité que des réflexions ont été menées pour pouvoir construire les indicateurs pertinents au sein des secteurs analysés.

#### Indicateur Climat-Air-Energie

Il y a plus de 40 variables Climat-Air-Energie dans l'inventaire. Une analyse générale du secteur afin de savoir quelles sont les variables Climat-Air-Energie les plus représentatives et pertinentes permettent de cibler plus précisément un comportement.

Analyses de la consommation énergétique ?

Analyses des GES ?

Analyses des Polluants ?

#### L'échelle d'analyse

L'échelle d'analyse peut avoir un grand impact sur les résultats. Plus l'échelle sera précise, plus il sera possible de détecter un comportement anormal. Il faut alors savoir si les variables Climat-Air-Energie du secteur sont soumises à des facteurs spatio-temporels qui pourraient alléger ou amplifier un phénomène.

Est-ce qu'il y a une ventilation spatiale ?

Est-ce qu'il y a une ventilation temporelle ?

Est-ce qu'il y a des phénomènes temporels périodiques ?

#### Normalisation de la distribution

L'objectif de la normalisation est de permettre une comparaison entre des entités différentes. (*Consommation énergétique de Nantes 300 000 habitants et de Panon 50 habitants*). La normalisation ne s'est pas faite par la méthode standard<sup>6</sup>, mais par l'analyse d'une variable explicative. Cependant, cette normalisation peut être impactée par d'autre facteurs (*Logement vacant*). C'est l'ensemble de ses phénomènes qu'il faut cerner.

Sur quels éléments contextuels est basé le facteur d'émission ?

Est-ce que les tendances sont différentes en fonction d'un élément contextuel ?

Représentons-nous bien un phénomène logique ?

#### L'impact des méthodes de calculs

Les méthodes de calculs permettent de savoir quels sont les différents éléments pris en compte, ce qui permet d'identifier plus rapidement les variables explicatives. Cette méthode peut aussi nous permettre de savoir si les variables Climat-Air-Energie sont corrélées ensemble afin d'éviter des analyses inutiles.

Echange autour des méthodes de calcul et analyse des documents référents des méthodologies de l'inventaires

Test de corrélation

#### Validation de l'indicateur

Enfin, il est nécessaire de vérifier si le comportement de l'indicateur est suffisamment stable pour être analysé statistiquement.

Si celui-ci est stable il permettra d'identifier des cas particuliers qu'il faudra alors faire vérifier par les ingénieurs d'études pour valider la présence de réelles erreurs.

Distribution statistiques normales ?

Ecart interquartiles assez précis ?

Nombre d'outliers raisonnable et/ou cohérent ?

<sup>6</sup> Méthode de normalisation standard (Webographie)

### 3.4 Conception de l'indicateur qualité temporelle

L'inventaire BASEMIS® montre différentes problématiques liées à la qualité de l'évolution temporelle des émissions GES et polluants atmosphériques ainsi qu'à la consommation énergétique. Effectivement les rapports annuels ou encore les analyses des plans climats se basent sur ces évolutions pour prévoir, agir et fixer des objectifs énergétiques et climatiques. Le besoin de cohérence des évolutions est alors primordial.

Pour analyser et évaluer la qualité des évolutions temporelles de consommation d'énergie ou d'émissions de GES ou de polluants, plusieurs indicateurs temporels ont été construits et analysés (schéma ci-dessous). Chaque indicateur permet d'avoir une vision spécifique sur l'évolutions temporelles. Nous utilisons l'indicateur de précision thématique réalisé précédemment, pour analyser l'évolution temporelle.

<b>Taux d'évolution annuel</b>	$\left[ \frac{\text{Indicateur}_{\text{annee } N}}{\text{Indicateur}_{\text{annee } N-1}} \right] - 1$	Permet d'évaluer un comportement des émissions à l'échelle annuelle. Dépendant du comportement de l'année précédente
<b>Taux d'évolution global</b>	$\left[ \frac{\text{Indicateur}_{\text{annee } N}}{\text{Indicateur}} \right]$	Permet d'évaluer la tendance d'évolution générale sur l'année. Sur les émissions sensibles au climat la précision peut être variable voir approximative
<b>Taux d'évolution référent</b>	$\left[ \frac{\text{Indicateur}_{\text{annee } N}}{\text{Indicateur}_{\text{annee } 2008}} \right]$	Permet d'évaluer l'évolution par rapport à une référence ciblée Dépend de l'année de référence
<b>La différence annuelle</b>	$[\text{Indicateur}_{\text{annee } N}]$	Permet de mesurer un écart des émissions sur une annuée. Sur les émissions sensibles au climat les résultats sont variables ce qui rend l'interprétation difficile.
<b>La différence à la référence</b>	$[\text{Indicateur}_{\text{annee } N}]$	Permet de mesurer une baisse d'émissions depuis la mise en place d'une action.
<b>La moyenne des évolutions annuelles</b>	$\text{taux annuelle}$	Permet d'évaluer la tendance d'évolution globale sur l'ensemble des années

Parmi les indicateurs, seuls les taux d'évolution annuel et le taux d'évolution global permettent de représenter une évolution annuelle précise (le taux d'évolution référent peut être biaisé par l'année de référence choisie).

Le taux d'évolution annuel permet d'évaluer la stabilité des comportements d'une année par rapport à l'année précédente. Ce qui permet d'analyser un comportement d'évolution annuel. Les facteurs climatologiques ne perturbent pas les analyses, car la tendance d'évolution sera la même pour l'ensemble des données. (Si l'année précédente présente plus de pic de froid alors la tendance générale des émissions montrera une diminution) ce qui n'est pas le cas du taux d'évolution global.

Cependant Le taux d'évolution annuel dépend de l'année antérieure, ce qui expose deux problématiques :

1. Si l'année antérieure présente une anomalie, le taux d'évolution peut être impacté. Il est susceptible d'être considéré en dehors de la distribution à l'année N alors que l'erreur provient de l'année N-1.
2. Si une consommation ou émission annuelle ne représente pas l'intégralité de l'année (ajout d'un chauffage en milieu d'année donc présence de la consommation d'une demi-année) alors la détection de ce cas particulier est systématique.

## 3.5 Détection d'anomalie

Une fois les indicateurs définis avec une distribution, notre objectif est de détecter les **valeurs aberrantes** aussi appelées « outliers ». Ces valeurs qui s'éloignent du comportement général sont soit des cas particuliers dont les ingénieurs d'études ont besoin d'avoir connaissance, soit de réelles erreurs liées aux fichiers sources des partenaires fournisseurs de données.

### 3.5.1 Les tests de détection de valeur aberrantes

La détection de valeurs aberrantes peut se faire de plusieurs manières. Le Schéma 17 présente les différentes méthodes de détection d'outliers.

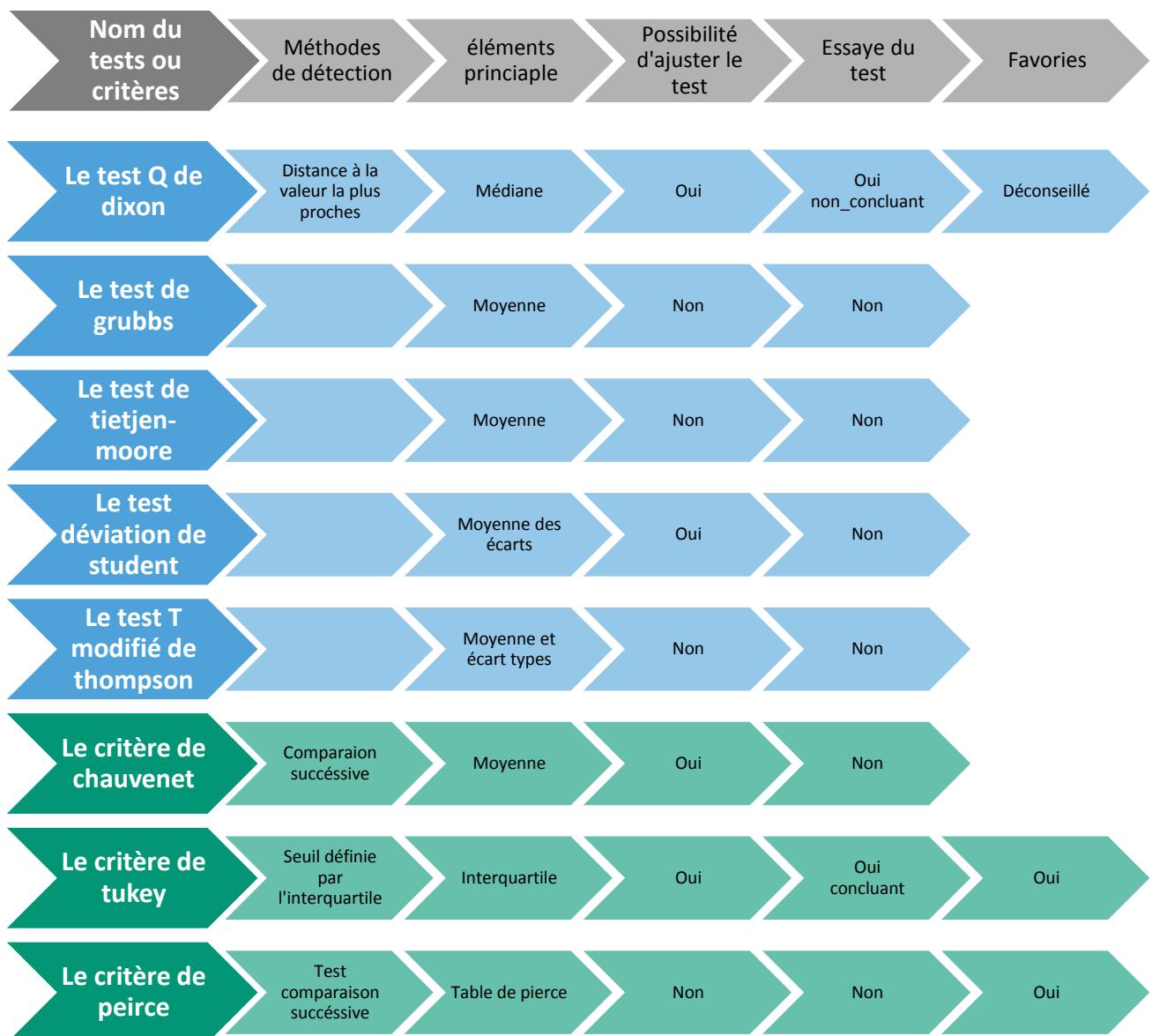
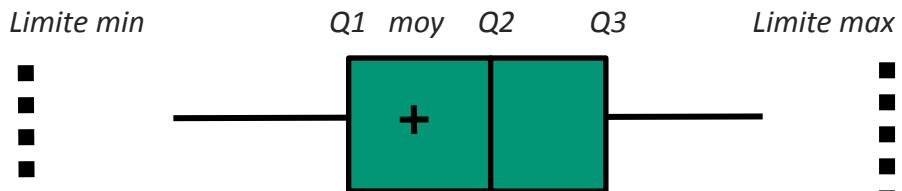


Schéma 17 Récapitulatif des différentes méthodes de détections d'outliers

### 3.5.2 Le critère de Tukey

Après une analyse des méthodes de détection d'outliers, le critère de Tukey a été retenu pour cette étude.

Le critère de Tukey détermine un seuil minimum et maximum qui est calculé à partir de l'écart interquartile. Il possède l'avantage de pouvoir déterminer un seuil à partir duquel les observations, en fonction de leur valeur, doivent être considérées comme aberrantes relativement à la distribution de  $X$ , permettant ainsi la détection d'un ou plusieurs outliers simultanément.



$$\text{Limite min} = Q1 - 1.5 \times (Q3 - Q1)$$

$$\text{Limite max} = Q3 + 1.5 \times (Q3 - Q1)$$

Schéma 18 Représentation graphiques du test de Tukey

L'application du critère de Tukey pour les besoins du projet a permis les observations suivantes :

- Le test s'effectue rapidement (les tests qui utilisent des récurrences sont longs)
- Les résultats détectent des anomalies ce qui s'avère concluant.

Suite à ces observations, ce test a été retenu pour l'ensemble des analyses.

#### Consommation d'énergie de Gaz tout usages (cuisson, eau chaude, ...)

Echantillon de 406 communes

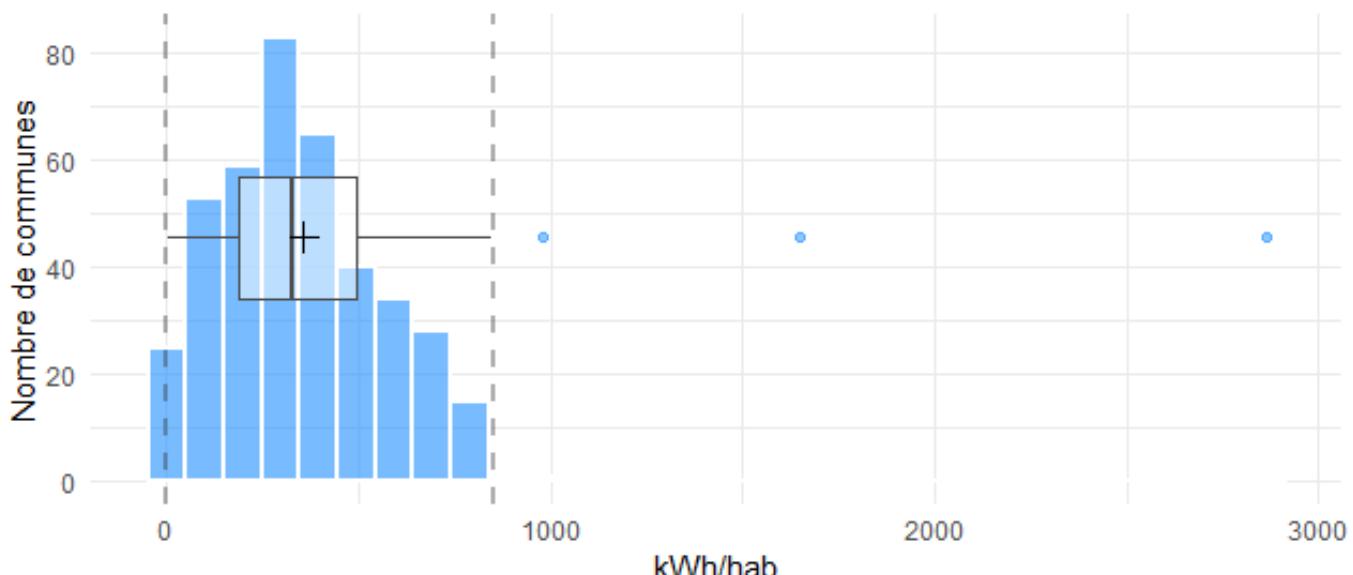
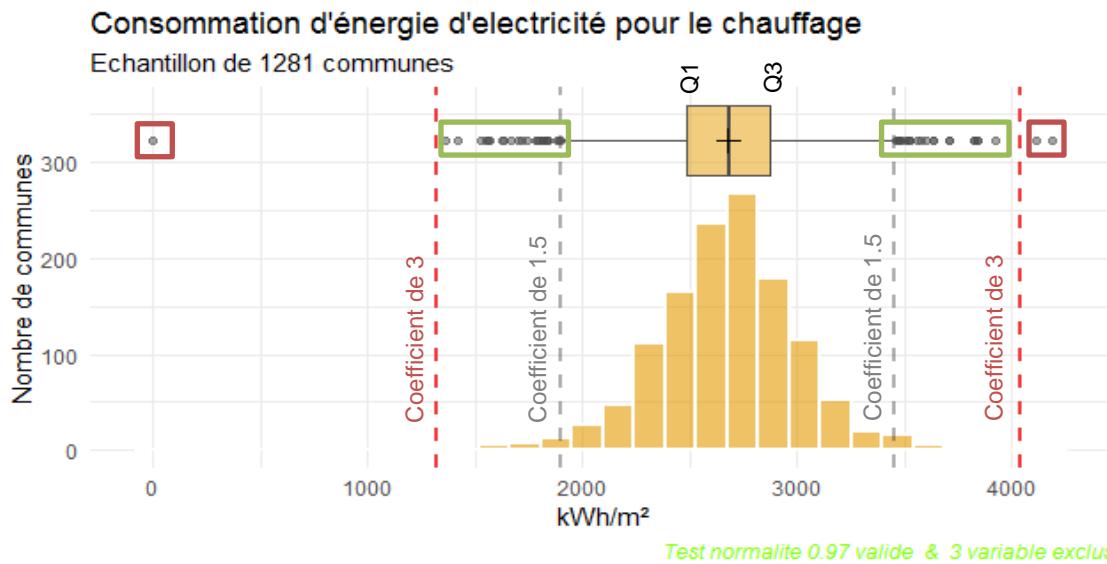


Schéma 19 Visualisation du test de Tukey sur l'indicateur de précision thématique du secteur résidentiel (kWh/hab.) sur les consommations d'énergies en gaz

Cependant le test présente des imperfections. Si la distribution a un faible écart interquartile, la valeur de notre limite min et max sera proche de Q1 et Q3 alors que l'amplitude min et max des données peut être plus large. Ce qui peut parfois exclure une série de valeurs en fin de distribution. (Valeurs encadrées en vert)



Comme le montre le schéma ci-dessus, une série de données est détectée en anomalie alors que ces valeurs sont correctes (points encadrés en vert).

Afin d'éviter cette problématique, une solution mise en œuvre a été la modification du coefficient de 1.5 (trait gris) en fixant le coefficient à 3 (trait rouge). On augmente ainsi notre critère de détection d'anomalies. Les valeurs détectées (valeurs encadrées en rouge) sont des cas anormaux dont la détection est nécessaire.

Cette modification du test de Tukey a fait l'objet d'une analyse (Schéma 20).

### 3.5.3 Approches et variation du test de Tukey

- Approche gestion de projet :** Il est possible de choisir le coefficient du test pour ressortir un nombre d'erreurs fixe. Avec cette méthode, l'objectif est de cibler les X valeurs les plus éloignées du comportement global afin de les corriger.
- Approche statistique :** Il est possible de détecter à partir de quel coefficient le nombre d'erreurs se stabilise. Ceci permet d'assurer que les valeurs ressorties appartiennent bien à des valeurs anormales.
- Approche arbitraire :** la méthode de base propose un coefficient à 1.5. Un seuil fixe permet d'établir une certaine stabilité du test pour assurer une fiabilité des résultats.

L'approche arbitraire a été retenue avec un ratio de 3. Ce ratio plus élevé que le coefficient proposé par la méthode permet de faire ressortir uniquement les données ayant un comportement anormal. Il a été testé sur de nombreuses distributions, les résultats ont permis de faire ressortir des anomalies dans la plupart des cas.

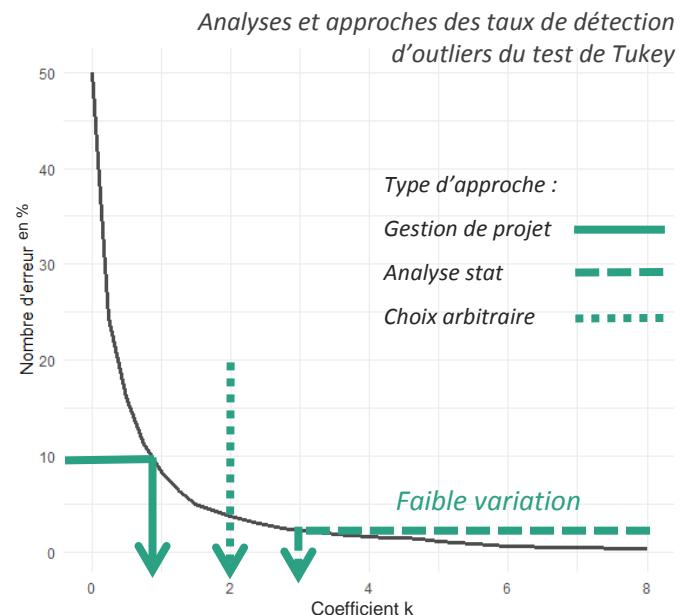


Schéma 20 Graphique d'analyse sur les données résidentielles

## 3.6 Indicateur de qualité des données

Pour résumer, nos critères d'évaluation de qualité des données ont été ciblés sur deux notions clés : la précision thématique et la qualité d'évolution temporelle. Nous avons abordé ces notions à l'aide de la création d'indicateurs de précision thématique et d'évolution temporelle spécifiques à chaque secteur pour apporter un diagnostic à l'échelle la plus fine. Grâce au test de Tukey, nous avons obtenu les valeurs aberrantes de ces deux indicateurs.

*Que faire du résultat de ses valeurs aberrantes ? Comment évaluer ce résultat ?*



Schéma 21 Processus d'analyse de la qualité des données

Les échanges avec les inventoriastes d'Air Pays de la Loire ont permis d'identifier deux attentes principales des travaux :

- Le besoin de détecter des anomalies dans la base de données afin de les corriger,
- Le besoin d'évaluer de manière générale la qualité des données de l'inventaire.

A travers ces deux finalités, deux approches ont été établies : l'évaluation de la qualité générale et le diagnostic des valeurs aberrantes.

**Pour évaluer la qualité générale des données** à partir des valeurs aberrantes détectées, trois raisonnements sont fondamentaux :

1. **Le nombre total de valeurs aberrantes détectées :**

*Combien de valeurs sont détectées ?*

$$NB \text{ erreur} = \text{count}(\text{valeur\_aberrantes})$$

2. **Le taux de valeur aberrantes présentes :**

*Est-ce que le nombre de valeurs détectées est significatif dans l'ensemble des données ?*

$$\text{Taux d'erreur} = \frac{NB \text{ erreur}}{\text{Totales des données}}$$

3. **L'amplitude des erreurs :**

*Est-ce que les valeurs détectées sont des erreurs minimes ou importantes ?*

$$\text{écart max} = \max\left(\frac{\text{valeur\_aberrantes} - \text{seuil}}{\text{seuil du critère de tukey}}\right)$$

**Pour diagnostiquer les valeurs aberrantes**, il est important de rétablir le contexte qui est associé à ces données. En fonction de chaque secteur les variables, les indicateurs, les méthodes de calcul, les niveaux d'analyse sont différents. Cependant il est nécessaire de visualiser :

- Le comportement général de l'indicateur de précision thématique et de qualité temporelle. (*Histogramme + boxplot + seuils établis*)
- Le comportement spatial de l'indicateur étudié (*cartographie*)
- Les éléments contextuels liés à la valeur aberrante (*nombre d'habitants, surface agricole, consommations d'énergie, émissions de GES et de polluants à l'atmosphère, ...*)

## 3.7 Outils de reporting

Ce stage vise également à développer des outils de traitement automatisés. Deux outils ont permis de répondre aux deux approches de l'évaluation de la qualité des données.

D'une part un rapport Rmarkdown, qui explique le comportement des données, avec l'ensemble des outils de data visualisation, cartographie et tableaux nécessaires à diagnostiquer un secteur, sur une version de l'inventaire et pour une année précise de l'inventaire.

D'autre part une application Shiny, qui permet de visualiser de manière interactive, les différents critères d'évaluation de la qualité des données sur l'ensemble des années, des secteurs et des versions de BASEMIS®. L'application permet de choisir l'année, le secteur et la version d'inventaire pour lancer un rapport d'analyse détaillé.

(Annexes pages 59)

### 3.7.1 Application surveillance de la qualité des données globale

L'application est composée de deux parties. Le premier encadré bleu permet de modifier les paramètres de visualisation. La deuxième partie encadrée en vert permet quant à elle de visualiser les résultats de manière interactive. Dans cette deuxième partie un menu de 4 onglets contenant des informations différentes pour chaque secteur est mis en place (menu, Information générale, ...).

Analyses de la qualité des données BASEMIS

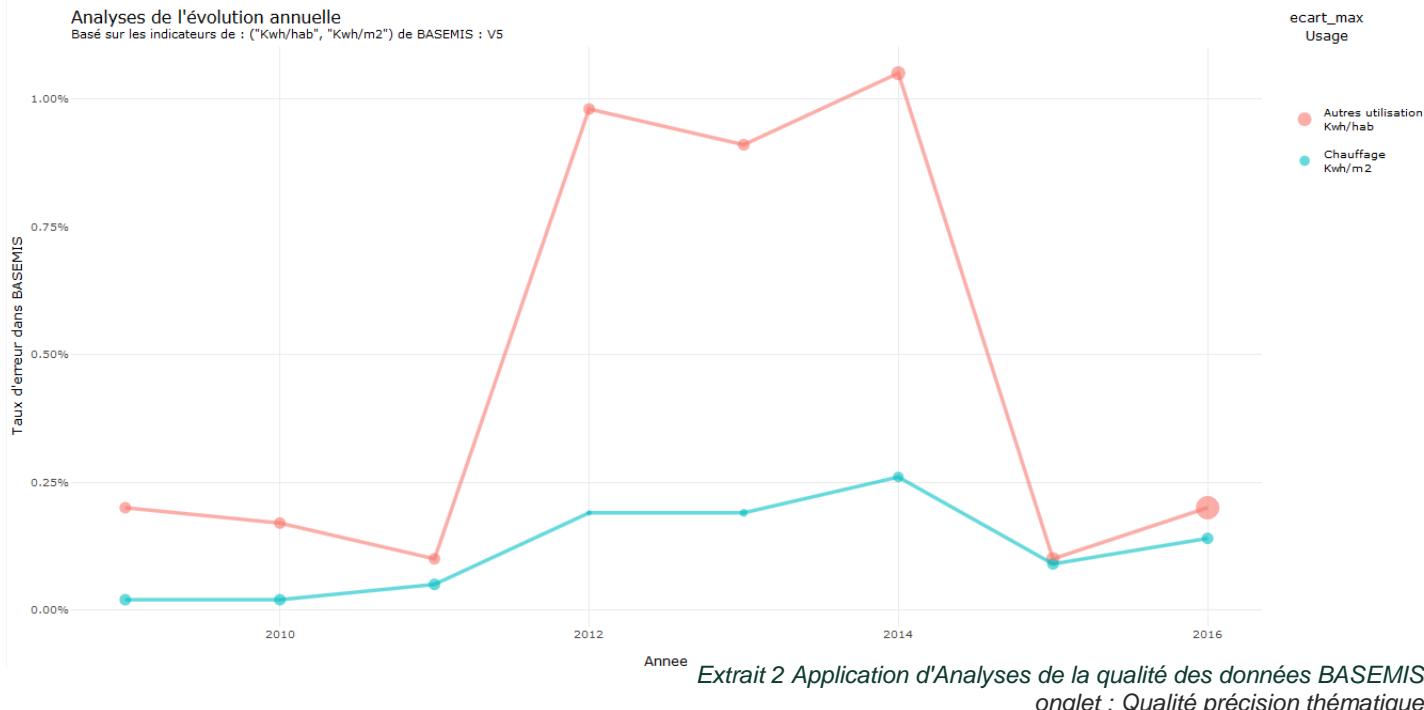
*Extrait 1 Application d'Analyses de la qualité des données BASEMIS  
Onglet : Information générale*

1. **Information générale :** Ce panel textuel (Extrait 1), permet de contextualiser les différentes informations nécessaires à l'utilisation de l'application. Une partie explique de manière générale nos indicateurs de qualité de la donnée, puis les variables et méthodes de calcul des indicateurs spécifiques au secteur d'activité sélectionné.

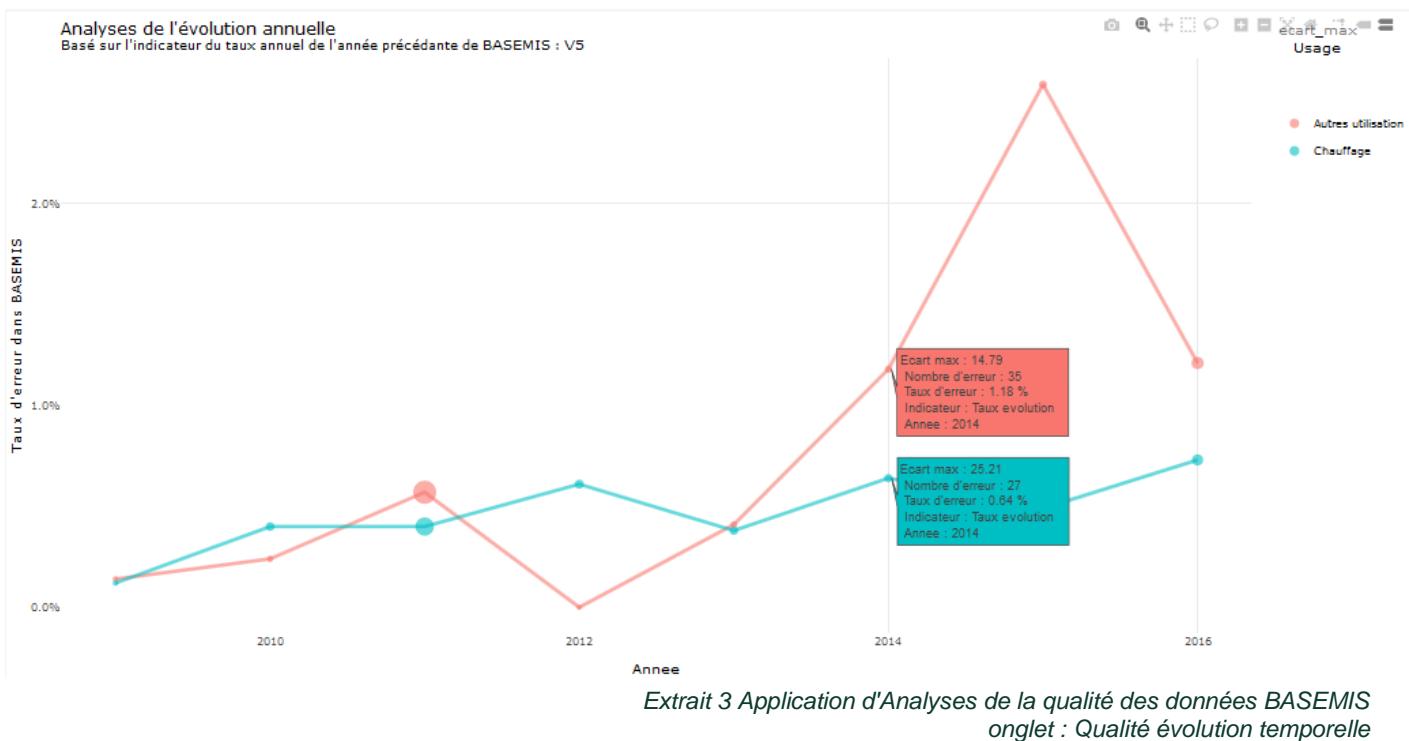
# Qualité des données

2. **Qualité précision thématique** : Ce panel de data visualisation (Extrait 2) permet de voir le taux d'anomalies détectées dans la base de données sur l'indicateur de précision thématique. Il nous permet aussi de voir l'amplitude des erreurs maximum chaque année à travers la largeur des points. Grâce à son interactivité il est possible de zoomer, se déplacer ou encore filtrer certaines variables.

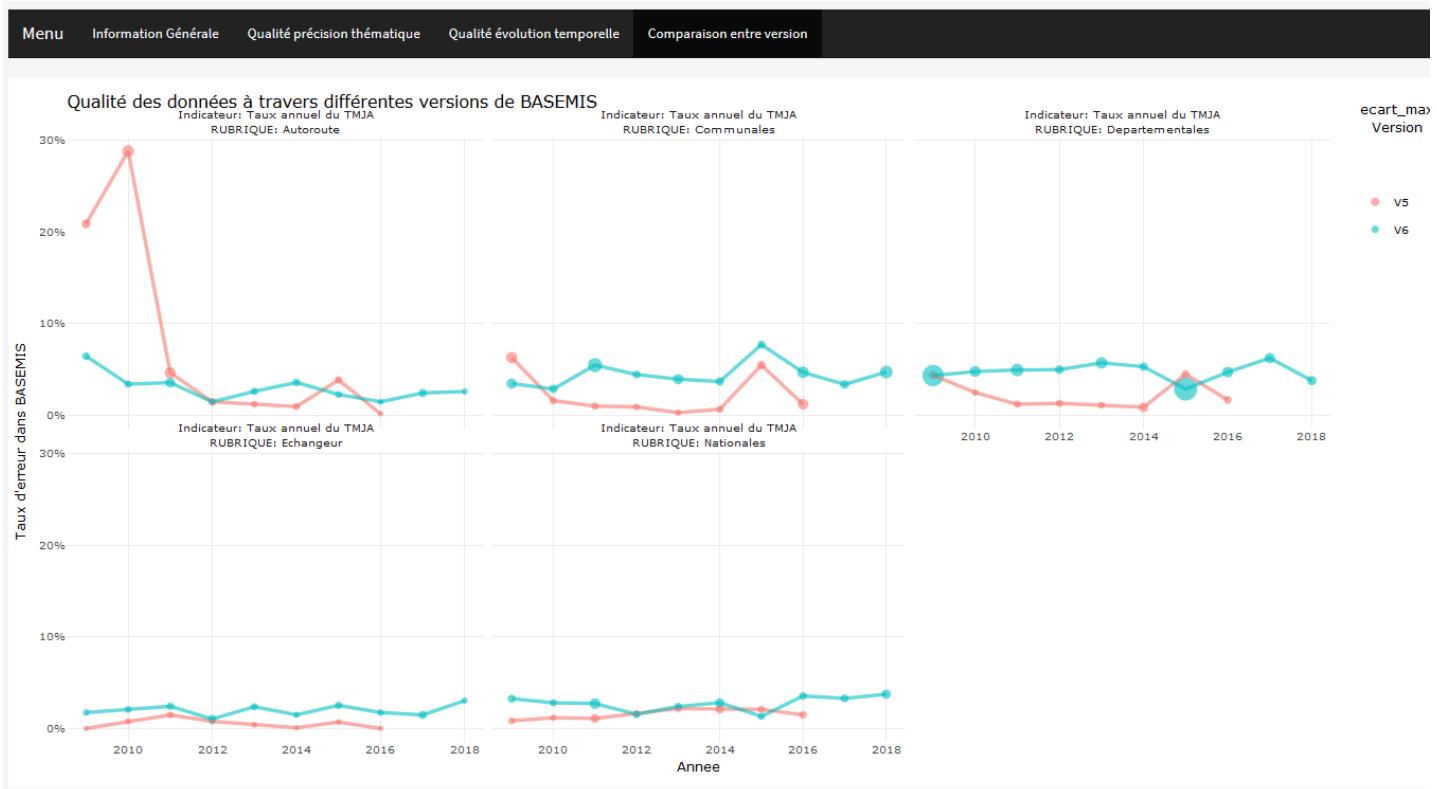
Ce panel permet d'avoir un premier aperçu du comportement de la qualité de la donnée sur la précision thématique. Cet aperçu permet de visualiser rapidement les comportements inhabituels (taux d'erreurs très important sur une année). Il permet aussi de voir la présence d'une erreur très éloignée des limites définies.



3. **Qualité évolution temporelle** : Similaire au panel précédent, une data visualisation interactive des taux de valeurs aberrantes présent dans les comportements d'évolution annuel. L'objectif ici est de voir si la qualité de l'évolution temporelle suit un comportement stable.



4. **Comparaison entre versions** : le dernier panel permet de visualiser de façon distincte pour chaque rubrique du secteur le comportement qualité entre les versions. Cette partie rassemble l'intégralité des indicateurs de précision thématique et d'évolution temporelle d'un secteur.



Extrait 4 Application d'Analyses de la qualité des données BASEMIS  
onglet : Comparaison entre version

A travers cette datavisualisation, l'objectif est de percevoir les améliorations ou les dégradations de la qualité des données d'une nouvelle version d'inventaire. Sur l' Extrait 4 ci-dessus, les données de la version 6 sont dans l'ensemble de moins bonne qualité (le taux d'erreur est plus important que la version 5), ce qui s'explique par le fait que les données n'ont pas encore été retravaillées et validées contrairement à la version 5 de l'inventaire BASEMIS.

Enfin l'application dispose d'un cinquième panel qui permet de générer un rapport détaillé en fonction des variables sélectionnées.

**Choix du secteur**

**Choix de l'année**

**Version Basemis**

### 3.7.2 Rapport de la qualité des données sur une thématique spécifique

Le rapport de qualité des données, permet de diagnostiquer de façon précise le comportement des indicateurs d'un secteur spécifique pour une année choisie et une version de l'inventaire choisie. Ce diagnostic se présente en 4 parties distinctes :

#### Partie 1 : Informations sur le secteur

## Analyses des données énergétiques BASEMIS du secteur résidentiel

23 juillet, 2020 - Air Pays de la Loire

- 1 - Le secteur résidentiel : Les indicateurs
- 2 - Résultats de l'année 2013: Comportements des indicateurs
  - 2.1 - Consommation d'énergie liée à la cuisson, l'eau chaude sanitaire et prises électriques
  - 2.2 - Consommation d'énergie liée au chauffage
- 3 - Résultats des valeurs abberantes
- 4 - Qualité des données

### 1 - Le secteur résidentiel : Les indicateurs

#### 1.1 Information sur les types de données

La base de données BASEMIS représente les données finales de l'inventaire, ses données sont calculées à partir des données INSEE, CITADEL. Pour vérifier la cohérence des données nous allons principalement utiliser les données suivantes

- Consommation (en kWh)
- Type d'énergie utilisé (gaz, électricité, ...)
- Type de résidence (principale, collective, ...)
- Nombre d'habitant par résidence
- Superficie des résidences (en m<sup>2</sup>)

#### 1.2 Méthode de calcul des indicateurs

Au sein de ce secteur, nous avons mis en place trois indicateurs permettant de voir le comportement général des consommations d'énergies :

$$Kwh/hab = \frac{EnergieConsommé}{NombreHabitant}$$

$$Kwh/m^2 = \frac{EnergieConsommé}{SurfaceLogement}$$

$$TauxAnnuelle = \left[ \frac{Kwh/hab_{anneeN}}{Kwh/hab_{anneeN-1}} - 1 \right]$$

#### 1.3 Méthode de traitements des données :

Jointures des données par :

- Numcom
- Année
- Rubrique

Point de vigilances à vérifier :

- Présence de gaz / réseaux de chaleur dans la ville possible ?

## [1] "Consommation possible valide"

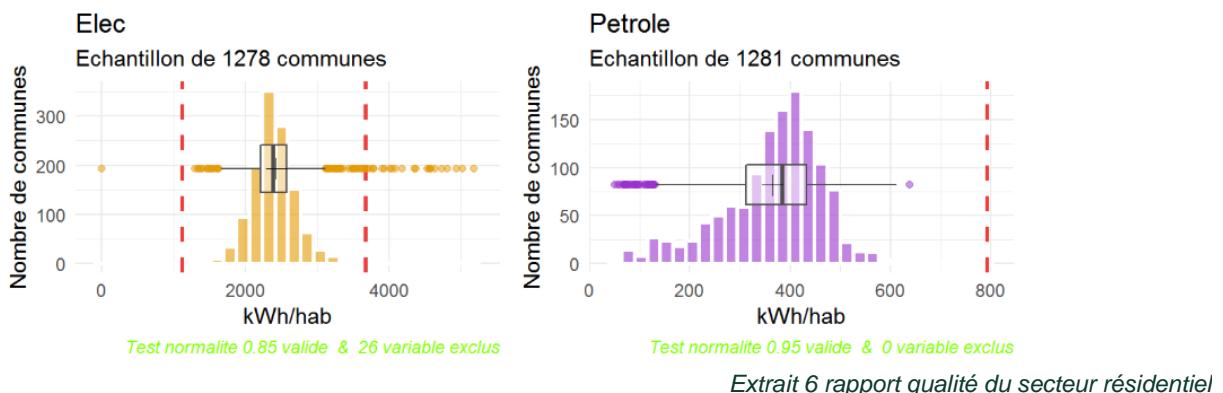
Extrait 5 rapport qualité du secteur résidentiel

Comme le montre l'exemple ci-dessus (Extrait 5), cette partie présente les informations utiles à l'analyse du secteur. En préambule de cette partie, sont indiqués le titre du rapport, avec la date de réalisation et le sommaire des différentes parties du rapport.

Dans un premier temps la partie d'information sur le secteur présente un listing des champs de données utilisées avec une précision des unités et des valeurs (1.1). Puis les formules de calcul de l'indicateur (1.2). Enfin les champs de jointures utilisés pour la jointure de différentes tables de données ainsi que certain test de cohérence logique des données (1.3). Ici la vérification permet de voir si les communes ayant une consommation d'énergie en gaz ont bien un réseau de gaz associé

## Partie 2 : Data visualisation des données

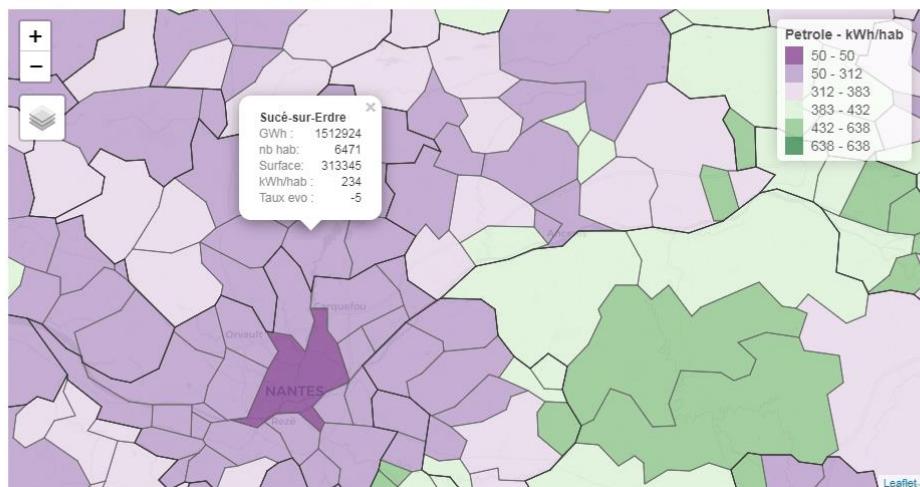
Graphique 1 : Distribution des consommations d'énergies en 2013



Dans cette deuxième partie (Extrait 6 rapport qualité du secteur résidentiel) sont regardés les différents histogrammes pour visualiser le comportement global des données, vérifier que la distribution se rapproche d'une loi normale, afficher les limites de détection d'anomalie et ainsi valider notre analyse.

Pour accompagner l'histogramme, une carte interactive (Extrait 7) est insérée. Elle permet de représenter spatialement la donnée étudiée. Grâce à son interactivité, il est possible remettre des informations contextuelles et les informations nécessaires à la compréhension des données. Ainsi lors du clic sur une commune il est possible de visualiser les informations contextuelles du secteur. Dans l'extrait 7 le pop-up lors du clic permet de voir le nom de la commune, sa consommation, le nombre d'habitants, la surface totale des logements,

Cartographie 1 : Cartographie des consommations d'énergies en 2013  
choix des indicateurs à cocher/décocher dans la gestion des couches



Extrait 7 rapport qualité du secteur résidentiel

l'indicateur kWh/hab et le taux d'évolution annuel ?).

La représentation sémiologique se base sur la distribution de l'indicateur de la façon suivante :

## Partie 3 : Tableau des valeurs anomalies détectées

### 3 - Résultats des valeurs aberrantes

**Table 1 :** Communes dont l'indicateur semble anormale

Nom de la commune	ANNEE	NAPFUE	Energie	indice	ratio
<b>Chaussage</b>					
SAINTE-HILAIRE-LA-FORET	2016	Gaz	9948559.47 kWh	266 kWh/m2	1.7
MOUILLERON-SAINT-GERMAIN	2016	Gaz	18623500.5 kWh	244 kWh/m2	1.6
NESMY	2016	Gaz	22876168.49 kWh	183 kWh/m2	1.2
SENONNES	2016	Elec	236047.9 kWh	15 kWh/m2	1
NOIRMOUTIER-EN-L'ILE	2016	Elec	3547138.93 kWh	16 kWh/m2	1
NAUVAY	2016	Elec	1.66 kWh	0.0024 kWh/m2	0

**Table 2 :** Communes dont l'évolution semble anormale

Nom de la commune	ANNEE	NAPFUE	Energie	indice	ratio
<b>Autres</b>					
NESMY	2016	Gaz	4663116.83 kWh	2723 %	131.8
MOUILLERON-SAINT-GERMAIN	2016	Gaz	5252690.31 kWh	2054 %	99.4
MONTREVAULT-SUR-EVRE	2016	Gaz	799496.39 kWh	779 %	37.7
BEAULIEU-SOUS-LA-ROCHE	2016	Gaz	427707.03 kWh	622 %	30.1

*Extrait 8 rapport qualité du secteur résidentiel*

Dans cette troisième partie (Extrait 8), un tableau qui liste l'ensemble des valeurs détectées comme aberrantes a été mis en place. Ce tableau donne lui aussi des valeurs contextuelles afin de comprendre si une spécificité explique le comportement anormal de notre indicateur. De plus ce tableau est ordonné par l'amplitude des erreurs (ratio), ce qui permet de visualiser prioritairement les erreurs s'éloignant le plus de notre modèle.

## Partie 4 : Récapitulatif des indicateurs de qualité

### 4 - Qualité des données

**Table 3 :** Estimation de la qualité des données

ANNEE	Indicateur kWh/m2			Taux d'évolution %		
	Taux err	NB err	Ecart max	Taux err2	NB err2	Ecart max2
<b>Autres</b>						
2016	0.2 %	6	7.0	1.21 %	36	131.8
<b>Chaussage</b>						
2016	0.14 %	6	1.7	0.73 %	31	96.3

*Extrait 9 rapport qualité du secteur résidentiel*

Pour conclure ce diagnostic, les indicateurs de qualité pour chaque catégorie d'analyse sont représentés (Extrait 9) afin de pouvoir estimer l'importance et l'impact de ces possibles erreurs sur l'inventaire.

C'est à travers l'ensemble de ces explications, visualisations et contextualisations qu'un inventariste peut s'appuyer pour analyser le comportement des données du secteur. Ceci lui permettra d'appréhender les incohérences, les erreurs, les cas spécifiques et particulier pour estimer la qualité des données de l'inventaire BASEMIS®.

## 4 Analyses thématiques

### 4.1 Les différents secteurs de l'inventaires BASEMIS®

#### 4.1.1 Les secteurs représentatifs de l'inventaire

L'inventaire BASEMIS® comprend 9 secteurs d'activité. Tous ces secteurs ont un impact sur la qualité de l'air de manière directe ou indirecte (Schéma 22).

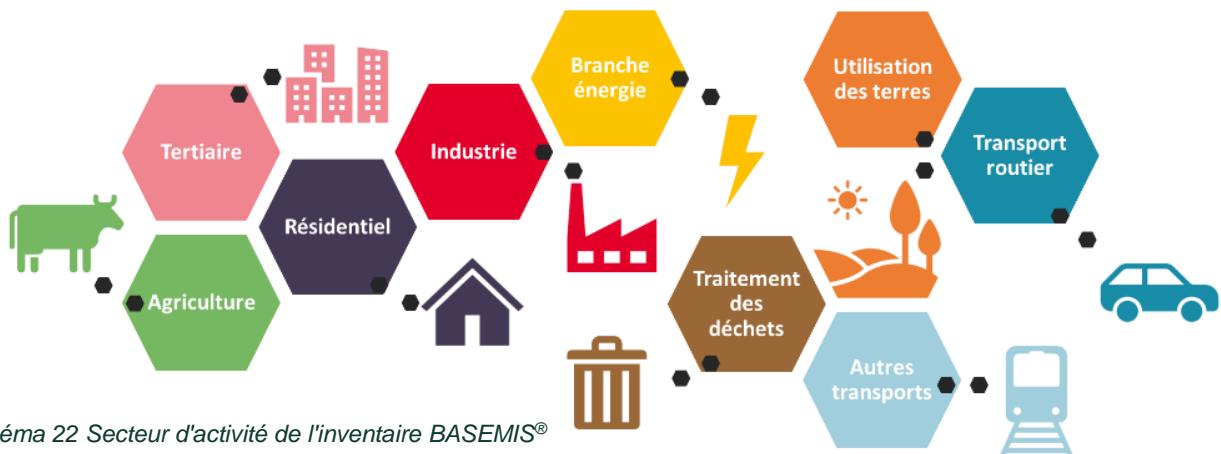


Schéma 22 Secteur d'activité de l'inventaire BASEMIS®

Lors de la prise en main de l'inventaire BASEMIS®, une première approche (GES, Polluants, Energie) a permis d'identifier les secteurs les plus représentatifs de l'inventaire (Schéma 23).

Les secteurs résidentiel, industriel et routier sont des secteurs représentatifs des consommations d'énergies, émissions de GES et polluants atmosphériques de l'inventaire. Le secteur agricole quant à lui est un secteur fortement émetteur de GES et d'ammoniac (NH3).

Le stage s'est focalisé sur l'analyse des secteurs Résidentiel, Transport routier et Agriculture. L'industrie étant un sujet complexe qui nécessite un regard particulier.

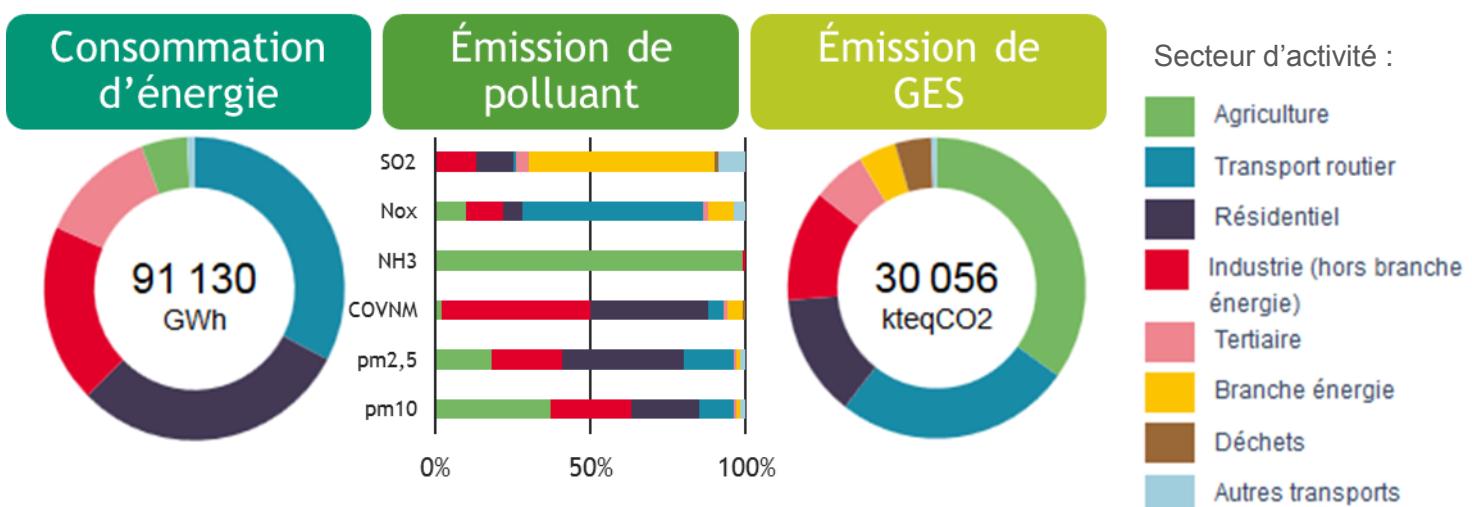


Schéma 23 Data visualisation Climat-air-Energie pour l'année 2016 – source BASEMIS® V5

## 4.2 Secteur Résidentiel

### 4.2.1 Analyses de l'impact Climat-Air-Energie



Le secteur résidentiel est le deuxième secteur le plus consommateur d'énergie. Les logements de la région ont en effet consommé en 2016 pas loin de **27 200 GWh** d'énergie. Par ailleurs, les émissions de gaz à effet de serre du secteur s'élèvent à **3 100 kteqCO<sub>2</sub>**. Ce qui fait du secteur résidentiel, l'un des secteurs représentatifs des données Climat-Air-Energie.

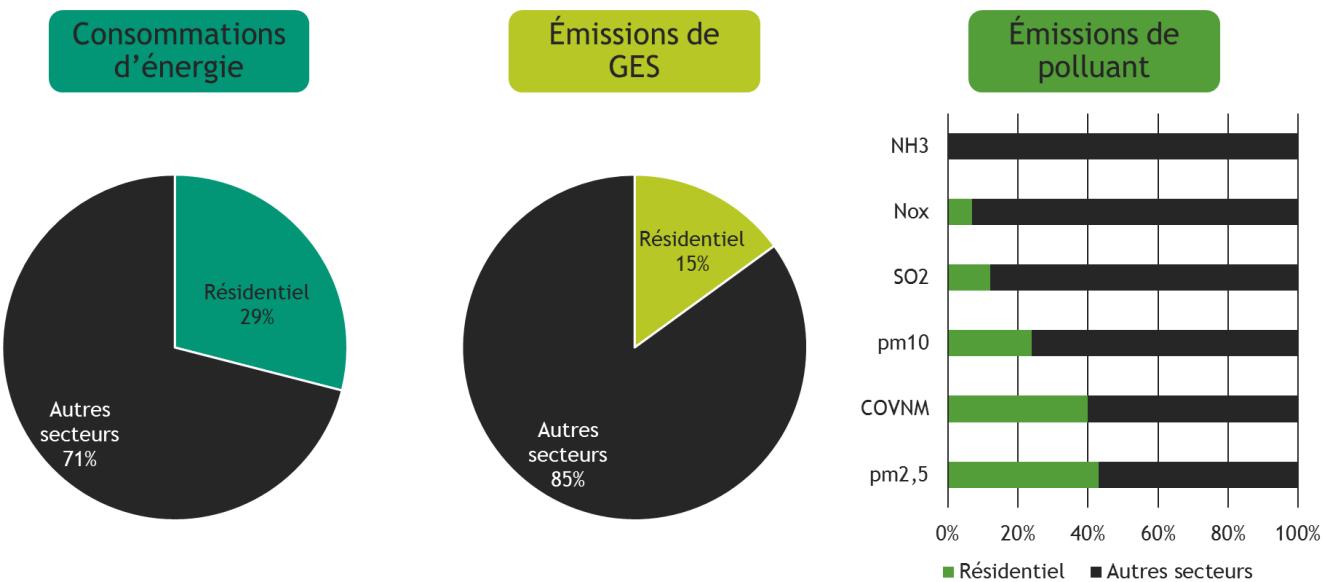


Schéma 24 Impact Climat-Air-Energie du secteur résidentiel – source BASEMIS® V5

### 4.2.2 Méthodologie de calcul des émissions

Le Schéma 25 permet de représenter la méthodologie de calcul des données Climat-Air-Energie du secteur résidentiel. Nous pouvons voir que celle-ci se base notamment sur les données de logement et de population afin d'estimer les consommations d'énergie qui serviront par la suite de référence pour le calcul des émissions de GES et de polluants.

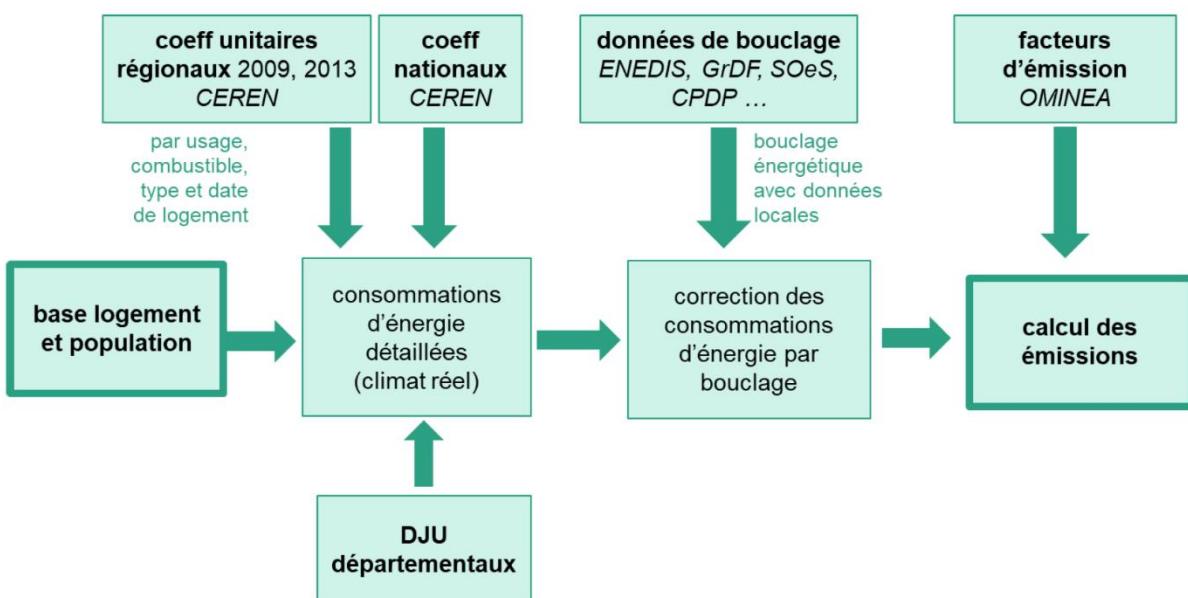


Schéma 25 méthodologie de calcul des consommations et émissions dans les bâtiments du secteur résidentiel



La détermination des d'émissions est directement dépendante (corrélées) des données de consommations d'énergie du secteur. Les facteurs d'émission du guide OMINEA, constituent une référence nationale et ne peuvent être modifiés. C'est pourquoi l'analyse du secteur résidentiel se fera par l'analyse des consommations d'énergie qui dépendent de trois variables :

- Le type de logement
- L'usage de l'énergie
- Le vecteur énergétique

L'analyse des types de logements ([Erreur ! Source du renvoi introuvable.](#)) permet de cibler uniquement les logements principaux qui représentent plus de 90% du parc de logements de la région. Un filtre sur les consommations d'énergie des logements principaux est mis en place.

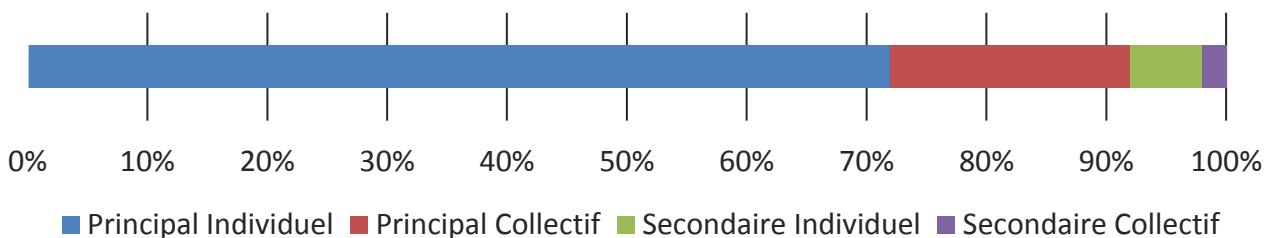


Schéma 26 Analyses des types de logements en Pays de la Loire en 2016 – source BASEMIS® V5

Concernant les usages et les sources d'énergie dans le secteur résidentiel (Schéma 27), l'usage principal est le chauffage. Cependant l'ensemble des autres usages reste non négligeable.

Pour les sources d'énergies, chaque forme d'énergie est représentée, même si la chaleur de réseau ne représente que 2 % des consommations d'énergie du secteur.

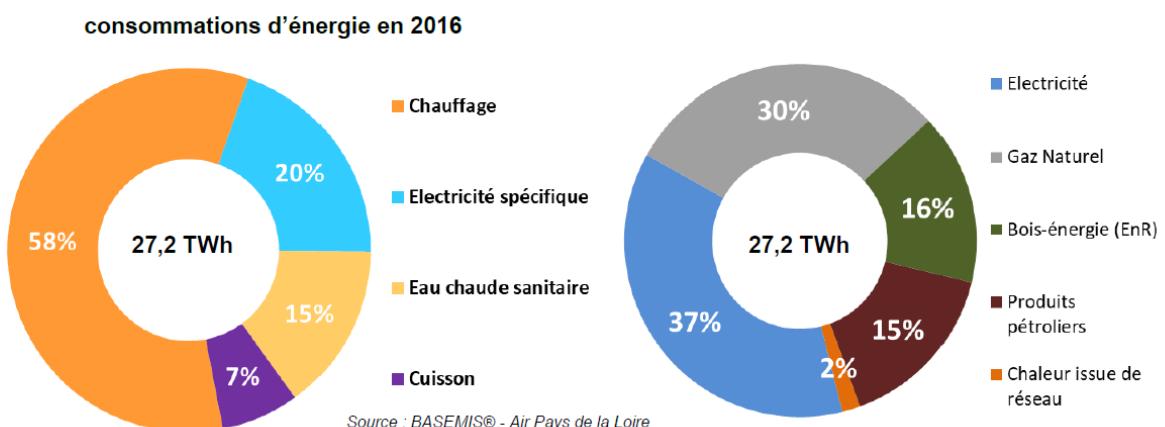


Schéma 27 Consommations d'énergie en 2016 par usage (gauche) et source d'énergie (droite)  
– source BASEMIS® V5

Il faut maintenant trouver la méthode de normalisation qui permettra d'analyser la consommation d'énergie à travers ces facteurs.

#### 4.2.3 Normalisation de la distribution



Grâce à la méthode de calcul, nous savons que les consommations d'énergie sont influencées par le nombre d'habitants, les types de logements, les usages de l'énergie et les sources d'énergies utilisées.

Deux indicateurs permettent de normaliser la distribution :

- Les kWh/habitant pour les usages énergétiques de type : électricité spécifique, eau chaude sanitaire et cuisson. Ces usages énergétiques varient en fonction du nombre de personnes occupant le logement.
- Les kWh/m<sup>2</sup> pour le chauffage. Dans ce cas-là, ce sont principalement la surface du logement, le mode de chauffage ou l'isolation qui vont influencer la consommation d'énergie.

En fonction de la source d'énergie utilisée (électricité, Gaz, Bois, ...) la consommation d'énergétique sera différente. A titre d'exemple, le bois est consommé à hauteur de 45 kWh/m<sup>2</sup> en moyenne pour le chauffage alors que l'électricité est consommée en moyenne à 10kWh/m<sup>2</sup> pour le même usage. Il est donc nécessaire de séparer individuellement chaque source d'énergies. Nous retrouvons alors cette grille d'analyse (Schéma 28), qui nous permet d'avoir

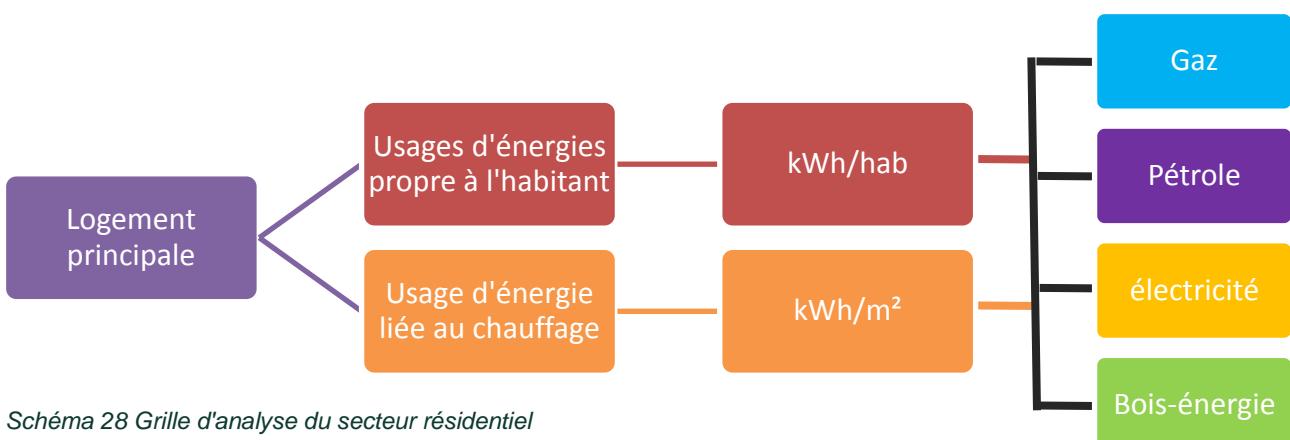


Schéma 28 Grille d'analyse du secteur résidentiel

une distribution normalisée et cohérente pour l'ensemble de la région

#### 4.2.4 L'échelle d'analyse spatio-temporelle

Les données issues de l'inventaire BASEMIS® permettent d'avoir une précision spatiale à l'échelle de la commune, des EPCI, des départements ou de la région. Pour la précision temporelle, nous sommes limités à l'échelle annuelle sur la durée 2008-2016. L'objectif étant de détecter une anomalie, nous avons besoin de la précision géographique la plus fine possible. Le choix de l'échelle spatiale s'est donc porté sur la commune, et l'échelle temporelle sur l'année.

De plus, les consommations d'énergie sont très sensibles aux conditions climatiques, qui peuvent varier d'une année à l'autre. Cette sensibilité climatique va donc créer une tendance d'évolution annuelle spécifique entre deux années. La comparaison du taux d'évolution entre deux années n'est donc pas possible.

## 4.2.5 Validation de l'analyse sectorielle

Une fois les paramètres du secteur définis et le rapport réalisé, il faut valider l'ensemble du rapport pour s'assurer de la cohérence et de la pertinence des résultats. Pour évaluer ces résultats, une méthode de validation a été mise en place en se basant sur 4 critères :

- La cohérence de l'indicateur** : Savoir si les notions sont utilisées ou cohérentes dans le métier. Et si les filtrages et découpages des analyses sont pertinents.

Dans notre cas l'indicateur des kWh/hab ou kWh/m<sup>2</sup> sont des notions très utilisées par les énergéticiens pour parler des consommations énergétiques. Sur la finesse de nos analyses, le filtre entre les logement principale et secondaire est essentiel car la consommation d'énergie n'est pas comparable. Un sous découpage entre logement individuel et collectif pourrait être envisagé dans la partie usage chauffage, ce qui permettrait de gagner un peu de précision.

- La stabilité de l'indicateur** : Confirmer que notre distribution de l'indicateur se rapproche d'une loi normale.

Comme nous le montre le rapport (Extrait 10) nos distributions ont des tendances très proche de la loi normale.

Le test de normalité n'allant pas en dessous de 0.80

- Le nombre d'erreurs en sorties** : Estimer la qualité des données du secteur, à travers le nombre et ce que cela représente dans la base de données.

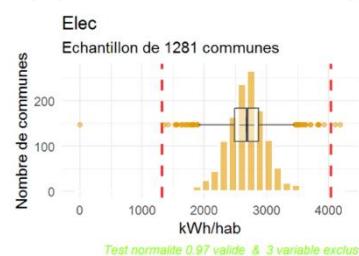
Sur la notion de la précision thématique il y a moins de 0.2% d'erreurs. Ce qui montre un comportement très uniforme ayant peu d'anomalie.

Sur la notion de la qualité d'évolution, on retrouve moins de 2% d'erreurs. Mais avec des amplitudes d'erreurs plus élevés (écart max).

- La confirmation des erreurs** : Valider que les erreurs détectées soient effectivement des anomalies ou cas spécifique présentes dans la base de données.

Certains cas ont interrogé les inventoriastes qui estiment une erreur dans les fichiers sources. Pour Nauvay on trouve un cas particulier car il moins de 12 habitants dans la commune. Il est possible que le mode de vie soit atypique.

Graphique 1 : Distribution des consommations d'énergies en 2016



Extrait 10 rapport résidentiel 2016

ANNEE	Indicateur kWh/m <sup>2</sup>			Taux d'évolution %		
	Taux err	NB err	Ecart max	Taux err2	NB err2	Ecart max2
<b>Autres</b>						
2016	0.2 %	6	7.0	1.21 %	36	131.8
<b>Chaussage</b>						
2016	0.14 %	6	1.7	0.73 %	31	96.3

Extrait 11 rapport résidentiel 2016

Nom de la commune	ANNEE	NAPFUE	Energie	indice		ratio
				Autres		
SAINTE-HILAIRE-LA-FORET	2016	Gaz	8139005.65 kWh	9794 kWh/hab	5.96	
MOUILLERON-SAINT-GERMAIN	2016	Gaz	5252690.31 kWh	2864 kWh/hab	1.03	
NAUVAY	2016	Elec	15.13 kWh	1 kWh/hab	1	

Extrait 12 rapport résidentiel 2016

1	Cohérence de l'indicateur	2	Stabilité de l'indicateur	3	Nombre d'erreurs en sorties	4	Confirmation des erreurs
	Très cohérent dans les analyses		Stabilité des distributions robuste		Le nombre d'erreurs est faible (max 30)		Erreurs détectées
	Notion très cohérente pour les énergéticiens				Le taux d'erreur dans la base ne dépasse pas 1%		Cas spécifiques identifiés

## 4.3 Secteur Routier

### 4.3.1 Analyse de l'impact Climat-Air-Energie



Les transports routiers représentent le secteur le plus consommateur d'énergie finale de la région. En 2016, il a consommé **30 000 GWh** et émis **7 700 kteqCO<sub>2</sub>** de gaz à effet de serre, essentiellement dû à la combustion des carburants. De plus le secteur routier est le principal émetteur de NOx<sup>7</sup>

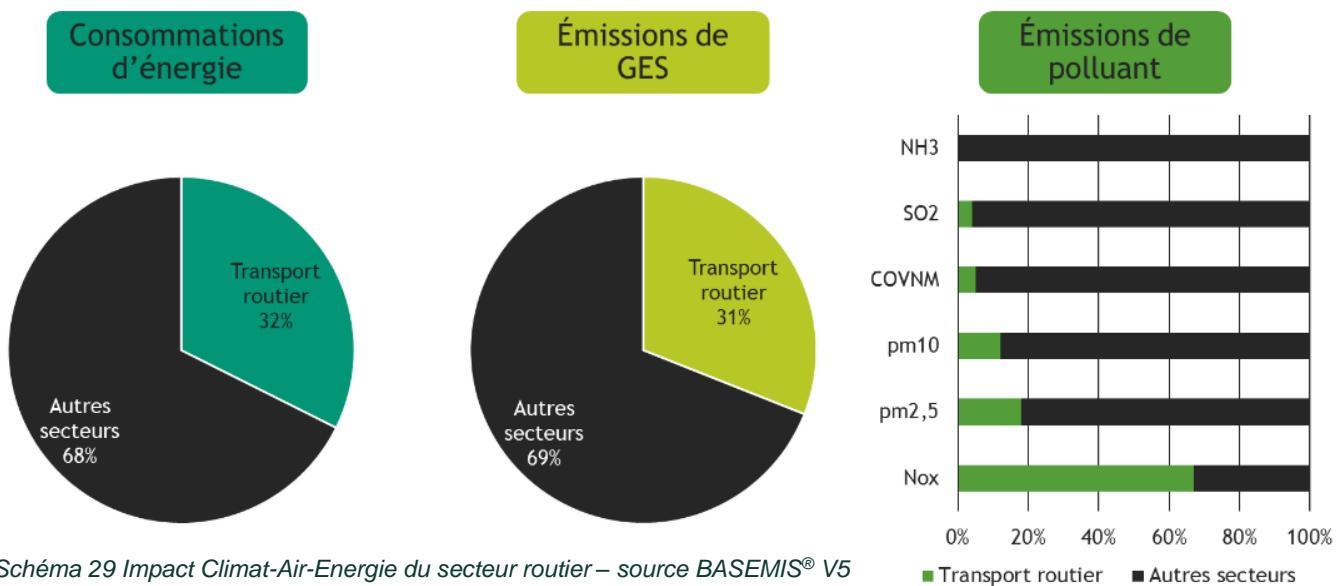


Schéma 29 Impact Climat-Air-Energie du secteur routier – source BASEMIS® V5

### 4.3.2 Méthodologie de calcul des émissions

La méthodologie de calcul des émissions est basée sur la méthode COPERT V (Schéma 30). Cette méthode utilise comme données primaires des comptages routiers fournis par le CEREMA. Elle nécessite l'utilisation d'un serveur dédié afin d'effectuer les modélisations d'émissions issues de ce secteur.

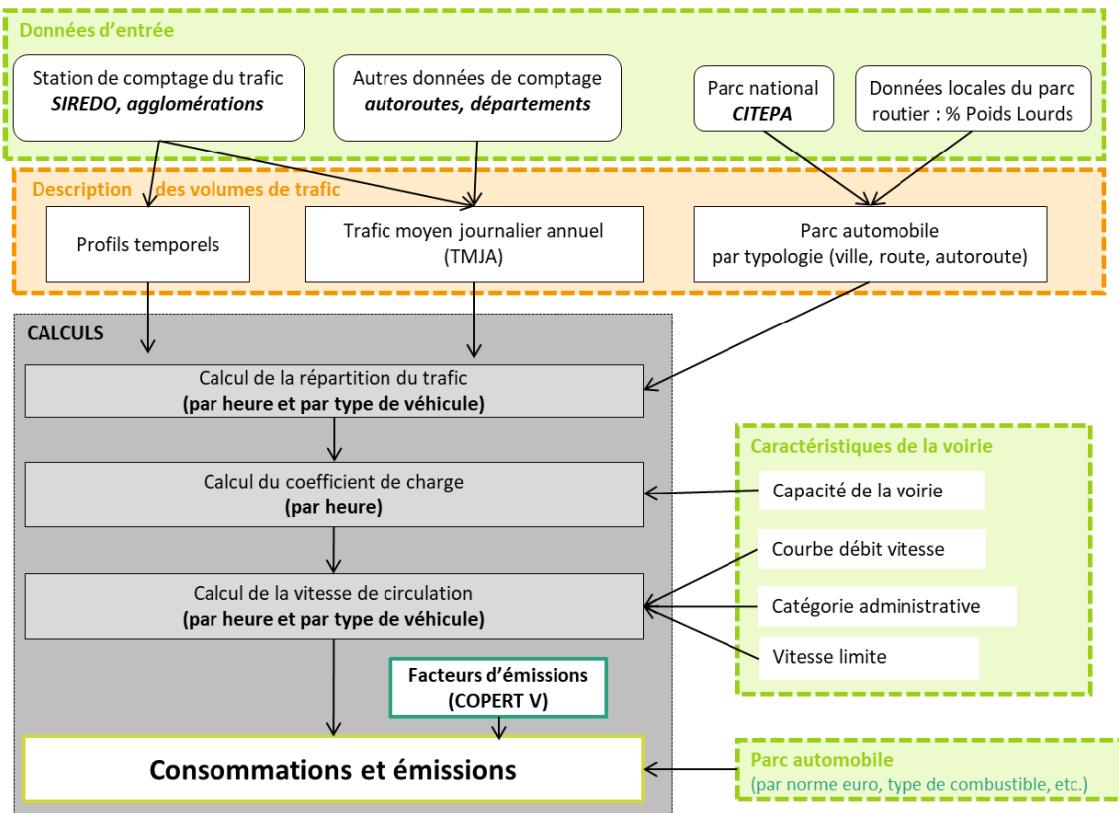


Schéma 30 méthodologie de calcul mise en œuvre pour la prise en compte des transports routiers

<sup>7</sup> 7.2Annexe 2 : Polluants et Gaz à effet de serre

Cette modélisation a déjà fait l'objet de multiples analyses quant à la qualité des résultats obtenus. L'objectif n'était donc pas de vérifier les données ressorties par le modèle mais les données d'entrée nécessaires au modèle (Schéma 10), en particulier les trafics moyens journalier annuel (TMJA).



Le trafic moyen journalier annuel est une donnée spécifique et particulière. Elle représente un flux de voiture roulant sur une voie. Les TMJA sont les variables nécessaires à la détermination des consommations d'énergie du secteur et des émissions de polluants et de GES.

L'analyse de la qualité temporelle entre deux années s'est basée sur le taux d'évolution annuelle des TMJA.

#### 4.3.3 L'échelle d'analyse spatio-temporelle

Deux échelles spatiales sont utilisées pour l'analyse des TMJA :

- Echelles communales
- Echelles des voiries

Dans un premier temps, l'échelle communale a été faite pour garder une cohérence avec les autres analyses sectorielles réalisées. Cependant ce premier résultat n'a pas été convainquant dans l'interprétation des résultats ni dans le ciblage des corrections.

L'analyse à l'échelle communale a permis de mettre en valeur certains territoires ayant eu des augmentations de trafic moyens journalier annuels inhabituels. L'analyse communale n'est pas suffisamment précise pour permettre aux ingénieurs de déterminer la cause des taux d'évolutions anormaux. Ce manque de précision nécessite de visualiser les données primaires.

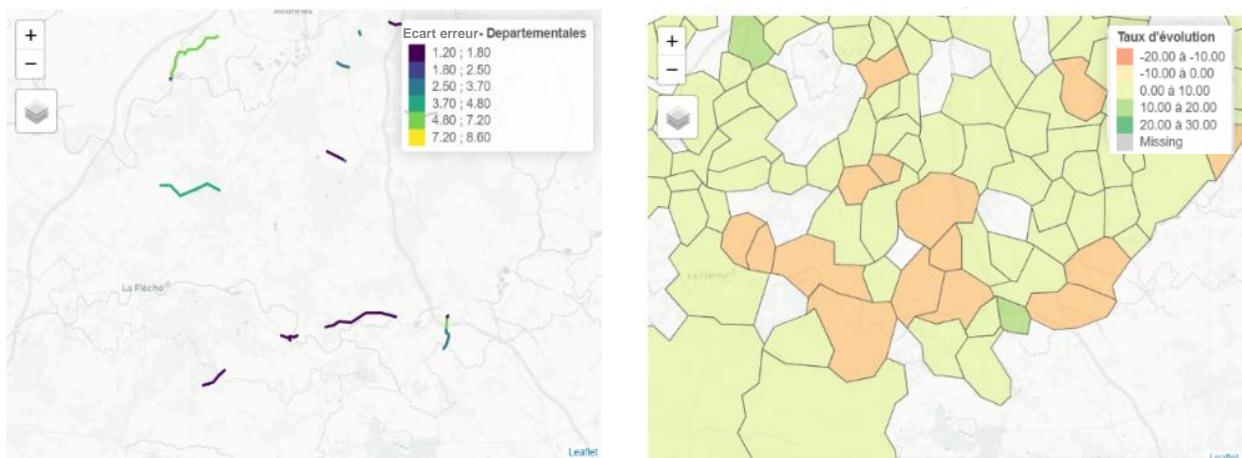


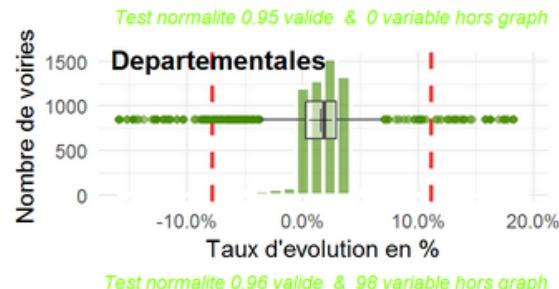
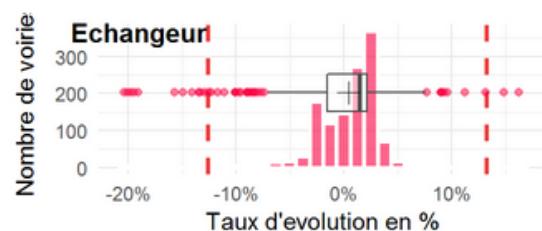
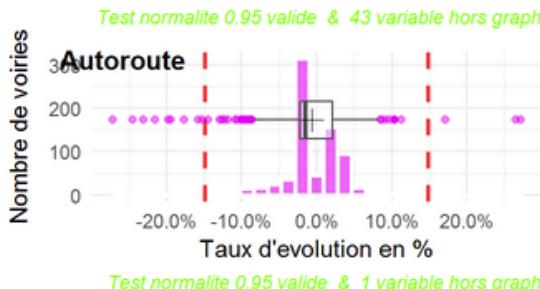
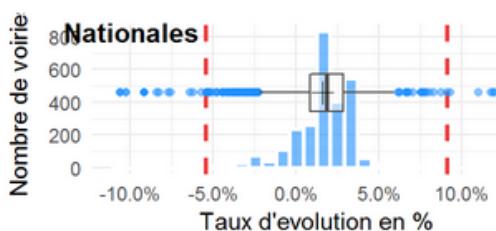
Schéma 31 Comparaison des méthodes d'analyses à différentes échelles (échelles communales à droites et échelles de la voirie à gauche)

C'est pourquoi dans un deuxième temps nous avons analysé l'évolution des TMJA pour chaque voiries (Schéma 31 à gauche). Cela a permis d'avoir un résultat plus ciblé. Le nombre de voirie étant trop important, seules les voiries étant considérés comme des valeurs aberrantes ont été affichées.

Pour permettre de cibler rapidement les routes s'éloignant de la distribution statistique, une sémiologie s'appuyant sur l'amplitude de l'erreur (l'écart à la limite) a été mis en place. Cette interprétation des résultats été plus accessible ce qui permis d'identifier des potentiels erreur du de trafic moyens journalier annuel (TMJA).

## 4.3.4 Validation de l'indicateur

- La cohérence de l'indicateur :** Ici l'analyse de la précision thématique ne peut pas aboutir, car nous analysons une valeur brute (le trafic moyen annuel). L'analyse a donc principalement été portée sur la qualité temporelle à travers son évolution annuelle. Cependant il a été important de séparer les typologies de route, car elles ne suivent pas la même évolution.
- La stabilité de l'indicateur :** Nous pouvons voir que l'évolution adopte un comportement assez stable dans chaque typologie de route (ci-dessous). La distribution ne suit pas une loi normale parfaite. Cependant, avec le nombre d'échantillon important (minimum 4000 voiries dans chaque typologie), l'écart interquartile devient plus robuste. Ce qui permet de définir des limites très cohérentes.



Extrait 13 rapport routier 2016

### 3. Le nombre d'erreurs en sorties :

Dans l'ensemble il y a moins de 1.5% d'erreur présente dans la base de données. La qualité de la version 5 de BASEMIS montre donc une qualité temporelle correct.

Cependant le nombre d'erreur reste assez important (plus de 100 erreurs) ce qui peut prendre beaucoup de temps aux inventoriastes pour les corriger.

Table 2 : Estimation de la qualité des données

RUBRIQUE	ANNEE	Taux erreur	NB erreur	Ecart max
Communales	2016	0.97 %	40	31.3
Departementales	2016	1.49 %	89	8.6
Nationales	2016	1.28 %	35	6.9

Extrait 14 rapport routier 2016

- La confirmation des erreurs :** Les erreurs détectées ont deux profils. Des augmentations ou diminutions très brutales (double, ou triple de la valeur initiale) ce qui est dû aux méthodes de modélisation du CEREMA. D'autres variations plus souples, liées à changements de circulation temporaires (travaux, ...)



Cohérence de l'indicateur



cohérent dans les analyses



Stabilité de l'indicateur



Stabilité des distributions robustes



Nombre d'erreurs en sorties



Le nombre d'erreurs est important (+100)



Confirmation des erreurs



Erreurs détectées



Le taux d'erreur dans la base ne dépasse pas 1.5%



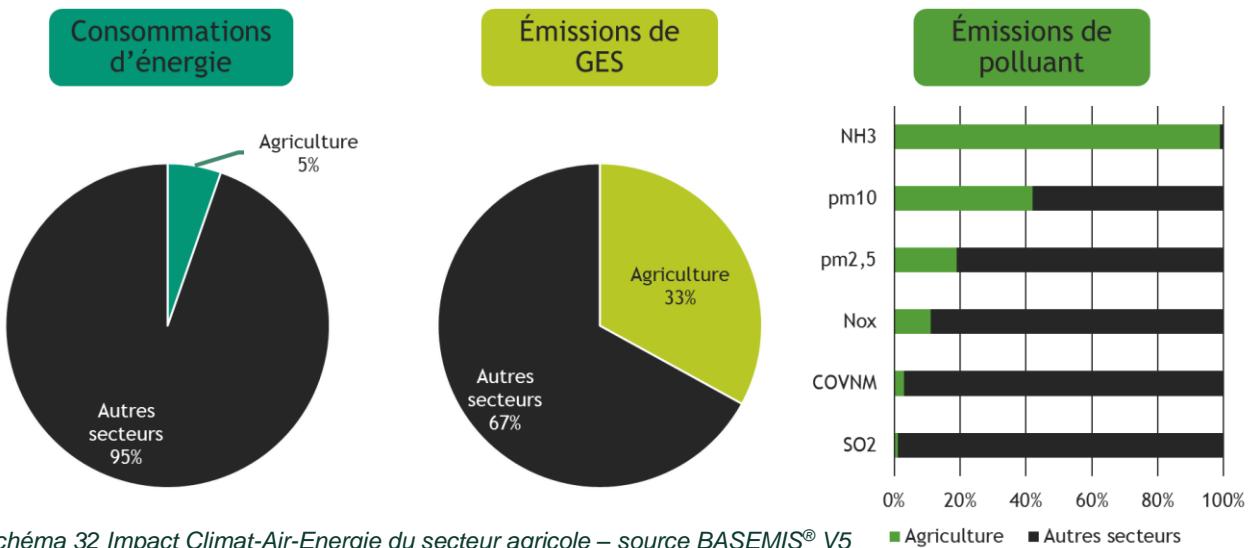
Cas spécifiques identifiés

## 4.4 Secteur Agricole

### 4.4.1 Analyses de l'impact Climat-Air-Energie

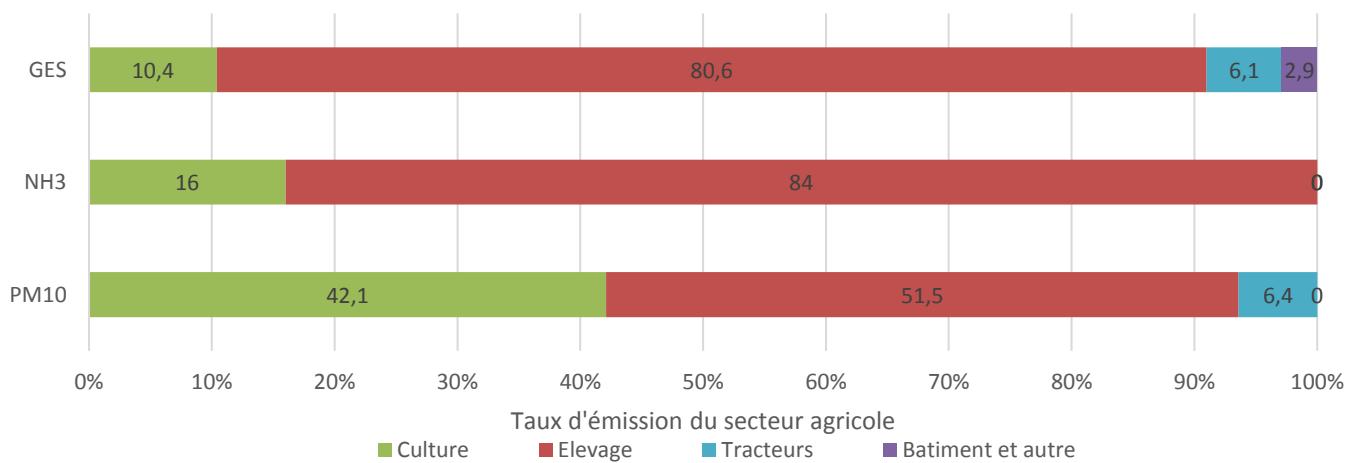


A l'échelle de la région Pays de la Loire, le secteur agricole consomme, en 2016, **4 500 TWh** d'énergie finale et émet **10 500 kteqCO<sub>2</sub>** de gaz à effet de serre, ce qui en fait le secteur le plus émetteur de GES en Pays de la Loire. De plus le secteur agricole est le principal émetteur de certains polluants atmosphériques (NH<sub>3</sub> et PM10). Il est donc important d'avoir un regard à travers les émissions de GES, des NH<sub>3</sub> et des PM10.



### 4.4.2 Méthodologie de calcul des émissions

La méthodologie de calcul d'émission dans le secteur agricole est différente pour chaque sous-secteur présent (élevage, agriculture, engins agricoles, bâtiments agricoles, sylviculture). Un premier regard (Schéma 33) sur les émissions Climat-Air-Energie s'est donc imposé pour comprendre la répartition.



Nous pouvons voir que seulement les sous-secteurs liés à la culture et l'élevage sont réellement représentatifs du secteur agricole.

#### 4.4.2.1 Le sous-secteur de l'élevage

Grâce à la méthode de calcul (Schéma 34), nous savons que les émissions de polluants et de GES sont influencées par le type de cheptels (bovins, caprins, ovins...) et le nombre de cheptels.

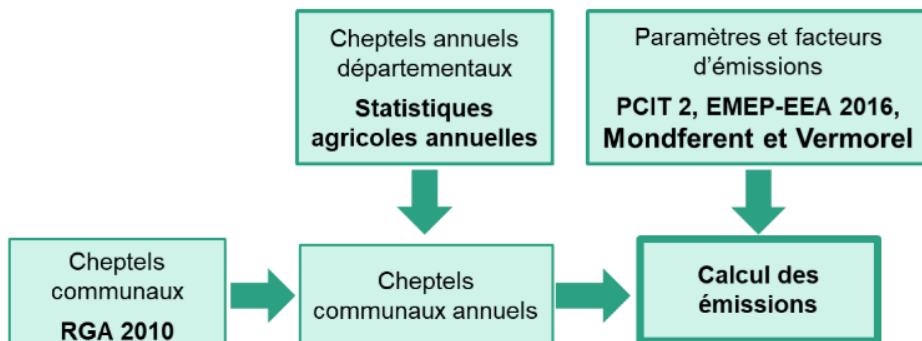


Schéma 34 méthodologie de calcul pour les émissions liées à l'élevage

Nous pouvons alors vérifier trois phénomènes :

- Les émissions de GES avec l'indicateur  $\text{GES}(\text{TeqCO}_2)$  /nombre de cheptels pour chaque catégorie de cheptels.
- Les émissions de NH<sub>3</sub> avec l'indicateur  $\text{NH}_3(\text{kg})$ / nombre de cheptels pour chaque catégorie de cheptels
- La corrélation indirecte entre les GES et les NH<sub>3</sub> avec l'indicateur  $\text{GES}(\text{TeqCO}_2)$  / $\text{NH}_3(\text{kg})$  pour chaque catégorie de cheptels.

Ces trois indicateurs permettent de vérifier le comportement des émissions avec une distribution normale. Cependant, la corrélation entre les NH<sub>3</sub> et les GES étant très importante, les comportements anormaux qui sont visibles dans les émissions de GES se retrouvent dans les émissions de NH<sub>3</sub>. Nous avons donc réduit les analyses à l'indicateur des  $\text{GES}(\text{TeqCO}_2)$  /cheptel et des  $\text{GES}(\text{TeqCO}_2)$  / $\text{NH}_3(\text{kg})$

#### 4.4.2.2 Le sous-secteur des cultures

Tout comme l'élevage, la méthodologie (Schéma 35) permet de voir que les émissions de polluants et de GES sont influencées par le type de cultures et leurs surfaces mais aussi le type de travail (engrais minérale, résidus de cultures, travail du sol et récolte)

Le type de travail influence les émissions. On retrouve alors : les résidus de culture, l'utilisation d'engrais minéral qui émettent l'intégralité GES et le travail du sol et la récolte qui émet l'intégralité des PM10.

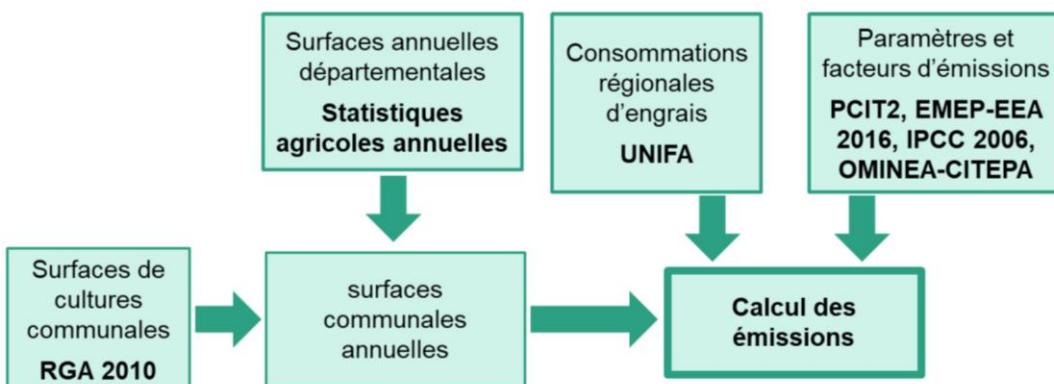


Schéma 35 méthodologie de calcul pour les émissions liées aux cultures

Deux indicateurs ont été mis en place pour l'analyse des émissions liées aux cultures :

- Les émissions de GES avec l'indicateur  $\text{GES}(\text{TeqCO}_2)$  /surface agricole( $\text{m}^2$ ) (Résidus de culture et engrais minéral)
- Les émissions de PM10 avec l'indicateur  $\text{PM10}(\text{kg})$ /surface agricole( $\text{m}^2$ ) (Le travail du sol et récolte)

Certains types de surface comme les jachères n'ont pas été prises en compte car aucun travail n'est réalisé sur ces surfaces agricoles.

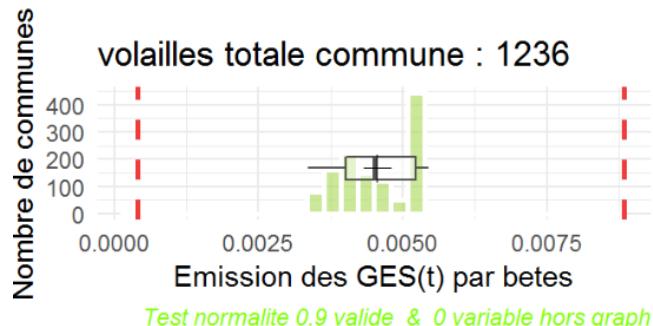
## 4.4.3 L'échelle d'analyse spatio-temporelle



Les données issues de l'inventaire BASEMIS® permettent d'avoir une précision spatiale à l'échelle de la commune, des EPCI, des départements ou de la région. Pour la précision temporelle, nous sommes limités à l'échelle annuelle sur la durée 2008-2016. Notre objectif étant de détecter une anomalie, nous avons besoins de la précision géographique la plus fine possible. Le choix de l'échelle spatiale s'est donc porté sur la commune, et l'échelle temporelle sur l'année.

## 4.4.4 Validation de l'indicateur

- La cohérence de l'indicateur :** Si les indicateurs du secteur agricole restent cohérents, le manque de précision dans BASEMIS sur la nature exacte des animaux (toutes volailles mélangées) rend parfois les indicateurs moins stables.
- La stabilité de l'indicateur :** la stabilité des indicateurs montre des comportements particuliers dans le sous-secteur d'élevage. Comme le montre le cas des volailles où l'on perçoit deux comportements distincts.



Extrait 15 rapport agricole 2016

On peut voir sur le schéma ci-contre, le taux de présence des poulets dans une commune et l'émission des GES pm10 de la commune.

On distingue un facteur d'émission très corrélé aux taux de présence de poulets mais aussi aux départements.

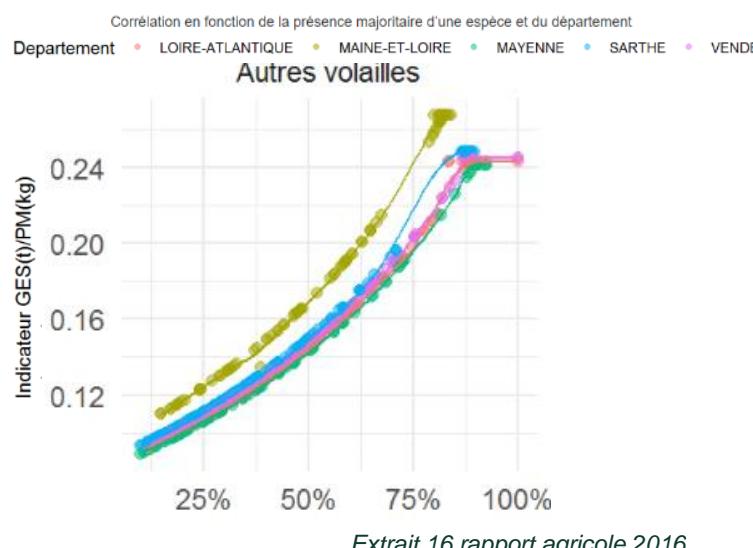
L'objectif de cette analyse est de déterminer une droite d'équation qui nous sert de référence et de mesurer l'écart à la droite pour détecter des anomalies.

Après avoir présenté ces résultats et conclusions aux inventoriastes d'Air Pays de la Loire, il semblerait que la version 6 de BASEMIS adopte un changement sur la méthode de calcul des émissions agricole. Il est donc possible que les indicateurs gagnent en cohérence.

Malgré cette problématique sur le sous-secteur d'élevage, les autres sous-secteurs montrent une stabilité des émissions de GES et de PM10.

Ces deux comportements s'expliquent par le fait que dans l'analyse des volailles il y a deux types de volailles (les poules pondeuses d'oeufs et celle des poulets)

Un début d'analyse statistique avancée a été réalisée pour mieux comprendre et essayer de prendre en compte ces différences.



Extrait 16 rapport agricole 2016

### 3. Le nombre d'erreurs en sorties :

Dans le secteur agricole peu d'anomalies sont détectées.

Pour le critère de la qualité temporelle on remarque une grande stabilité qui avoisine les 2% d'erreurs.

Pour le critère de précision thématique, ici l'émission des GES et des PM10, on peut voir que seulement deux sous-rubriques (ovins et engrais minéral) contiennent des erreurs.

Pour le cas des ovins, chaque année, 58 ( $\pm 2$ ) erreurs sont détectées. Tout comme le cas des volailles, deux comportements sont distincts, (émission des brebis et moutons). La différence d'émission de GES entre une brebis et un mouton est élevée. De plus le nombre de brebis est aussi plus important. Ces deux paramètres, font que lors de notre détection d'erreurs, les émissions de GES des moutons ne sont pas représentatives et donc sont considérées comme un cas d'anomalie.

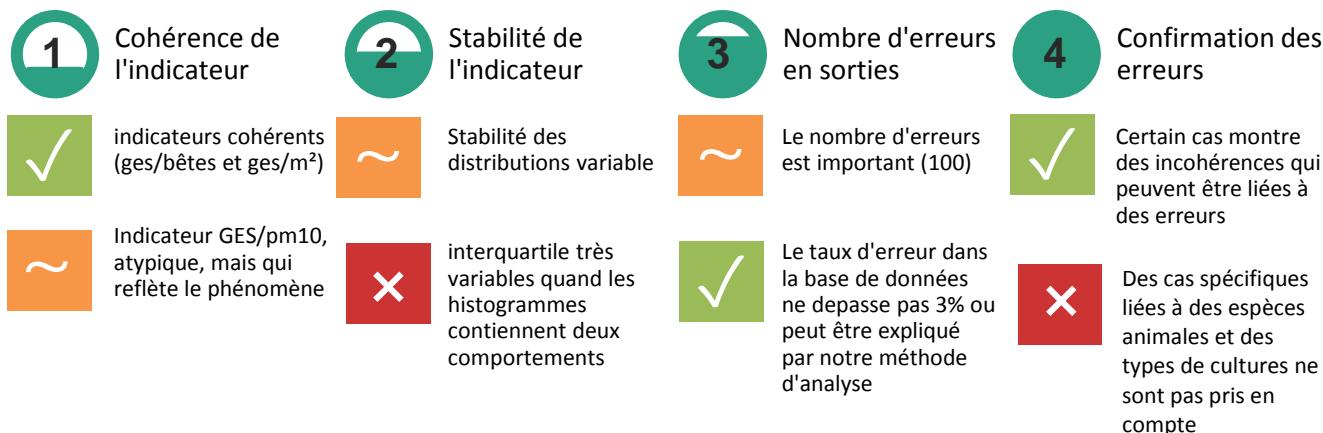
4. **La confirmation des erreurs** : Comme expliqué précédemment, nous détectons un nombre important de cas d'anomalies qui sont en fait des spécificités d'émissions de sous-espèces animales. Il est donc important de regarder si le taux d'erreurs reste stable d'une année à l'autre. Si l'erreur est stable, ce n'est pas une erreur, c'est une « non prise en compte » d'un comportement dans une distribution.

Dans ce cas-là, il est nécessaire de changer la méthode d'analyse pour aboutir à des résultats plus pertinents. Dans les temps impartis le développement de nouvelles méthodes d'analyses n'a pas pu être mener.

Du côté des erreurs sur les engrais minéraux, il semblerait que le type de cultures (maïs, horticultures, ...) joue un rôle important qui n'est pas encore intégré. Une amélioration peut être également amenée sur ce point.

ANNEE	Indicateur émissions			Taux d'évolution %		
	Taux err	NB err	Ecart max	Taux err2	NB err2	Ecart max2
<b>bovins</b>						
2016	0.08 %	1	0.20	0 %	NA	NA
<b>caprins</b>						
2016	0 %	NA	NA	1.58 %	7	1.3
<b>ovins</b>						
2016	7.07 %	58	0.77	1.83 %	15	3.6
<b>porcins</b>						
2016	0 %	NA	NA	0.3 %	2	3.3
<b>engrais mineral</b>						
2016	2.91 %	37	0.58	1.65 %	21	2.5
<b>résidus de culture</b>						
2016	0 %	NA	NA	2.15 %	27	1.5
<b>travail du sol et recolte</b>						
2016	0 %	NA	NA	0.96 %	11	5.2

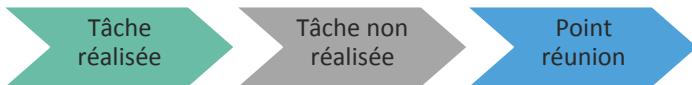
Extrait 17 rapport agricole 2016



# 5 Gestion de projet

## 5.1 Planification des objectifs

Vous trouverez un planning (Schéma 36) qui a permis de planifier mes objectifs pendant ses 6 mois de stages (du 14 mars au 10 septembre) avec les codes couleurs suivants :



Des points hebdomadaires avec mes tuteurs ont été réalisés durant mon stage pour les informer des différentes tâches réalisées, découvertes faites ou valider mes futures tâches et objectifs.

En plus de ces points, des réunions ont été organisées avec des personnes travaillant sur l'inventaire BASEMIS® pour me faire découvrir chaque secteur et pour valider le travail réalisé sur certain secteur.

Enfin à partir de fin juin, suite à une reprise de travail en présentiel, une présentation de mon travail a été faite à l'ensemble du service pour restituer les objectifs de mon stage, les outils apportés, les méthodes utilisées ainsi que les futurs outils afin de valider ce qui a été mis en place.

Dans cette deuxième partie du stage réalisée en présentiel, j'ai pu favoriser les échanges avec les futurs utilisateurs de l'application. Ces échanges ont permis de valider le travail réalisé, les aider à prendre l'outil en main, leur transmettre les compétences pour faire évoluer l'outil.

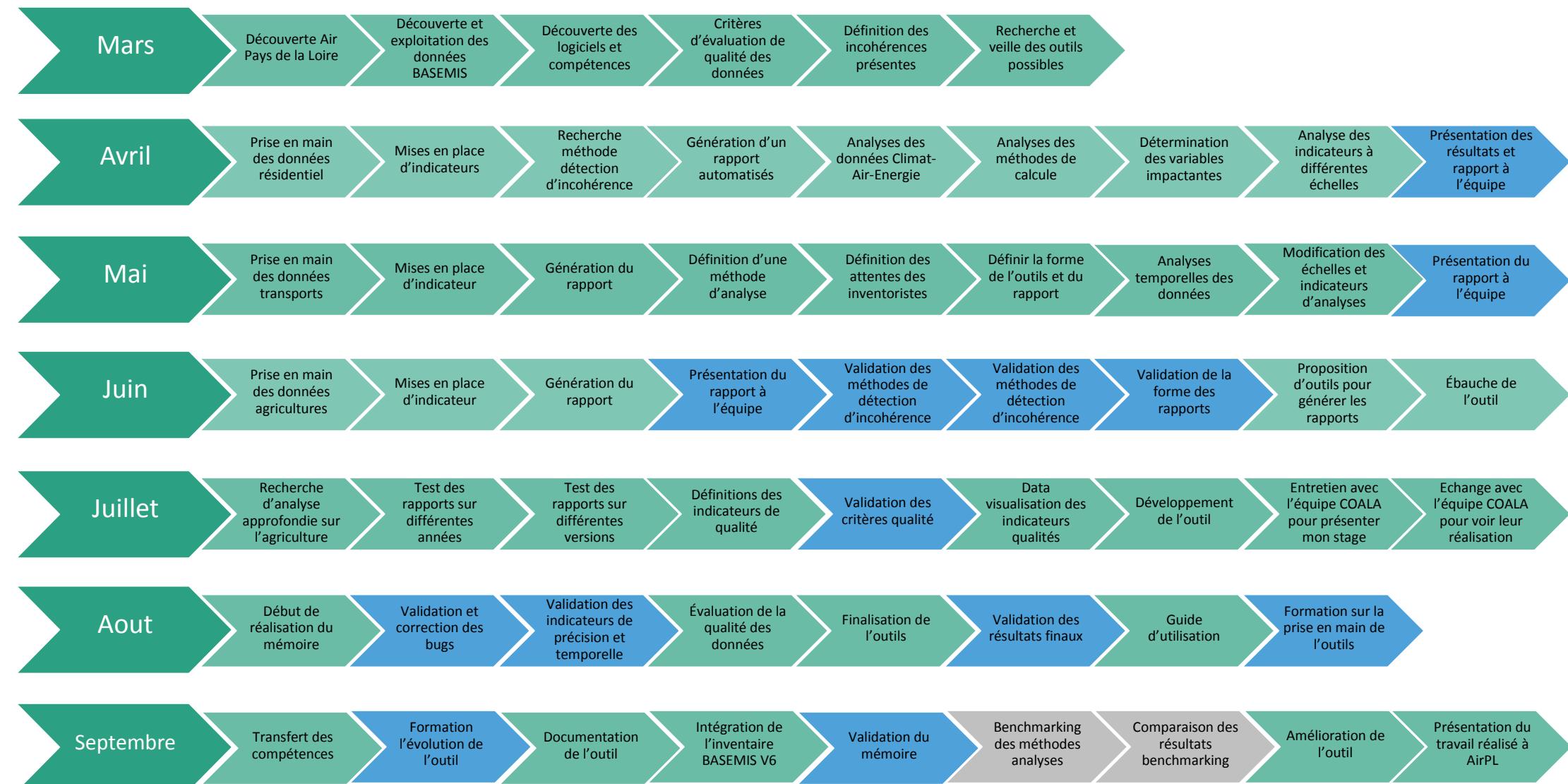


Schéma 36 Planning du stage

## 5.2 Méthodes de travail et organisation

### 5.2.1 Analyses et définition des besoins

Afin de définir mon projet stage, il été important de comprendre le contexte situationnel, les attentes, les besoins et l'environnement du projet. Afin d'obtenir l'ensemble de ses informations, l'utilisation de la méthode QQQQCCCP : Quoi, Qui, Où, Quand, Comment, Combien, Pourquoi s'est vu pertinente. C'est un outil adaptable à diverses problématiques permettant la récolte d'informations précises et exhaustives d'une situation.

#### **Pourquoi évaluer la qualité de la donnée ?**

- Assurer la fiabilité de l'inventaire
- Permettre un ciblage des corrections
- Surveiller la stabilité de l'inventaire

#### **Quel critère des données faut-il évaluer ?**

- La précision thématique
- La qualité d'évolution temporelle
- La cohérence logique

#### **Comment réaliser ces évaluations ?**

- En réalisant un diagnostic des données qui s'appuie sur des analyses statistiques.
- Avec un outil permettant d'automatiser ce diagnostic
- Avec un outil permettant de visualiser les résultats des diagnostics

#### **Qui seront les utilisateurs de l'outil ?**

- Les ingénieurs d'études qui réalise l'inventaire BASEMIS
- Les membres de l'équipes COALA si l'outils viens à être diffusé.

#### **Où seras enregistré les résultats ?**

- Possibilité d'enregistrer les scripts d'automatisation en local ou sur le server
- Possibilité d'enregistrer les résultats qualité sur server ou dans une BDD

#### **Quand diagnostiquer et à quel fréquence doit-il être fait ?**

- Avant la publication d'un inventaire afin de s'assurer de la qualité
- Lors d'un audit de qualité des données
- Dès qu'une modification est apporter à l'inventaire

#### **Quel sont les moyens matériels et logiciels possible ?**

- Mise a disposition du server
- Tout les logiciels sont envisageables, mais de préférences les logiciel libre de droit pour respecter la charte numérique de l'association

### 5.2.2 Méthode de travail

La méthode de travail adoptée pour la réalisation du projet s'est beaucoup appuyée sur les modèles AGILE (Schéma 37).

Le stage s'est déroulé dans des conditions atypiques avec les circonstances du COVID-19. D'une part la mise en place du télétravail de mon côté et d'autre part la mise en place du chômage partiel pour les employés d'Air Pays de la Loire.

Ces conditions montrent des désavantages ; communication réduites, présence des employés réduites... Pour pallier ces éventuelles difficultés, nous avons mis en place une organisation permettant de favoriser des échanges hebdomadaires (Un point par semaine de 15-20min sur les tâches réalisées et les futures tâches à faire).

Ces échanges hebdomadaires ont permis :

- De présenter les résultats des recherches ou développements réalisés
- De discuter des attentes, volontés et enjeux face aux résultats
- De définir les objectifs ou améliorations pour la semaine suivante

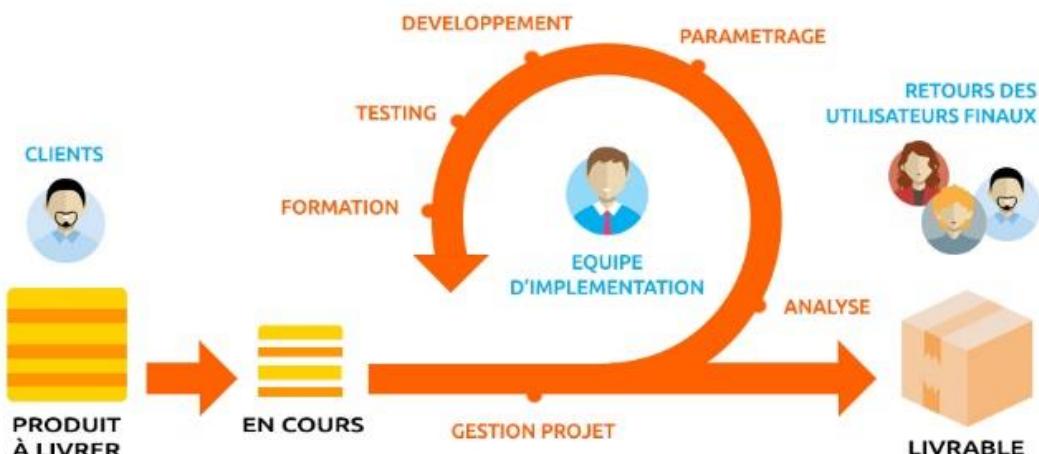


Schéma 37 Méthode de travail AGILE : <https://www.mercator.eu/fr/la-methode-agile-comme-methode-de-travail-chez-mercator-explications.shtml>

Sur le planning (Schéma 36), l'ensemble des objectifs définis au cours du projet est présenté. Nous avons fixé ces objectifs de façon à pouvoir réaliser plusieurs tâches en parallèle pour éviter de se retrouver bloqués.

De plus, afin de mieux définir les attentes sur le fond et la forme de l'outil, la stratégie a été de présenter les résultats des objectifs hebdomadaires à travers des outils de data visualisation et reportings automatisés.

De cette façon en très peu de temps et de façon progressive, il a été possible de :

- Concrétiser les attentes du projet
- Affiner les outils
- Confirmer des résultats
- Valider des méthodes

### 5.2.3 Analyses et constats

Après avoir planifié le projet mais aussi examiné les attentes et besoins du projet, la pérennité du projet dans le temps s'est posée. Effectivement l'objectif de réaliser un diagnostic sur la qualité des données ne doit pas se faire seulement sur le 5<sup>ème</sup> et 6<sup>ème</sup> inventaire BASEMIS, mais aussi sur les suivants, avec peut-être des nouvelles questions, hypothèses, méthodes, critères et objectifs.

C'est dans cette dynamique que j'ai réalisé un constat et proposer des solutions (Schéma 38) pour répondre à cette question de pérennité.

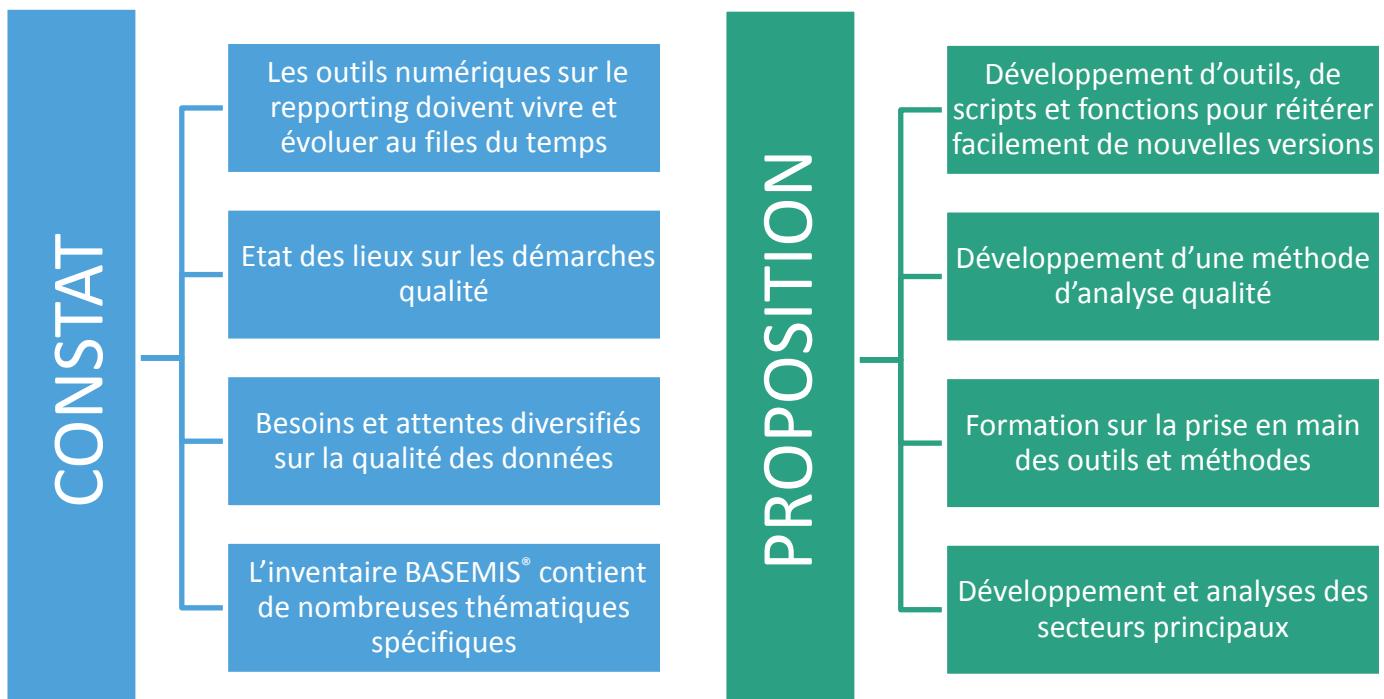
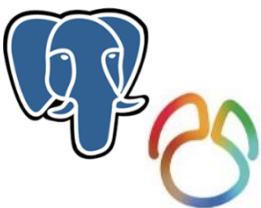


Schéma 38 Constat et proposition de solution pour le projet

La solution s'est donc orientée vers un outil évolutif, facile de prise en main par les utilisateurs. Pour que la prise en main et l'amélioration de l'outil soit possible, il faut alors réaliser une documentation explicative, mais aussi une formation auprès des utilisateurs. (5.3 Formation et documentation des outils mis en place)

## 5.2.4 Définition des logiciels et outils du projet

Lors de ce projet de nombreux logiciels (Schéma 39) ont été utilisés. Le choix de ces logiciels a été basé principalement sur le respect de la charte numérique d'Air Pays de la Loire qui vise à prioriser les logiciels libres et sur les compétences internes à utiliser ces logiciels et langues de programmation associée. Les compétences et habitudes externes des ASSAQA partenaires ont également été pris en compte.



### PostgreSQL et NAVICAT

Gestion de base de données de l'inventaire BASEMIS.  
Les logiciels ont principalement été utiles pour préparer les requêtes SQL et visualiser le résultat.

### R et Rstudio

Traitements des données automatisé pour réaliser des : requêtes sur base de données, analyses statistiques, des data visualisations, la génération de rapport automatique, le développement d'application.



### Github

Archiver et versionner les scripts réalisés.  
En mode privée pour ne pas diffuser l'application.  
Cependant la réflexion de partage peut être envisagé.

### Trello

Gestion et organisation du projet de manière personnel afin de pouvoir réaliser différentes tâches en parallèle tout en permettant de rappeler les priorités du projet.



### Skype et Zoom

Communication et présentation des résultats via des réunions hebdomadaires ce qui a permis de maintenir la bonne orientation du sujet mais la validation des résultats.

Schéma 39 : Descriptif des logiciels utilisés

## 5.3 Formation et documentation des outils mis en place

« Il ne suffit pas de ‘voir’ un objet jusque-là invisible pour le transformer en objet d’analyse .... Il faut encore qu’une théorie soit prête à l’accueillir. »

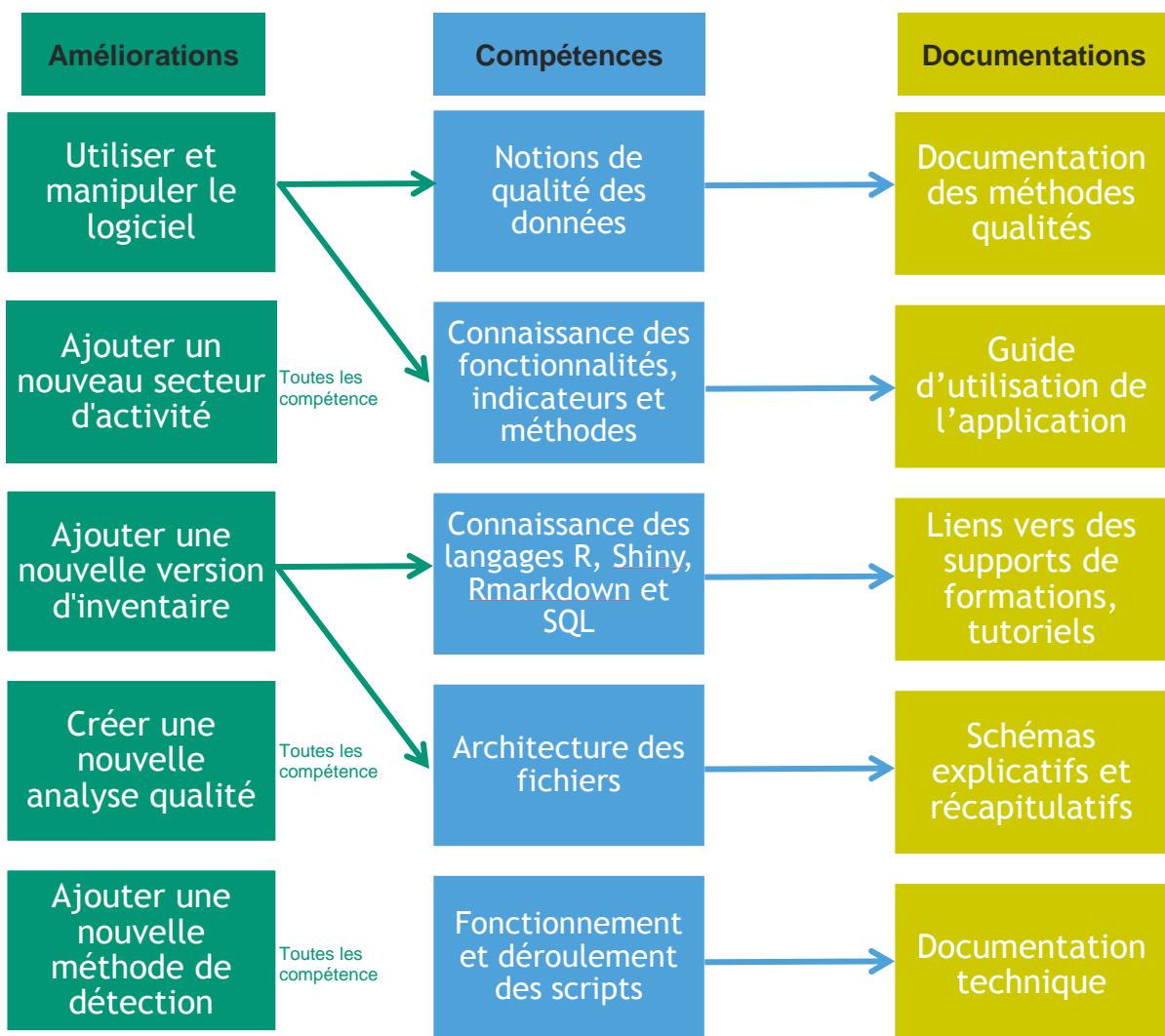
François Jacob, *la logique du vivant* (1976)

La mise en place de l'outil applicatif permet de donner une interprétation de la qualité des données. Il est également nécessaire que cet outil soit compris et utilisé par les inventoriistes. C'est dans cette volonté qu'une partie du stage s'est concentrée sur la formation des utilisateurs à l'outil mais aussi à la documentation qui accompagne l'outil.

### 5.3.1 Stratégie et organisation

Nous avons décidé de mettre en place un outil évolutif afin que celui-ci puisse être complété et amélioré face à de nouvelles attentes dans le futur.

Pour rendre l'outil manipulable et modifiable par tous, il a été nécessaire d'identifier les améliorations possibles ainsi que les compétences à développer pour modifier l'outil. Une fois l'ensemble identifié, la mise en place d'une documentation pour répondre à ses besoins devait être réalisée.



### 5.3.2 Des formations adaptés et variées

Pour apporter l'ensemble des compétences utiles à la prise en main de l'outil et à son évolution, une succession de formation a été mise en place :

1. **Une présentation générale** : destinée à l'ensemble de l'équipe, l'objectif a été de présenter les méthodes et critères d'évaluation de la qualité des données d'un point de vue théorique. Celle-ci s'est accompagnée d'une présentation des premiers résultats pour montrer comment ces méthodes et critères ont été mis en place.
2. **Présentation sectorielle et formation** : spécifique à la présentation d'un secteur d'activité avec la ou les personnes concernées, l'objectif était de présenter les résultats obtenus du secteur afin de valider l'ensemble du travail. Cette présentation s'est suivie par la prise en main des outils finaux (application et rapports) avec une explication sur le fonctionnement et déroulement des scripts ainsi que leur architecture.
3. **Prise en main personnelle** : l'idée était que chaque membre de l'équipe essaye de lancer l'application sur leur ordinateur de façon autonome avec le guide d'utilisation afin de voir si celui-ci était complet et suffisant.
4. **Phase d'amélioration** : suite à la première prise en main, plusieurs remarques ont été faites pour améliorer l'application. L'objectif est de montrer comment apporter ces modifications.
5. **Réunion d'évaluation générale** : une réunion va avoir lieu avec les personnes utilisatrices de l'application dans l'objectif de cerner les informations pouvant être manquantes, des documents nécessaires mais aussi de discuter de l'évolution de l'application.

C'est grâce à ces méthodes d'apprentissages variées (expositive, démonstrative et expérimentuelle) que j'espère avoir pu apporter ou consolider les compétences nécessaires à l'évolution de l'outil de validation des données.

### 5.3.3 Les objectifs des documentations

Pour compléter les informations apportées dans les formations, mais aussi, partager plus facilement l'outil, différentes documentations ont été mises en place. On retrouve alors :

1. **Les supports de présentation** : pour refaire des formations ou revoir les méthodes d'analyses utilisés
2. **Une guide d'utilisation** : pour lancer l'application depuis un nouvel ordinateur (Annexe 10 : Guide d'utilisation)
3. **Des schémas d'architectures** : pour comprendre où se situe les scripts (Annexe 9 : Documentation architecture des fichiers)
4. **Une documentation des fonctions et script** : pour comprendre comment sont utilisés les scripts, les variables et lesquels sont indispensables. (Annexe 11 : Documentation – logigrammes (extrait))
5. **Des liens de formation et tutoriels** : pour avoir des éléments d'aides de programmation (Annexe 12 : Liens de formation et tutoriels)
6. **Un mémoire de stage** : pour recenser l'intégralité des méthodologies, des bibliographies et des stratégies réalisées.

# 6 Conclusions

## 6.1 Evaluation des outils apportés

L'objectif d'apporter un outil automatique permettant d'évaluer la qualité des données de l'inventaire BASEMIS® est pratiquement atteint.

### Pourquoi l'objectif initial n'est pas complément atteint ?

L'outil ne contient pas l'intégralité des secteurs d'activité ce qui ne permet pas de balayer l'ensemble de l'inventaire. Il ne contient pas non plus l'intégralité des versions de BASEMIS®.

Cependant, dans la durée de temps imparti, la stratégie ne s'est pas orientée sur un outil final mais sur l'élaboration conceptuelle d'un outil d'analyse, avec une méthodologie et une transmission des compétences afin de rendre l'outils évolutif et que chaque membre de l'équipe BASEMIS puisse s'approprier cet outil.

De plus, l'intégration de trois secteurs importants de l'inventaire a été réalisée sur la version 5 de l'inventaire, ainsi que sur la future version à venir en 2020.

### Un outil qui répond aux besoins.

L'outil permet de visualiser les critères de qualité de données souhaités à travers des indicateurs de qualité pertinents se basant sur des diagnostics et analyses thématiques annuels précis.

Grâce à cet outil, Air Pays de la Loire peut visualiser facilement la qualité de ses données, cibler des anomalies et corriger de possibles erreurs avant de publier un inventaire en ligne. Il contribue ainsi à améliorer la qualité et l'expertise des données d'Air Pays de la Loire.

### Elaboration conceptuelle, méthode aboutie et outil fonctionnel.

Pour conclure sur la partie conceptuelle de l'outil, on note de nombreuses possibilités de démarches qualités en lien avec les inventaires des émissions. La partie Qualité des données de ce mémoire permet d'avoir un recensement de ces autres démarches possibles.

Sur les méthodes d'analyse, une démarche a été conçue pour faciliter l'analyse thématique des secteurs pris en compte. L'intégration de différents secteurs et versions a été réalisée, avec une transmission des possibilités de l'outil à l'ensemble de l'équipe émissions d'Air Pays de la Loire.

### Une qualité des données vérifiée dans l'inventaire BASEMIS.

Sur la partie analyse de la qualité des données des trois secteurs réalisés, nous avons pu évaluer la précision thématique et la qualité temporelle.

Dans l'ensemble, les résultats montrent que l'inventaire BASEMIS® V5 a une précision thématique satisfaisante. Chaque secteur a moins de 3% de cas anormaux s'éloignant d'un comportement générale stable.

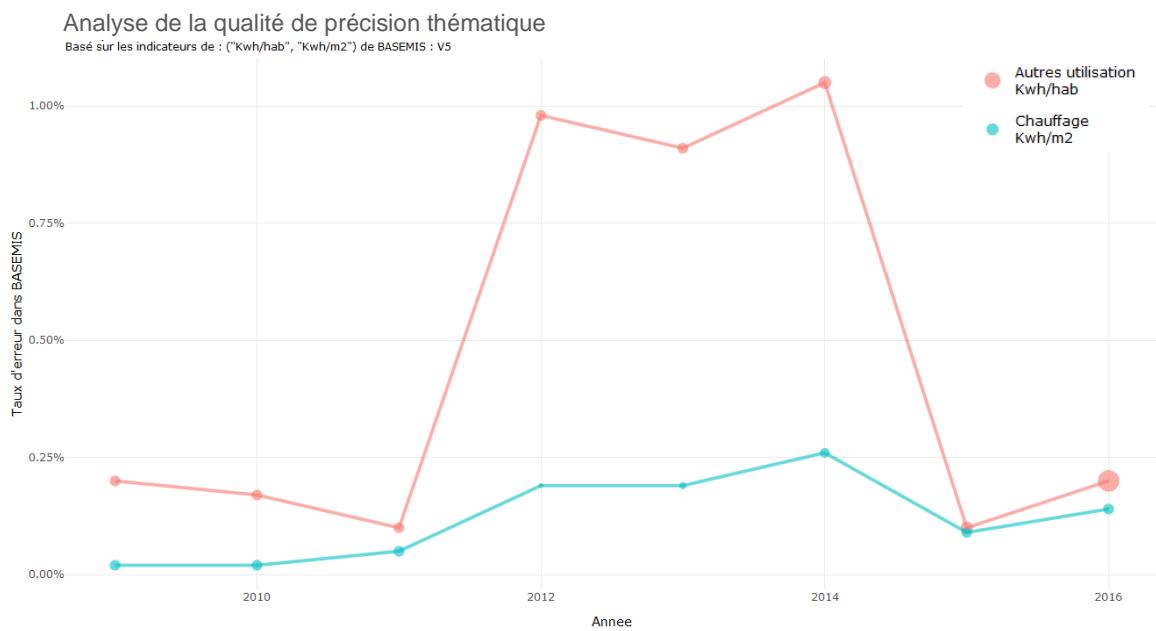
Sa qualité d'évolution temporelle est convenable. On perçoit une stabilité des évolutions annuelles, avec moins de 5% des données au sein des secteurs qui s'éloignent des tendances d'évolutions régionales.

## 6.1 Qualité des données des secteurs analysés

### 6.1.1 Secteur résidentiel

Les indicateurs de précision thématique du secteur résidentiel permettent parfaitement d'identifier un comportement des consommations d'énergétiques général dans la région. Ce comportement général permet d'identifier les communes ayant des consommations énergétiques anormales pouvant être des erreurs liées aux données primaires. Ces anomalies peuvent également être retrouvées à travers notre indicateur de qualité d'évolution temporelle.

Les résultats de l'application sur l'inventaire BASEMIS V5 nous montrent que dans ce secteur très peu de cas anormaux de précision thématique surviennent. En moyenne 10 anomalies annuelles sont présentes chaque année sur un échantillon de 1280 données.



Du côté de la qualité des évolutions temporelles les anomalies sont plus fréquentes. En moyenne 20 anomalies sont présentes (toujours sur un échantillon de 1280 données), ce qui représente peu d'anomalie.

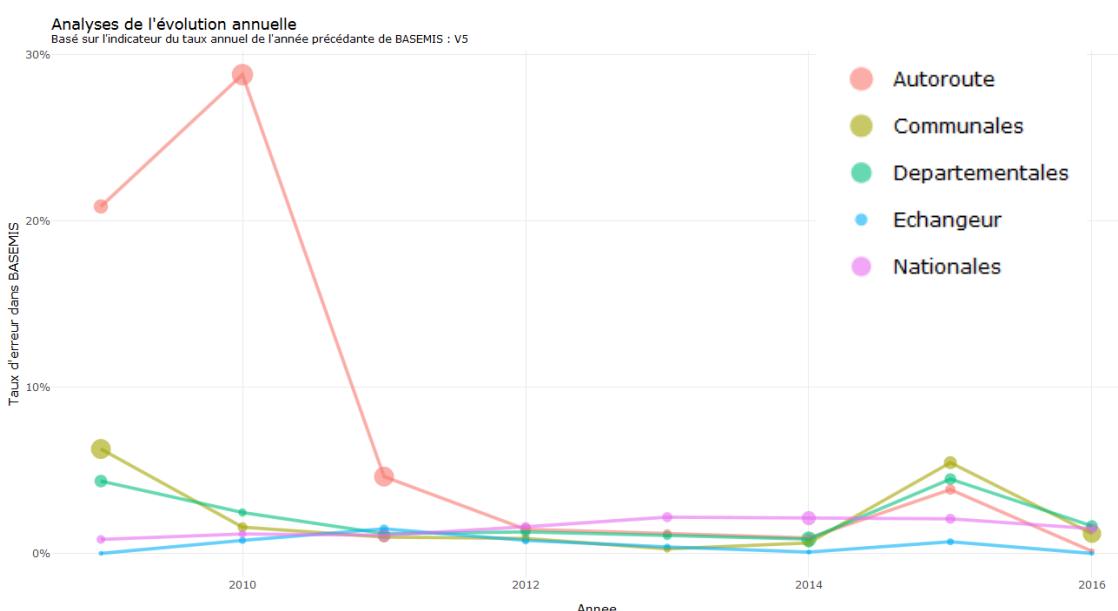
Ce nombre faible d'anomalies montre que les comportements généraux sont très stables et donc que la qualité de nos données est satisfaisante.

## Conclusions

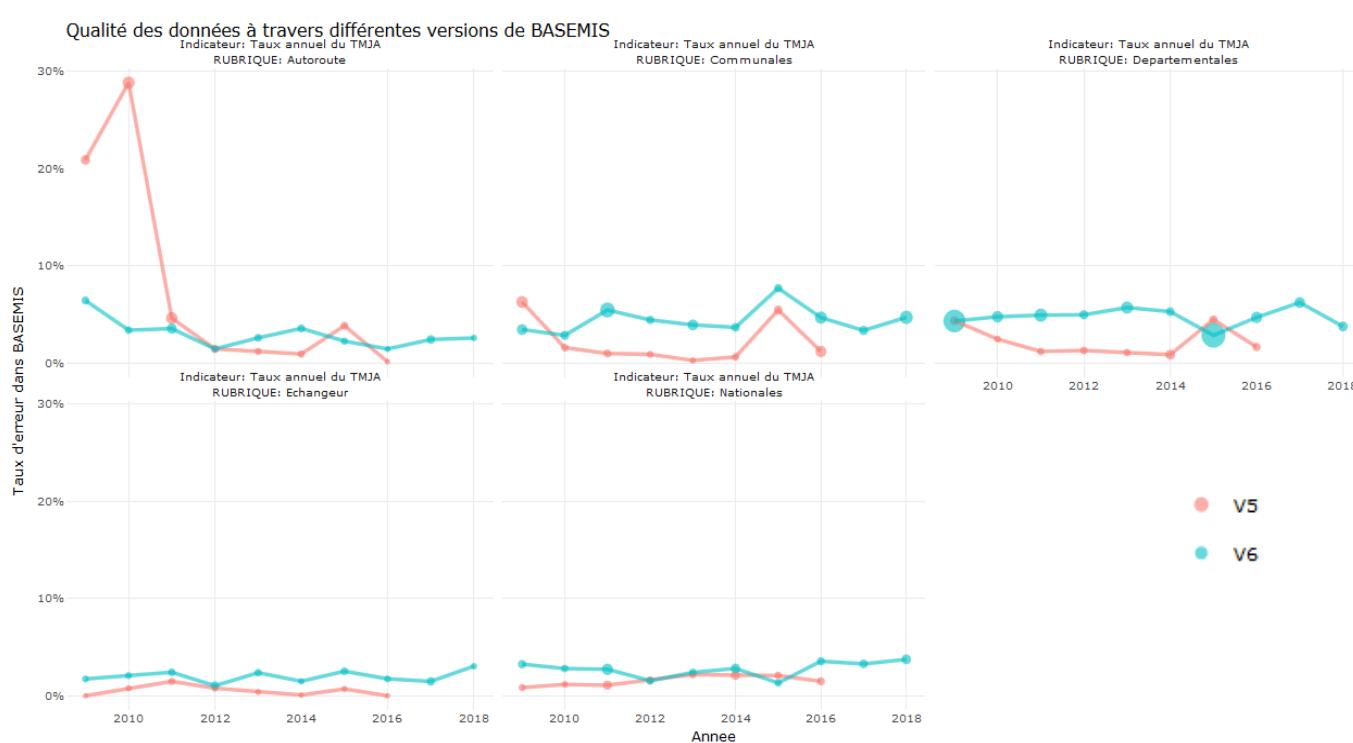
## 6.1.2 Secteur routier

Sur le secteur routier seul l'indicateur de qualité d'évolution temporelle est évalué. Celle-ci montre une stabilité avec très peu de variations d'erreur pour l'ensemble des types de routes (hors le cas des autoroutes avant 2010 qui s'explique par la méthode de comptage du CEREMA).

Le nombre d'anomalies ressorties reste faible (moins de 2% des données sont considérées comme anormales). Cependant une amplitude importante des erreurs (30 fois supérieur à notre limite) est constaté ce qui peut avoir un impact sur l'utilisation des données si cette évolution n'est pas justifiée.



La comparaison entre la version 5 et la version 6 montre une dégradation des données de la version 6 (le taux d'erreurs est plus important que la version 5). En effet, les données n'ont pas encore été retravaillées et validées contrairement à la version 5 de l'inventaire BASEMIS.

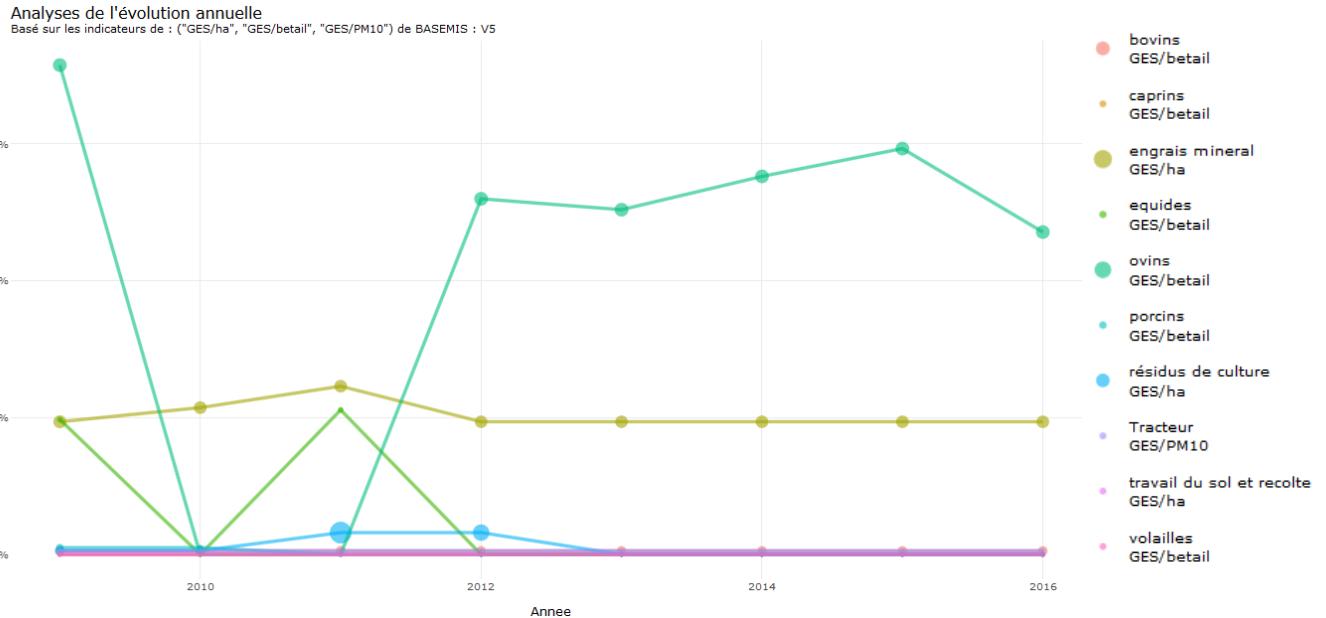


# Conclusions

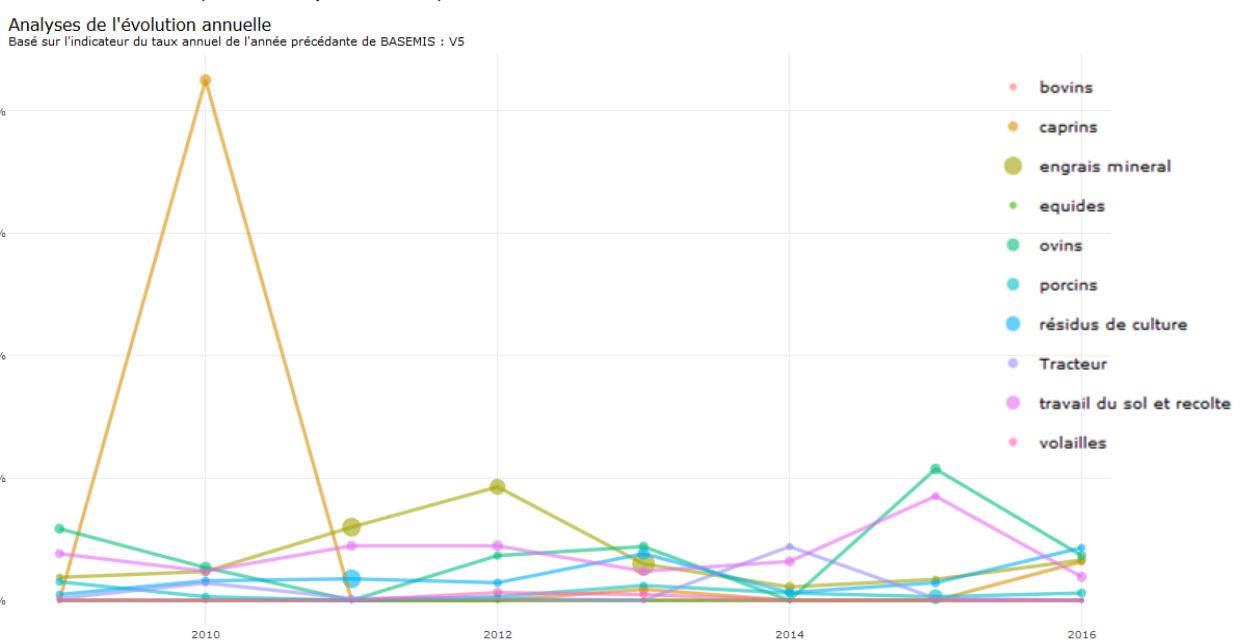
## 6.1.3 Secteur agricole

Le secteur agricole présente de nombreuses sous catégories. Les indicateurs de précision thématique des GES/ha, GES/bétail et GES/PM10 permettent de représenter des comportements généraux des émissions de chaque catégorie. Cependant certains comportements montrent que certaines analyses devraient être réalisées à un niveau de détail plus fin (analyse par sous-espèces).

Les résultats des indicateurs de précision thématique montrent deux comportements :



- Les catégories d'analyse des ovins et engrais minéraux montrent un nombre important d'erreurs (en moyenne plus de 3% d'erreurs pour les engrais et 8% pour les ovins). Cependant il y a une grande stabilité dans la détection de ces anomalies. Ceci montre que la méthode d'analyse détecte comme abnormal des cas normaux.
- L'ensemble des autres catégories détecte peu d'anomalies (3 erreurs en moyenne). Cependant parmi ces anomalies certaines ont une amplitude assez éloignée de notre limite (3 fois supérieures).



Sur le point de la qualité d'évolution temporelle on note des résultats satisfaisants ne dépassant pas les 5% d'anomalies. Un pic important d'anomalies est détecté pour l'analyse des caprins suite à la mise à jour d'un référentiel.

## 6.2 L'ensemble du projet de validation de données

Le stage réalisé n'est qu'une partie du projet de validation de données. L'outil permet d'évaluer la qualité des données sur deux thématiques principalement.

Une problématique reste les sources des données primaires, qui ne concerne pas seulement Air Pays de la Loire mais plus globalement les inventoriastes de manières générales. Les inventaires sont totalement dépendant des données primaires ce qui peut entraîner des changements brutaux dans la qualité et le type de données si le fournisseur de ces données décide de modifier sa méthodologie de création des données primaires.

Il est difficile, voire impossible, d'agir sans créer une perturbation face à cette situation. Cependant l'outil mis en place pourra permettre de faire prendre conscience de ces changements. D'autres améliorations ([Méthode d'analyse descriptive VS machine learning ?](#)) peuvent permettre de proposer des solutions face à cette problématique.

Une autre problématique présente est la correction des erreurs. Actuellement l'outil permet de réaliser un diagnostic descriptif du comportement des données mais il ne permet pas d'identifier les types d'erreurs rencontrés afin de proposer une solution de correction automatique. C'est aussi une amélioration envisageable ([La source des erreurs ?](#)).

## 6.3 Perspectives d'amélioration et préconisation

L'outil de qualité de données nous permet d'évaluer la qualité des données de manière générale sur l'inventaire BASEMIS®. La visualisation de la qualité des données et les diagnostics annuels associés nous permettent de remplir plusieurs objectifs :

- Des outils permettant de faciliter les audits sur la qualité des données
- La justification de la qualité des données de l'inventaire aux près des partenaires
- S'assurer de la qualité des données avant une publication de l'inventaire
- Ciblé des corrections à apporter

Cependant d'autres objectifs répondant peuvent nous permettre d'améliorer l'évaluation de la qualité des données.

### 1. La source des erreurs ?

**Point de vigilance :** Si l'on parle de qualité des données ce n'est pas juste la correction de données dégradées une fois les processus méthodologiques appliqués, c'est aussi veiller à ce que : d'une part les données primaires soient d'une qualité correcte, d'autre part que la détection et la correction soient faites pendant les processus méthodologiques de l'inventaire, voire trouver, la source génératrice d'erreurs afin d'intervenir sur celle-ci.

*L'outil mis en place ne permet pas d'identifier automatiquement les types d'erreurs, ni d'appliquer une correction adaptée aux erreurs. C'est sûrement l'un des points d'améliorations les plus importants à mettre en place dans la suite de ce processus.*

### 2. Méthode d'analyse descriptive VS machine learning ?

**Perspectives d'amélioration :** L'analyse statistique montre parfois certaines limites. D'autres méthodes existent et peuvent se montrer plus efficaces dans l'analyse évolutive des données. Le machine learning ou le deep learning sont des méthodes d'analyses qui ne s'appuient pas sur une hypothèse statique mais plutôt un apprentissage grâce aux quantités de données. Ce sont des pistes de réflexion à mener qui peuvent se montrer efficaces lorsque les quantités de données sont importantes, mais aussi lorsque les données évoluent.

*Dans le cadre de l'inventaire l'évolution des données est très réduite et contrôlée. Cependant cette évolution peut survenir des données fournies par les partenaires (notamment sur les futures évolutions des données énergétiques de l'article 179 de la LTECV). Si ces évolutions impactent les analyses climat-air-énergie mises en place alors un ajustement par deep learning peut aider à réduire cet impact.*

### 3. « To be error or not to be »?

**Point de vigilance :** L'outil permet de cibler des potentielles erreurs, toutes les erreurs ressorties n'en sont pas. Des cas particuliers sont parfois présents (possibilité de les voir grâce à leur stabilité et leur récurrence dans le temps). Mais il est aussi possible que certaines erreurs ne soient pas détectées.

*Il est aussi possible d'avoir des cas particuliers (interdiction de circuler sur une voirie pendant 1 an donc un trafic routier de 0 par exemple) ce qui doit ressortir dans nos analyses. Cependant si les données fournies par le partenaire, ne prennent pas en compte cette spécificité et qu'il remplace ce trafic qui est réellement de 0 par une valeur de trafic en se basant sur l'année précédente, alors cette erreur n'est pas détectée.*

### 4. Evaluation de la cohérence logique

**Préconisation :** aujourd'hui la cohérence logique est vérifiée sur de multiples points à travers différents outils (script python, table excel, requête sql). L'inventaire final n'accepte aucune incohérence logique (aucune émission de trafic routier en mer par exemple, ...). Pour cela les inventoriastes traitent, améliorent et perfectionnent les données.

*Un rapport détaillant et mesurant l'ensemble des incohérences trouvées peu permettre de faciliter l'identification d'une potentielle source d'erreur.*

## 7 Annexes

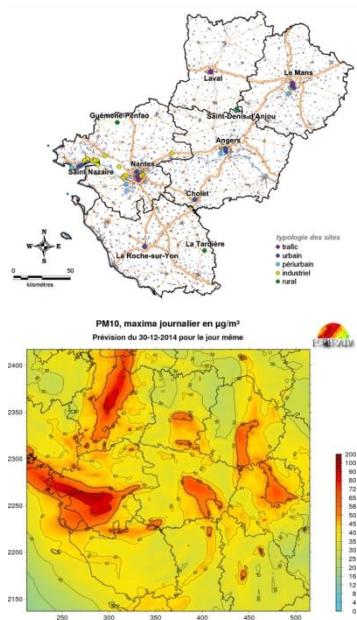
- Annexe 1 : Air Pays de la Loire
- Annexe 2 : Polluants
- Annexe 3 : Carte ATMO France
- Annexe 4 : Equipe BASEMIS®
- Annexe 5 : Rapport secteur résidentiel
- Annexe 6 : Rapport secteur routier
- Annexe 7 : Rapport secteur agricole
- Annexe 8 : Application qualité de la donnée
- Annexe 9 : Documentation – architecture des fichiers
- Annexe 10 : Support de formation et guide d'utilisation
- Annexe 11 : Documentation – Logigrammes des scripts
- Annexe 12 : Liens de formation et tutoriels
- Annexe 13 : Table des schémas, cartes et illustrations

## 7.1 Annexe 1 : Air Pays de la Loire

Dotée d'une solide expertise riche de trente ans d'expérience, Air Pays de la Loire est agréée par le Ministère de la Transition écologique et solidaire pour surveiller la qualité de l'air de la région des Pays de la Loire. Air Pays de la Loire regroupe de manière équilibrée l'ensemble des acteurs de la qualité de l'air : services de l'État et établissements publics, collectivités territoriales, industriels et associations et personnalités qualifiées.

Air Pays de la Loire mène deux missions d'intérêt général : surveiller et informer.

### Surveiller pour savoir et comprendre



### Informier pour prévenir



#### L'air de la région sous haute surveillance

Fonctionnant 24 heures sur 24, le dispositif permanent de surveillance est constitué d'une trentaine de sites de mesure, déployés sur l'ensemble de la région : principales agglomérations, zones industrielles et zones rurales.

#### Mesurer où et quand c'est nécessaire

Air Pays de la Loire s'est doté de systèmes mobiles de mesure (laboratoires mobiles, prélevageurs...). Ces appareils permettent d'établir un diagnostic complet de la qualité de l'air dans des secteurs non couverts par le réseau permanent. Des campagnes de mesure temporaires et ciblées sont ainsi menées régulièrement sur l'ensemble de la région.

#### La fiabilité des mesures garantie

Les mesures de qualité de l'air consistent le plus souvent à détecter de très faibles traces de polluants. Elles nécessitent donc le respect de protocoles très précis. Pour assurer la qualité de ces mesures, Air Pays de la Loire dispose d'un laboratoire d'étalonnage, airpl.lab accrédité par le Cofrac et raccordé au Laboratoire National d'Essais.

#### Simuler et cartographier la pollution

Pour évaluer la pollution dans les secteurs non mesurés, Air Pays de la Loire utilise des logiciels de modélisation. Ces logiciels simulent la répartition de la pollution dans le temps et l'espace et permettent d'obtenir une cartographie de la qualité de l'air. La modélisation permet par ailleurs d'estimer l'impact de la réduction, permanente ou ponctuelle, des rejets polluants. Elle constitue un outil d'aide à la décision pour les autorités publiques compétentes et les acteurs privés.

#### Prévoir la qualité de l'air

Si le public souhaite connaître la pollution prévue pour le lendemain afin de pouvoir adapter ses activités, les autorités politiques ont, elles, besoin d'anticiper les pics de pollution pour pouvoir prendre les mesures adaptées. En réponse à cette attente, Air Pays de la Loire réalise des prévisions de la pollution atmosphérique grâce à la plateforme interrégionale ESMERALDA.

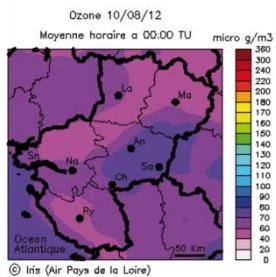
#### Pics de pollution : une vigilance permanente

En cas d'épisode de pollution, une information spécifique est adressée aux autorités publiques, aux médias et à tous les internautes inscrits gratuitement. Suivant les concentrations de pollution atteintes, le préfet de département prend, si nécessaire, des mesures visant à réduire les émissions de polluants (limitations de vitesse, diminution d'activités industrielles...)

#### Sur Internet : tous les résultats, tous les dossiers

Le site Internet [www.airpl.org](http://www.airpl.org) donne accès à de très nombreuses informations sur la qualité de l'air des Pays de la Loire. Elles sont actualisées toutes les heures. On y trouve les cartes de pollution et de vigilance, les communiqués d'alerte, les indices de la qualité de l'air, les mesures de pollution heure par heure, les actualités, toutes les publications d'Air Pays de la Loire...

## Modélisation



### Le système de prévision IRIS

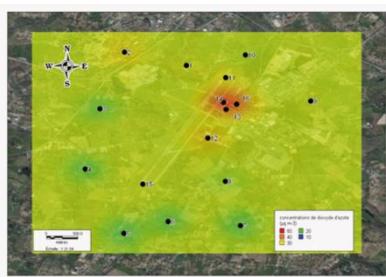
La plateforme inter-régionale ESMERALDA réalise quotidiennement une prévision de la qualité de l'air sur la région des Pays de la Loire.

Elle comprend un ensemble de modèles déterministes (météo, émissions, chimie-transport) permettant d'obtenir des cartes horaires d'ozone, de dioxyde d'azote et de particules fines pour les 4 échéances allant de la veille au surlendemain.



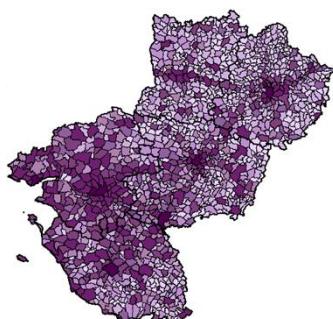
### Modélisation de la pollution de fond et industrielle : ADMS

ADMS est un système de modélisation qui permet de simuler la pollution provenant de sources urbaines (trafic automobile, chauffage, ...) et des sources fixes (établissements industriels). Ce système nécessite la connaissance de paramètres géophysiques (relief, occupation des sols, ...), de la météorologie et des émissions. ADMS est utilisé pour la production de cartographies ou bien pour évaluer des scénarii.



### Les systèmes d'interpolation

Ces techniques permettent de calculer la pollution entre les points de mesure dans le but de réaliser des cartographies. Air Pays de la Loire utilise des techniques d'interpolation géostatistique basées sur la variation des concentrations en fonction de la distance entre les sites de mesures. Ces méthodes permettent également d'intégrer des données auxiliaires (émissions, ...). Elles peuvent être ainsi employées pour des approches hybrides combinant les données de mesure et de modélisation pour une représentation fidèle de la réalité comme cela est le cas au sein de la plateforme inter-régionale ESMERALDA.



### BASEMIS®

Inventaire régional à résolution communale des émissions de gaz à effet de serre et des consommations d'énergie élaboré et mis en œuvre par Air Pays de la Loire. Développé avec le soutien de l'ADEME, de la Région des Pays de la Loire et de l'Union européenne, cet inventaire est utilisé par les services de l'État et des collectivités dans le cadre de la réalisation de leur Plan Climat Énergie Territorial, pour l'élaboration et le suivi du Schéma Régional Climat Air Energie, et pour le suivi de plans régionaux (PPA, PDU, etc.). BASEMIS® est également un élément d'entrée essentiel dans le cadre de la modélisation de la qualité de l'air.

## 7.2 Annexe 2 : Polluants et Gaz à effet de serre

### Les oxydes d'azote (NOx)

Les NOx comprennent essentiellement le monoxyde d'azote (NO) et le dioxyde d'azote (NO<sub>2</sub>). Ils résultent de la combinaison de l'azote et de l'oxygène de l'air à haute température. Environ 95 % de ces oxydes sont la conséquence de l'utilisation des combustibles fossiles (pétrole, charbon et gaz naturel). Le trafic routier (53 %) en est la source principale. Ils participent à la formation des retombées acides. Sous l'action de la lumière, ils contribuent à la formation d'ozone au niveau du sol (ozone troposphérique).

Le monoxyde d'azote présent dans l'air inspiré passe à travers les alvéoles pulmonaires, se dissout dans le sang où il limite la fixation de l'oxygène sur l'hémoglobine. Les organes sont alors moins bien oxygénés.

Le dioxyde d'azote pénètre dans les voies respiratoires profondes. Il fragilise la muqueuse pulmonaire face aux agressions infectieuses, notamment chez les enfants. Aux concentrations rencontrées habituellement, le dioxyde d'azote provoque une hyperréactivité bronchique chez les asthmatiques.

### Les particules fines (PM10, PM2.5)

Les particules fines ou poussières constituent en partie la fraction la plus visible de la pollution atmosphérique (fumées). Elles ont pour origine les différentes combustions, le trafic routier et les industries. Elles sont de nature très diverse et peuvent véhiculer d'autres polluants comme des métaux lourds ou des hydrocarbures. De diamètre inférieur à 10 µm (PM10), elles restent plutôt en suspension dans l'air. Supérieures à 10 µm, elles se déposent, plus ou moins vites, au voisinage de leurs sources d'émission. Les particules plus fines, appelées PM2,5 (diamètre inférieur à 2,5 µm) pénètrent plus profondément dans les poumons. Celles-ci peuvent rester en suspension pendant des jours, voire pendant plusieurs semaines et parcourir de longues distances.

La profondeur de pénétration des particules dans l'arbre pulmonaire est directement liée à leurs dimensions, les plus grosses étant arrêtées puis éliminées au niveau du nez et des voies respiratoires supérieures. Le rôle des particules en suspension a été montré dans certaines atteintes fonctionnelles respiratoires, le déclenchement de crises d'asthme et la hausse du nombre de décès pour cause cardio-vasculaire ou respiratoire, notamment chez les sujets sensibles (enfants, bronchitiques chroniques, asthmatiques...).

### Les gaz à effets de serre (GES)

Plus d'une quarantaine de gaz à effet de serre ont été recensés par le Groupe Intergouvernemental d'Experts sur l'Evolution du Climat (GIEC) parmi lesquels figurent : la Vapeur d'eau (H<sub>2</sub>O), le Dioxyde de carbone (CO<sub>2</sub>), le Méthane (CH<sub>4</sub>), l'Ozone (O<sub>3</sub>), le Protoxyde d'azote (N<sub>2</sub>O), l'Hydrofluorocarbures (HFC), le Perfluorocarbures (PFC) et l'Hexafluorure de soufre (SF<sub>6</sub>). Dans les analyses réalisées, les GES sont principalement calculés à partir du CO<sub>2</sub>, CH<sub>4</sub> et N<sub>2</sub>O dans le secteur agricole, les autres gaz étant presque non présent.

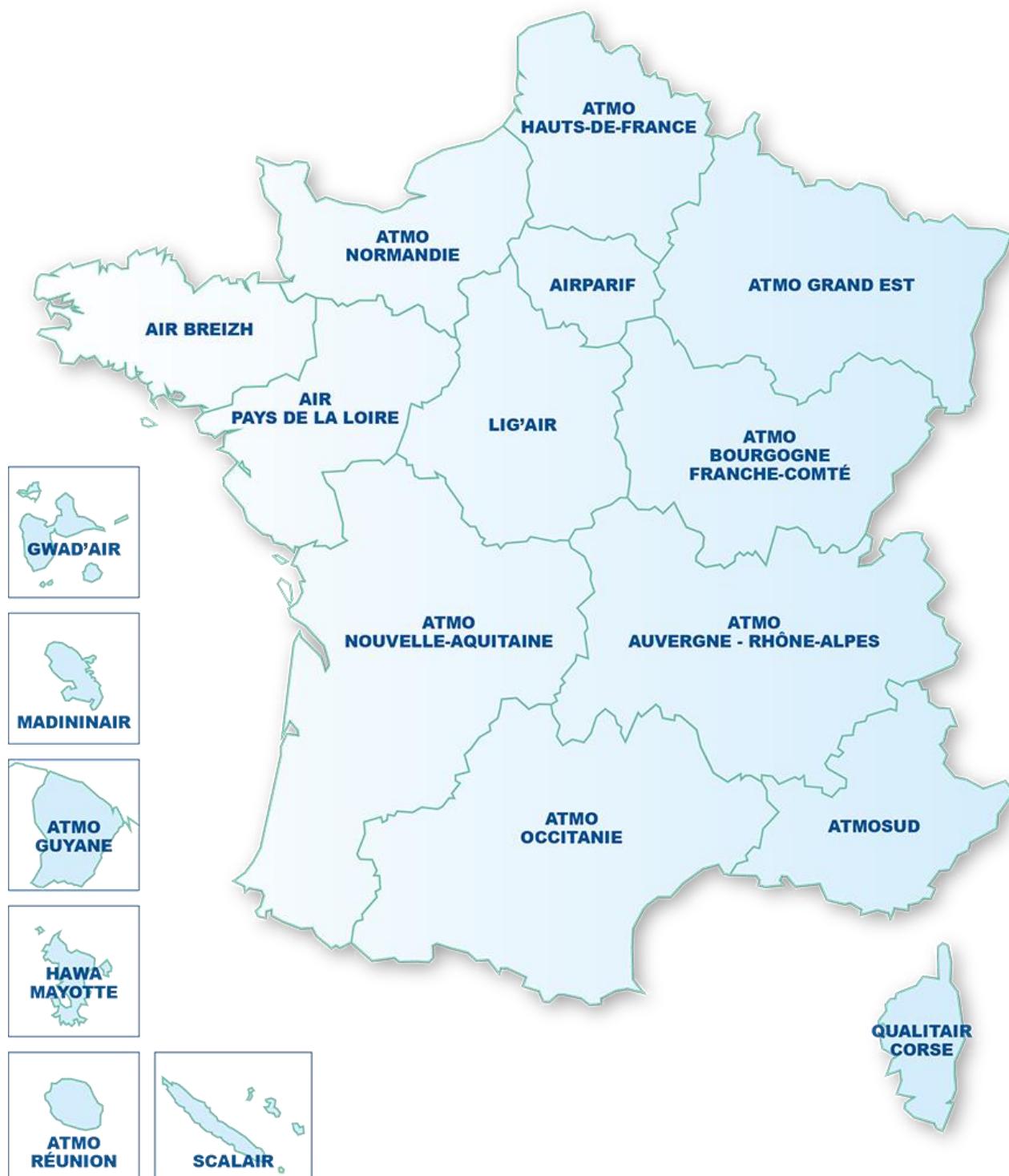
### Le méthane (CH<sub>4</sub>)

Le méthane est le constituant principal du gaz naturel, combustible d'origine fossile. Il est libéré dans l'atmosphère quand la matière organique se décompose dans des environnements avec de faibles niveaux d'oxygène. Il contribue fortement à l'effet de serre tandis que sa durée de vie dans l'atmosphère est de l'ordre de la décennie. Les sources synthétiques incluent l'exploitation et la brûlure des combustibles fossiles, les processus digestifs chez les ruminants tels que les bétails et les sites d'enfouissement des déchets. La plus grande partie du méthane émis est décomposé dans l'atmosphère par les réactions avec les radicaux d'hydroxyle (OH).

### Le protoxyde d'azote (N<sub>2</sub>O)

Le Protoxyde d'azote (également appelé Oxyde nitreux ou proto) de formule chimique N<sub>2</sub>O est un puissant gaz à effet de serre qui subsiste longtemps dans l'atmosphère (environ 120 ans.) Il est en partie responsable de la destruction de l'ozone. Le sol et les océans sont les principales sources naturelles de ce gaz, mais il est également produit par l'utilisation d'engrais azotés, la combustion de matière organique et de combustibles fossiles, la production de nylon... En France, l'agriculture contribuerait aux 3/4 des émissions de N<sub>2</sub>O provenant essentiellement de la transformation des produits azotés (engrais, fumier, lisier, résidus de récolte) dans les sols agricoles. Le N<sub>2</sub>O est un gaz incolore et ininflammable, stable dans les basses couches de l'atmosphère mais décomposé dans les couches plus élevées (stratosphère) par des réactions chimiques impliquant la lumière du soleil.

## 7.3 Annexe 3 : Cartes ATMO France



Carte 1 Découpages des AASQA issue de ATMO France

## 7.4 Annexe 4 : équipe BASEMIS®



## 7.5 Annexe 5 : Rapport secteur résidentiel

### Analyses des données énergétiques BASEMIS du secteur résidentiel

23 juillet, 2020 - Air Pays de la Loire

- 1 - Le secteur résidentiel : Les indicateurs
- 2 - Résultats de l'année 2013: Comportements des indicateurs
  - 2.1 - Consommation d'énergie liée à la cuisson, l'eau chaude sanitaire et prises électriques
  - 2.2 - Consommation d'énergie liée au chauffage
- 3 - Résultats des valeurs abbréviées
- 4 - Qualité des données

#### 1 - Le secteur résidentiel : Les indicateurs

##### 1.1 Information sur les types de données

La base de données BASEMIS représente les données finales de l'inventaire, ses données sont calculées à partir des données INSEE, CITADEL. Pour vérifier la cohérence des données nous allons principalement utiliser les données suivantes

- Consommation (en kWh)
- Type d'énergie utilisé (gaz, électricité, ...)
- Type de résidence (principale, collective, ...)
- Nombre d'habitant par résidence
- Superficie des résidences (en m<sup>2</sup>)

##### 1.2 Méthode de calcul des indicateurs

Au sein de ce secteur, nous avons mis en place trois indicateurs permettant de voir le comportement général des consommations d'énergies :

$$Kwh/hab = \frac{EnergieConsommé}{NombreHabitant}$$

$$Kwh/m^2 = \frac{EnergieConsommé}{SurfaceLogement}$$

$$TauxAnnuelle = \left[ \frac{Kwh/hab_{anneeN}}{Kwh/hab_{anneeN-1}} - 1 \right]$$

##### 1.3 Méthode de traitements des données :

Jointures des données par :

- Numcom
- Année
- Rubrique

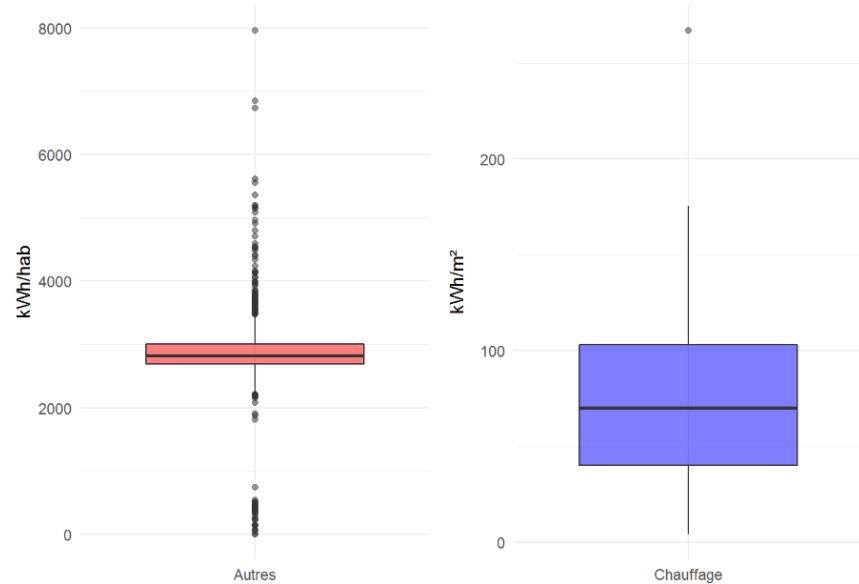
Point de vigilance à vérifier :

- Présence de gaz / réseaux de chaleur dans la ville possible ?

## [1] "Consommation possible valide"

##### 1.4 Information sur les consommations moyennes

Graphique 0 : Consommations moyenne d'énergie

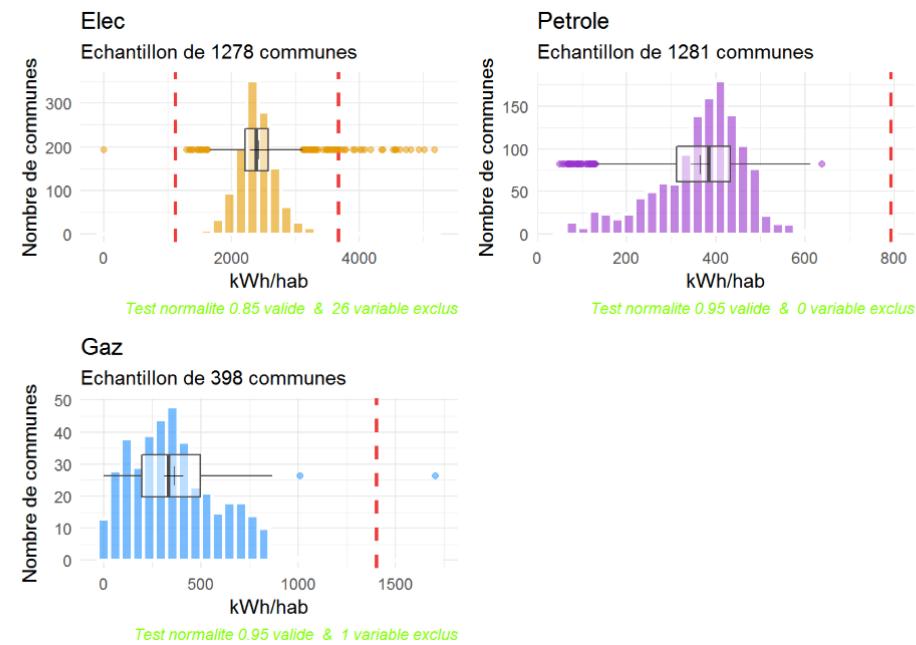


## 2 - Résultats de l'année 2013: Comportements des indicateurs

### 2.1 - Consommation d'énergie liée à la cuisson, l'eau chaude sanitaire et prises électriques

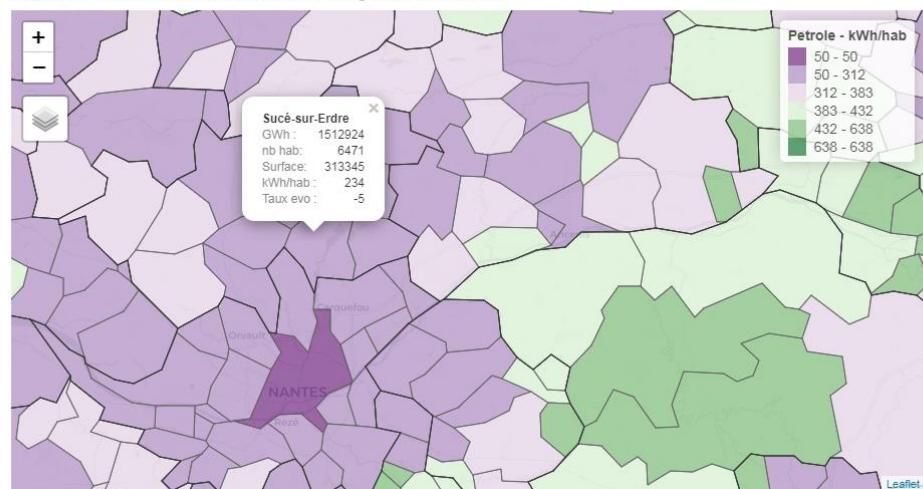
#### 2.1.1 - Consommation énergétique par type d'énergie

Graphique 1 : Distribution des consommations d'énergies en 2013



Cartographie 1 : Cartographie des consommations d'énergies en 2013

choix des indicateurs à cocher/décocher dans la gestion des couches



### 3 - Résultats des valeurs aberrantes

Table 1 : Communes dont l'indicateur semble anormale

Nom de la commune	ANNEE	NAPFUE	Energie	indice	ratio
<b>Chauffage</b>					
SAINTE-HILAIRE-LA-FORET	2016	Gaz	9948559.47 kWh	266 kWh/m2	1.7
MOUILLERON-SAINT-GERMAIN	2016	Gaz	18623500.5 kWh	244 kWh/m2	1.6
NESMY	2016	Gaz	22876168.49 kWh	183 kWh/m2	1.2
SENONNES	2016	Elec	236047.9 kWh	15 kWh/m2	1
NOIRMOUTIER-EN-L'ILE	2016	Elec	3547138.93 kWh	16 kWh/m2	1
NAUVAY	2016	Elec	1.66 kWh	0.0024 kWh/m2	0

Table 2 : Communes dont l'évolution semble anormale

Nom de la commune	ANNEE	NAPFUE	Energie	indice	ratio
<b>Chauffage</b>					
NESMY	2016	Gaz	4663116.83 kWh	2723 %	131.8
MOUILLERON-SAINT-GERMAIN	2016	Gaz	5252690.31 kWh	2054 %	99.4
MONTREVAULT-SUR-EVRE	2016	Gaz	799496.39 kWh	779 %	37.7
BEAULIEU-SOUS-LA-ROCHE	2016	Gaz	427707.03 kWh	622 %	30.1

### 4 - Qualité des données

Table 3 : Estimation de la qualité des données

ANNEE	Indicateur kWh/m2			Taux d'évolution %		
	Taux err	NB err	Ecart max	Taux err2	NB err2	Ecart max2
<b>Autres</b>						
2016	0.2 %	6	7.0	1.21 %	36	131.8
<b>Chauffage</b>						
2016	0.14 %	6	1.7	0.73 %	31	96.3

## 7.6 Annexe 6 : Rapport secteur routier

### Analyses linéaires des données primaires de de 'CEREMA' dans le secteur routier

22 juillet, 2020 - Air Pays de la Loire

- 1 - Le secteur routier : Les indicateurs de qualité
- 2 - Résultats de l'année 2016 : Comportements de l'évolution temporelle des voiries
- 3 - Résultats des 100 valeurs abberantes ayant le plus haut ratio
- 4 - Visualisation cartographique
- 5 - Qualité des données

#### 1 - Le secteur routier : Les indicateurs de qualité

##### 1.1 Information sur les types de données

La base de données CIRCULAIRE de l'année V5 contient les données primaires qui permettent de calculer l'inventaire des émissions du secteur routier. Dans cette base de données, nous allons analyser la qualité des champs suivants :

- TMJA (moyenne de véhicules circulant par jours dans une année)
- Geom (géométrie de la route)
- Long (longueur de la géométrie en km)
- nb\_voies / RUBRIQUE (type de voiries, autoroute, départementales, ...)

##### 1.2 Méthode de calcule des indicateurs

Au sein de ce secteur, nous souhaitons voir l'évolution du TMJA entre deux années

$$Taux\ evolution = \frac{TMJA_{anneeN}}{TMJA_{anneeN-1}} - 1$$

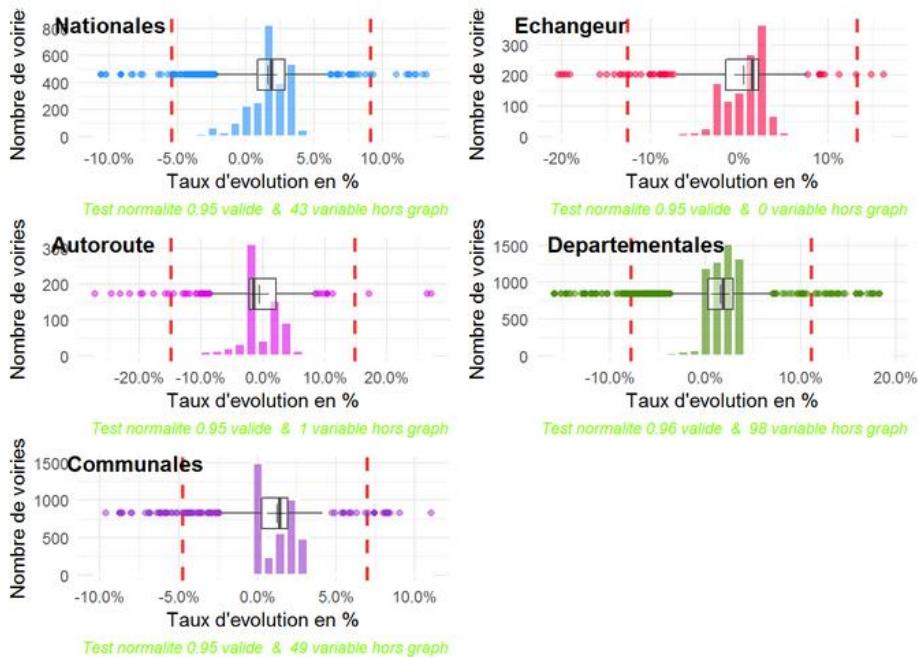
##### 1.3 Méthode de traitements spatiaux :

- Sélection des géométries identiques
- Calcule du taux d'évolution entre années
- Calcule de la différence de TMJA entre années

*Point de vigilance : Si la route est séparé en deux tel que A = A1 + A2 d'une année à l'autre, le taux d'évolution n'est pas calculé, car la méthode de traitement ne considère pas les géométries identiques*

#### 2 - Résultats de l'année 2016 : Comportements de l'évolution temporelle des voiries

Graphique 1 : Taux d'évolutions des voiries



## 3 - Résultats des 100 valeurs abbérantes ayant le plus haut ratio

**Table 1 :** Voiries dont l'évolution semble anormale

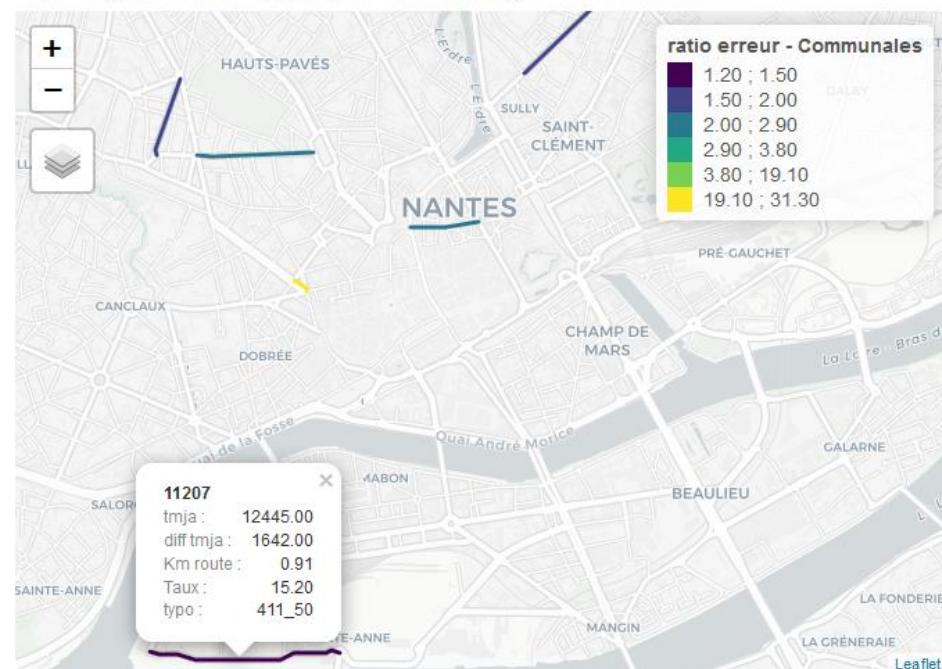
ATTENTION visualisation des 100 variables sur les 164

	id	TMJA	diff_tmja	long	indice	ratio
<b>Communales</b>						
	14929	6356	5016	0.08 km	374 %	31.3
	7392	1157	-2389	0.68 km	-67 %	6.9
	1664	1241	-807	0.81 km	-39 %	4
<b>Departementales</b>						
	7352	3374	2104	1.31 km	166 %	8.6
	13547	3374	2104	0.86 km	166 %	8.6

## 4 - Visualisation cartographique

**Cartographie 1 :** Voiries dont l'évolution est anormale

choix du type de voirie à cocher/décocher dans la gestion des couches



## 5 - Qualité des données

**Table 2 :** Estimation de la qualité des données

RUBRIQUE	ANNEE	Taux erreur	NB erreur	Ecart max
Communales	2016	0.97 %	40	31.3
Departementales	2016	1.49 %	89	8.6
Nationales	2016	1.28 %	35	6.9

## 7.7 Annexe 7 : Rapport secteur agricole

### Analyses des données BASEMIS dans le secteur agricole

10 août, 2020 - Air Pays de la Loire

- 1 - Le secteur agricole : Les indicateurs de qualité
- 2 - Résultats de l'année 2016: Comportements des indicateurs spécifiques pour chaque sous secteurs
  - 2.1 - Sous Secteur : Elevage
  - 2.2 - Sous Secteur : Culture
  - 2.3 - Sous Secteur : Tracteur
- 3 - Résultats des valeurs abbréantes
- 4 - Qualité des données

#### 1 - Le secteur agricole : Les indicateurs de qualité

##### 1.1 Information sur les types de données

La base de données BASEMIS représente les données finales de l'inventaire, ses données, pour le secteur agricole sont calculées à partir des données partenaires de la DRAAF et AGRESTE. Pour vérifier la cohérence des données nous allons principalement utiliser les données suivantes :

- CO2 (en kg)
- CH4 (en kg)
- N2O (en kg)
- GES (Total en Tonne)
- PM10 (en kg)
- Type d'activité (Culture, Elevage, Tracteurs, ...)
- Nombre de bêtes
- Type de bêtes (Volailles, Bovins, ...)
- Surface culture (ha)

##### 1.2 Méthode de calcul des indicateurs

Au sein de ce secteur, nous avons mis en place 3 indicateurs permettant de voir le comportement général des émissions de **GES** et **PM10** dans les communes :

$$GES = (CO2 * 0.001) + (CH4 * 0.028) + (N2O * 0.265)$$

$$Ind\_GES\_Betes = \frac{GES}{nombre\_bete}$$

$$Ind\_GES\_surf = \frac{GES}{surface\_ha}$$

$$Ind\_GESPM = \frac{GES}{PM10}$$

##### 1.3 Méthode de traitements :

- Séparation des données en fonction des sous secteurs
- Jointures des données spécifiques aux secteurs
- Calculs des indicateurs

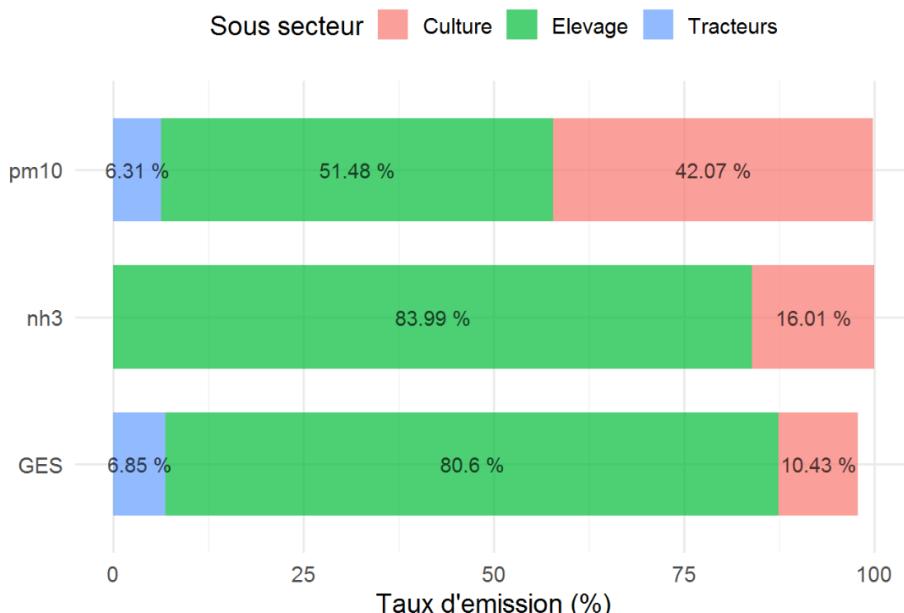
Point de vigilance à vérifier :

- S'assurer que pour chaque émission de GES ou PM10 les données de contexte (Surface, nombre de bête) soient présentes.

##### 1.4 Information sur les émissions moyennes

Graphique 0 : Emission du secteur agricole sur l'année 2016

Représentation des sous-catégories principales uniquement

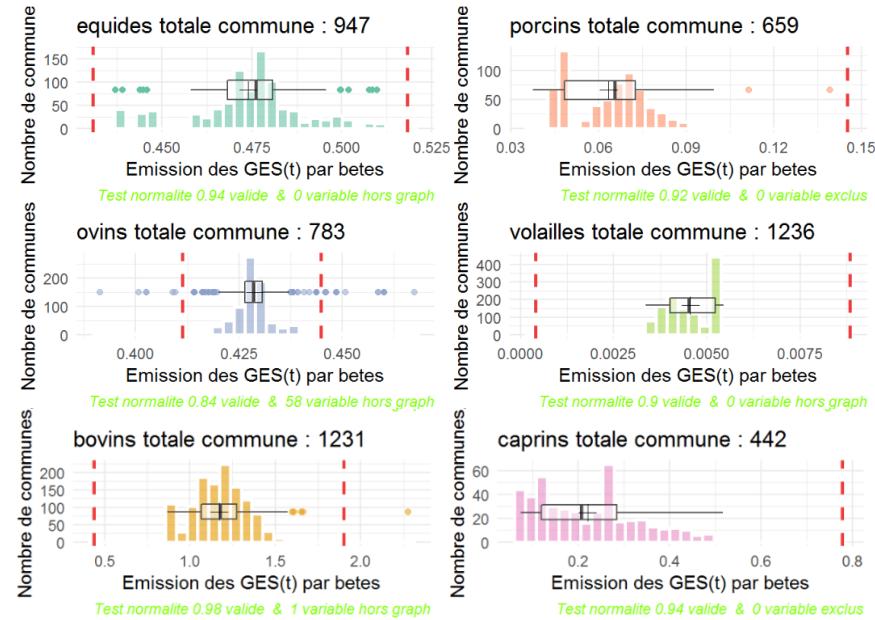


## 2 - Résultats de l'année 2016: Comportements des indicateurs spécifiques pour chaque sous secteurs

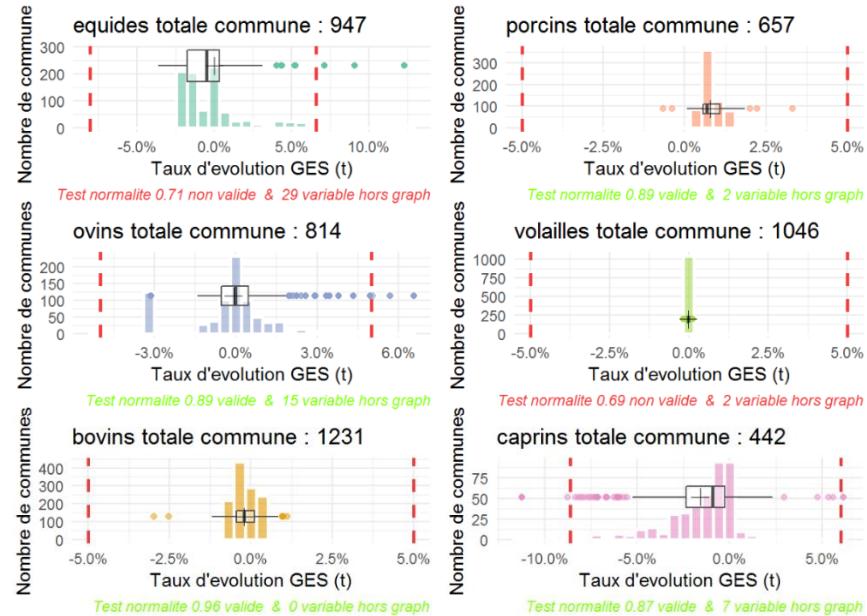
### 2.1 - Sous Secteur : Elevage

#### 2.1.1 Les émissions de GES

Graphique 1 : Comportements des émissions de GES par cheptels

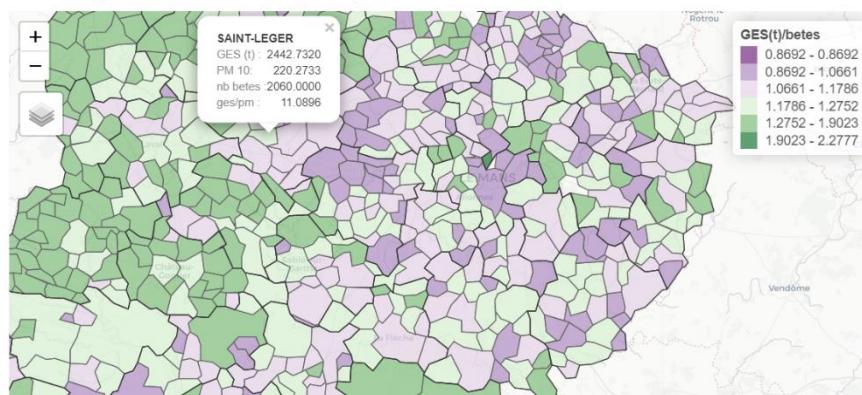


Evolution annuelle



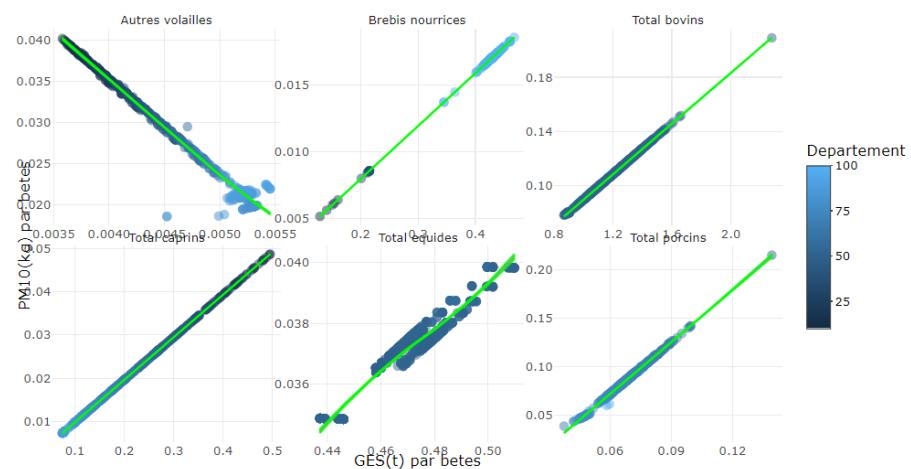
Cartographie 1 : Comportements des émissions de GES par cheptels

choix des catégories animales à cocher/décocher dans la gestion des couches



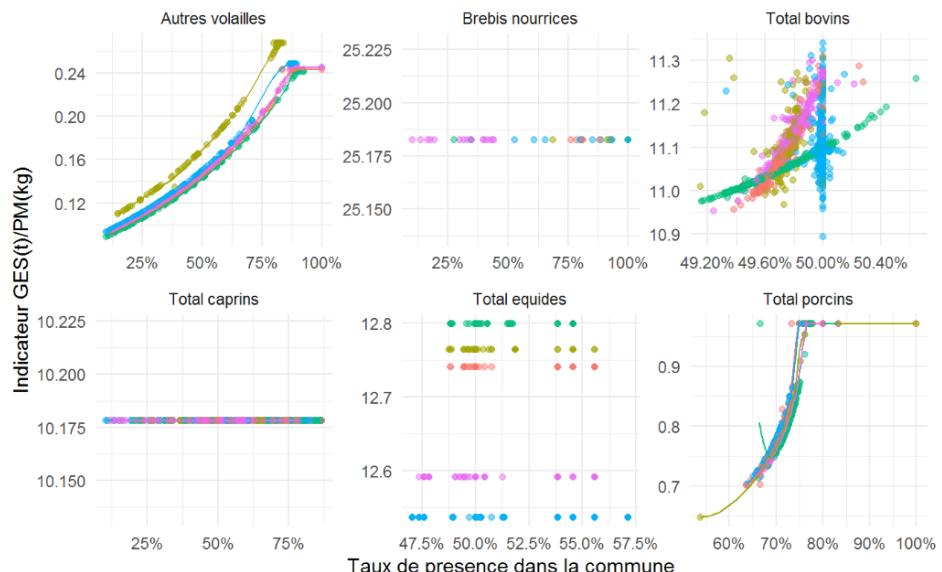
## 2.1.2 Rapport entre GES et PM10

Graphique 2 : Corrélation entre les émissions de GES et PM10

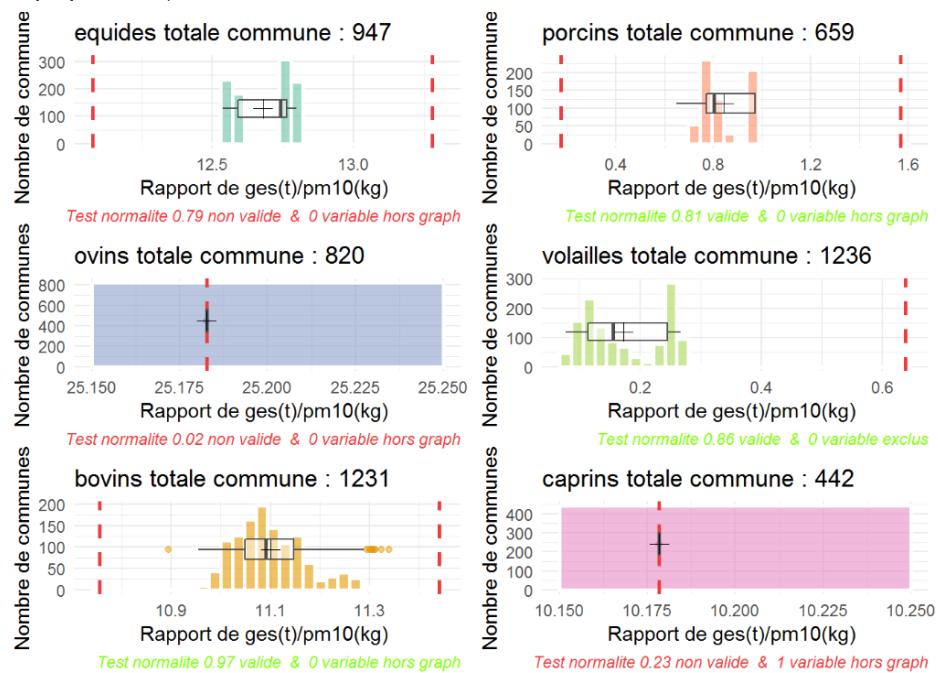


Corrélation en fonction de la présence majoritaire d'une espèce et du département

Departement ● LOIRE-ATLANTIQUE ● MAINE-ET-LOIRE ● MAYENNE ● SARTHE ● VENDEE



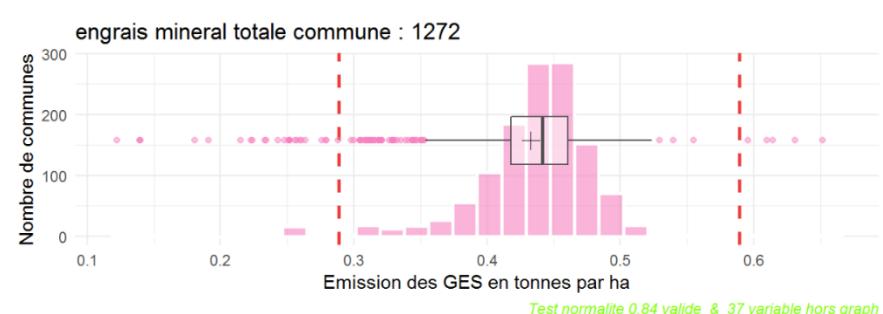
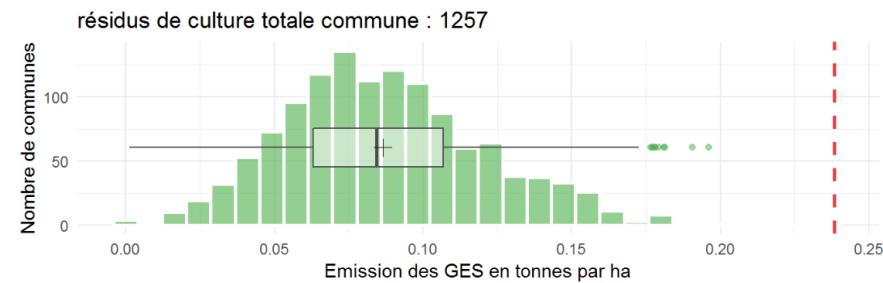
Graphique 3 : Comportements des GES/PM10



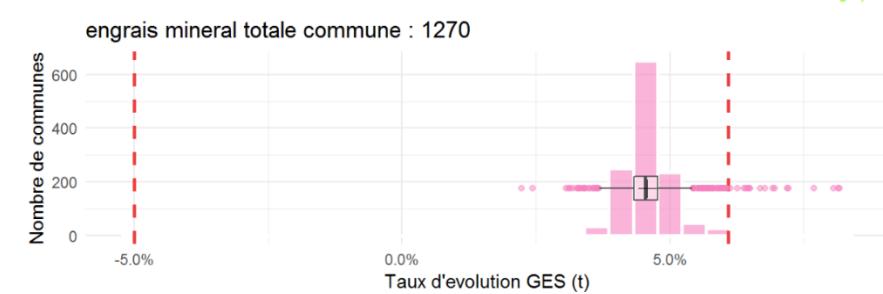
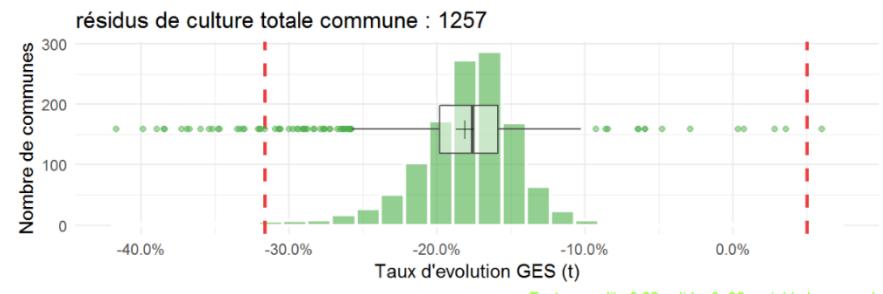
## 2.2 - Sous Secteur : Culture

### 2.2.1 Les émissions de GES

Graphique 3 : Comportements des émissions de GES par surface agricole



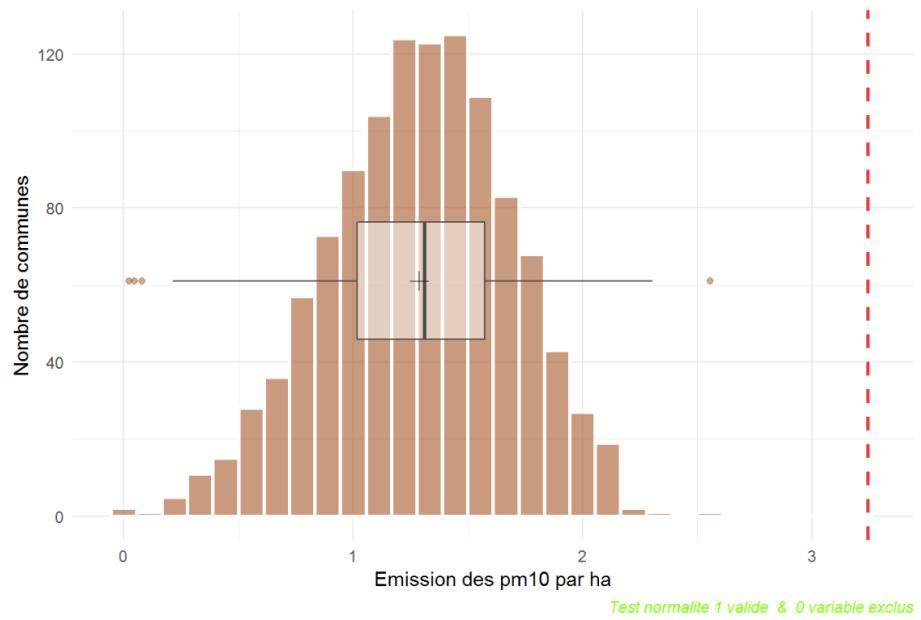
Evolution annuelle



## 2.2.2 Les émissions de PM10

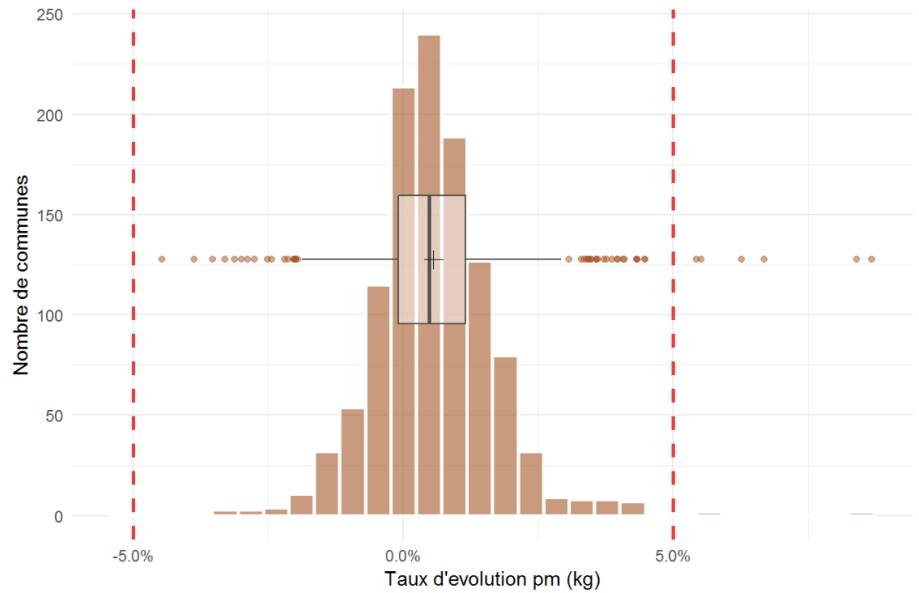
Graphique 3 : Comportements des émissions de pm10 par surface agricole

travail du sol et recolte totale commune : 1147



Evolution annuelle

travail du sol et recolte totale commune : 1144



## 3 - Résultats des valeurs aberrantes

Table 1 : Résultat du sous secteurs Elevage

numcom	ANNEE	tot_tetes	tot_ges	tot_pm	ind_GESPM	indicateur	ratio
<b>bovins</b>							
72310	2016	14	31.8879	2.9270	10.8945	2.2777 ges/b	0.2
<b>ovins</b>							
85260	2016	830	79.1992	3.1450	25.1826	0.0954 ges/b	0.77
85261	2016	1020	97.3369	3.8652	25.1826	0.0954 ges/b	0.77
85304	2016	620	59.1826	2.3501	25.1826	0.0955 ges/b	0.77

numcom	ANNEE	tot_tetes	tot_ges	tot_pm	ind_GESPM	indicateur	ratio
<b>caprins</b>							
72074	2016	32	3.1598	0.3104	10.1783	-11.25 Taux GES	1.3
72129	2016	26	2.5245	0.2480	10.1783	-11.25 Taux GES	1.3
72186	2016	31	3.0577	0.3004	10.1783	-11.25 Taux GES	1.3
72355	2016	16	1.5855	0.1558	10.1783	-11.25 Taux GES	1.3
53003	2016	12	3.4218	0.3362	10.1783	6.15 Taux GES	1
53185	2016	14	3.9921	0.3922	10.1783	6.15 Taux GES	1
72218	2016	36	3.6137	0.3550	10.1783	-8.78 Taux GES	1
<b>ovins</b>							
72062	2016	6	2.7610	0.1096	25.1826	13.03 Taux GES	3.6
72077	2016	6	2.7610	0.1096	25.1826	13.03 Taux GES	3.6
72185	2016	6	2.7610	0.1096	25.1826	13.03 Taux GES	3.6
44023	2016	7	3.2111	0.1275	25.1826	10.72 Taux GES	3
numcom	ANNEE	tot_surf	tot_ges	tot_pm	indicateur	ratio	
<b>engrais mineral</b>							
44117	2016	17.7876588	4.4592	NA	-7.6 % GES	2.5	
49274	2016	722.6517887	261.4505	NA	8.42 % GES	1.4	
49053	2016	301.8729950	95.3404	NA	8.17 % GES	1.3	
49237	2016	3.5850762	1.1507	NA	7.69 % GES	1.3	
<b>travail du sol et recolte</b>							
44117	2016	17.7876588	0.0000	11.1311	-19.9 % PM	5.2	
49237	2016	3.5850762	0.0000	1.7199	16.19 % PM	3.3	
72006	2016	125.5184185	0.0000	31.3400	12.56 % PM	2.6	
49054	2016	29.1007715	0.0000	13.2974	8.68 % PM	1.8	

## 4 - Qualité des données

**Table 4 : Estimation de la qualité des données**

ANNEE	Indicateur emissions			Taux d'évolution %		
	Taux err	NB err	Ecart max	Taux err2	NB err2	Ecart max2
<b>bovins</b>						
2016	0.08 %	1	0.20	0 %	NA	NA
<b>caprins</b>						
2016	0 %	NA	NA	1.58 %	7	1.3
<b>ovins</b>						
2016	7.07 %	58	0.77	1.83 %	15	3.6
<b>porcins</b>						
2016	0 %	NA	NA	0.3 %	2	3.3
ANNEE	Indicateur emissions			Taux d'évolution %		
	Taux err	NB err	Ecart max	Taux err2	NB err2	Ecart max2
<b>engrais mineral</b>						
2016	2.91 %	37	0.58	1.65 %	21	2.5
<b>résidus de culture</b>						
2016	0 %	NA	NA	2.15 %	27	1.5
<b>travail du sol et recolte</b>						
2016	0 %	NA	NA	0.96 %	11	5.2

## 7.8 Annexe 8 : Application qualité de la donnée

### Analyses de la qualité des données BASEMIS

**Choix du secteur**

- Résidentiel

**Choix de l'année**

- 2016

**Version Basemis**

- Version 5

**Générer le rapport**

**Le secteur résidentiel : Les indicateurs**

**1 Information sur les types de données**

La base de données BASEMIS® représente les données finales de l'inventaire, ses données sont calculées à partir des données INSEE, CITADEL. Pour vérifier la cohérence des données nous allons principalement utiliser les données suivantes :

- Consommation (en kWh)
- Type d'énergie utilisé (gaz, électricité, ...)
- Type de résidence (principale, collective, ...)
- Nombre d'habitants par résidence
- Surface des résidences (en m<sup>2</sup>)

**2 Méthode de calcul des indicateurs**

Au sein de ce secteur, nous avons mis en place trois indicateurs permettant de voir le comportement général des consommations d'énergies :

$$Kwh/hab = \frac{EnergieConsommé}{NombreHabitant}$$

$$Kwh/m^2 = \frac{EnergieConsommé}{SurfaceLogement}$$

$$TauxAnnuel = \left[ \frac{Kwh/hab_{annéeN}}{Kwh/hab_{annéeN-1}} - 1 \right]$$

**3 L'analyse qualité**

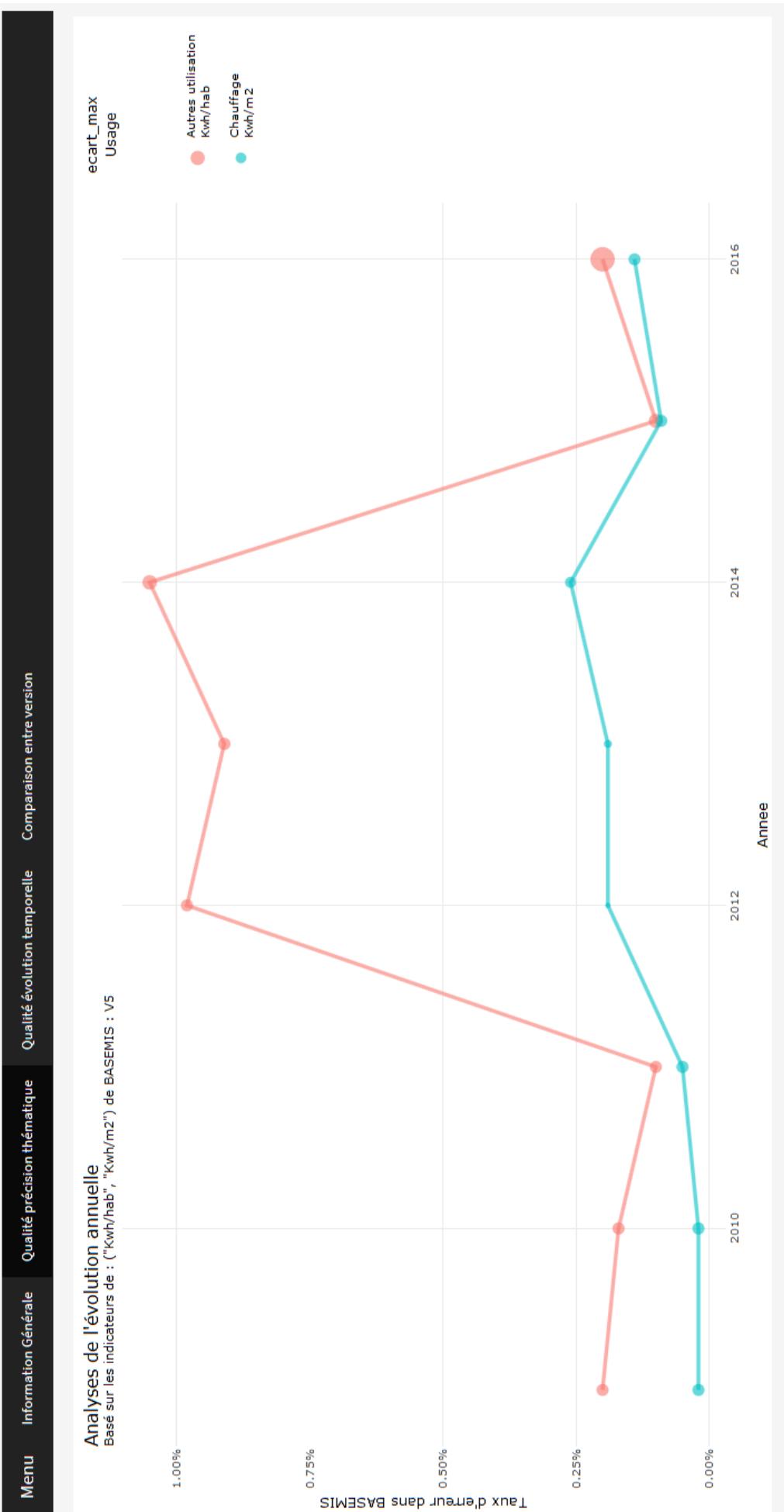
Grâce à ces indicateurs, nous réalisons des analyses statistiques nous permettent de détecter des valeurs aberrantes afin d'identifier des possibles erreurs. Nous avons alors estimé pour chaque sous-secteur la qualité des données en regardant :

- Le nombre total de valeurs aberrantes détecté : Est-ce qu'il y a beaucoup de valeur détecté ?
- Le taux de valeur aberrante présentes dans un sous-secteur : Est-ce que nombre de valeur détectée sont significative dans l'ensemble des données ?
- La gravité des erreurs : Est-ce que les valeurs détectées sont minimales ou très importantes ?

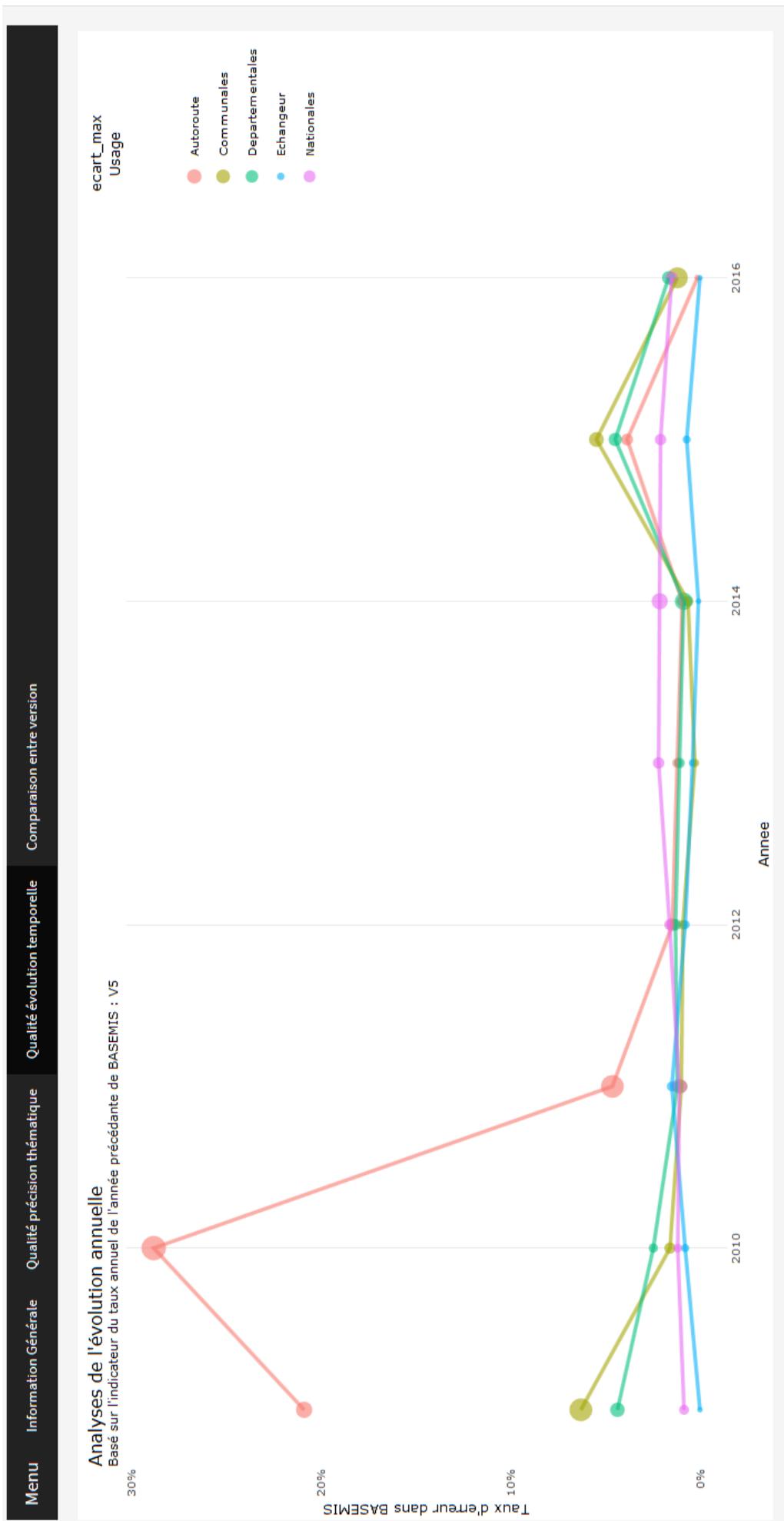
L'interface shiny nous permet alors de visualiser ses trois questions à l'aide de graphique dynamique et interactif. Cependant, si vous souhaitez avoir plus d'information sur le comportement des données sur une année, dans un secteur, il faudra alors lancer la génération d'un rapport.

**4 Méthode de traitements des données :**

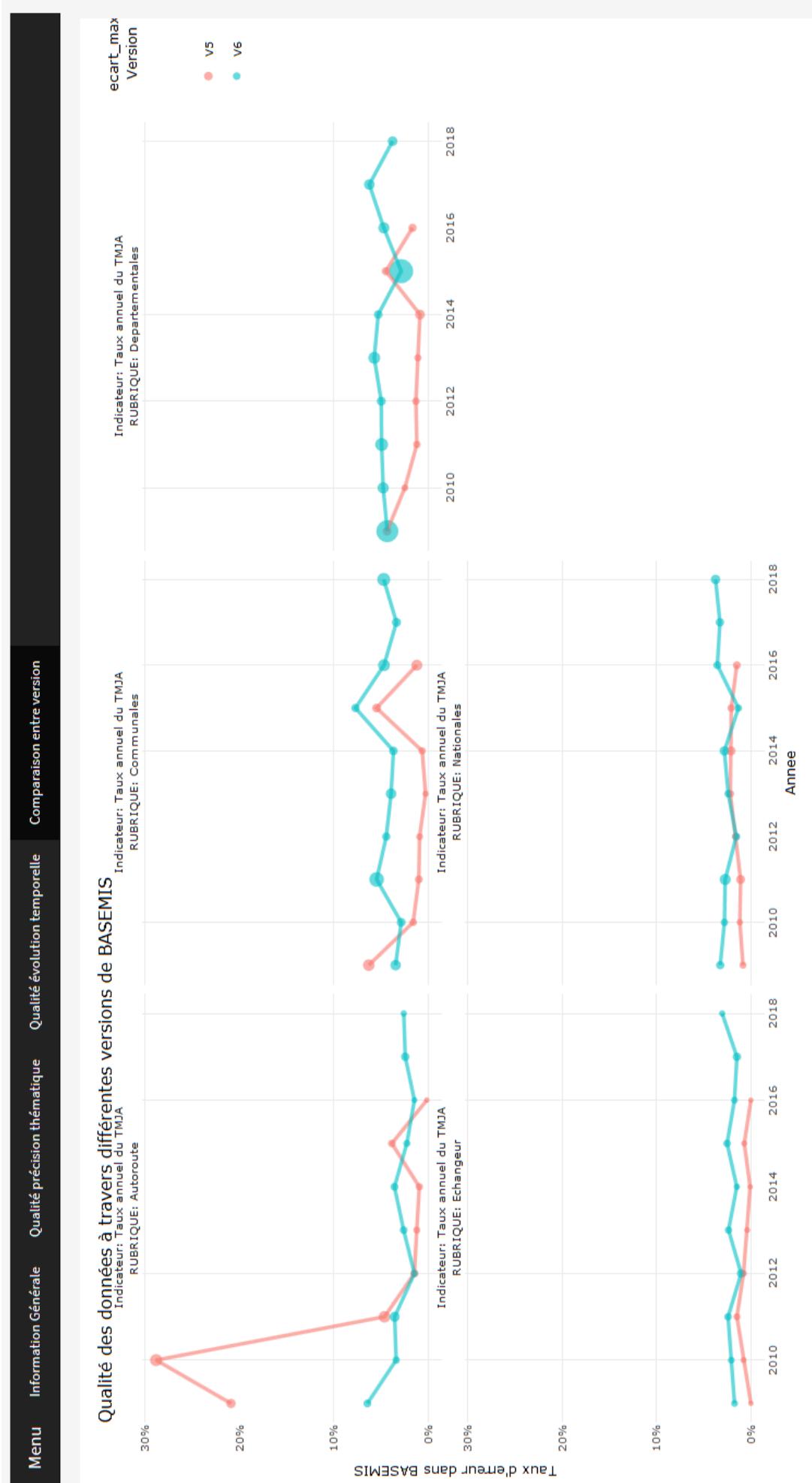
# Annexes



# Annexes



## Annexes

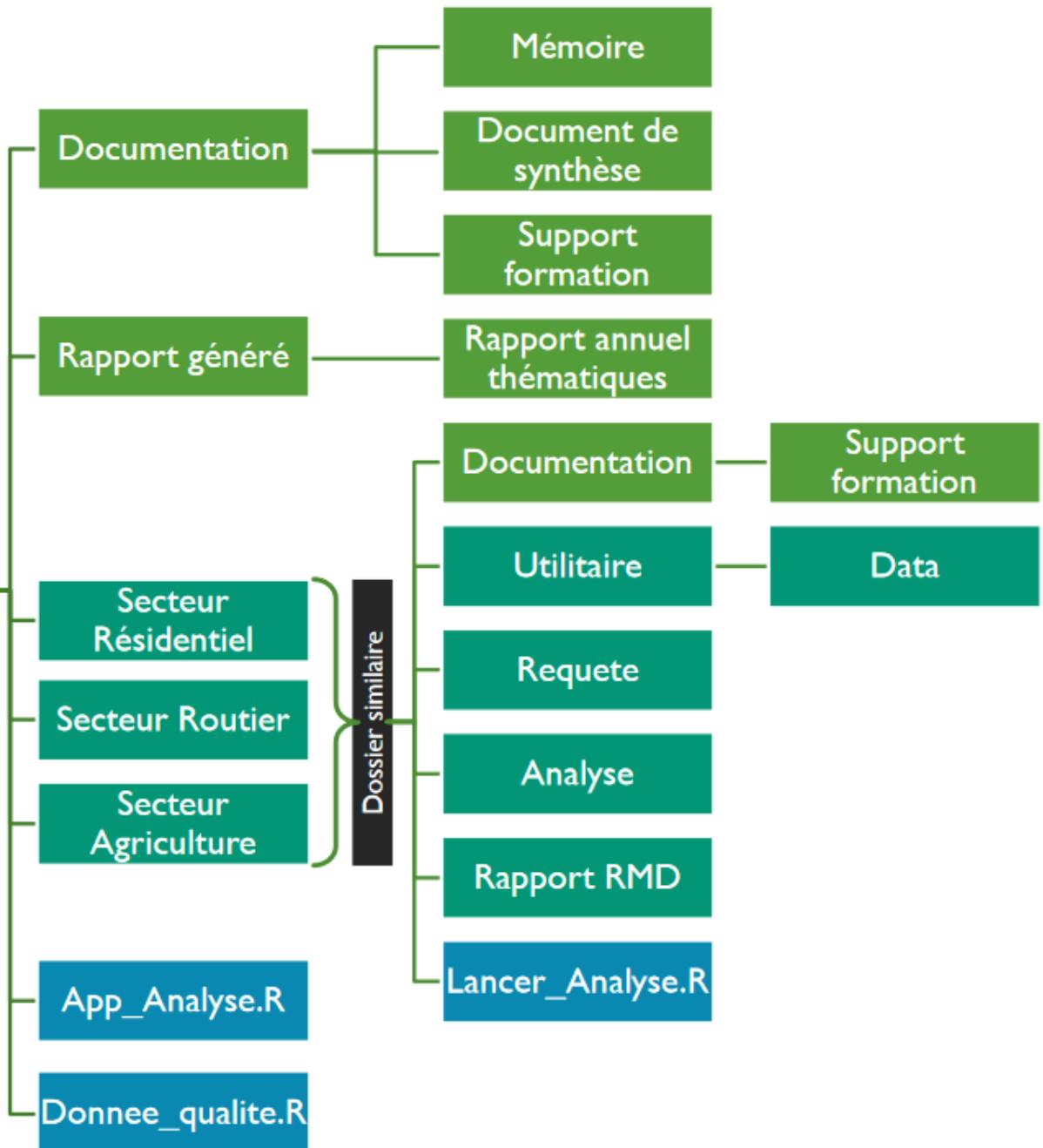


## 7.9 Annexe 9 : Documentation architecture des fichiers

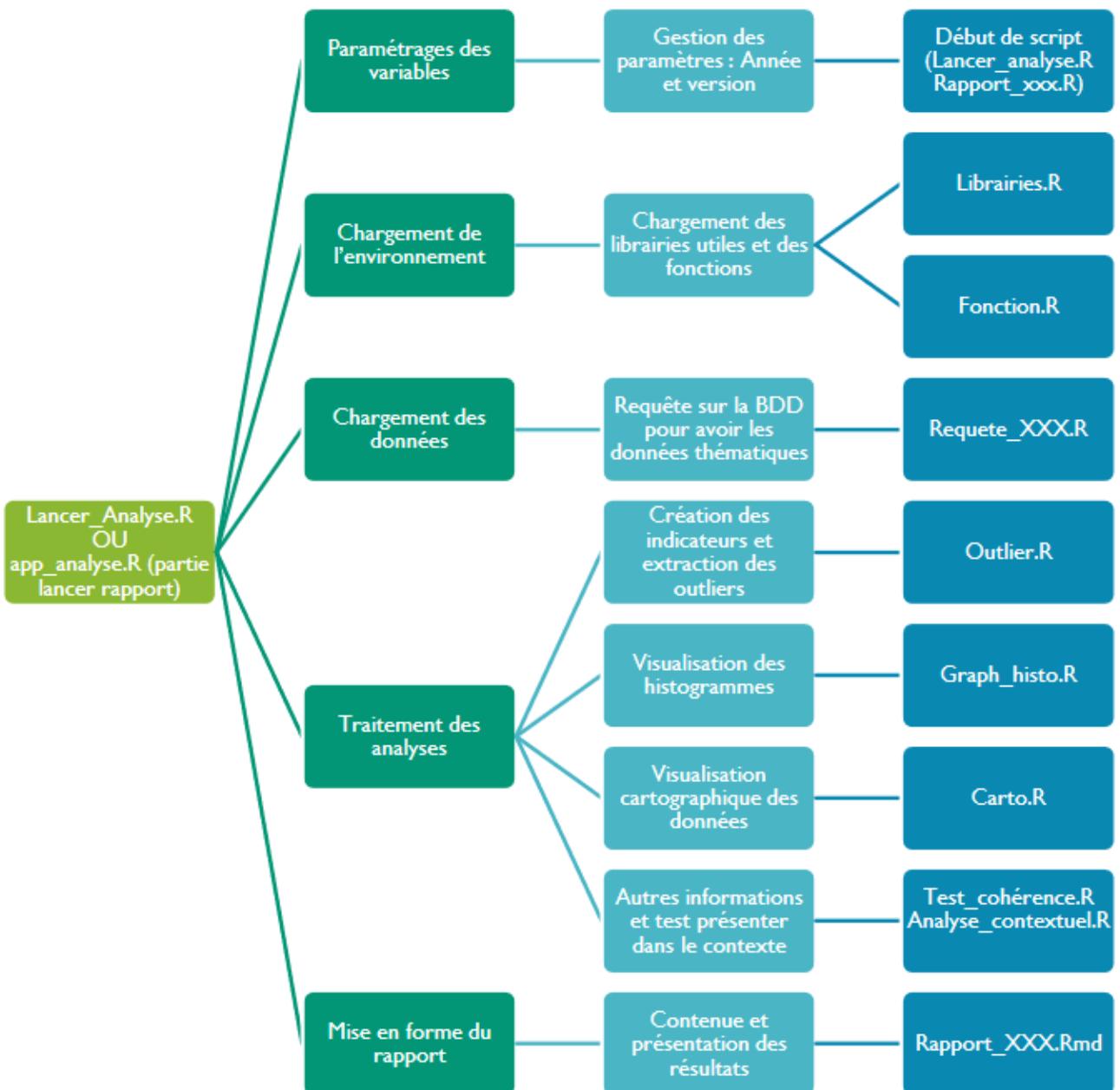
# Annexes



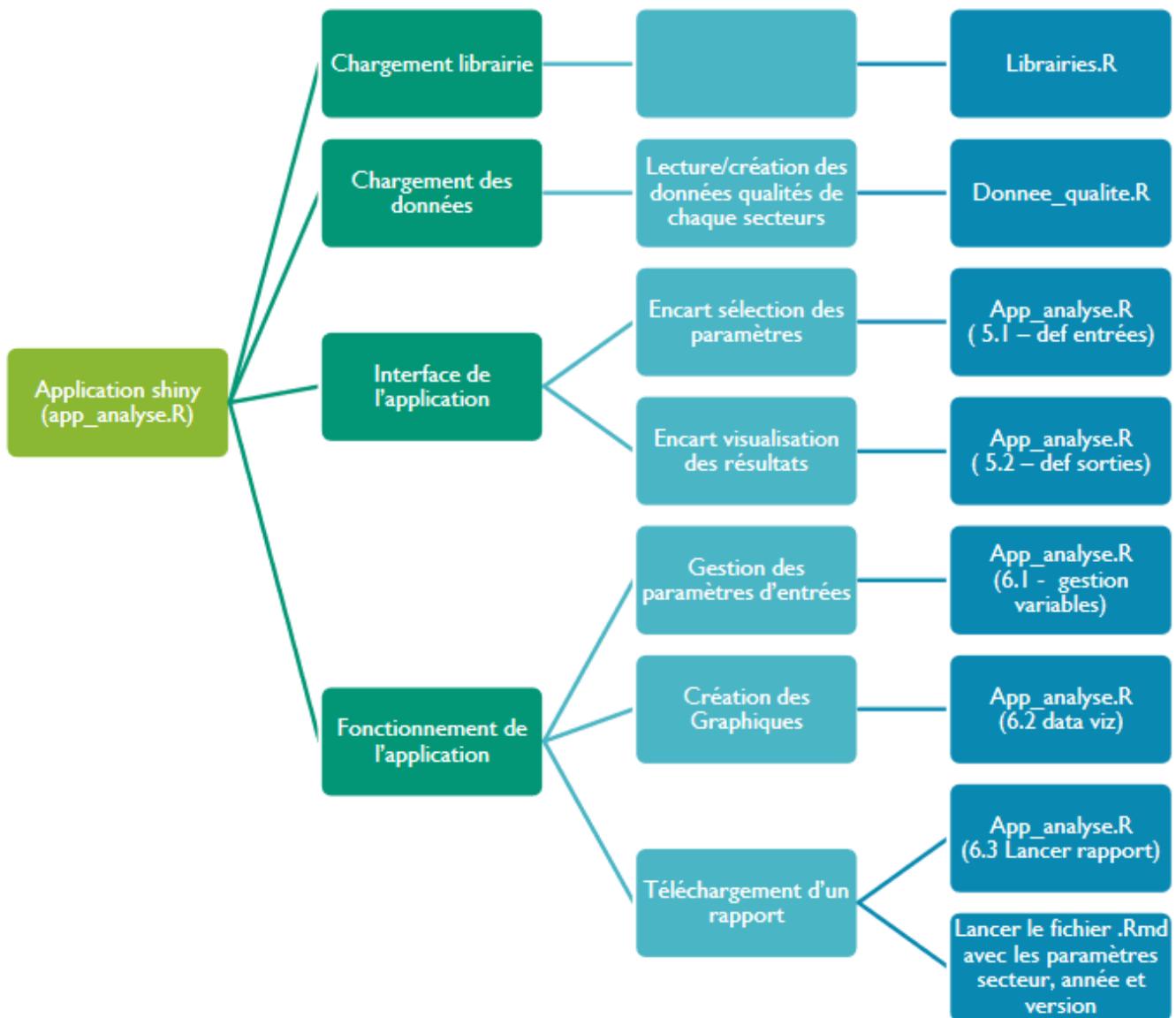
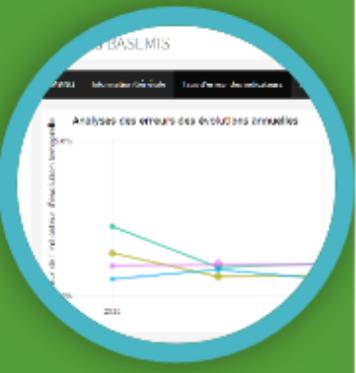
19 Validation Stage 2020



# FONCTIONNEMENT D'UNE GÉNÉRATION DE RAPPORT



## FONCTIONNEMENT GÉNÉRALE DE L'APPLICATION SHINY



## 7.10 Annexe 10 : Guide d'utilisation

# GUIDE DES PRÉCONISATIONS POUR UNE PREMIÈRE UTILISATION DE L'APPLICATION : **VALIDATION DES DONNÉES**

Réalisé Par Axel Fourneyron

Du 18 Mars Au 11 Septembre 2020

Analyse De La Qualité Des Données Climat-air-énergie

## Analyses de la qualité des données BASEMIS

**Choix du secteur**  
Transport Routier

**Choix de l'année**  
2016

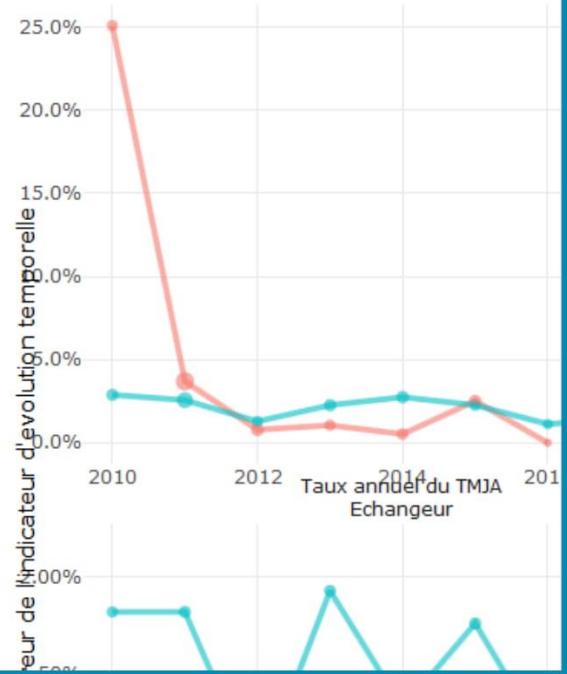
**Version Basemis**  
Version 5

**Générer le rapport**

Menu    Information Générale    Taux d'erreur des indi

### Qualité des données à travers diffé

Taux annuel du TMJA  
Autoroute



## LOGICIELS ET CONNEXION RÉSEAUX NÉCESSAIRES POUR LANCER L'APPLICATION

*L'installation des logiciels ne requiert pas de paramètres particulier ou spécifique (le simple suivant, suivant, suivant, accepter, installer est suffisant)*

### Logiciels

#### R version 4.0

Savoir quelle version est installé ?

Lancer le logiciel R  
Puis utiliser la ligne de code :  
**sessionInfo()**

R est un logiciel libre, sous licence GPL. Depuis la version 4.0 des nouvelles librairies sont disponibles. Les scripts d'analyses de la qualité de la donnée ont besoins de cette version

Lien de téléchargement certifié sans virus :  
<https://cran.r-project.org/bin/windows/base/R-4.0.2-win.exe>

#### Pandoc

Savoir si pandoc est installé ?

Dans démarrer faites une recherche Pandoc si un fichier Pandoc s'affiche c'est tout bon

Logiciel libre qui permet de convertir des fichiers LaTeX et markdown en format html, pdf ou word.

Lien de téléchargement certifié sans virus :  
[https://github.com/jgm/pandoc/releases/download/2.10/pandoc-2.10-windows-x86\\_64.msi](https://github.com/jgm/pandoc/releases/download/2.10/pandoc-2.10-windows-x86_64.msi)

### Connexion réseaux

#### Accès aux fichiers serveurs

Accès aux fichiers scripts est indispensable pour les lancer. Ils sont situés dans le répertoire suivant :

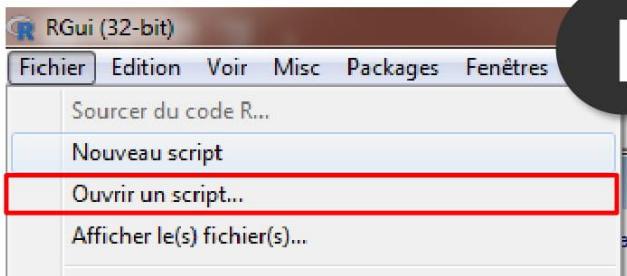
P:\2- expertise - aide decision\2- methodes - outils\basemis\4- projets\19-Validation-Stage2020

#### Accès aux serveurs de base de données

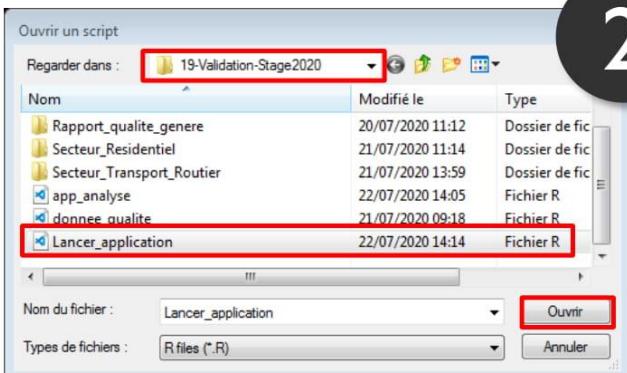
Accès aux bases de données de AIRPL est indispensable pour charger les données, il faut donc s'assurer que ses connexions soient établies.

En cas de problématique réseaux voir avec julien.

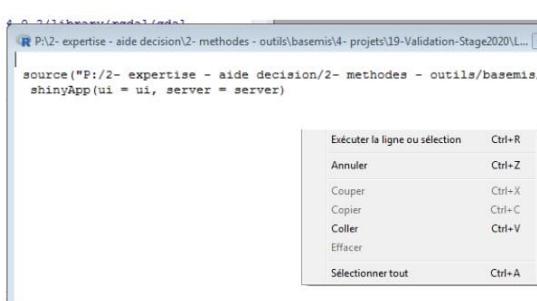
## LANCER POUR LA PREMIÈRE FOIS UN SCRIPT ? AVEC LE LOGICIEL R



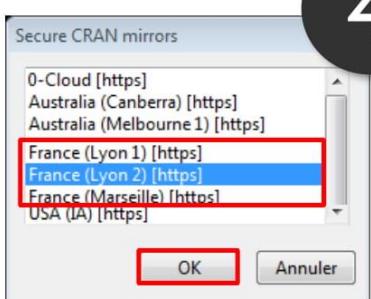
**ETAPE 1 :**  
Lancer le logiciel R  
Aller dans l'onglet fichier  
Puis dans Ouvrir un script



**ETAPE 2 :**  
Chercher le fichier R  
**Lancer\_application.R**  
situé dans le dossier :  
P:\2- expertise - aide decision\  
2- methodes - outils\basemis\  
4- projets\19-Validation-Stage2020



**ETAPE 3 :**  
Lancer l'exécution du  
script :  
Raccourci ( CTRL+A puis CTRL +R)  
Clic droit (sélectionner tout puis  
Exécuter la sélection)

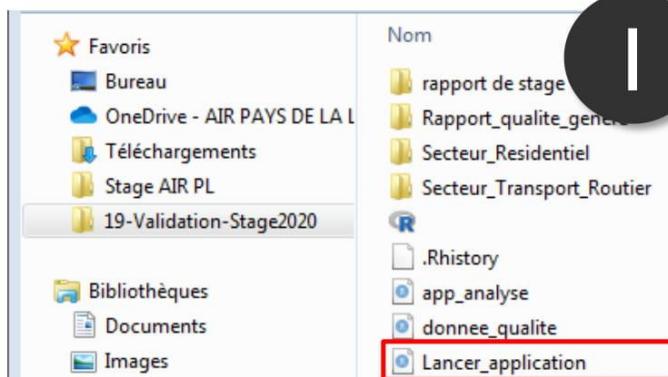


**ETAPE 4 :** Nécessaire uniquement quand le script est exécuté pour la première fois après l'installation du logiciel. (il télécharge toutes les librairies nécessaires cette étape peut durer 5 à 10 min, mais une fois installée cette étape n'est plus à faire)  
**Sélectionner le lieu de téléchargement des librairies puis appuyer sur OK**

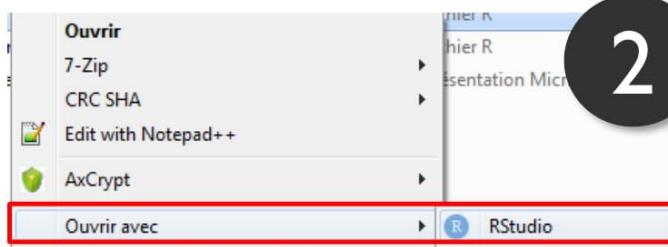
## LANCER POUR LA PREMIÈRE FOIS UN SCRIPT ? AVEC LE LOGICIEL Rstudio



Installer Rstudio ? <https://download1.rstudio.org/desktop/windows/RStudio-1.3.1056.exe>

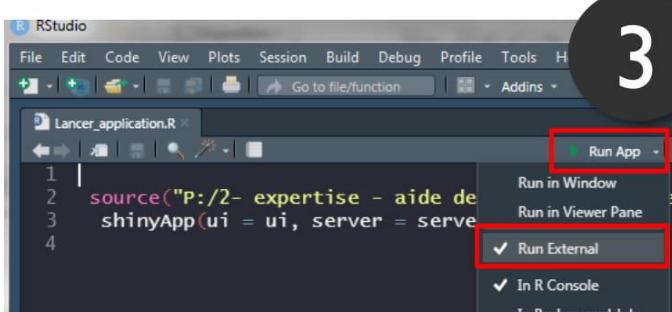


**ETAPE 1 :**  
Chercher le fichier R  
Lancer\_application.R  
situé dans le dossier :  
P:\2- expertise - aide decision\  
2- methodes - outils\basemis\  
4- projets\19-Validation-Stage2020



**ETAPE 2 :**  
Clic droit ouvrir avec  
Rstudio

(n'hésitez pas à choisir Rstudio comme programme par défaut)



**ETAPE 3 :**  
Cliquer sur Run App pour  
lancer l'exécution.

Vous pouvez paramétriser dans quel type de fenêtre l'application doit s'ouvrir grâce à la petite flèche.

Attention : il est possible que lors de la première exécution du script Rstudio installe toutes les librairies ce qui peut prendre entre 5 à 10 min et un redémarrage du logiciel.

(une fois installée cette étape n'est plus à faire et le script sera bien plus rapide au lancement)

## 7.11 Annexe 11 : Documentation – logigrammes (extrait)

### FONCTION

Les fonctions développer en R sont présente pour faciliter certain processus répétitif d'une analyse à l'autre. Ses fonctions ont pour but de faciliter et optimiser le développement des analyses, cependant elle nécessite plusieurs conditions d'utilisations :

- Le respect des données d'entrées
- L'ordre de déclaration des fonctions

Certaines fonctions sont imbriquées dans d'autres fonctions, ont retrouvé alors un ordre de déclaration nécessaire pour le bon déroulement des opérations. Je représenterais ici seulement les fonctions utiles (et non pas les sous fonctions utiliser indirectement).

Voici les schémas récapitulatifs pour expliquer les restrictions

#### Légende des schémas utilisés

Données d'entrées

Variable 1 →  
Variable 2 →

Données de sorties

→ Variable sortie

Fonctions imbriquées

**Variable 1** : type de données et explication

**Variable 2** : type de données et explication

#### Fonction Int\_Q

Aucune dépendance

*Objectifs : cette fonction permet de renvoyer les limites minimum et maximum d'une série à ne pas dépasser (en cas de dépassement nous pouvons considérer ses valeurs comme anormale)*



**VAR** = liste de la variable à tester, dans un tableau ciblé le champs analyse ex : TAB\$CHAMPS

**Ratio** = ratio interquartile fixé (test de tukey ratio = 1,5) (dans nos analyses ratio = 3)  
(peut être modifier pour avoir un nombre d'erreur fixe, ou un stabilisation des erreurs)

**lim\_bas** = valeur minimum autorisé avant outlier

**lim\_haut** = valeur maximum autorisé avant outlier

Exemple d'utilisation :

```
limit_minimum = int_Q(TAB_DONNEE$Indicateur)[1]  
limit_max = int_Q(TAB_DONNEE$Indicateur)[2]
```

## FONCTION

### Fonction extraction\_outlier\_CAT

Int\_Q0

Extraction  
outlier()

*Objectifs : cette fonction permet de renvoyer les valeur aberrantes en dehors des limites, tout en prenant en compte un champs de catégorie.*



**TAB** = Tableau des données ATTENTION : ce tableau nécessite un format standard où l'on a les champs numcom, rubrique, année et indice dans cet ordre.

**Ratio** = ratio interquartile fixé pour la détection d'outlier

**TEST** = possibilité d'indiquer sur quel indicateur les tests d'outliers sont réalisés

**Outlier** = Tableau des variables qui sont considérées comme outliers

Exemple d'utilisation :

```
out2_2 <- ENER_FINAL %>%  
  select(numcom, categorie, ANNEE, ind_enerhab) %>%  
  extraction_outlier_CAT(3, "hab")
```

### Fonction all\_histo\_cat

Int\_Q0

Histogramme  
ecart type()

*Objectifs : cette fonction permet d'avoir un graphique comprenant l'ensemble des histogrammes + boxplot et limite des outliers pour chaque catégories.*



**TAB** = Tableau des données ATTENTION : ce tableau nécessite un format standard où l'on a les champs numcom, rubrique, indice et info\_complémentaire dans cet ordre.

**Ratio** = ratio interquartile fixé pour la détection d'outlier qui sera affiché en ligne rouge

**TEST** = permet d'indiquer l'unité de l'indicateur pour afficher la bonne unité sur le graphique

**Graph** = objet ggplot qui permet l'affichage de graphique dans le rapport

Exemple d'utilisation :

```
GRAPH_ANALYSE_N2C_3 <- ENER_FINAL %>%  
  select(numcom, categorie, ind_enersurf, TYPO) %>%  
  all_histo_cat("kWh/m2")
```

## 7.12 Annexe 12 : Liens de formation et tutoriels

### Programmation en R

- Restructuration des données et mise en format des données qualités : <http://larmarange.github.io/analyse-R/manipuler-les-donnees-avec-dplyr.html>
- Outils cartographiques bases : <https://cran.r-project.org/web/packages/tmap/vignettes/tmap-getstarted.html>
- Outils carto plus avancé : <https://bookdown.org/robinlovelace/geocompr/adv-map.html#static-maps>
- Outils graphique : <https://juba.github.io/tidyverse/08-ggplot2.html>
- Outils graphique animé : <https://gganimate.com/index.html>
- Outils graphique interactif : <https://plotly.com/r>
- Outils tableau : [https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome\\_table\\_in\\_html.html](https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_html.html)
- R de manière générale approfondie : <https://r4ds.had.co.nz/model-basics.html>

### Programmation Rmarkdown

- Les bases : <https://statistique-et-logiciel-r.com/guide-de-demarrage-en-r-markdown/>
- L'ensemble des notions : <https://bookdown.org/yihui/rmarkdown/>

### Programmation R / SQL

- La base : <https://db.rstudio.com/databases/postgresql>

### Programmation Shiny

- Première prise en main : <https://superstatisticienne.fr/r-sur-le-web-le-package-shiny>
- Shiny de manières générales :  
[https://geoviz.sciencesconf.org/data/pages/GeoViz2018\\_R\\_shiny.pdf](https://geoviz.sciencesconf.org/data/pages/GeoViz2018_R_shiny.pdf)  
<https://mastering-shiny.org>
- Lancer des Rmarkdown depuis shiny :  
<https://beta.rstudioconnect.com/content/2671/Combining-Shiny-R-Markdown.html>
- Mise en ligne : [https://rstudio-pubs-static.s3.amazonaws.com/440403\\_2fe4b00a09dd4b268efd6efb353ccad7.html](https://rstudio-pubs-static.s3.amazonaws.com/440403_2fe4b00a09dd4b268efd6efb353ccad7.html)

### Autres références utiles sur le logiciel R

- Analyse spatiale R : <https://bookdown.org/robinlovelace/geocompr>
- Analyse spatiale SQL : <http://www.postgis.fr/chrome/site/docs/workshop-foss4g/doc/PostGISIntro.pdf>
- Gestion librairie : <https://pkgrverse.mikewk.com>

## 7.13 Annexe 13 : Table des schémas

Schéma 1 Organisation d'Air Pays de la Loire .....	7
Schéma 2 Panel des activités d'Air Pays de la Loire.....	7
Schéma 3 distinction entre l'émission et la concentration de pollution .....	8
Schéma 4 Les différentes utilisations de l'inventaire BASEMIS® .....	9
Schéma 5 Descriptions des détails de précision de l'inventaire.....	10
Schéma 6 Méthode de calcul de l'inventaire BASEMIS® .....	11
Schéma 7 Chronologies des versions de BASEMIS® .....	11
Schéma 8 Représentation de l'architecture des données sollicitées .....	13
Schéma 9 Modèles conceptuelles des tables de données sollicités.....	14
Schéma 10 Méthode de calcul de l'inventaire BASEMIS® .....	15
Schéma 11 Coordinations et utilisations des données .....	15
Schéma 12 : Critères d'évaluation de la qualité des données .....	17
Schéma 13 Inventaires des approches qualités avec les méthodes associées .....	18
Schéma 14 : exemple de recherche du secteur agricole .....	19
Schéma 15 Datavisualisation des émissions de GES des bovins .....	20
Schéma 16 Datavisualisation des émissions de GES des porcins .....	20
Schéma 17 Récapitulatif des différentes méthodes de détections d'outliers.....	23
Schéma 18 Représentation graphiques du test de Tukey .....	24
Schéma 19 Visualisation du test de Tukey sur l'indicateur de précision thématique du secteur résidentiel (kWh/hab.) sur les consommations d'énergies en gaz .....	24
Schéma 20 Graphique d'analyse sur les données résidentielles .....	25
Schéma 21 Processus d'analyse de la qualité des données .....	26
Schéma 22 Secteur d'activité de l'inventaire BASEMIS® .....	33
Schéma 23 Data visualisation Climat-air-Energie pour l'année 2016 – source BASEMIS® V5 .....	33
Schéma 24 Impact Climat-Air-Energie du secteur résidentiel – source BASEMIS® V5.....	34
Schéma 25 méthodologie de calcul des consommations et émissions dans les bâtiments du secteur résidentiel.....	34
Schéma 26 Analyses des types de logements en Pays de la Loire en 2016 – source BASEMIS® V5.....	35
Schéma 27 Consommations d'énergie en 2016 par usage (gauche) et source d'énergie (droite) – source BASEMIS® V5 .....	35
Schéma 28 Grille d'analyse du secteur résidentiel .....	36
Schéma 29 Impact Climat-Air-Energie du secteur routier – source BASEMIS® V5.....	38
Schéma 30 méthodologie de calcul mise en œuvre pour la prise en compte des transports routiers .....	38
Schéma 31 Comparaison des méthodes d'analyses à différentes échelles (échelles communales à droites et échelles de la voirie à gauche) .....	39
Schéma 32 Impact Climat-Air-Energie du secteur agricole – source BASEMIS® V5 .....	41
Schéma 33 Analyses des sous-secteurs d'activité agricole en 2016 – source BASEMIS® V5 .....	41
Schéma 34 méthodologie de calcul pour les émissions liées à l'élevage .....	42
Schéma 35 méthodologie de calcul pour les émissions liées aux cultures .....	42
Schéma 36 Planning du stage .....	46
Schéma 37 Méthode de travail AGILE : <a href="https://www.mercator.eu/fr/la-methode-agile-comme-methode-de-travail-chez-mercator-explications.chtml">https://www.mercator.eu/fr/la-methode-agile-comme-methode-de-travail-chez-mercator-explications.chtml</a> .....	48
Schéma 38 Constat et proposition de solution pour le projet.....	49
Schéma 39 : Descriptif des logiciels utilisés .....	50

# 8 Références

## 8.1 Bibliographie

La référence, tant pour les statistiques que pour l'analyse de données :

« Probabilités, analyse des données et statistiques » (Gilbert Saporta) TECHNIP 2011.

Guide explicatif sur les enjeux de la qualité des données :

« Data Management, qualité des données et compétitivité » (Christophe Brasseur) ©LAVOISIER 2005.

Document informatif sur les données de manières globales :

« Base de données - Technologies logicielles Architectures des systèmes » (Technique de l'ingénieur) Technologies de l'information 2018.

Réflexion sur la détection et gestion des valeurs aberrantes :

« Outliers ou données extrêmes, comment les détecter ? que faire de ces observations ? » (Claire Chabanet et Fabrice Dessaint) 10 juillet 2015.

## 8.2 Webographie

1. Organisation internationale de normalisation, (Norme ISO), La qualité des données géographiques : <https://www.iso.org/obp/ui/#iso:std:iso:19157:ed-1:v1:fr>
2. Ministère de l'environnement et du développement durable de l'énergie, Guide pour l'élaboration des inventaires territoriaux :  
[https://www.lcsqa.org/system/files/rapport/MTES\\_Guide\\_methodo\\_elaboration\\_inventaires\\_PCIT\\_mars2019.pdf](https://www.lcsqa.org/system/files/rapport/MTES_Guide_methodo_elaboration_inventaires_PCIT_mars2019.pdf)
3. Ministère de l'environnement et du développement durable de l'énergie, Organisation et méthodes des inventaires nationaux des émissions atmosphériques en France :  
[https://www.citepa.org/wp-content/uploads/Citepa\\_Ominea\\_ed2020.pdf](https://www.citepa.org/wp-content/uploads/Citepa_Ominea_ed2020.pdf)
4. Atmo France, portails des données open data des AASQA : <https://atmo-france.org/les-donnees/>
5. Air Pays de la Loire, guide méthodologique de l'inventaire BASEMIS® :  
[http://www.airpl.org/content/download/862/3916/file/airpl\\_guide\\_methodo\\_basemisV5\\_2018-11-16-f.pdf](http://www.airpl.org/content/download/862/3916/file/airpl_guide_methodo_basemisV5_2018-11-16-f.pdf)
6. Balicchi Julien, Responsable du service Études et Statistiques, Agence régionale de santé de Mayotte, références et inventaire des tests de détection de valeur aberrante : <https://lemakistatheux.wordpress.com/2016/04/01/les-tests-pour-la-detection-doutliers/>
7. Norman Matloff, enseignant en statistique et informatique à l'université de Californie à Davis, Comparatif des avantages des langages de programmation R et Python : <https://github.com/matloff/R-vs.-Python-for-Data-Science>
8. Data Novia, créé par [Alboukadel Kassambara](#), méthode de normalisation et standardiser des données : <https://www.datanovia.com/en/fr/blog/comment-normaliser-et-standardiser-les-donnees-dans-r-pour-une-visualisation-en-heatmap-magnifique/>

## 9 Glossaire

<b>AASQA</b>	Association Agréée de Surveillance de la Qualité de l'Air
<b>ADEME</b>	Agence De l'Environnement et de la Maîtrise de l'Energie
<b>AFNOR</b>	Agence Française de Normalisation
<b>BASEMIS®</b>	BASE des EMISsions de polluants et GES
<b>DREAL</b>	Direction Régionale de l'Environnement, de l'Aménagement et du Logement
<b>GES</b>	Gaz à Effet de Serre
<b>PCET</b>	Plan Climat Energie Territorial
<b>PRQA</b>	Plan Régional pour la Qualité de l'Air
<b>SCOT</b>	Schéma de COhérence Territoriale
<b>SRCAE</b>	Schéma Régional Climat Air Energie
<b>TMJA</b>	Traffic Moyen Journalier Annuel
<b>CEREMA</b>	Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement
<b>EPCI</b>	Établissement public de coopération intercommunale
<b>PLU</b>	Plan Local d'Urbanisme
<b>PDU</b>	Plan de Déplacements Urbains
<b>PCIT</b>	Pôle de Coordination des Inventaires Territoriaux
<b>PCAET</b>	Plan Climat Air Energie Territorial
<b>teqCO2</b>	tonne équivalent CO2
<b>µm</b>	Micromètre (= 1 millionième de mètre)
<b>PPA</b>	Plans de Protection de l'Atmosphère
<b>CCNUCC</b>	Convention Cadre des Nations Unies sur les Changements Climatiques
<b>INSEE</b>	Institut National de la Statistique et des Etudes Economiques
<b>OMINEA</b>	Organisation et Méthodes des Inventaires Nationaux des Emissions Atmosphériques
<b>CITEPA</b>	Centre Interprofessionnel Technique d'Etudes de la Pollution Atmosphérique
<b>SDES</b>	Service de la Donnée et des Etudes Statistiques
<b>ENEDIS</b>	Entreprise d'Électricité Réseau Distribution France
<b>GRDF</b>	Entreprise distribution de gaz naturel en France
<b>GRT gaz</b>	Entreprise distribution de gaz naturel
<b>RTE</b>	Entreprise Réseau de transport d'électricité
<b>DRAAF</b>	Direction Régionale de l'Agriculture de l'Alimentation et de la Forêt
<b>AGRESTE</b>	Service de la statistique, de l'évaluation et de la prospective agricole
<b>IGN</b>	Institut Géographique National
<b>LTECV</b>	Loi de Transition Energétique pour la Croissance Verte
<b>ISO</b>	Organisation Internationale de normalisation
<b>MRV</b>	Monitoring, Reporting, Verification
<b>CEREN</b>	Centre d'Etude et de Recherche Economique sur l'éNergie
<b>SOeS</b>	Service de l'Observation et des Etudes Statistiques
<b>CPDP</b>	Commission Particulière du Débat Public
<b>SIREDO</b>	Soins, Innovation, Recherche, en oncologie de l'Enfant, de l'aDOlescent
<b>RGA</b>	Recensement Général de l'Agriculture
<b>UNIFA</b>	Union des industries de la fertilisation
<b>COALA</b>	Coopération entre Air Breizh, Lig'Air et Air Pays de la Loire

# air pays de la loire

5 rue Édouard-Nignon – CS 70709 – 44307 Nantes cedex 3  
Tél + 33 (0)2 28 22 02 02  
Fax + 33 (0)2 40 68 95 29  
[contact@airpl.org](mailto:contact@airpl.org)

**air** | **pays de**  
**la loire**  
[www.airpl.org](http://www.airpl.org)