



# Évaluation de la qualité des données de l'inventaire BASEMIS

*Axel Fourneyron*

*16 mars - 11 septembre  
Master  
Géographie Numérique*

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

**UNIVERSITÉ  
JEAN MONNET**  
SAINT-ÉTIENNE

**GÉO  
NUM**

**ENS DE LYON**

**UNIVERSITÉ  
LUMIÈRE  
LYON 2**

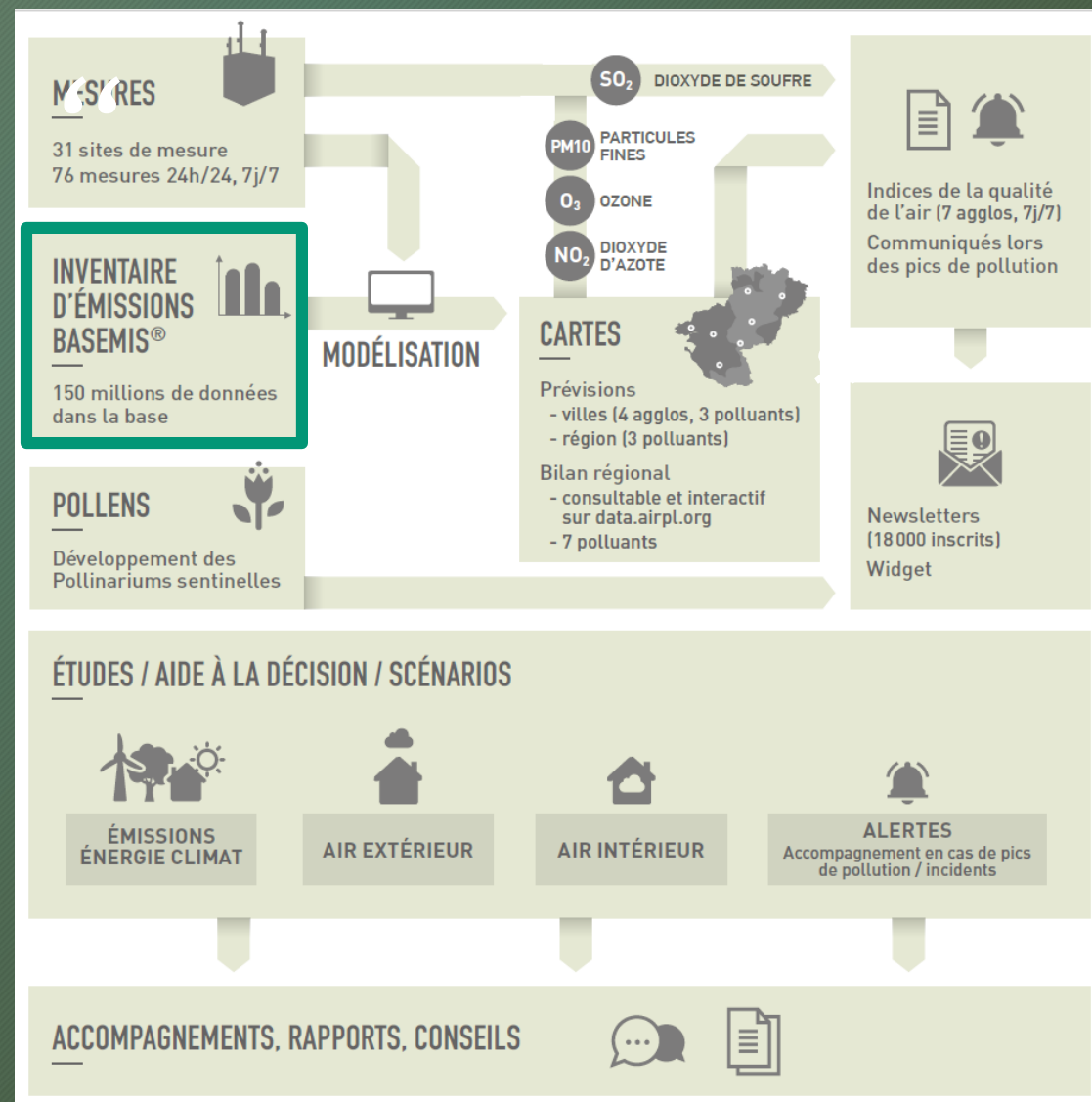
# Projet de validation des données de l'inventaire BASEMIS

2

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

Surveillance  
de la qualité  
de l'air

Informier le  
public et les  
autorités





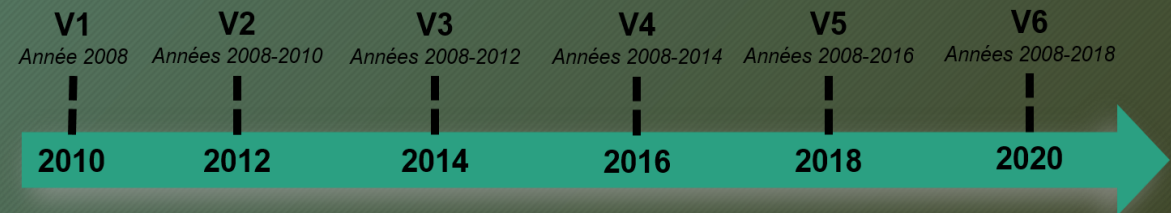
*“ Inventaire en  
amélioration  
continue ”*

Surveillance  
de la qualité  
de l'air

Informier le  
public et les  
autorités



BASEMIS

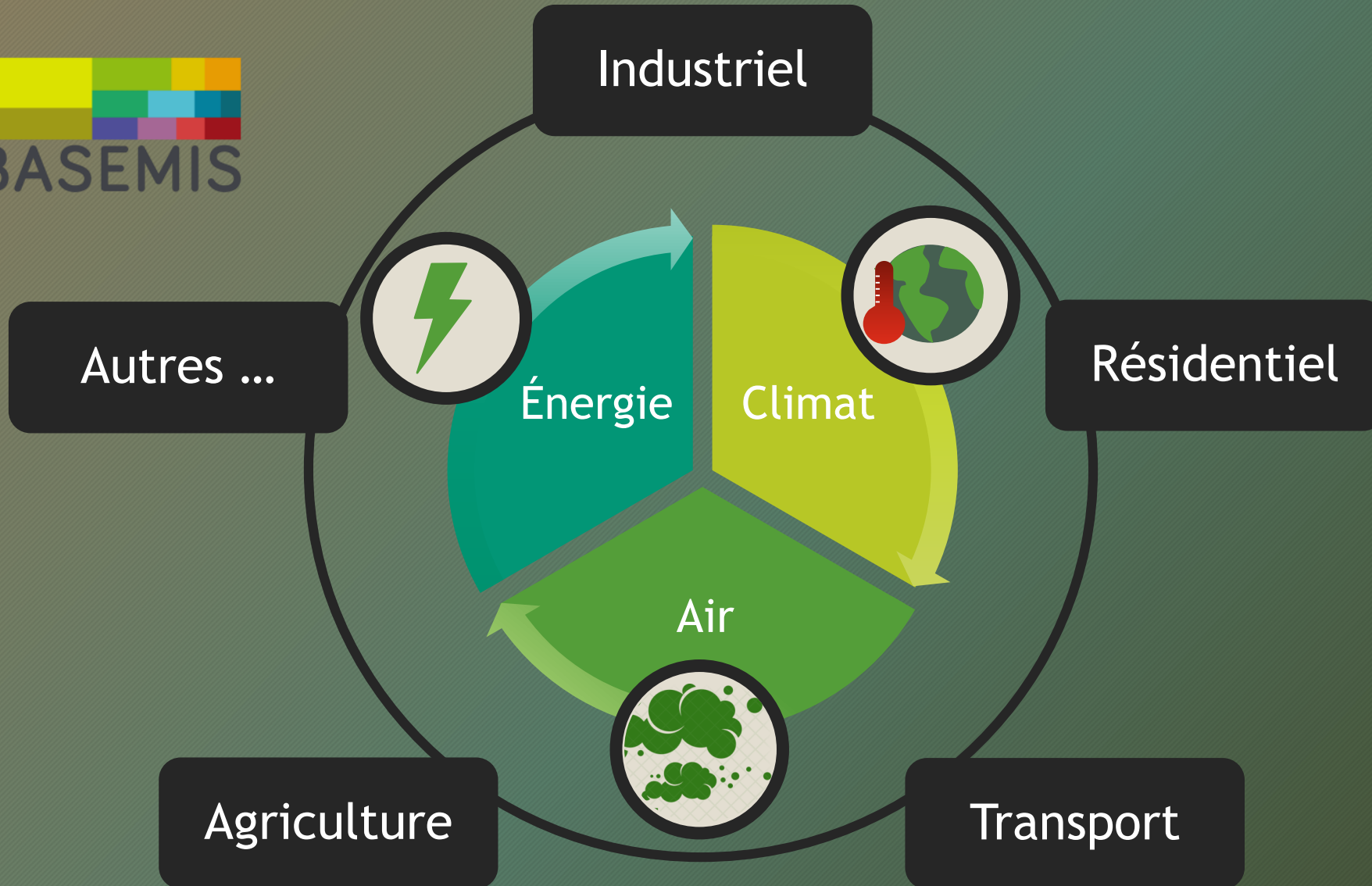


# Projet de validation des données de BASEMIS

4



- 1 Découvrir les données de l'inventaire
- 2 Définir une méthode d'analyse de qualité des données
- 3 Développer les outils d'analyse, visualisation et d'interprétation
- 4 Former à l'utilisation des outils
- 5 Résultat de l'évaluation de qualité des données



Consommations  
d'énergie

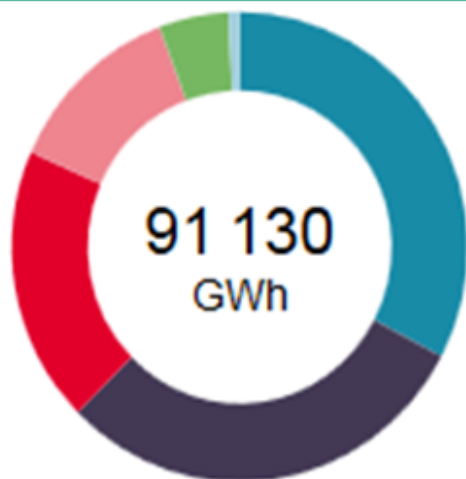
Émissions de  
GES

Émissions de  
polluant

Secteur  
d'activité

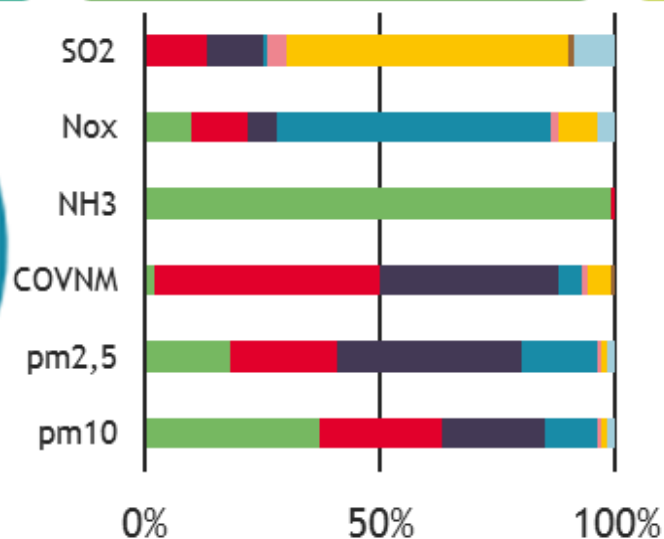


### Consommation d'énergie

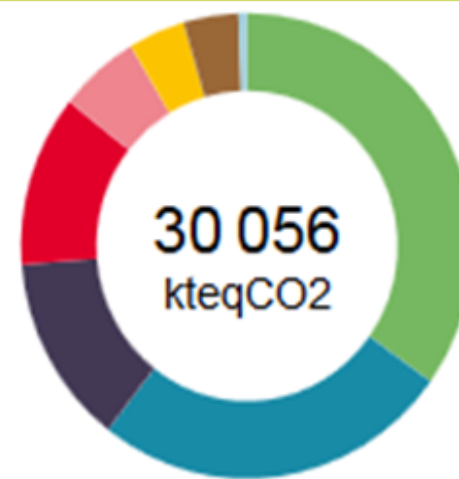


**BASEMIS**

### Émission de polluant

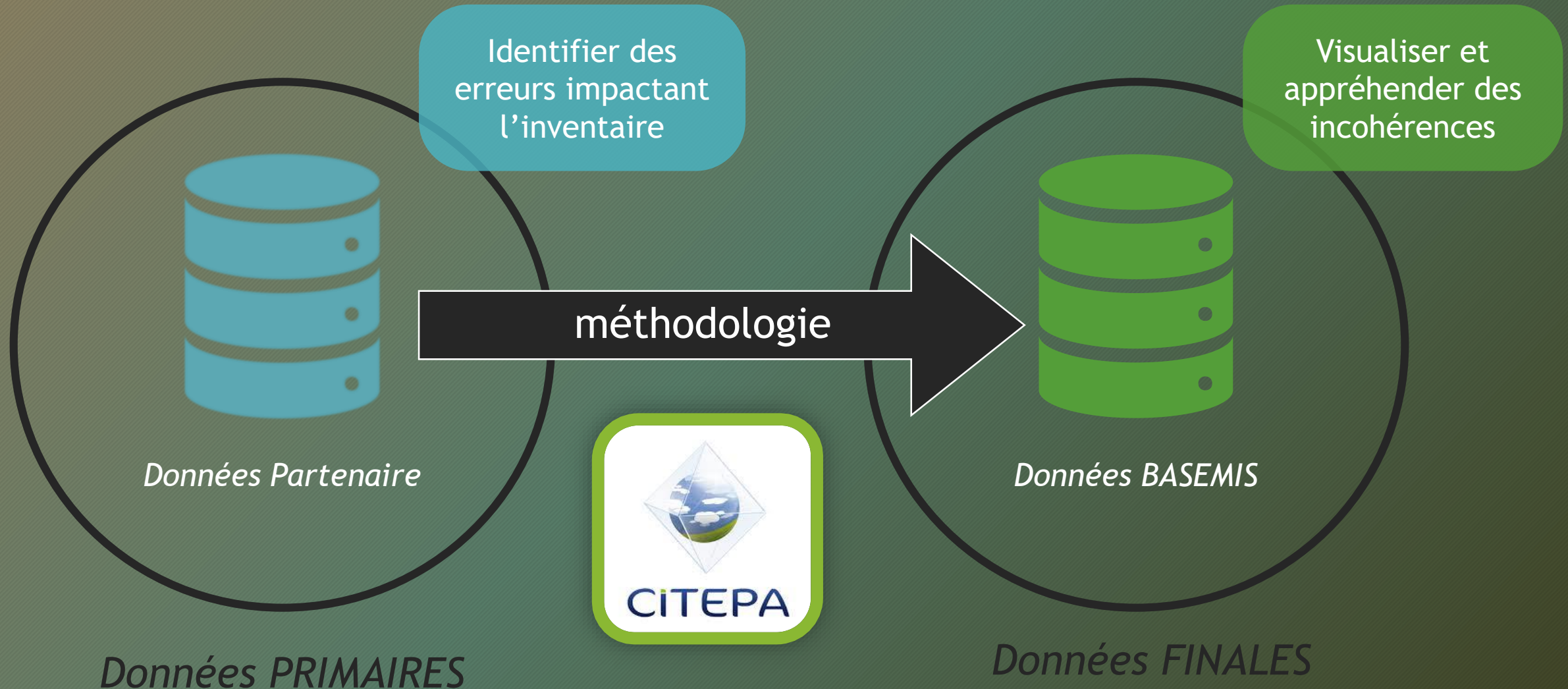


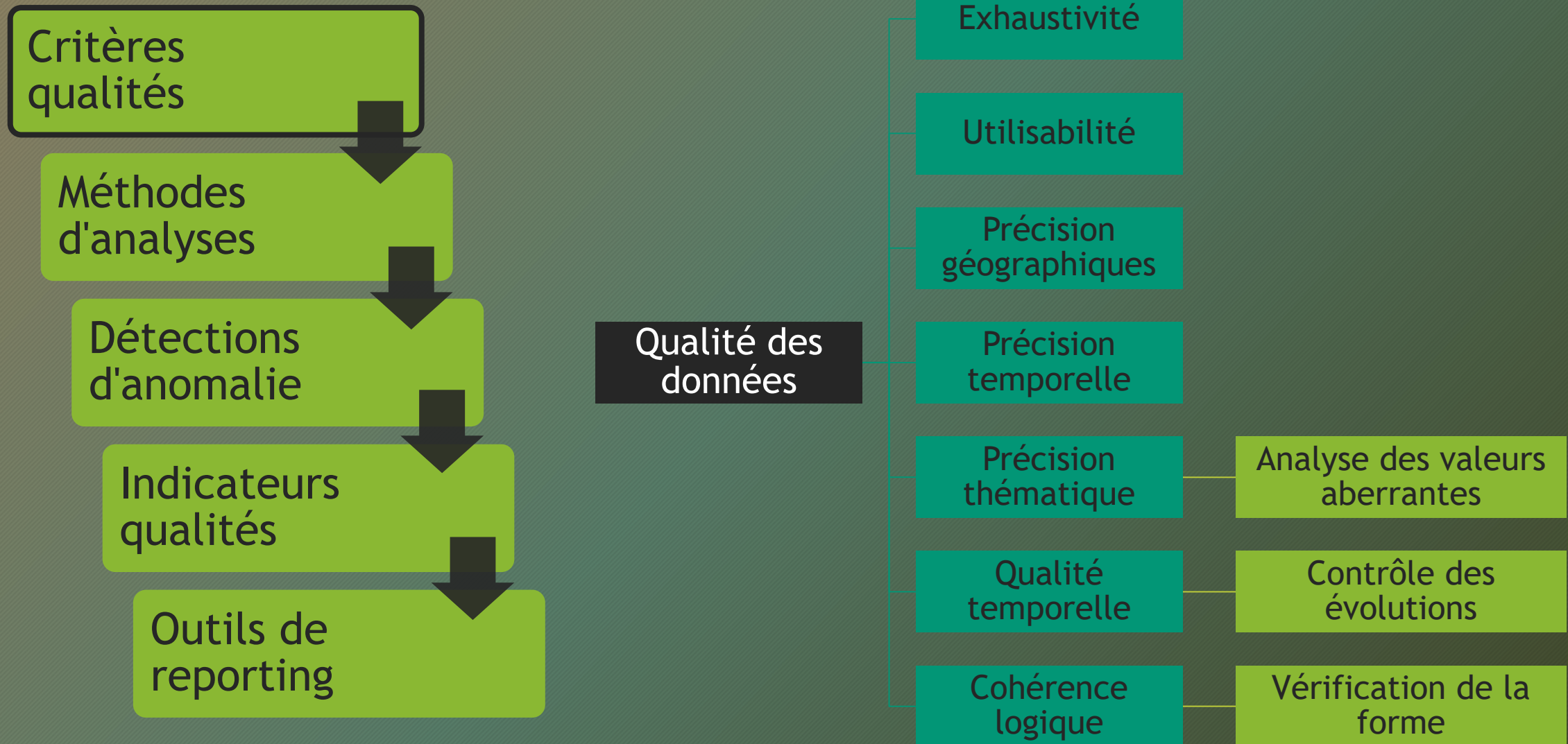
### Émission de GES



Secteur d'activité :









Critères  
qualités

Méthodes  
d'analyses

Détections  
d'anomalie

Indicateurs  
qualités

Outils de  
reporting

Analyse  
statistique

Secteur  
résidentiel

Indicateur  
climat-air-  
énergie

Consommation  
d'énergie

Niveaux de  
détails

Usage  
énergétique

Source  
d'énergie

Indicateur(s)  
de  
normalisation

kWh/m<sup>2</sup>

kWh/hab

Indicateur(s)  
d'évolution  
temporelle

Taux  
d'évolution  
annuel

Taux  
d'évolution  
annuel

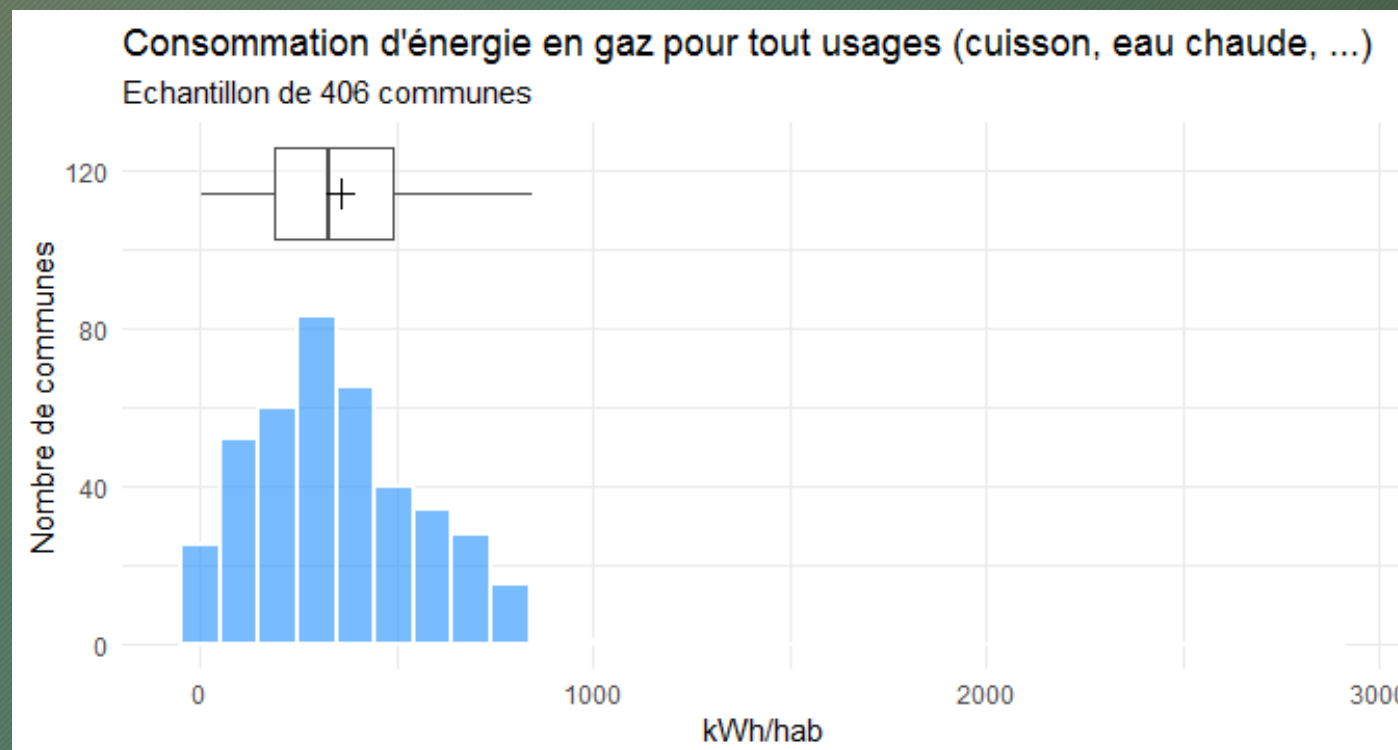
Critères  
qualités

Méthodes  
d'analyses

Détections  
d'anomalie

Indicateurs  
qualités

Outils de  
reporting



## 2 Méthode d'analyse de qualité des données

11

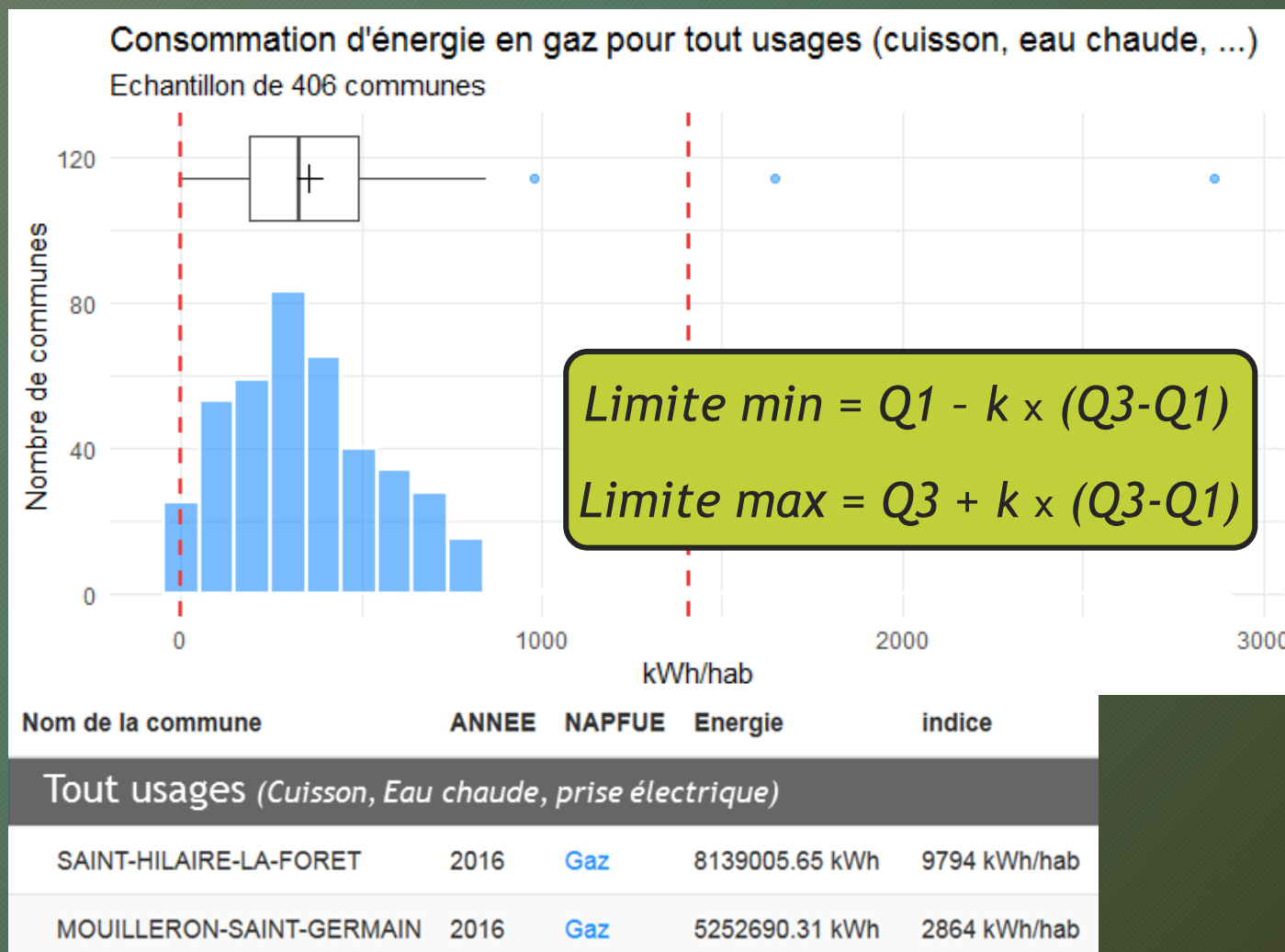
Critères  
qualités

Méthodes  
d'analyses

Détections  
d'anomalie

Indicateurs  
qualités

Outils de  
reporting



## 2 Méthode d'analyse de qualité des données

12

Critères  
qualités

Méthodes  
d'analyses

Détections  
d'anomalie

Indicateurs  
qualités

Outils de  
reporting

Le taux  
d'erreur  
dans la base

Le nombre  
d'erreurs

L'amplitude  
des erreurs

Evaluation  
qualité  
des  
données

Critères  
qualités

Méthodes  
d'analyses

Détections  
d'anomalie

Indicateurs  
qualités

Outils de  
reporting

*Outil de diagnostic des  
comportements des données*

*Outil d'évaluation de la  
qualité des donnée*

# 3 Développer les outils d'analyse et visualisation

Information Générale    Qualité précision thématique    Qualité évolution temporelle

## 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 de l'inventaire.

- 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>)

### 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 consommateurs.

$$Kwh/hab = \frac{EnergieConsommé}{NombreHabitant}$$
$$Kwh/m^2 = \frac{EnergieConsommé}{SurfaceLogement}$$
$$Taux Annuelle = \left[ \frac{Kwh/hab_{anneeN}}{Kwh/hab_{anneeN-1}} - 1 \right]$$

### 3 L'analyse qualité

À ces indicateurs, nous réalisons des analyses statistiques nous permettant de détecter des valeurs aberrantes.

Application qualité  
Tout secteur et toutes années

## Analyses des données énergétiques BASEMIS secteur résidentiel

15 juillet, 2020 - Air Pays de la Loire

- 1 - Le secteur résidentiel : Les indicateurs
- 2 - Résultats de l'année 2012: 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 aberrantes
- 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 de l'inventaire. 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 consommateurs.

$$Kwh/hab = \frac{EnergieConsommé}{NombreHabitant}$$
$$Kwh/m^2 = \frac{EnergieConsommé}{SurfaceLogement}$$
$$Taux Annuelle = \left[ \frac{Kwh/hab_{anneeN}}{Kwh/hab_{anneeN-1}} - 1 \right]$$

Rapport qualité  
1 secteur et 1 année

# Analyses de la qualité des données BASEMIS

## Choix du secteur


Résidentiel

## Choix de l'année

2016

## Version Basemis

Version 5

 Générer le rapport

Menu

Information Générale

Qualité précision thématique

Qualité évolution temporelle

Comparaison entre version

## 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'habitant par résidence
- Superficie 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 :

$$Kwh/hab = \frac{EnergieConsommé}{Nombre.Habitant}$$

$$Kwh/m^2 = \frac{EnergieConsommé}{Surface.Logement}$$

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

### 3 L'analyse qualité

Grâce à ses 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 génération d'un rapport.

# Analyses de la qualité des données BASEMIS

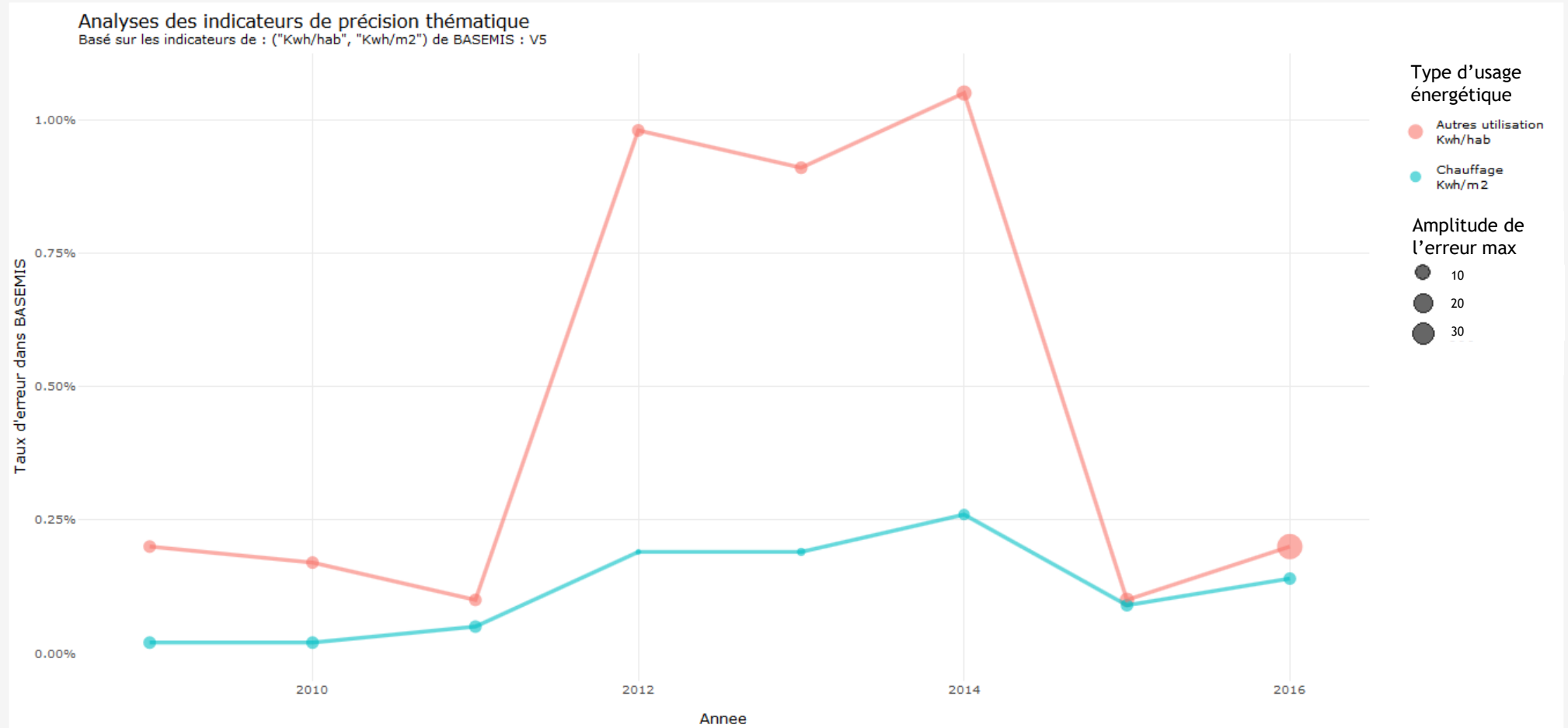
**Choix du secteur**  
Résidentiel

**Choix de l'année**  
2016

**Version Basemis**  
Version 5

**Générer le rapport**

- Menu
- Information Générale
- Qualité précision thématique
- Qualité évolution temporelle
- Comparaison entre version





# Analyses de la qualité des données BASEMIS

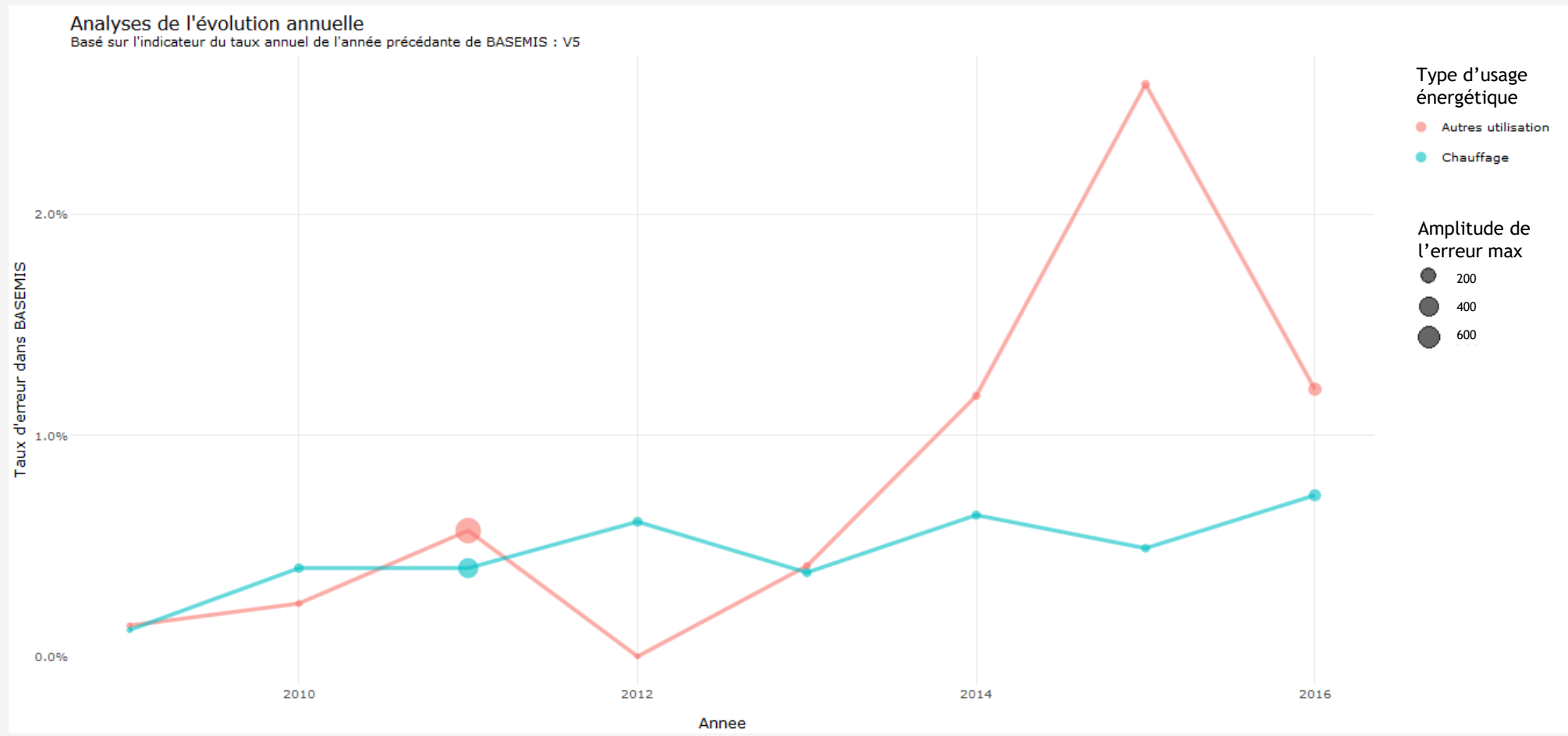
**Choix du secteur**  
Résidentiel

**Choix de l'année**  
2016

**Version Basemis**  
Version 5

**Générer le rapport**

- Menu
- Information Générale
- Qualité précision thématique
- Qualité évolution temporelle**
- Comparaison entre version



# Analyses de la qualité des données BASEMIS

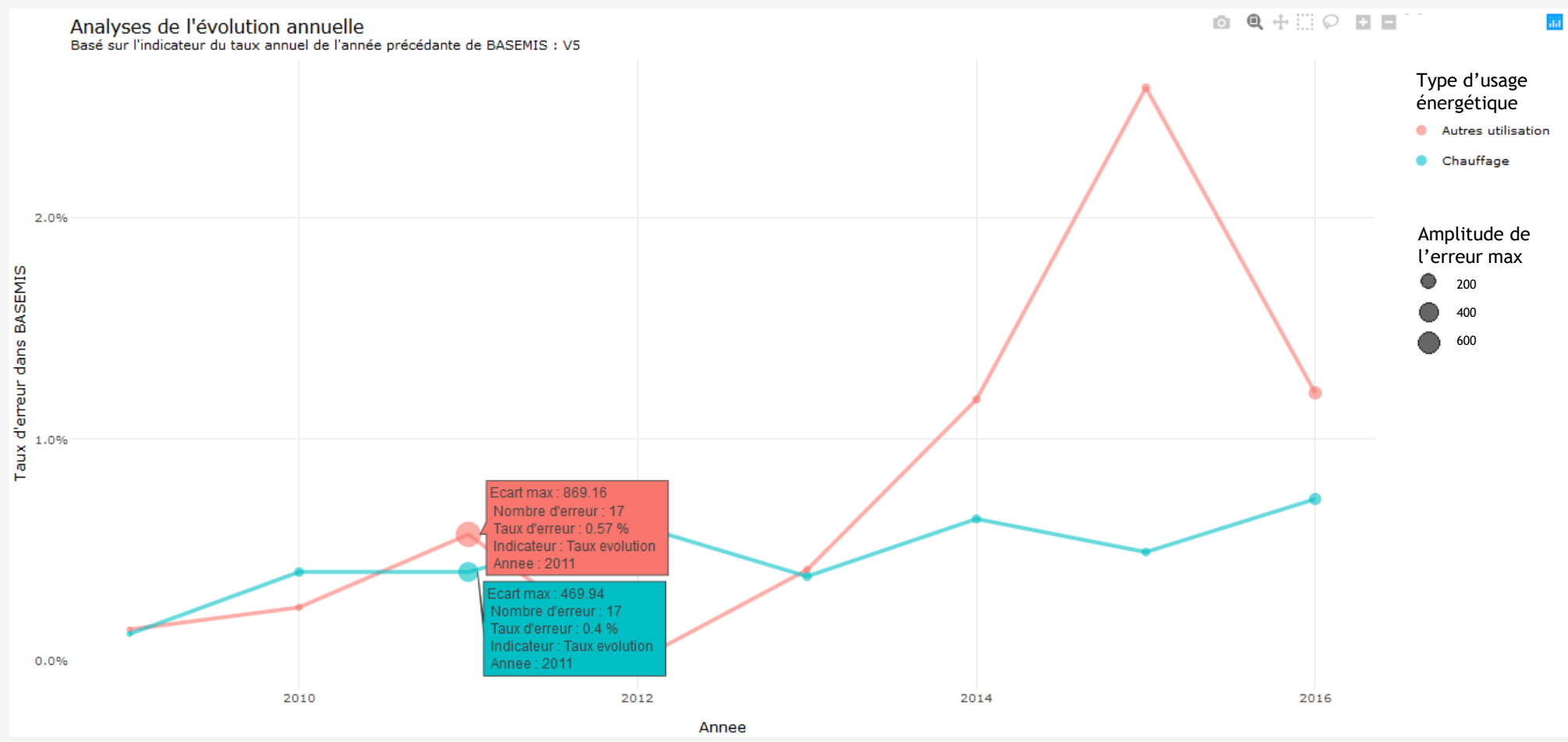
**Choix du secteur**  
 Résidentiel

**Choix de l'année**  
 2016

**Version Basemis**  
 Version 5

**Générer le rapport**

- Menu
- Information Générale
- Qualité précision thématique
- Qualité évolution temporelle
- Comparaison entre version



# Analyses de la qualité des données BASEMIS

**Choix du secteur**

Transport Routier ▲

Résidentiel

Agriculture

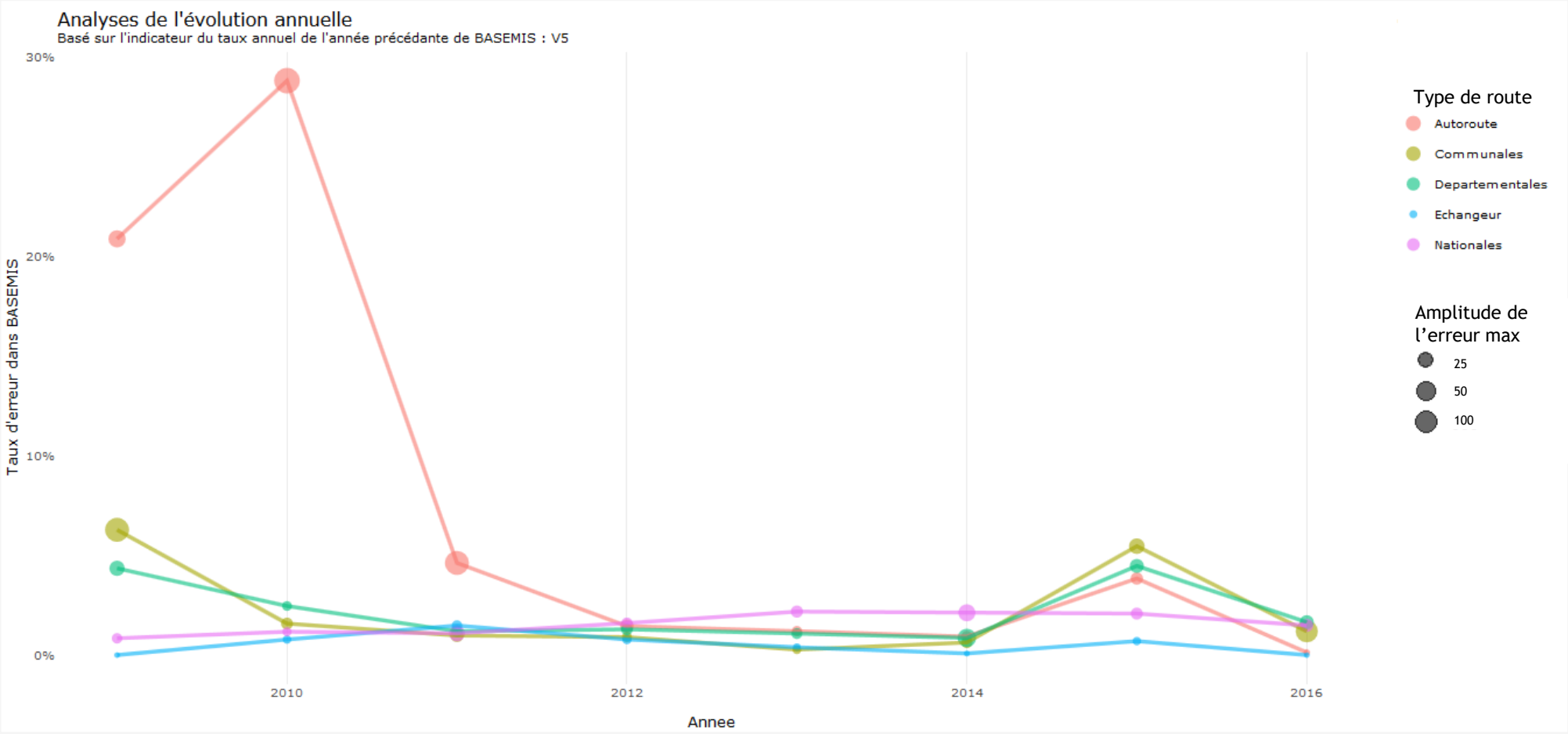
Transport Routier

**Version Basemis**

Version 5 ▼

**Générer le rapport**

- Menu
- Information Générale
- Qualité précision thématique
- Qualité évolution temporelle
- Comparaison entre version



# Analyses de la qualité des données BASEMIS

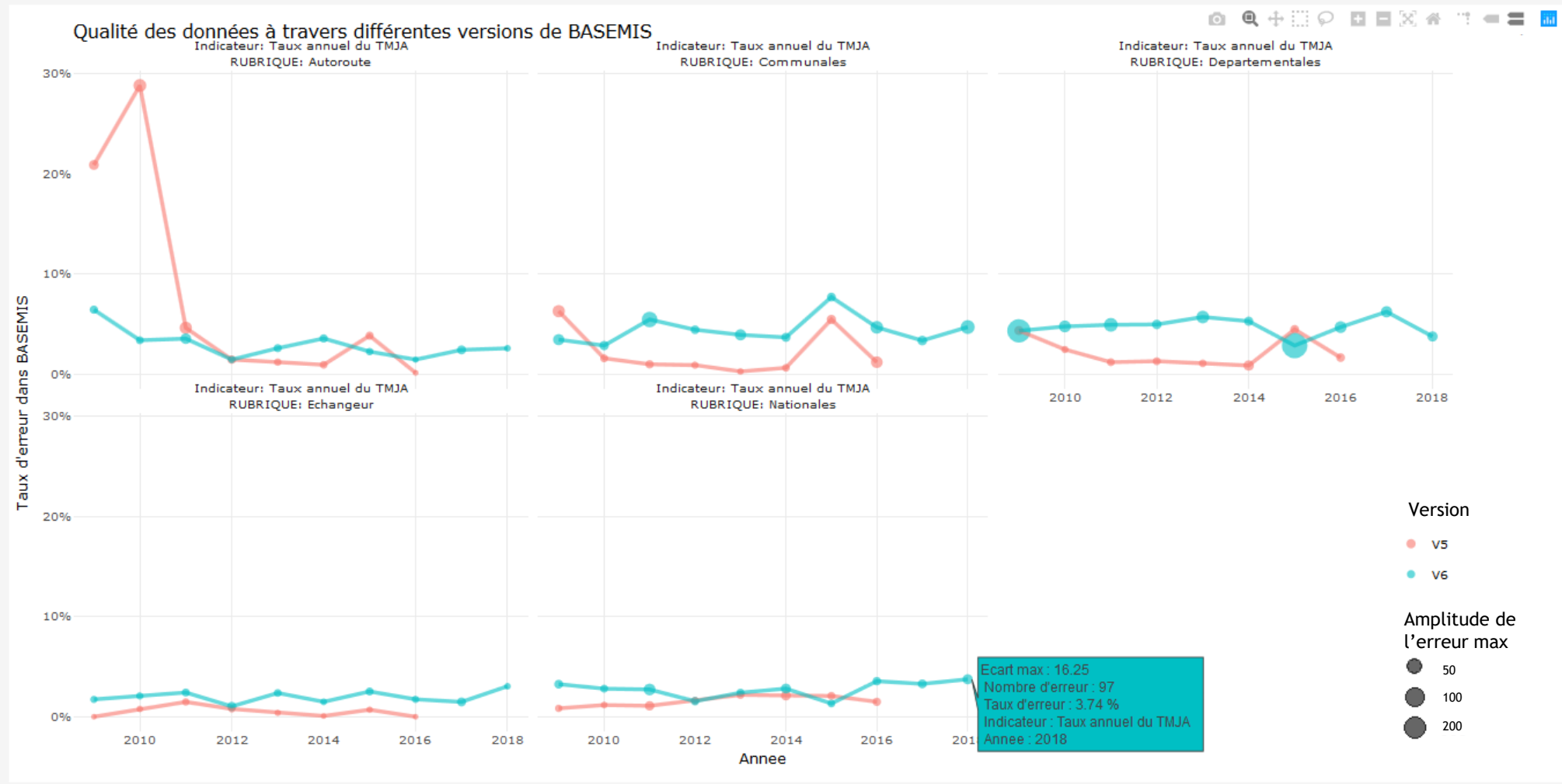
**Choix du secteur**  
Transport Routier

**Choix de l'année**  
2016

**Version Basemis**  
Version 6

[Générer le rapport](#)

- Menu
- Information Générale
- Qualité précision thématique
- Qualité évolution temporelle
- Comparaison entre version**



# 3 Développer les outils d'analyse et visualisation

## Analyses des données énergétiques BAS secteur résidentiel

15 juillet, 2020 - Air Pays de la Loire

- 1 - Le secteur résidentiel : Les indicateurs
- 2 - Résultats de l'année 2012: 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 aberrantes
- 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 de la base de données 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
- Surface des résidences (en m²)

#### 1.2 Méthode de calcul des indicateurs

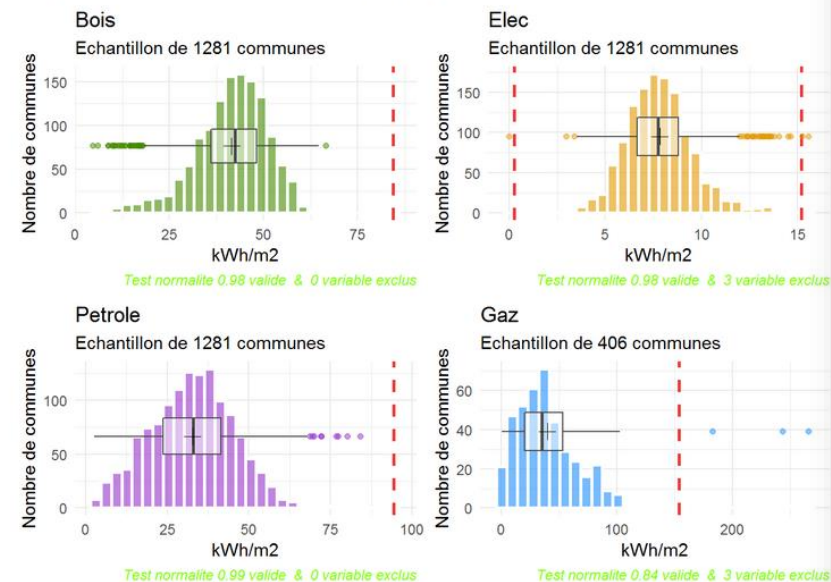
Au sein de ce secteur, nous avons mis en place trois indicateurs permettant de voir le comportement général de

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

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

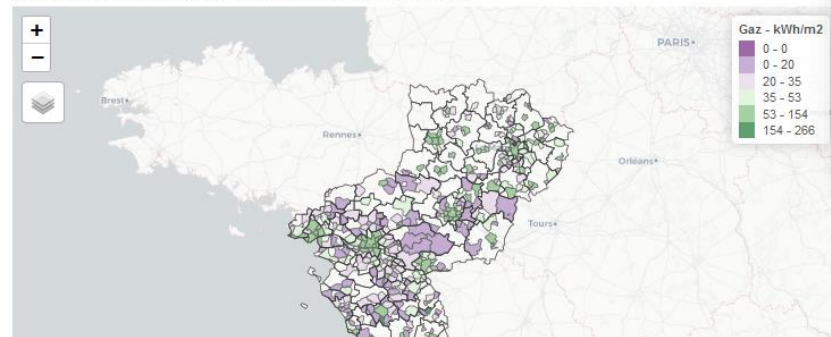
$$TauxAnnuelle = \left[ \frac{Kwh/hab_{anneeN}}{Kwh/hab_{anneeN-1}} - 1 \right]$$

Graphique 3 : Distribution des consommations d'énergies en 2012



Cartographie 3 : Cartographie des consommations d'énergies en 2012

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>					
SAINT-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

### 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

Support de formation

Guide technique

Guide d'utilisation

Guide de formation

Mémoire

**Nouveaux secteur d'activité**

Recherche  
Analyses appropriées d'un thème

Visualisation des résultats et des valeurs aberrantes

5

Mise en place de limite, test, pour évaluer les valeurs incohérentes

Analyse des comportements différents

**Nouvelle version Basemis**

2\_Requete  
donnee\_analyse\_NZ

1 Requetes SQL

2 Mise en forme des données

3 Assemblage et création

Loge\_POP\_ALL doit avoir les champs suivants :

- ANNEE
- Numcom
- Type\_residence
- Type

**ARCHITECTURE DES FICHIERS**

Fichier PDF ou HTML  
Dossier contenant des scripts  
Fichier Script

**FONCTION**

Fonction extraction\_outier\_CAT

Objets: cette fonction permet de renvoyer les valeurs aberrantes en dehors des limites, présent en compte un champs de catégorie.

TAB → extraction\_outier\_CAT() → TAB (OUTIER)

ratio

TEST

TAB = Tableau des données ATTENTION : ce tableau nécessite un format standard ou le champs numerique, numerique, annee et indice dans cet ordre.  
Ratio = ratio integrable fixé pour la détection d'outier.  
TEST = possibilité d'indiquer sur quel indicateur les tests d'outiers sont réalisés  
Outier = Tableau des variables qui sont considérées comme outier

Exemple d'utilisation  
out2\_2 <- ENER\_FINAL %>%

**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

universit  Lumini re LYON 2  
UNIVERSIT  JEAN MONNET SAINT ETIENNE  
G O N UM  
air pays de la Loire www.airpl.org

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

Installer Rstudio ? <https://r4ds.had.co.nz/model-basics.html>

1 ETAPE 1 : Chercher le fichier R situé dans le dossier : P12 - expertise - aide decision / 2 - methodes - outils(basemis) / 4 - projet / 9 - Validation - Stage

2 ETAPE 2 : Clic droit ouvrir Rstudio (en choisissant le dossier R)

3 ETAPE 3 : Cliquer sur "lancer" et lancer le script

Attention : il est possible que lors de la première installation de Rstudio, il faille attendre entre 5 à 10 min et un redémarrage du serveur.

(une fois installée cette étape n'est plus à faire et le script s'exécute)

**Programation en R**

- Restructuration des données et mise en format des données qualifiées : <http://la.marange.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 avancés : <https://bookdown.org/robinlovelace/geocompr/adv-map.html#static-maps>
- Outils graphique : <https://luba.github.io/fidyverse/08-geomplot2.html>
- Outils graphique animé : <https://gganimale.com/index.html>
- Outils tableau interactif : <https://plotly.com/r/>
- R de manière générale approfondie : <https://r4ds.had.co.nz/model-basics.html>

**Programation Rmarkdown**

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

**Programation R / SQL**

- La base : <https://db.rstudio.com/data/bases/postgres/>

**Programation Shiny**

- Première prise en main : <https://superstatisticienne.fr/r-sur-le-web-le-package-shiny/>
- Shiny de manière générale : [https://geoviz.sciencesconf.org/data/pages/GeoViz2018\\_R\\_shiny.pdf](https://geoviz.sciencesconf.org/data/pages/GeoViz2018_R_shiny.pdf)
- 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\\_2fe4b00a09dd4b268ef6fb353cad7.html](https://rstudio-pubs-static.s3.amazonaws.com/440403_2fe4b00a09dd4b268ef6fb353cad7.html)

**Autres références utiles sur le logiciel R**

- Analyse spatiale R : <https://bookdown.org/robinlovelace/geocompr/08-4g/doc/postgisIntro.pdf>
- Analyse spatiale SQL : <http://www.postgis.fr/chrome/site/docs/workshop->
- Gestion librairie : <https://pkgsverse.mikew.com>

# 5 Résultat de la qualité des données

## Secteur d'activité



Résidentiel



Transport routier



Agriculture

## Précision thématique



## Qualité temporelle



# 5 Résultat du projet

Outil évolutif  
fonctionnel



Formation  
sur l'utilisation



Formation  
sur l'amélioration



Partage de l'outil aux  
autres AASQA



## Préconisation

Prise en main  
intégration V6  
Ajout sectoriel  
Puis partage si pertinent et efficace

Panel gestion de projet :  
Cartographier les anomalies détectées, afin de  
valider une anomalie ou corriger l'anomalie.

Détermination d'une typologie des erreurs pour  
permettre une correction :  
à la source du problème  
à la sortie de l'inventaire



Merci de votre attention

**air** | pays de  
la Loire  
*www.airpl.org*

*Axel Fourneyron*  
06 26 32 44 67  
*fourneyron.axel.lucas@gmail.com*