2. Notion de pollution lumineuse

La pollution lumineuse, terme initialement répandu par les astronomes, représente pour Cinzano et al. (2001, p. 689. cité dans Gallaway, Olsen et Mitchell, 2010, p. 1) « *one of the most rapidly increasing alterations to the natural environment* ». Elle se définit comme « *the adverse effects artificial outdoor illumination can have on ecosystems and human well-being, on the aesthetics qualities of town- and landscapes, or on the visibility of the star-filled night sky* » (Meier et al., 2014, p. 2) ou plus simplement comme « *any adverse effect cause by artificial light* » (Meier et al., 2014, p. 103). D'après l'association internationale Dark-Sky, la pollution lumineuse s'exprimerait selon une multitude de formes telles que la sur-illumination (usage exagéré de la lumière), la lumière intrusive (pénétration non désirée de lumières dans un lieu), l'éblouissement (luminosité excessive provoquant des gênes visuelles), la luminescence du ciel nocturne au-dessus des régions urbanisées (halo lumineux) ou encore le regroupement désordonné de lumières créant la confusion.

Il en ressort ainsi que toutes sources de lumière artificielle est « *potentiellement source de lumière indésirable* » et que son « *origine* […] *est facilement et clairement identifiable »* (Klaus et al., 2005, p. 12). Par conséquent, une utilisation inadaptée de l'éclairage public (surévaluation des besoins, non canalisation du flux lumineux, mauvaise orientation, forte intensité, durée d'éclairage non modulée) (Klaus et al., 2005) ou privé (illumination des jardins, des maisons, vitrines et des enseignes publicitaires) représente la cause principale de la pollution lumineuse. Bien que le droit suisse régisse les émissions en tout genre (loi sur la protection de l'environnement), une estimation des émissions lumineuses et de leur développement se révèle indispensable pour évaluer l'ampleur du phénomène en Suisse. En conséquence, l'utilisation de la télédétection apparaît comme une solution adaptée pour mesurer les rayonnements lumineux nocturnes.

La figure 2 illustre de manière simplifiée l'observation nocturne à partir d'un satellite (Cao et Bai, 2014, p. 11919).

Propagation de la lumière

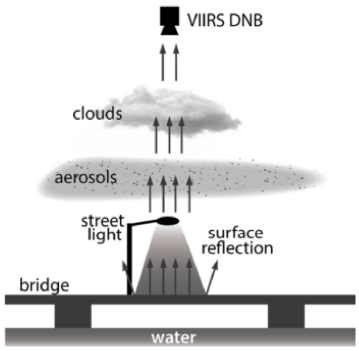

Figure 2 : Propagation de la lumière jusqu'au capteur du satellite. Sources : Cao et Bai (2014, p. 1191)

Une source lumineuse émet des rayons lumineux vers le sol. En fonction de la capacité réfléchissante (albédo) de ce dernier, une partie du rayonnement lumineux est propagée et transformée dans l'atmosphère, puis enregistré par les capteurs du satellite en fonction de leur bande spectrale (Cao et Bai, 2014, p. 11918). Il est important de mentionner qu'en l'absence de déflecteurs/réflecteurs canalisant les rayons vers le sol, les émissions lumineuses peuvent s'effectuer directement en direction du ciel.

Ainsi l'intensité de la source lumineuse (candela), son orientation (haut vers le bas ou inversement), sa canalisation (mécanisme d'occultation pour diriger le flux lumineux) sa durée d'éclairement (il faut bien entendu que le satellite passe sur la zone au moment où la lampe soit allumée) ou encore son type (lampes led, sodium,…) influencent grandement la détection du rayonnement lumineux par les capteurs du satellite. Ces derniers doivent donc posséder une sensibilité aux rayonnements de très faible magnitude émis par l'éclairage public et une largeur de bande adaptée aux longueurs d'onde des différents dispositifs d'éclairage (Figure 3). Par conséquent, il s'agit de capteurs spécifiques tels que l'Operational Linescan System (OLS) du Defense Meteorological Satellite Program (DMSP) ou le Visible/Infrared Imager Radiometer Suite (VIIRS) du Suomi National Polar-orbiting Partnership (NPP) et conçus pour des applications précises (militaires, météorologiques).

Longueur d'onde pour les lampes LED et sodium et largeur spectrale de la bande Day/Night du capteur VIIRS

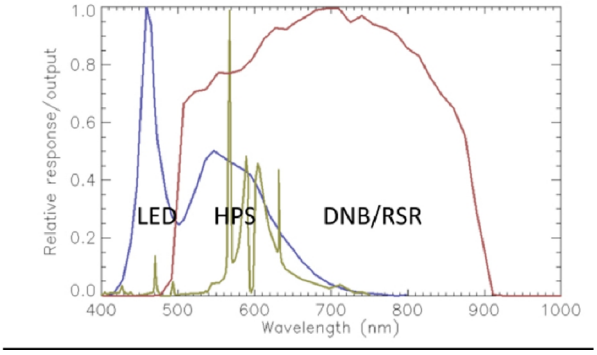

Figure 3: longueur d'onde pour quelques dispositifs et largeur spectrale du canal DNB-VIIRS. Sources : Cao et Bai (2014, p. 1191)

Alors que le capteur OLS du DMSP était initialement utilisé dans les années 70 pour « *detect clouds illuminated by moonlight for military bombing operations at night* » (Zhang et al., 2015, p. 291), sa sensibilité aux faibles luminosités[1] lui permet de détecter les lumières artificielles en l'absence de pleine lune, lui décernant ainsi des applications nouvelles telles que la détection des émissions lumineuses artificielles (pollution lumineuse). Quant au capteur VIIRS du satellite Suomi NPP météorologique devant succéder aux satellites DMSP, la haute sensibilité aux basses luminances de la bande day/night (DNB) lui confère, tout comme l'OLS, la possibilité de détecter les lumières artificielles.

Ces deux capteurs, dont les caractéristiques sont résumées dans le tableau 1, se différencient en certains points. Le capteur VIIRS offre de multiples avantages tels qu'une résolution spatiale au sol de 742 mètres et constante vers les bords (5km pour l'OLS à l'origine, 3km après lissage mais distorsion vers les bords du capteur), une meilleure détection des lumières faibles, une absence de saturation et une calibration interne (diffuseur solaire). Quant à la largeur de leur bande spectrale, le capteur OLS dispose d'une bande de 400 à 1100 nm et le VIIRS de 505 à 890. Ainsi ce dernier est moins apte à détecter les lumières de type LED (Light Emitting Diode) car le « *blue "pump" (peak near 450 nm) of the white LED is out of the spectral reponse of the DNB* » (Cao et Bai, 2014, p. 11918). En outre, aucun des deux ne couvre véritablement le pic lumineux aux alentours de 22 heures (19h30 pour l'OLS et 1h30 pour le VIIRS) (Elvidge et al., 2013, p. 63).

Quant à leur couverture temporelle, les archives des données remontent jusqu'en 1992 pour les images satellitaires du DMSP-OLS et 2011 pour celles du NPP-VIIRS. Par conséquent, les images issues des différents satellites[2] OLS du DMSP seront utilisées pour effectuer une analyse comparative spatio-temporelle. Néanmoins, elles ne peuvent « *be used directly for temporal analyses due to the lack of inflight calibration* » (Wu et al., 2013, p. 7356). Ainsi, les différents satellites DMSP possèdent leur propre performance radiométrique mais aussi une usure au fil du temps provoquant ainsi des « *differences between data from the same year obtenained by different satellites, as well as random fluctuations in data for consecutive years obtenained by the same satellite* » (Wu et al., 2013, p. 7357). Pour permettre l'analyse temporelle des images satellitaires DMSP, des méthodes d'inter-calibration ont été développées et consistent à « *convert data values from individual satellite products into a common range defined by reference year* » (Elvidge et al., 2014, p. 98). Quant aux images NPP-VIIRS, elles serviront à la mise en œuvre d'une analyse comparative spatiale. Dès lors, il est possible d'effectuer des hypothèses guidant la suite du travail.

Tableau 1: caractéristiques des capteurs OLS et VIIRS. Sources : (Elvidge et al., 2013, p. 64) et (Zhang et al., 2015, p. 294)

Capteur	***DMSP-OLS***	***NPP-VIIRS***
Constructeur - opérateur	U.S. Air Force	NASA – NOAA (JPSS)
Années opérationnelles	1972 – présent - archives depuis 1992	2011 – présent
Résolution spatiale	5000 ou 2700 (smoothed)	742 mètres
Résolution temporelle	Couverture globale chaque 24h	Quotidiennes
Heure de passage	~19h30	~1h30
Unité	6 bit	14 bit
Bandes spectrales	Panchromatique 400 à 1100 nm	Panchromatique 505 à 890 nm
Saturation	fréquente dans les centres urbains	aucune
Limite de détection des basses lumières	~5^{E}-10 Watts/cm^2/sr	~2^{E}-10 Watts/cm^2/sr
Calibration	aucune	Diffuseur solaire
Bandes supplémentaires	Infrarouge thermique (10 um)	21 bandes de 0,4 à 13 um.

[1] Luminances très faibles (10^{E}-9 W/sr/µm))

[2] F10, F12, F14, F15, F16 et F18

3. Hypothèses de travail

Les objectifs de travail ont été matérialisés en hypothèses explorant l'absence de différences entre les grandes régions suisses, l'existence de disparité entre les villes de tailles différentes, l'opposition entre l'urbain et le rural et la variation selon un gradient centre-périphérie. Ainsi, le phénomène ne devrait pas varier entre les grandes régions mais plutôt en fonction de la taille de la ville et d'un éloignement par rapport aux communes centrales.

Bien que l'hypothèse alternative (H_1) représente l'hypothèse de recherche, il a été choisi de mentionner l'hypothèse nulle (H_0) car « *le travail statistique est fondé sur les deux* » (Guay, 2014, p. 8). Ainsi :

- H_0 : Les cantons présentent des taux de croissance et une répartition des émissions lumineuses différents.
  H_1 : Les taux de croissance et la répartition des émissions lumineuses entre les cantons sont similaires.
- H_0 : Les aires métropolitaines connaissent des taux de croissance et une répartition des émissions lumineuses différents.
  H_1 : Les taux de croissance et la répartition des émissions lumineuses sont similaires entre les aires métropolitaines.
- H_0 : Les régions linguistiques possèdent des taux de croissance et une distribution du phénomène différents.
  H_1 : Les taux de croissance et la distribution des émissions lumineuses sont similaires entre les régions linguistiques.
- H_0 : Les taux de croissance et les émissions lumineuses sont identiques entre l'urbain et le rural.
  H_1 : Les régions urbaines connaissent des taux de croissance et des émissions lumineuses plus élevés que les régions rurales.
- H_0 : Les communes centres (CEN), périurbaines (PERI) et suburbaines (SUB) connaissent des taux de croissance et une répartition des émissions équivalents.
  H_1 : Les communes centres (CEN) possèdent des taux de croissance et une répartition des émissions plus élevés que les communes périurbaines (PERI) et suburbaines (SUB).
- H_0 : les taux de croissance et la répartition des émissions entre les communes intégrant une ville sont similaires.
  H_1 : Les communes de grande taille (>100'000 habitants) et incorporant une ville possèdent des taux de croissance et une répartition des émissions plus élevés que les communes de taille moyenne (50'000 à 99'999 habitants) et petite (20'000 à 49'999) incluant une ville.

L'expérimentation de ces hypothèses passe par l'acquisition d'une multitude de données dont la présentation est indispensable.

4. Présentation des données

Tableau 2: données utilisées

Nom	*Type de données*	*Format*	*Période temporelle*	*Autres*	*Sources*
DMSP-OLS Nighttime Lights Time Series	Images satellitaires	Tiff	1992 à 2012 (annuelle)	33 images	Earth Observations Group (EOG), National Oceanic and Atmospheric Administration (NOAA)
Nighttime VIIRS Day/Night Band Composites	Images satellitaires	Tiff	Avril 2015 à mars 2016 (mensuelle)	12 images	Earth Observations Group (EOG), National Geophysical Data Center, NOAA
Les niveaux géographiques de la Suisse	Table	Excel	Janvier 2015	2324 Communes	Office fédéral de la statistique, GEOSTAT (OFS)
Limites communales généralisées CH (G1g15)	Vectoriel (polygones)	Shapefile	2015	2324 Communes	Office fédérale de la statistique, GEOSTAT (OFS), Swisstopo
Statistique de la superficie 2004/09 (NOAS04)	Table	Excel	2004 - 2009	Hectares	Statistique suisse de la superficie (AREA)
PIB de la Suisse en $ constant (2005)	Table	Excel	1991 à 2014	Suisse	World Bank national accounts data, and OECD National Accounts data files.

Le tableau 2 résume l'ensemble des données utilisées lors de cette étude. Pour retracer l'évolution du phénomène, 33 images satellitaires DMSP-OLS Nighttime Lights Time Series en format tiff ont été récupérées gratuitement sur le site de l'Agence américaine d'Observation Océanique et Atmosphérique (NOAA)[3]. Lors de la création des images composites annuelles, la NOAA utilise uniquement des observations de grande qualité sans couverture nuageuse, ni d'éclairage solaire et lunaire. Ces images prennent des valeurs entre 1 et 63 et couvrent la période 1992 – 2012. Il a été choisi pour cette étude d'utiliser les données « *F1?YYYY_v4b_stable_lights.avg_vis.tif* » car elles sont exemptes de tout évènement éphémère (feu) et représentent uniquement les lumières persistantes ou stables. Cependant, l'inter-calibration des images annuelles provenant de multiples satellites DMSP-OLS[4] est essentielle en l'absence de calibration interne.

Quant à la distribution du phénomène, 12 images mensuelles d'avril 2015 à mars 2016 en format tiff ont été téléchargées[5]. Leur construction repose sur la combinaison de données issues de la bande DBN et exemptes de couverture nuageuse et de lumières parasites (foudre, lumière de la lune) (NOAA).

Pour effectuer des analyses statistiques sur les données, il est nécessaire d'agréger les pixels selon une unité d'analyse. La commune semble l'unité la plus adaptée par son nombre restreint (2324 en 2015) et la possibilité d'y joindre facilement les différents niveaux géographiques de la Suisse (unités d'observation). Cette jointure rend ainsi possible l'élaboration d'analyses en fonction de l'appartenance à un type de commune (urbain, rural, périurbain, etc.) ou à une région (linguistique, métropolitaine, etc.).

En outre, différents éléments utiles au bon déroulement de l'étude ont été collectés (couches de base en *shapefile* des lacs, des frontières administratives nationales et communales, une table contenant le PIB de la Suisse entre 1991 et 2014 indispensable lors du choix de la méthode d'inter-calibration des images DMSP-OLS, …).

[3] http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html
[4] F10, F12, F14, F15, F16 et F18
[5] http://ngdc.noaa.gov/eog/viirs/download_monthly.html

5. Méthodologie

La méthodologie suivie s'appuie sur quatre phases. Les données utilisées pour quantifier le phénomène sont, tout d'abord, extraites des images satellitaires DMSP-OLS et NPP-VIIRS pour chaque commune (SOL – Sum Of Lights). Elles sont ensuite traitées sous forme de tables, jointes aux niveaux géographiques de la Suisse et combinées pour former une base de données. Après une analyse exploratoire, les données sont transformées si nécessaire puis finalement analysées par des tests statistiques en fonction du caractère paramétrique ou non des données.

I. Collecte des données nécessaires sous Arcgis

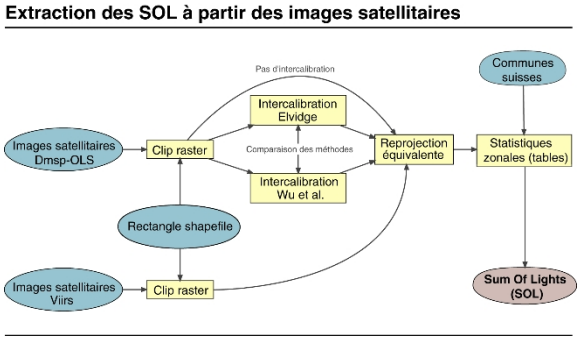

Figure 4 : traitements effectués pour extraire la somme des valeurs des pixels par communes

Les traitements résumés à la figure 4 ont été automatisés par un script python disponible à l'annexe 15 pour DMSP-OLS et à l'annexe 14 pour NPP-VIIRS. Dans la mesure où l'inter-calibration des images DMSP-OLS selon la méthode d'Elvidge a été automatisée lors d'un précédent travail (Haenni, 2016), il n'a pas été jugé nécessaire d'intégrer son script en annexe.

a. Prétraitements

Découpage

Les images satellitaires couvrant l'entièreté du globe (DMSP-OLS) ou une partie (NPP-VIIRS), il a été nécessaire de les découper en fonction d'un rectangle incluant la Suisse. Ce procédé permet d'éviter la manipulation de fichiers volumineux ralentissant les processus et traitements futurs.

Inter-calibration des images DMSP-OLS

Les images NPP-VIIRS étant calibrées entre elles, seules les images satellitaires DMSP-OLS ont été inter-calibrées selon les méthodes d'Elvidge et al. (2014) et Wu et al. (2013). Ces deux méthodes reposent sur l'invariance des valeurs des pixels d'une région pour Elvidge (Sicile) (2014, p. 101) ou de trois pour Wu (Mauritus, Puerto Rico, Okinawa) (Wu et al., 2013, p. 7360) afin d'obtenir un modèle de régression. La méthode de Wu, se basant sur une image de référence non saturée, permet en outre de corriger la saturation des images satellitaires DMSP-OLS(Wu et al., 2013, p. 7359). Dès lors, il est possible d'appliquer les équations suivantes à l'ensemble des images satellitaires :

Pour la méthode d'Elvidge (2014, p. 102) :

$$Y=C_0+XC_1+X^2C_2$$

Pour la méthode de Wu et al. (2013, p. 7360) :

$$Y + 1 = a*(X+1)^b$$

X représente les images non inter-calibrées, C_o, C_1, C_2, a et b sont les coefficients issus des différents modèles et indiqués par Elvdige et al. (2014, p. 102) ou Wu et al. (2013, p. 7362).

Vérification des résultats

Pour contrôler les résultats de l'inter-calibration, les valeurs des pixels pour la Suisse ont été sommées avant et après l'application des deux méthodes et rassemblées dans les figures disponibles à l'annexe 1. Les deux méthodes permettent une meilleure continuité des données entre les années et une diminution des écarts entre des images d'années similaires mais de satellites DMSP différents. Des analyses complémentaires quant à la vérification de l'inter-calibration peuvent, en outre, être consultées à l'annexe 1.

Comparaison des méthodes

Les deux méthodes harmonisant les valeurs des données, il est dès lors nécessaire de choisir celle la plus optimale pour la Suisse. Pour ce faire, l'indice de la différence normalisée (NDI – normalized difference index) proposé par Wu et al (2013, p. 7361) et reposant sur la comparaison d'images de

même année mais de capteurs différents, a été calculé pour les deux méthodes.

$$NDI = \frac{|SOL_1 - SOL_2|}{SOL_1 + SOL_2}$$

Un NDI faible représente ainsi un inter-calibration de qualité satisfaisante. La méthode la plus adaptée correspondrait donc à celle dont la moyenne des NDI est la plus faible. La figure 5, rassemblant les différents NDI selon les deux méthodes, illustre la difficulté à départager les deux méthodes car les moyennes (0.0261 pour Elvidge et 0.0283 pour Wu), et les écarts sont similaires entre les méthodes.

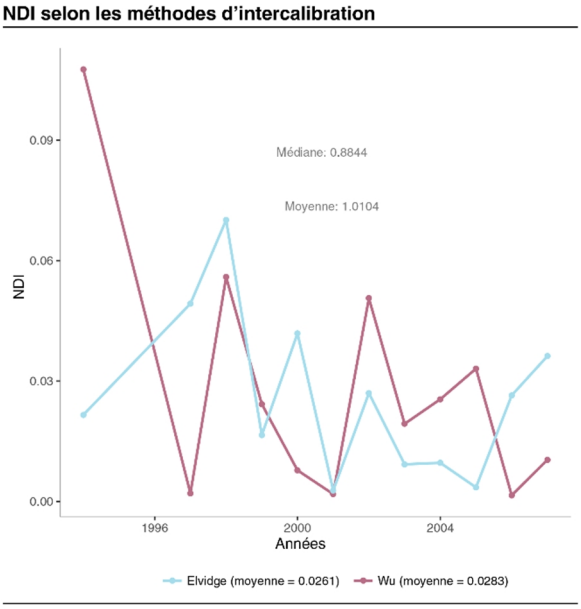

Figure 5 : NDI selon les méthodes d'intercalibration - utile à la vérification

Une autre démarche soumise par Wu et al (2013, p. 7361) consiste à analyser la relation entre la somme des valeurs des pixels (SOL) pour chaque année et le produit intérieur brut (PIB – en $ constant 2005). Ces deux variables étant liées par une relation linéaire (corrélation) selon certaines études (Wu et al., 2013, p. 7366), la méthode la plus adéquate possède ainsi le R^2 le plus proche de 1. La figure 6 révèle une corrélation forte (R^2 de 0.747) entre le PIB et la somme des lumières pour la méthode de Wu et faible (R^2 de 0.409) pour la méthode d'Elvidge. Par conséquent la méthode de Wu semble la plus adaptée pour cette étude.

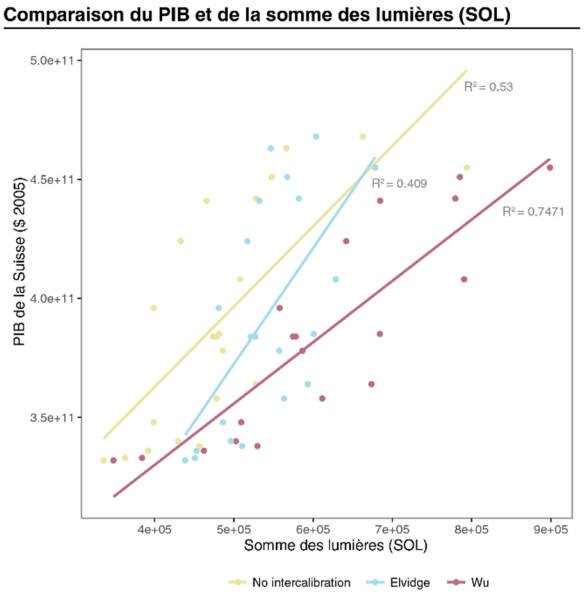

Figure 6 : lien entre le PIB en $ constant de 2005 et la somme des valeurs des pixels (SOL)

Calcul d'une composite annuelle (moyenne) – NPP-VIIRS
Les images mensuelles du NPP-VIIRS d'avril 2015 à mars 2016 ont été combinées en une seule image couvrant ainsi une année complète et permettant une analyse plus robuste.

b. Projection équivalente

Les images satellitaires présentant un système de coordonnées géographiques GCS_WGS_1984, il a été indispensable de les projeter selon une projection équivalente conique d'Albers conservant les surfaces et dont les caractéristiques sont résumées ci-contre. Pour obtenir une projection optimale, cette dernière est centrée sur la Suisse.

Projection	Albers
False_Easting	0
False_Northing	0
Central_Meridian	8.15
Standard_Parallel_1	45
Standard_Parallel_2	48
Latitude of origin	46.41
Linear Unit	Mètre

c. Extraction et agrégation des valeurs des pixels par communes

La somme des valeurs des pixels (SOL) de chaque commune a été extraite par l'outil « *statistique zonale (table)* » avec une résolution de 250 mètres pour correspondre à l'unité spatiale d'analyse (la commune). Dès lors, un travail de recombinaison des tables est essentiel pour obtenir une base de données exploitable.

II. Création d'une base de données sous Excel

L'étape précédente a permis la collecte d'une table contenant la somme des lumières issue des images satellitaires NPP-VIIRS et de 31 tables correspondant aux images DMSP-OLS pour les 2324 communes suisses. Lorsque deux satellites DMSP couvrent la même année, une moyenne des deux images est utilisée, réduisant ainsi les tables à 19 (1992 à 2010). L'étape suivante consiste à joindre, en utilisant le numéro des communes comme géocode (identifiant), les différents niveaux géographiques (*Les niveaux géographiques de la Suisse – OFS*) ainsi que la superficie des communes en hectares avec la table intégrant les données agrégées issues des satellites. À ce propos, l'annexe 2 résume les différentes nomenclatures retenues pour cette étude, leurs valeurs et leurs intitulés.

Pour retracer l'évolution du phénomène entre 1992 et 2010, un taux de croissance (tc) a été calculé :

$$tc = \frac{(Présent - Passé)}{Passé}$$

Etant donnée la grande variabilité des données issues des satellites DMSP-OLS, ce taux de croissance a été calculé à partir de la moyenne de deux périodes (1992-1995 à 2007-2010), réduisant ainsi passablement le risque d'erreur dû à l'inter-calibration.

Quant aux différentes données issues de l'agrégation par commune des valeurs des pixels des images satellitaires (SOL), elles sont divisées par la superficie en hectares de la commune correspondante pour minimiser les effets de taille et ainsi rendre possible leur comparaison.

$$SOLc\ commune\ i = \frac{SOL\ commune\ i}{Surface\ commune\ i}$$

La base de données rassemblant les deux variables quantitatives (taux de croissance entre les périodes 1992-1995 et 2007-2010 et la somme des valeurs des pixels corrigée par la superficie en hectares) et les différents niveaux géographiques pour chacune des 2324 communes suisses, il est dès lors possible d'effectuer des comparaisons spatio-temporelles (1992 et 2010 – DMSP-OLS) et spatiales (NPP-VIIRS, moyenne d'avril 2015 à mars 2016).

III. Analyse exploratoire des données sous R

La base de données recueillant ainsi des variables quantitatives (tc et SOLc) et qualitatives (niveaux géographiques de la Suisse), il est judicieux de partager l'analyse exploratoire entre les données spatio-temporelles (taux de croissance) et spatiales (SOLc) ainsi qu'entre les différentes variables qualitatives). L'analyse exploratoire consiste principalement à vérifier la normalité de la distribution (test de Shapiro-Wilk, Skewness, kurtosis, histogramme et boxplot), l'homogénéité des variances (Test de Levene et de Fligner Killeen) et la présence d'outliers (Boxplot, z-score). Les deux premières assomptions sont fondamentales car elles orientent le choix des futurs tests statistiques qui serviront à évaluer la significativité des différences au sein des divers sous-groupes. Dans le cas de leur violation, une transformation (logarithme, racine) est appliquée sur les données et permet parfois de résoudre certains problèmes de distribution et d'hétérogénéité des variances.

Toujours est-il que, lorsqu'aucune transformation ne fonctionne, le recours à des tests non-paramétriques ne faisant aucune hypothèse sur la distribution des données s'impose. Après une brève initiation théorique aux tests statistiques mentionnés précédemment et à leur seuil de rejet α, ces derniers seront, comme en témoigne la figure 7, appliqués au taux de croissance (tc) et à la somme des valeurs des pixels corrigée par la surface en hectares (SOLc) pour l'ensemble des données et pour chaque sous-groupe (niveaux géographiques suisses). De plus, différents exemples de commandes sous *R* en lien avec l'analyse exploratoire sont disponibles aux annexes 3 et 4.

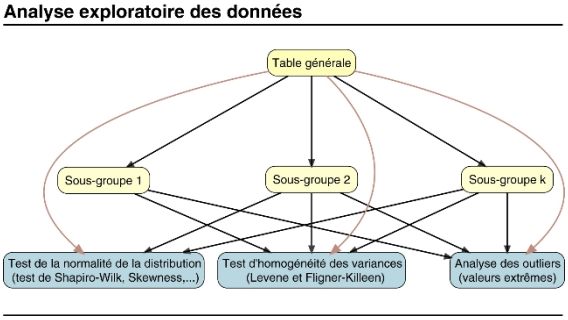

Figure 7: résumé des traitements effectués lors d'une analyse exploratoire des données

a. Généralité des tests et traitements effectués

Seuil de rejet a

La plupart des tests statistiques reposent sur l'acceptation d'une hypothèse nulle (H_0) ou son rejet pour une alternative (H_1) en fonction d'un seuil de rejet α. Ce dernier correspond à « *une valeur de p value choisie avant de réaliser le test statistique et au-dessous de laquelle on rejette H_0* » (Millot, 2014, p. 205). Comme dans la plupart des travaux, un seuil de rejet $\alpha = 0.05$ a été jugé adéquat et sera retenu pour l'ensemble des tests utilisés dans cette étude.

Ainsi une *p-value* inférieure ou égale à α ($p \leq 0.05$) implique le rejet de l'hypothèse nulle et donc l'acceptation de l'alternative H_1. Au contraire une *p-value* supérieure au seuil de rejet ($p > 0.05$) entraine la non possibilité de rejeter H_0 (Millot, 2014, p. 205).

Hypothèse de normalité de la distribution

Dans la mesure où la plupart des tests font l'hypothèse d'une distribution normale des données, il est nécessaire de vérifier cette dernière avec le test de Shapiro-Wilk, un histogramme et les coefficients de dissymétrie (Skewness) et d'aplatissement (kurtosis).

D'après Millot (2014, p. 637), le test de Shapiro-Wilk vise « *l'ajustement d'une distribution observée à une loi normale* ». Ainsi, le rejet au seuil $\alpha = 0.05$ de l'hypothèse nulle (suit une loi normale) implique la non-conformité à une loi normale.

Quant aux coefficients d'asymétrie et d'aplatissement, ils prendraient des valeurs quasi nulles lorsque la distribution est normale (Field, Field et Miles, 2013, p. 174). Un coefficient d'asymétrie positif « *indicate a pile-up of scores on the left of the distribution, whereas negative values indicate a pile up on the right* » (Field et al., 2013, p. 174). Quant au kurtosis, plus sa valeur est positive et plus la distribution présente un pic vers la moyenne (Field et al., 2013, p. 174). Il est alors possible d'estimer la forme de la distribution en fonction de la valeur de ces deux coefficients.

Hypothèse de l'homogénéité des variances

Condition primordiale pour la plupart des tests paramétriques, l'homogénéité des variances est estimée par le test paramétrique de Levene ou son équivalent non paramétrique Fligner-Killeen. Ils permettent, selon Millot (2014, p. 564) de comparer k variances ($k \geq 2$) et prennent comme hypothèse nulle « *les k variances sont identiques dans la population visée* » et alternative « *l'une au moins des variances diffère d'an moins une autre dans la population visée* » (Millot, 2014, p. 564).

Influence et traitement des outliers

Un outlier ou individu extrême se définit d'après Millot (2014, p. 128) « *par le fait que dans la distribution d'une variable quantitative, il se retrouve éloigné, par sa valeur, des autres individus de l'échantillon* ». Ainsi, les graphiques en boîtes à moustaches (Boxplot) sont particulièrement adaptés pour détecter les outliers.

Etant donnée leur tendance à influencer drastiquement la moyenne, il est primordial d'effectuer une analyse minutieuse de tous les cas dont le z-score (valeur standardisée et centrée) est supérieur ou égal à |3.29| (Field et al., 2013, p. 146). Néanmoins, un score de |3| a été retenu permettant une analyse plus minutieuse. Selon ce critères, les communes possédant des valeurs extrêmes ont été cartographiées (annexe 5).

Après la vérification, pour chaque valeur extrême, de l'absence d'erreur lors de l'extraction et du traitement des données, il convient de se poser la question de leur suppression ou non. A ce sujet, Millot mentionne à propos d'un outlier qu' « *il pourrait être risqué de l'écarter de l'étude car il est représentatif de la population visée, dans son hétérogénéité naturelle. Supprimer cet individu équivaudrait d'une certaine façon à biaiser l'échantillonnage.* » (Millot, 2014, p. 129).

En l'absence donc de critères justifiables (erreurs de mesure ou de traitement), l'élimination des valeurs extrêmes « *modifie les moyennes et biaise la variance résiduelle, qui est diminuée de façon systématique* » (Chabanet et Dessaint, 2015, p. 5) et doit donc être exceptionnelle.

b. Analyse exploratoire des taux de croissance (annexe 6 et 8)

Communes

L'exploration des données pour l'ensemble des 2324 communes (figure 8) a permis les constatations suivantes :

- L'asymétrie de la distribution vers la gauche (Skew de 2.52), sa forme pointue (kurtosis de 12.51), la différence entre la moyenne (1.0104) et la médiane (0.8844), l'absence de correspondance entre l'histogramme et la courbe de distribution normale théorique $N(1.0104, 0.6696)$ et la significativité ($p < 0.05$) du test de Shapiro-Wilk, montrent une distribution des taux de croissance par communes non conforme à une loi normale N (1.0104, 0.6696).
- La présence de 38 observations possédant un z-score supérieur à|3|, a nécessité un travail de vérification des traitements effectués précédemment pour détecter des erreurs de saisi ou de manipulation.

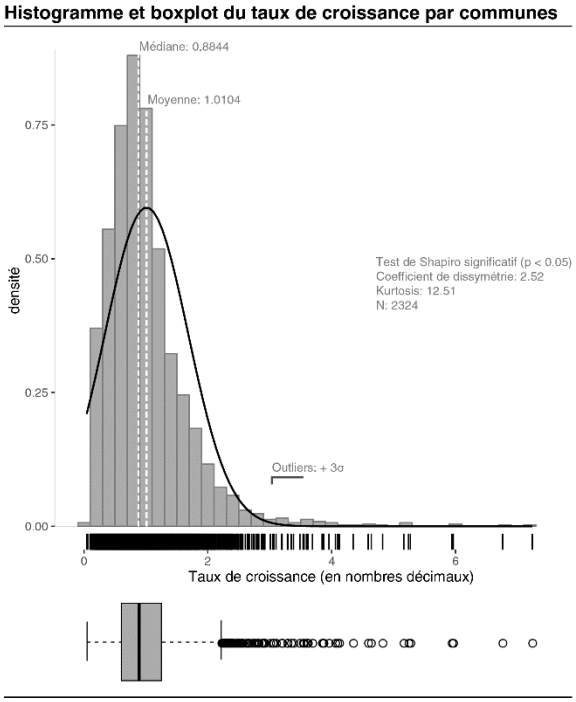

Figure 8 : histogramme et boxplot du taux de croissance par communes suisses. Une distribution fortement asymétrique.

Pour diminuer l'importance des valeurs extrêmes et l'asymétrie de la distribution, il a été choisi de transformer en logarithme les taux de croissance (log(x+1)). Avec cette transformation, la distribution coïncide en partie avec la courbe de distribution normale théorique $N(0.2842, 0.1236)$ (figure 9), les coefficients d'asymétrie (0.74) et d'aplatissement (1.39) tendent vers 0 et les outliers ne représentent plus que 26 observations soit 1% de l'échantillon. Cependant, la significativité ($p < 0.05$) du test de Shapiro-Wilk contredit partiellement les observations visuelles et illustre ainsi la limitation de ce test car « *in large samples, this test can be significant even when the scores are only slightly different from a normal distribution.* » (Field et al., 2013, p. 185).

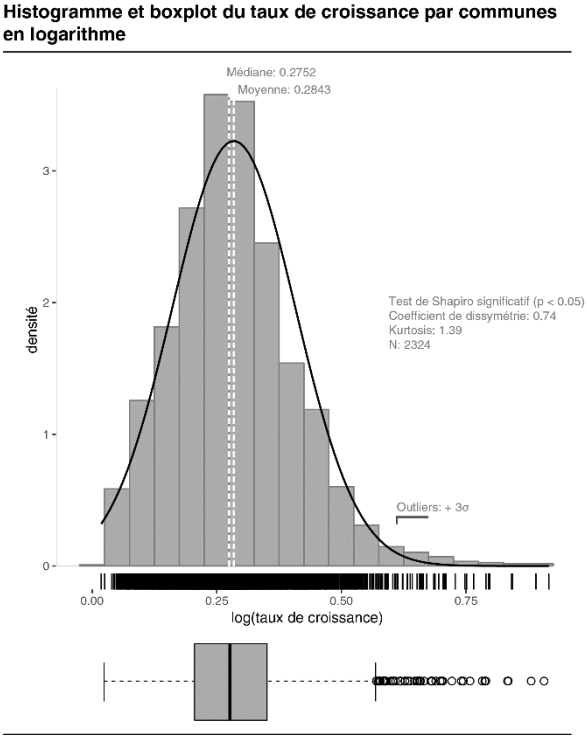

Figure 9 : histogramme et boxplot du taux de croissance par communes suisses en logarithme. Une courbe proche d'une distribution normale avec une minimisation des outliers.

Par conséquence, l'utilisation d'un qqplot (quantile-quantile plot) permettant de comparer l'ajustement d'une distribution observée avec une loi de distribution théorique (normale dans ce cas) (Millot, 2014, p. 633), apparait primordiale pour interpréter les résultats partiellement contradictoires.

Le QQplot de la figure 10 démontre ainsi la non-conformité de la distribution à une loi normale quelconque. En effet, les points forment une

courbe en S et diffèrent donc considérablement de la droite représentant la normalité.

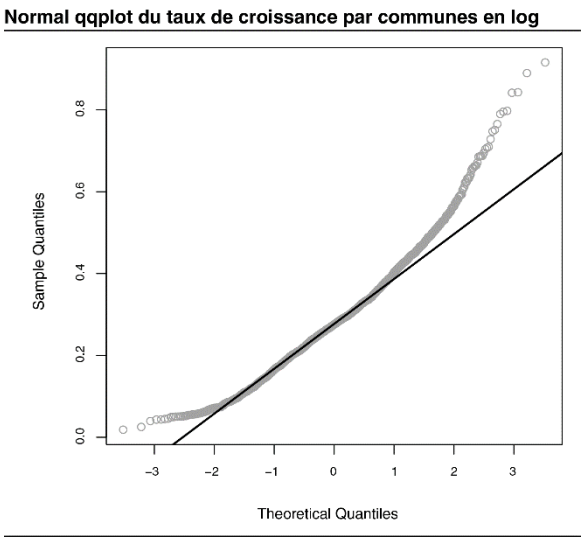

Figure 10 : qqplot du taux de croissance par communes en log. Les points ne s'alignant pas le long de la droite théorique d'une distribution normale, la distribution des taux de croissance en log ne suit pas une loi normale.

L'analyse précédente s'effectuant sur l'ensemble des 2324 communes suisses, il est nécessaire d'appliquer les mêmes procédés aux différents sous-groupes des variables qualitatives. Bien que différents résultats (données originales, en logarithme, sans outliers) aient été calculés, il a été choisi de présenter uniquement les résultats de l'analyse exploratoire des données transformées en logarithme pour éviter toute surcharge d'informations.

Cantons

La Suisse comptant 26 cantons, l'analyse visuelle est rendu difficile par le nombre élevé de sous-groupes. Ainsi, seuls les résultats statistiques seront présentés.

Normalité de la distribution

11 cantons suivent une loi normale selon le test de Shapiro-Wilk. Pour la plupart, les faibles valeurs des coefficients d'asymétrie et d'aplatissement confirment les résultats précédents.

Homogénéité des variances

Le test de Fligner-Killeen présente une *p-value* (2.2^{e}-16) inférieure au seuil de rejet α. Par conséquent, au moins une des variances se différencie d'au moins une autre.

Présence d'outliers

Une seule valeur extrême est constatée dans les cantons de Lucerne, Soleure, Bâle-Campagne, Schaffhouse, Saint-Gall, des Grisons et du Tessin, deux dans le canton de Vaud et Fribourg et trois pour le canton de Berne.

Régions linguistiques

La Suisse allemande, romande, italienne et romanche comptent respectivement 1459, 685, 152 et 28 communes.

Normalité de la distribution

Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) ainsi que l'analyse de la figure 11 montrent que les régions suisses-allemandes, romandes et italiennes ne suivent pas une loi normale. Malgré l'acceptation de l'hypothèse nulle du test de Shapiro-Wilk (p-value = 0.127), il semblerait que la région romanche s'éloigne d'une distribution normale. En effet, les points sur le qqplot s'alignant faiblement au bas de la droite normale, il est difficile de conclure quant à la distribution des données pour la Suisse romanche.

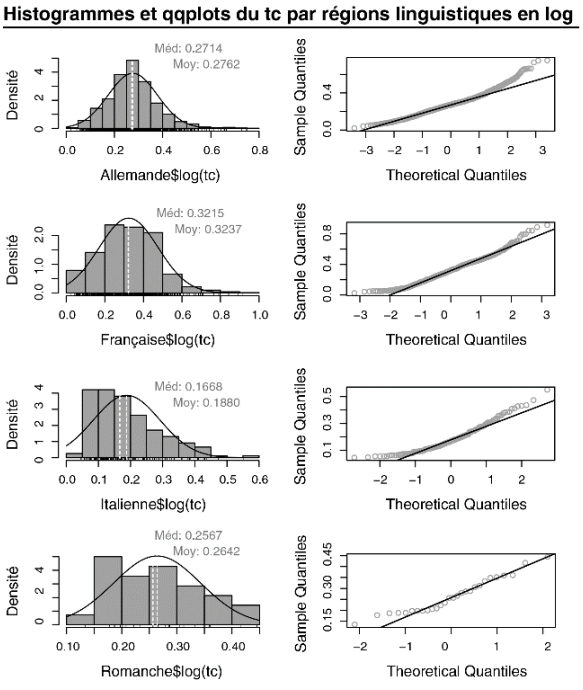

Figure 11 : histogrammes et qqplots du taux de croissance (tc) par régions linguistiques en log. Un éloignement général de la droite de la normalité théorique.

Homogénéité des variances

Le test de Fligner-Killeen dont la *p-value* (2.2^{e}-16) est inférieure au seuil de rejet α, décèle qu'au moins une des variances se différencie d'une autre, au moins.

Présence d'outliers

Les communes suisses-allemandes, romandes et italiennes contiennent respectivement 14, 7 et une communes aux valeurs extrêmes. La partie romanche se caractérise par l'absence de valeurs extrêmes.

Aires métropolitaines

La Suisse contient cinq aires métropolitaines. 216, 38, 74, 161 et 12 communes appartiennent respectivement aux aires métropolitaines de Zürich, Berne, Bâle, Genève-Lausanne et Ticino-Urbano.

Normalité de la distribution

Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) pour les aires métropolitaines de Zürich, Bâle et Genève-Lausanne ainsi que l'analyse de la figure 12 démontrent que ces régions ne suivent pas une loi normale. Quant à Berne et Ticino-Urbano, les résultats du test de Shapiro-Wilk sont en partie confirmés par l'analyse visuelle pour l'aire métropolitaine de Berne (coefficient d'asymétrie faible, les points suivent la droite normale sur le qqplot) et partiellement pour Ticino-Urbano (quelques faibles valeurs sont éloignées de la droite normale).

Homogénéité des variances

Le test de Fligner-Killeen présente une *p-value* (8.429^{e}-12) inférieure au seuil de rejet α et donc l'hypothèse nulle doit être rejetée. Par conséquent, au moins une des variances se différencie d'une autre, au moins.

Présence d'outliers

Seule l'aire métropolitaine de Genève-Lausanne contient des valeurs extrêmes (2).

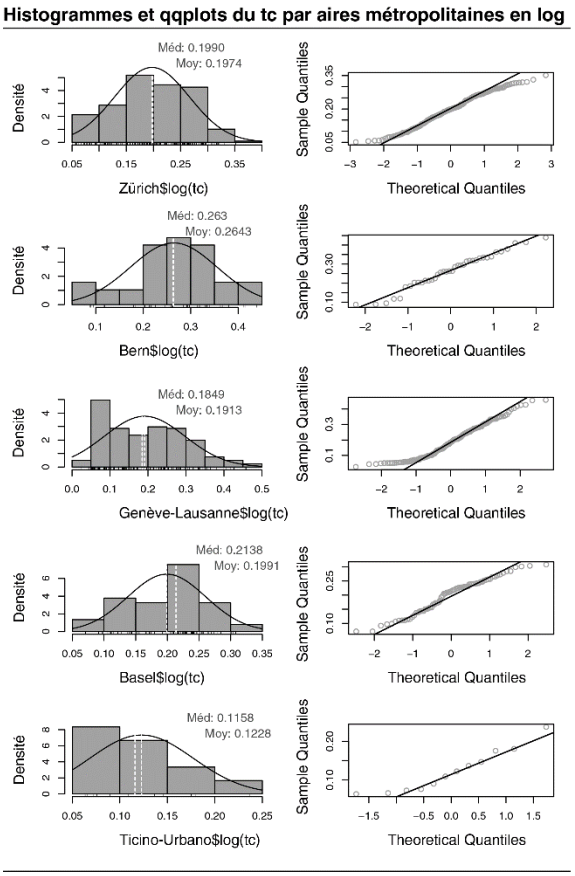

Figure 12 : histogrammes et qqplots du taux de croissance par aires métropolitaines. Des distributions bien éloignées d'une loi normale.

Urbain – rural

À partir de la typologie territoriale « *espace à caractères urbains* 2012 », 1504 communes ont été classées dans le type « *urbain* » et 820 dans le « *rural* ».

Normalité de la distribution

Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) ainsi que l'analyse de la figure 13 montrent que les deux sous-groupes ne suivent pas une loi normale. Bien que les histogrammes suivent quasiment une courbe normale, les deux qqplots présentent des points sensiblement éloignés de la droite de distribution normale.

Homogénéité des variances

Les deux sous-groupes ne suivant pas une loi normale, l'homogénéité des variances est testée par Fligner-Killeen. La *p-value* étant égale à 0.2274, l'hypothèse nulle ne peut pas être rejetée et donc les différences de variances ne sont pas significatives.

Présence d'outliers

13 communes rurales et 14 urbaines sont considérées comme des valeurs extrêmes.

Histogrammes et qqplots du tc par communes urbaines et rurales en logarithme

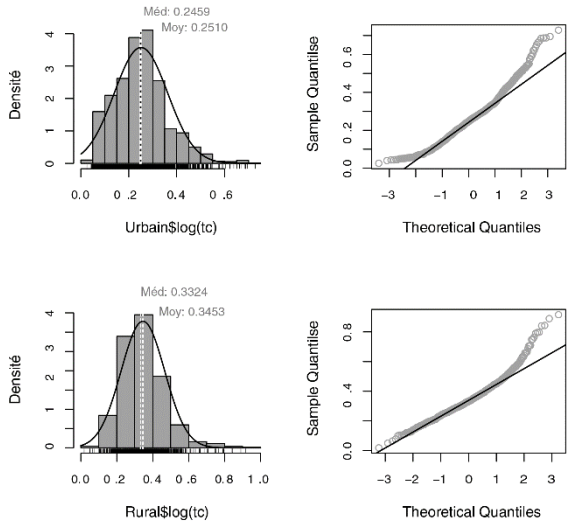

Figure 13 : histogrammes et qqplots du taux de croissance par communes urbaines et rurales en log. Des courbes en S vers les extrémités.

Communes incluant une ville de grande, moyenne et petite taille

Les communes intégrant une ville sur son territoire ont été classées en fonction de leur taille. Ainsi, la Suisse comprend 6 communes disposant d'une grande ville, 4 d'une ville moyenne et 33 d'une petite ville.

Normalité de la distribution

Les résultats du test de Shapiro-Wilk montrent, par le rejet de l'hypothèse nulle au seuil α (*p-value* > 0.5), que la distribution suit une loi normale pour les trois sous-groupes. Néanmoins, l'analyse visuelle de la figure 14 démontre, pour les villes moyennes, la difficulté de confirmer les résultats du test de Shapiro-Wilk par manque d'observations.

Boxplot et qqplots du tc selon la taille des communes incluant des villes de plus de 20'000 habitant(e)s, en log

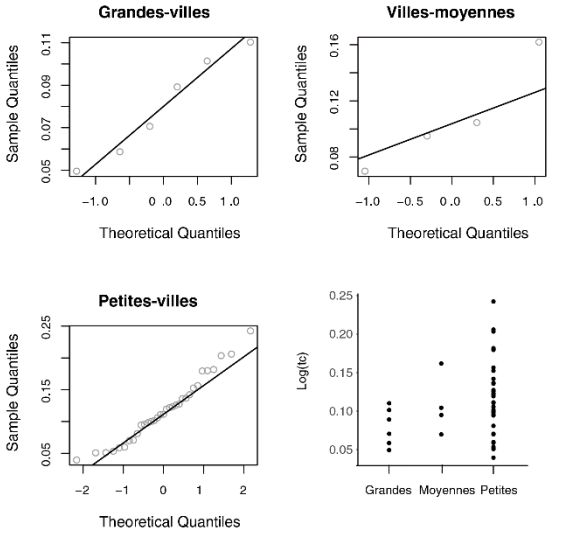

Figure 14 : histogrammes et qqplots du taux de croissance selon la taille des communes incluant une ville en log. Une interprétation difficile de la normalité.

Homogénéité des variances

Le test de Levene disposant d'une *p-value* (0.2814) supérieure au seuil de rejet α, décèle que les différences entre les variances ne sont pas significatives.

Présence d'outliers

Les trois sous-groupes ne présentent aucun outlier.

Gradient centre-périphérie

Le gradient centre-périphérie rassemble 71 communes centrales, 316 suburbaines et 396 périurbaines.

Normalité de la distribution

Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) ainsi que l'analyse de la figure 15 montrent que les trois sous-groupes ne suivent pas une loi normale. Quoique cette dernière confirme les résultats du test de Shapiro-Wilk pour les communes de type « *centre* » et « *suburbain* », le type « *périurbain* » se rapproche fortement d'une distribution normale par un coefficient d'asymétrie faible (0.3047) et des points suivant la droite de distribution normale (qqplot).

Homogénéité des variances

Le test de Fligner-Killeen souligne, par le non rejet de l'hypothèse nulle (*p-value* = 0.2077) que les différences entre les variances ne sont pas significatives.

Présence d'outliers

Alors que les communes centrales ne possèdent pas de valeurs extrêmes, le suburbain et le périurbain se caractérisent par la présence respective de quatre et un outliers.

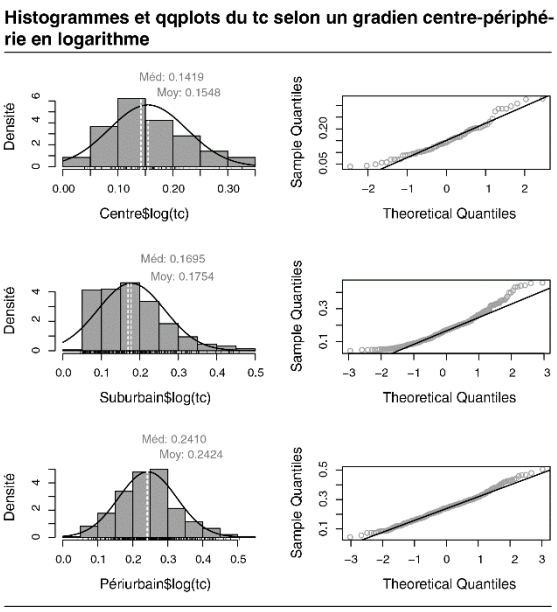

Figure 15 : histogrammes et qqplots du taux de croissance selon un gradient centre-périphérie. Une symétrie, une courbe en cloche et une relative droite laissant présager une distribution normale pour les communes périurbaines.

c. *Analyse exploratoire des SOLc (annexes 7 et 9)*

Communes

L'exploration des données pour l'ensemble des 2324 communes (Figure 16) a permis les constatations suivantes :

- L'asymétrie très forte de la distribution vers la gauche (Skew de 4.26), sa forme pointue (kurtosis de 25.90), la différence entre la moyenne (34.35) et la médiane (16.87), la très forte non-conformité entre l'histogramme et la courbe de distribution normale théorique $N(34.35, 51.26)$ et la significativité ($p < 0.05$) du test de Shapiro-Wilk, montrent une distribution des SOLs par communes non conforme à une loi normale N (34.35, 51.26).
- La présence de 56 observations possédant un z-score supérieur à|3|, a nécessité une vérification des traitements effectués auparavant pour éviter des erreurs

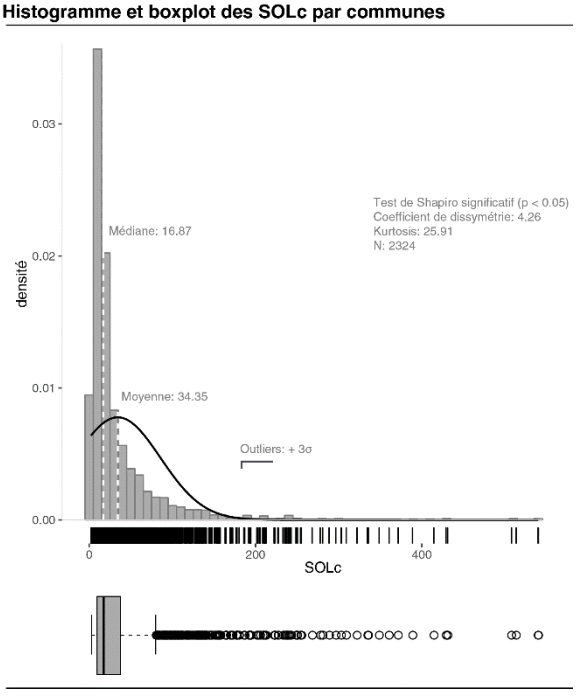

Figure 16 : histogramme et boxplot des SOLc par communes suisses. Une distribution fortement asymétrique, un pic vers la moyenne et une influence des outliers.

La transformation en logarithme base 10 (log(x)) des données a permis de réduire drastiquement les valeurs extrêmes (7 outliers) et l'asymétrie de la distribution (0.45) ainsi que d'aplatir sa forme (-0.20). La distribution tend à prendre une forme en cloche proche de la courbe de distribution normale théorique N (1.27, 0.45) (figure 17). Tout comme le taux de croissance, la significativité ($p < 0.05$) du test de Shapiro-Wilk conteste en partie les observations visuelles. Par conséquence, l'utilisation d'un qqplot se révèle nécessaire.

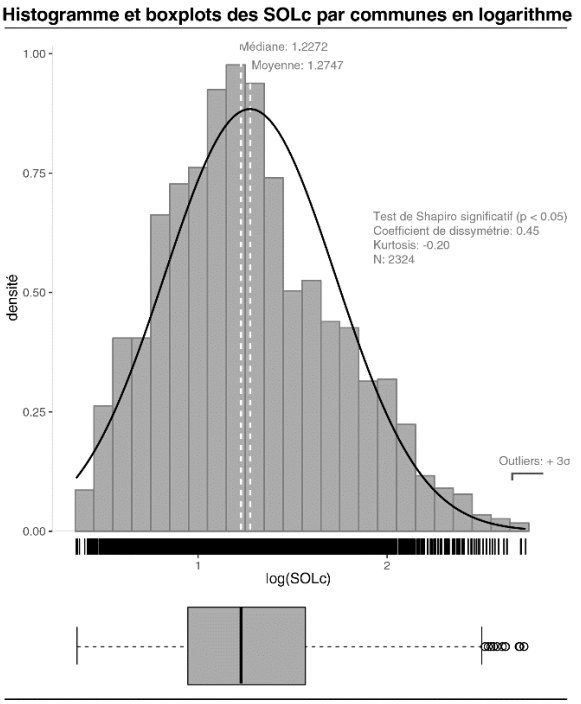

Figure 17 : histogramme et boxplot après transformation en logarithme des SOLc par communes. Une réduction de l'asymétrie et des outliers, un rapprochement d'une courbe en cloche et d'une distribution quasi normale.

L'analyse visuelle du qqplot (figure 18) montre une déviation forte de la droite représentant la normalité pour les valeurs faibles. Les points tendent à former une courbe en S. Ainsi, la distribution des SOLs diffère d'une distribution normale $N(34.35, 51.26)$.

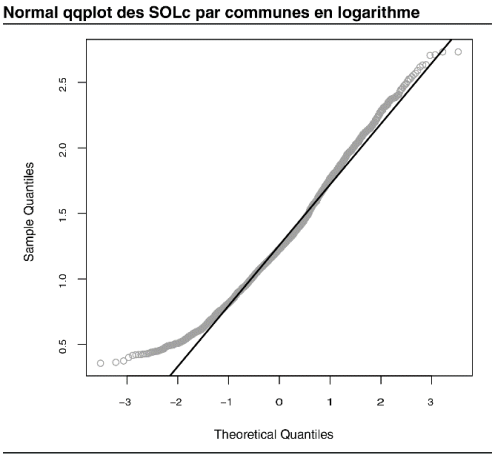

Figure 18 : qqplot SOLc par communes en log. Un éloignement dans les extrêmes.

Il est, dès lors, nécessaire d'effectuer des analyses exploratoires pour les différents sous-groupes de chaque variable qualitative.

Cantons

Normalité de la distribution
10 cantons suivent une loi normale selon le test de Shapiro-Wilk (l'hypothèse nulle ne peut être rejetée). Pour la plupart, les faibles valeurs des coefficients d'asymétrie et d'aplatissement confirment les résultats précédents.

Homogénéité des variances
Le test de Fligner-Killeen présente une *p-value* (2.2^{e}-16) inférieure au seuil de rejet α. Par conséquent, au moins une des variances se différencie d'au moins une autre.

Présence d'outliers
Une seule valeur extrême est constatée dans les cantons de Lucerne, Schaffhouse et du Jura, deux dans le canton de Vaud et Bern et trois pour le canton de Fribourg et des Grisons.

Régions linguistiques

Normalité de la distribution
Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) ainsi que l'analyse de la figure 19 montrent que les régions suisses-allemandes, romandes, italiennes et romanches ne suivent pas une loi normale. En effet, les points sur les différents qqplots prennent une forme en S et s'éloignent donc de la droite de normalité théorique.

Homogénéité des variances
Le test de Fligner-Killeen dont la *p-value* (2.2^{e}-16) est inférieure au seuil de rejet α, décèle qu'au moins une des variances se différencie d'au moins une autre.

Présence d'outliers :
Aucun outlier n'a été détecté.

Histogrammes et qqplots des SOLc par régions linguistiques en logarithme

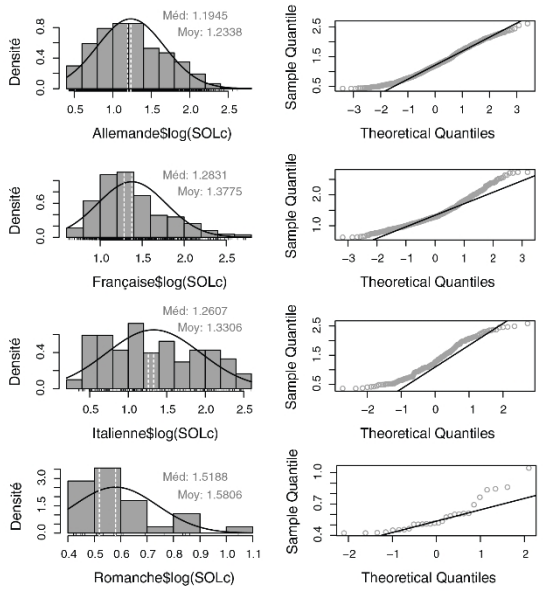

Figure 19 : histogrammes et qqplots des SOLc par régions linguistiques en logarithme. Présence de courbes en S.

Aires métropolitaines

Normalité de la distribution

Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) pour les aires métropolitaines de Berne, Bâle et Genève-Lausanne ainsi que l'analyse de la figure 20 montrent que ces régions ne suivent pas une loi normale. Quant à Zürich et Ticino-Urbano, les résultats du test de Shapiro (distribution normale) sont en partie confirmés par l'analyse visuelle pour l'aire métropolitaine de Zurich (coefficient d'asymétrie faible, les points suivent la droite normale sur le qqplot) et partiellement pour Ticino-Urbano (les points prennent une forme en S).

Histogrammes et qqplots des SOLc des aires métropolitaines en logarithme

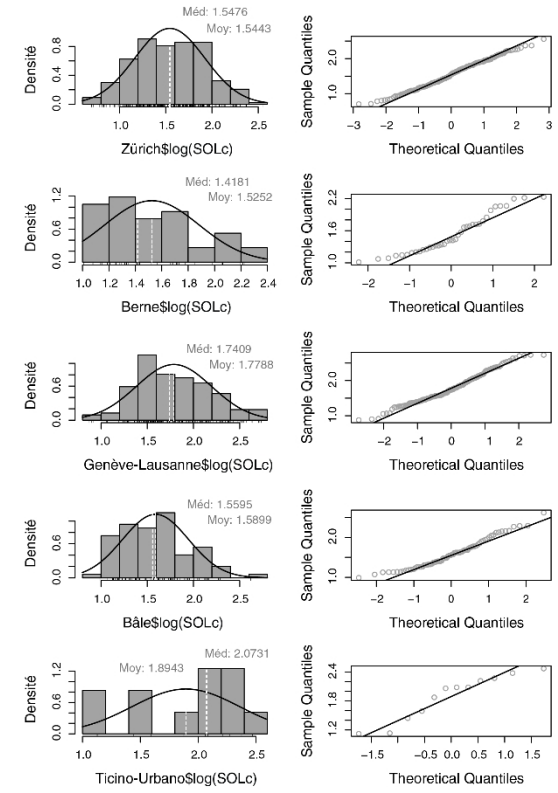

Figure 20 : histogrammes et qqplots des SOLc selon les aires métropolitaines en logarithme. Des difficultés pour interpréter la distribution du Ticino-Urbano.

Homogénéité des variances

Le test de Fligner-Killeen présente une *p-value* (0.6557) supérieure au seuil de rejet α. Par conséquent, les différences entre les variances ne sont pas significatives.

Présence d'outliers

Les aires métropolitaines suisses allemandes et romandes disposent respectivement de deux et quatre outliers.

Urbain – rural

Normalité de la distribution

Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) ainsi que l'analyse de la figure 21 confirment que les deux sous-groupes ne suivent pas une loi normale. Les deux qqplots témoignent d'une courbe en forme de S s'éloignant donc de la droite de normalité théorique.

Homogénéité des variances

La *p-value* du test de Fligner-Killeen étant égale à 0.1613, l'hypothèse nulle ne peut être rejetée et donc les différences de variances ne sont pas significatives.

Présence d'outliers

Quatre communes rurales et trois urbaines sont jugées comme des valeurs extrêmes.

Histogrammes et qqplots des SOLc par communes urbaines et rurales en logarithme

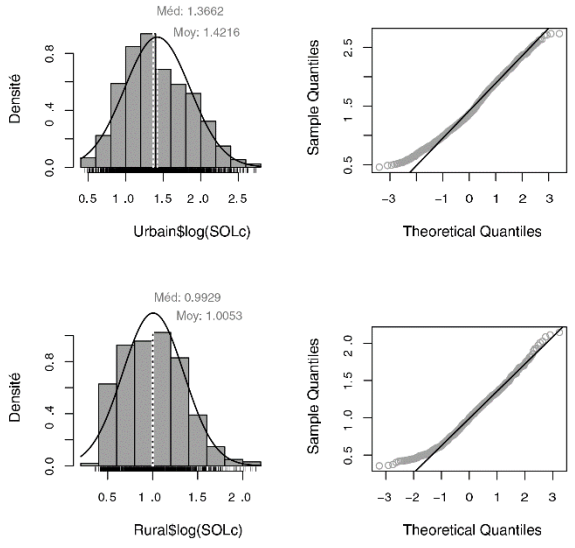

Figure 21 : histogrammes et qqplots des SOLc selon le caractère urbain ou rural des communes. La présence de courbes en s démontre la non-normalité des distributions.

Communes incluant une ville de grande, moyenne et petite taille

Normalité de la distribution

Les résultats du test de Shapiro-Wilk démontrent, par le rejet de l'hypothèse nulle au seuil α, que la distribution suit une loi normale pour les deux sous-groupes « *grandes-villes* » (*p-value* de 0.6087) et « *villes-moyennes* » (0.9607). Quant aux petites-villes, la *p-value* de 0.04352 indique que la distribution n'est pas normale. Néanmoins, l'analyse visuelle de la figure 22 démontre une difficulté pour interpréter les résultats des grandes et moyennes villes.

Boxplot et qqplots des SOLc selon la taille des communes incluant un ville de plus de 20'000 habitant(e)s, en log

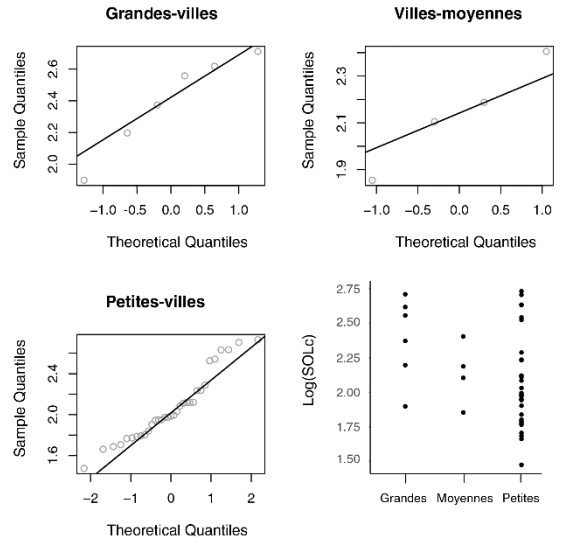

Figure 22 : histogrammes et qqplots des SOLc selon la taille des communes incluant une ville. Le manque d'observation rend difficile une confirmation des résultats du test de Shapiro-Wilk pour les grandes et moyennes villes.

Homogénéité des variances

Le test de Fligner-Killeen disposant d'une *p-value* (0.859) supérieure au seuil de rejet α, décèle que les différences entre les variances ne sont pas significatives.

Présence d'outliers

Aucun outlier n'a été détecté.

Gradient centre-périphérie

Normalité de la distribution

Le rejet de l'hypothèse nulle du test de Shapiro-Wilk (*p-value* < 0.05) ainsi que l'analyse de la figure 23 montrent que le sous-groupe « *centres* » ne suit pas une distribution normale. En effet, l'histogramme et le qqplot illustrent un trou dans la distribution (valeurs < 1.5) qui se répercute par un non alignement des points le long de la droite normale pour les faibles valeurs. Quant aux deux autres sous-groupes (périurbain et suburbain), ils suivent une distribution normale car l'hypothèse nulle du test de Shapiro-Wilk est acceptée (*p-value* de 0.9079 pour le suburbain et de 0.1746 pour le périurbain), leurs coefficients d'asymétrie et d'aplatissement tendent vers 0 et les points s'alignent parfaitement sur les qqplots.

Homogénéité des variances

Le test de Fligner-Killeen dont la *p-value* (0.03376) est légèrement inférieure au seuil de rejet α, détecte qu'au moins une des variances se différencie d'au moins une autre.

Présence d'outliers

Aucun outlier n'a été détecté.

Histogramme et qqplots des SOLc selon un gradient centre-périphérie en logarithme

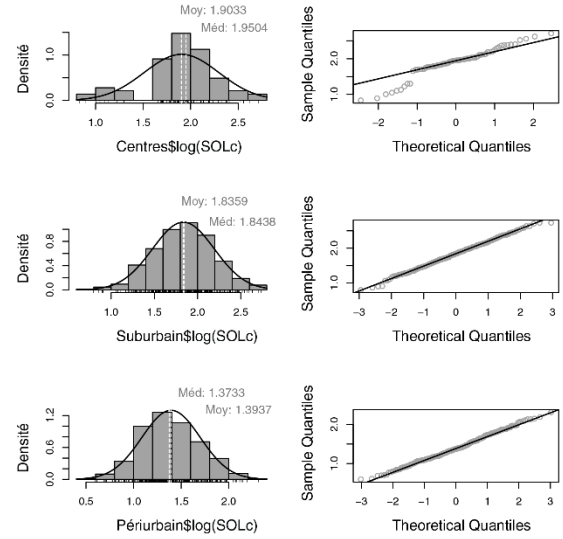

Figure 23 : histogrammes et qqplots des SOLc selon un gradient centre-périphérie. L'alignement des points sur la droite et les courbe en cloche témoignent d'une distribution normale pour les communes suburbaines et périurbaines.

Le tableau 3 résume l'ensemble des conclusions des analyses exploratoires précédentes. La plupart des données par sous-groupes ne répondant ni à l'assomption de normalité ni à celle de l'homogénéité des variances, l'analyse comparative consistera principalement en l'application de tests non paramétriques (Kruskal-Wallis, Mann-Whitney-Wilcox).

IV. Analyse comparative sous R

Tableau 3 : aspect paramétrique des données et type de tests à appliquer.

Taux de croissance en log	*N sous-groupes*	*Normalité de la distribution*	*Homogénéité des variances*	*Type de test*	*Test de comparaison*
Cantons	26	non	non	non paramétrique	Kruskal-Wallis
Aires métropolitaines	5	non	non	non paramétrique	Kruskal-Wallis
Urbain-rural	2	non	oui	non paramétrique	Mann-Whitney-Wilcoxon
Gradient centre-périphérie	3	non	oui	non paramétrique	Kruskal-Wallis
Taille des villes	3	oui	oui	paramétrique	Anova à un facteur
Régions linguistiques	4	non	non	non paramétrique	Kruskal-Wallis

SOLc en log	*N sous-groupes*	*Normalité de la distribution*	*Homogénéité des variances*	*Type de test*	*Test de comparaison*
Cantons	26	non	non	non paramétrique	Kruskal-Wallis
Aires métropolitaines	5	non	oui	non paramétrique	Kruskal-Wallis
Urbain-rural	2	non	oui	non paramétrique	Mann-Whitney-Wilcoxon
Gradient centre-périphérie	3	non	non	non paramétrique	Kruskal-Wallis
Taille des villes	3	non	oui	non paramétrique	Kruskal-Wallis
Régions linguistiques	4	non	non	non paramétrique	Kruskal-Wallis

Les tests de cette section visent à comparer l'homogénéité de certains paramètres de la population (Millot, 2014, p. 288), tels que k moyennes observées pour l'anova (analyse de variance à un facteur, $k \geq 2$) ou k médianes observées pour les tests de Mann-Withney-Wilcoxon ($k = 2$) et de Kruskal Wallis ($k \geq 2$).

L'avantage de ces deux derniers tests réside dans une minimisation de l'influence des valeurs extrêmes par la comparaison des médianes. Il est, néanmoins, possible de retranscrire les résultats des tests aux moyennes sous certaines conditions (distributions identiques, égalité de la différence entre la moyenne et la médiane) (Millot, 2014, p. 483).

Le tableau 3 indique les tests à appliquer aux différents niveaux géographiques de la Suisse en fonction du nombre k de sous-groupes, de la normalité de la distribution et de l'homogénéité des variances. Bien que la variable « taille des villes » demande l'utilisation d'une anova pour le taux de croissance, la majorité des analyses consisteront en un test de Kruskal-Wallis suivi d'une comparaison par paires (test unilatéral de Mann-Withney-Wilcoxon) ou d'un test de tendance (Jonckheere–Terpstra).

Le test de Kruskal Wallis permettant une première estimation de l'existence de différences au sein de chaque sous-groupe, il est donc nécessaire d'effectuer des analyses complémentaires par paires, c'est-à-dire pour les $k(k-1)/2$ combinaisons possibles, ou selon un ordre préétabli pour évaluer les tendances au sein des différents sous-groupes de chaque variable qualitative. La figure 24 résume ainsi les tests appliqués en fonction de l'aspect paramétrique ou non de la distribution et du nombre k de sous-groupes.

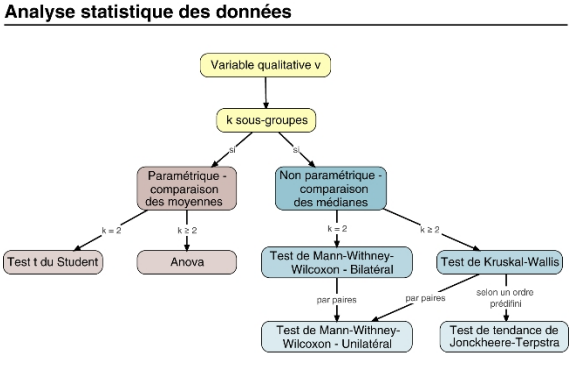

Figure 24 : schéma récapitulatif de l'analyse comparative des données.

Quant aux hypothèses, l'anova et le test de Kruskal-Wallis postulent comme hypothèse nulle « *l'égalité des k moyennes (ou médianes pour Kruskal-Wallis)* » et « *une moyenne (respectivement médiane) au moins se différencie d'une autre* » pour l'alternative.

Pour le test non paramétrique de Mann-Withney-Wilcoxon, « *l'égalité des deux médianes* $Méd_1 = Méd_2$ » représente l'hypothèse nulle, « *la différence des deux médianes* $Méd_1 \neq Méd_2$ » l'alternative pour un test bilatéral, « *la supériorité de la médiane* $Méd_1 > Méd_2$ » l'alternative pour un test unilatéral droit et « *l'infériorité de la médiane* $Méd_1 < Méd_2$ » pour un test unilatéral gauche. Une *p*-value inférieure à 0.05 implique le rejet de l'hypothèse nulle et ainsi l'existence d'une différence significative entre les sous-groupes.

De plus, des graphiques en boites à moustaches (boxplot) ont été compilés pour confirmer les résultats des tests et éviter toutes conclusions contradictoires (Annexe 11) et des exemples de codes sous *R* en lien avec l'analyse comparative sont disponibles à l'annexe 10. Quant aux différents résultats, ils sont disponibles aux annexes 12 pour le taux de croissance et 13 pour les SOLc.