



**UNIVERSITÉ
DE GENÈVE**

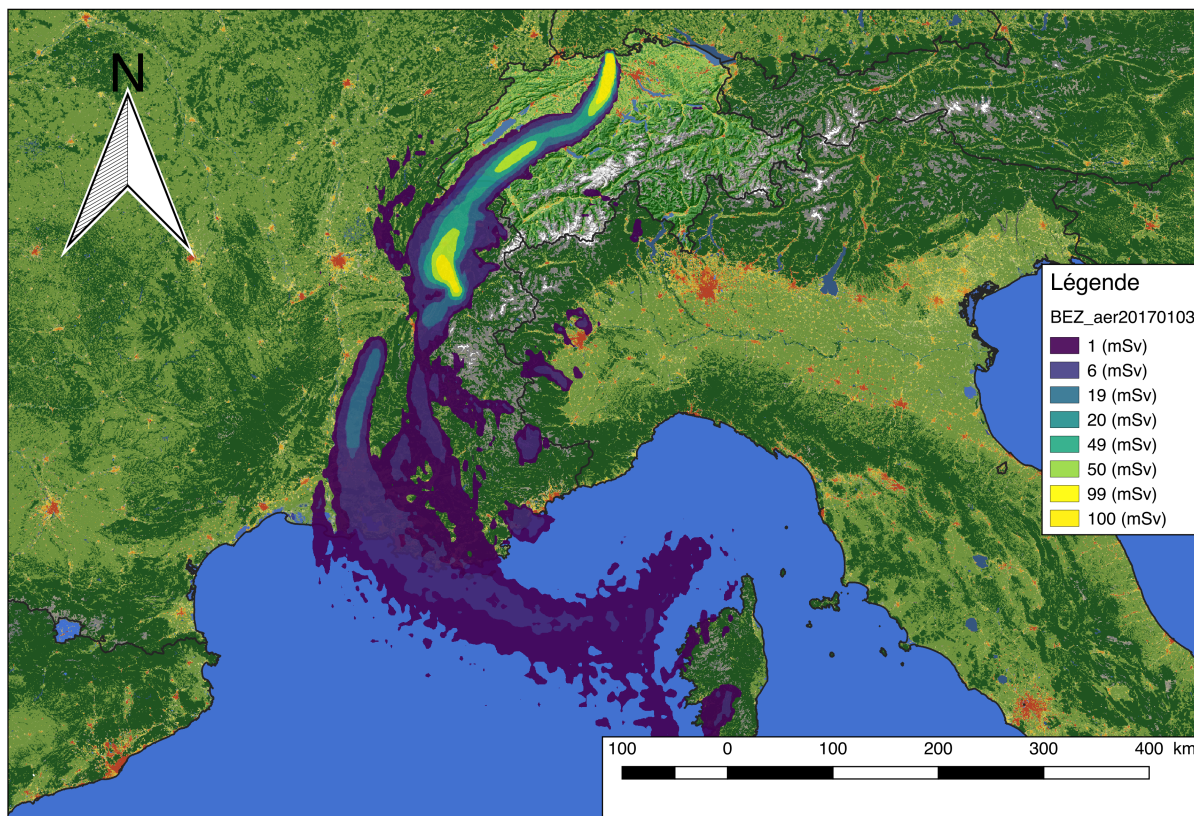


Institut Biosphère
Savoirs pour les Humains et la Planète

Rapport de stage

Certificat complémentaire en Géomatique

Cartographie des impacts d'un accident nucléaire majeur en Suisse



Candidat : Bastien Deriaz

Superviseur : Dr. Gregory Giuliani

Janvier 2019

Table des matières

1. Résumé	5
2. Introduction	7
2.1 Sortir du nucléaire	7
2.2 But de l'étude	7
2.3 Les centrales.....	8
2.4 Les données.....	8
2.5 But du stage	10
3. Méthode	11
3.1 Programmes utilisés.....	11
3.2 Stratégie.....	11
3.3 Les données statistiques.....	12
3.3.1 <i>Population</i>	12
3.3.2 <i>Couverture du sol</i>	12
3.4 Les scripts	13
3.4.1 <i>Conversion des KML en SHP</i>	15
3.4.2 <i>Production de fichiers raster</i>	18
3.4.3 <i>Cumule des données nuage et déposition</i>	19
3.4.4 <i>Statistiques</i>	20
3.4.5 <i>Organisation des résultats</i>	21
4. Résultats	25
4.1 Points forts	27
4.2 Points faibles.....	28
5. Bilan du stage	31
5.1 Perspectives.....	31
6. Conclusion	33
7. Références	35
8. Annexes	37

1. Résumé

Ce stage mandaté par l'association *Sortir du nucléaire* a pour but de faire une étude statistiques sur la population et les surfaces touchées lors d'un hypothétique accident nucléaire majeur en Suisse. Quatre sites nucléaires suisses (Beznau, Goesgen, Leibstadt et Mühleberg) ainsi que le site du Bugey en France ont été sélectionnés pour cette étude et des calculs de dispersion atmosphérique ont été effectués avec le programme *HYSPLIT* et les données météorologiques de NOAA (National Oceanic Atmospheric Administration) pour une période allant de mi-2016 à fin 2018. La mission a alors été de convertir de façon automatisée ces fichiers journaliers de dispersion (*KML*) en fichiers (*SHP* et *TIFF*) utilisables dans le logiciel *QGIS* afin de cumuler différents types de rejets et d'effectuer une statistique de zone en croisant les fichiers obtenus avec une couche de densité de population du GHS (Global Human Settlement, European Commission) et une couche de couverture du sol (*CLC, Corine Land Cover* du programme Copernicus). Un script *R* a été développé afin d'organiser les résultats sous une forme utilisable où les données journalières séparées sont regroupée par type de rejet et par année. L'automatisation des différents processus et traitements a été un succès grâce à l'utilisation de *Python* dans *QGIS*. Cependant, des ralentissements dus au programme lui-même ont été découverts et certains scripts ne peuvent pas être utilisés sur une trop long période de temps car les outils du programme font saturer la mémoire de la machine hôte.

2. Introduction

Des accidents graves tels ceux de Tchernobyl (1986) et Fukushima (2011) soulèvent des questions quant à la sécurité des centrales nucléaires dans le monde. À ce jour dépendante de l'énergie nucléaire, la Suisse compte quatre sites qui ont produit en 2017 31.7%¹ de l'électricité du pays. Malgré une légère diminution de la production depuis le début du millénaire (~25'000 GWh en 2000 contre ~19'500 GWh en 2017) le nucléaire reste en seconde place (après l'énergie hydraulique) vis à vis de la production d'électricité en Suisse, alors que la plupart de ces centrales ont été mises en service il y plus de 40 ans. La Suisse n'étant pas à l'abri d'un accident majeur il est nécessaire d'évaluer les impacts qu'un tel événement aurait sur la population (suisse et européenne) ainsi que sur les différentes surfaces de couverture du sol (suisse et européennes).

2.1 Sortir du nucléaire

Cette étude a été mandatée par l'association *Sortir du nucléaire* et a été menée en collaboration avec l'*institut Biosphère*. *Sortir du nucléaire* est une association à but non lucratif dont les buts sont :

- « L'interdiction de toute installation nucléaire de production d'énergie,
- La promotion de l'efficacité énergétique et des énergies renouvelables »²

Cette association fait partie de l'Alliance Suisse « Non au nucléaire », ainsi que de Réseau Sortir du Nucléaire en France. Elle regroupe des membres de différents cantons. Le mandat vise à produire une étude qui permettrait de sensibiliser la population aux enjeux de l'énergie nucléaire.

2.2 But de l'étude

Le but de cette étude est d'évaluer les impacts d'un accident nucléaire majeur en Suisse sur la population ainsi que sur la couverture du sol. Pour ce faire, il a été choisi de considérer un accident du type A5 (selon l'AIEA³) lors duquel de la matière radioactive est rejetée par la centrale et des contre mesures sont à appliquer. Cependant, le rejet reste dans ce cas limité contrairement aux accidents les plus importants de l'histoire (Tchernobyl (1986) et Fukushima (2011)) auxquels a été attribué le niveau maximal A7.

Lors de rejet dans l'atmosphère les zones impactées dépendent de la situation météorologique durant l'émission. Afin d'avoir une approche statistique l'étude est faite sur plusieurs années (2016 - 2018). L'évolution des nuages contenant les éléments radioactifs a été intégrée sur 72h en utilisant le programme de dispersion

¹ Selon l'Office Fédérale de l'Énergie (<http://www.bfe.admin.ch>)

² Statuts de l'association (<http://sortirdunucleaire.ch>)

³ IAEA (<https://www.iaea.org/>)

*HYSPLIT*⁴. Les données météorologiques permettant les calculs de trajectoire ont été récupérées à la NOAA (National Oceanic Atmospheric Administration) avec une résolution de 0.05°.

2.3 Les centrales

Les centrales prises en compte pour cette étude sont les centrales suisses ainsi que celle du Bugey qui présente un intérêt de part sa proximité géographique (voir Figure 2-1). Il existe différents types de centrales nucléaires. Les différences de fonctionnement sont caractérisées par des variations de puissance pouvant modifier les scénarios de rejet en cas d'accident. Les centrales étudiées sont soit du type REP (Réacteur à Eau Pressurisée) soit du type REB (Réacteur à Eau Bouillante), voir Table 1.

Table 1 : Propriétés des centrales nucléaires étudiées^{3,5}

Centrale	Réacteur	Type de réacteur	Mise en service commerciale	Puissance électrique nette (MWe)
BEZNAU	BEZNAU-1	REP	1969	365
	BEZNAU-2	REP	1971	365
BUGEY	Bugey-2	REP	1979	910
	Bugey-3	REP	1979	910
	Bugey-4	REP	1979	880
	Bugey-5	REP	1980	880
GOESGEN	GOESGEN	REP	1979	970
LEIBSTADT	Leibstadt	REB	1984	1165
MUHLEBERG	Mühleberg	REB	1972	373

2.4 Les données

Les données produites par le programme de dispersion (*HYSPLIT*) ont été converties au format KML afin de permettre leur ouverture dans un GIS. Celles-ci se présentent sous la forme de nuages composés de plusieurs isolignes représentant les niveaux de concentration des éléments radioactifs. Une distinction a été faite entre la dispersion des radioéléments dans l'atmosphère et leur déposition sur le sol. Les isolignes ont été directement traduites en $Bq*s/m^3$ pour les nuages et en Bq/m^2 pour la déposition.

Les rejets sont séparés en trois différents types :

⁴ NOAA/Air Resources Laboratory, College Park, Maryland

⁵ IFSN (<https://www.ensi.ch/fr>)

- les aérosols, qui contiennent des éléments volatiles ou partiellement volatiles (les isotopes utilisés dans cette étude sont ^{140}Ba , ^{58}Co , ^{60}Co , ^{134}Cs , ^{136}Cs , ^{137}Cs , ^{131}I , ^{86}Rb , ^{103}Ru , ^{106}Ru , ^{127}Sb , ^{89}Sr , ^{90}Sr , $^{127\text{m}}\text{Te}$, $^{129\text{m}}\text{Te}$ et ^{132}Te).
- les éléments réfractaires, qui contiennent des éléments du groupe de métaux réfractaires (les isotopes utilisés dans cette étude sont ^{241}Am , ^{141}Ce , ^{144}Ce , ^{242}Cm , ^{244}Cm , ^{95}Nb , ^{147}Nd , ^{143}Pr , ^{238}Pu , ^{239}Pu , ^{240}Pu , ^{241}Pu , ^{91}Y et ^{95}Zr).
- les gaz rares qui contiennent des éléments du groupe des gaz nobles (les isotopes utilisés dans cette étude sont ^{85}Kr et ^{133}Xe).

À noter que la déposition a été calculée à partir de la dispersion des nuages d'aérosols et réfractaires uniquement car la déposition de gaz est négligeable. Le but final étant d'étudier les risques pour la population, les isolignes des nuages (en $\text{Bq}\cdot\text{s}/\text{m}^3$) peuvent être converties en mSv (milli sieverts), unité permettant d'évaluer l'impact des rayonnements sur l'Homme. Il existe différentes façons de faire cette conversion, pour cette raison différents fichiers ont été créés pour chaque type de rejet avec des limites d'isolignes spécifiques suivant les directives de l'IFSN (Inspection Fédérale de la Sécurité Nucléaire), les recommandations de l'ECRR (European Committee on Radiation Risk) et l'ORaP (Ordonnance sur la Radioprotection). La déposition est quant à elle caractérisée par l'activité radioactive sur une surface donnée (Bq/m^2). Les codes d'identification/production des données (fichiers *KML*) sont présentés dans la Table 2.

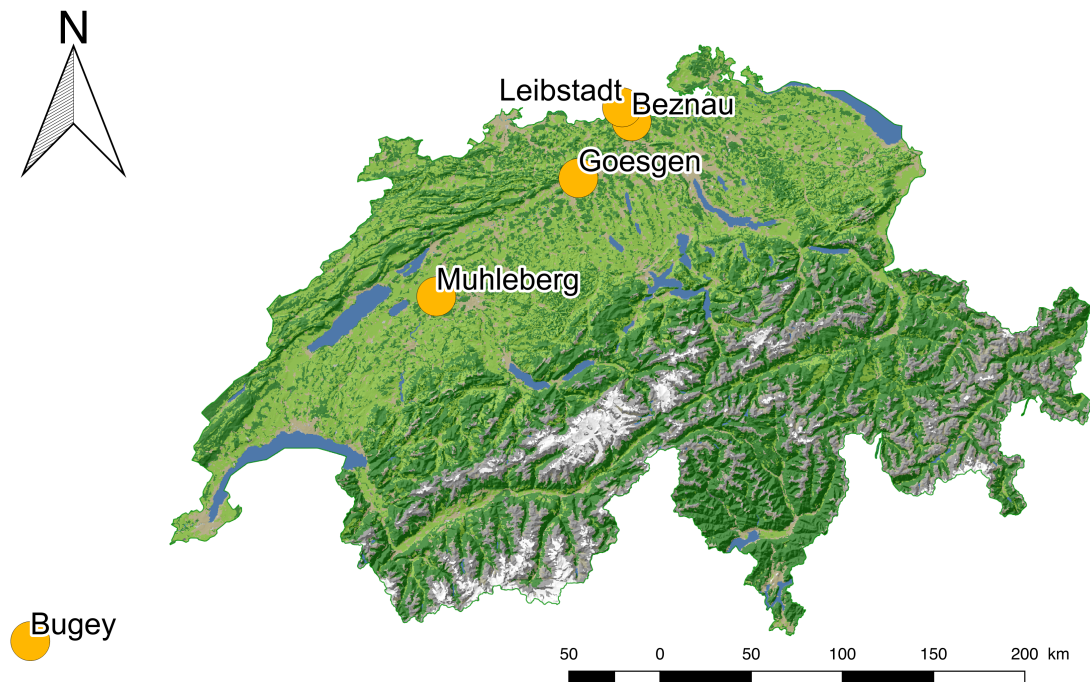


Figure 2-1 : Localisation des centrales nucléaires étudiées

Table 2 : Codes pour les différentes données

	Type de rejet	Directives de conversion (à mSv)	Code identification des données
Nuage	Gaz rare	IFSN	*1GAS*
	Aérosols	IFSN	*2AER*
		ORaP	*2AERHPRESS*
		ECRR	*2AERECRR*
	Réfractaires	IFSN	*3REF*
		ORaP	*3REFHPRESS*
Déposition	Aérosols	-	*2AERDEPO*
	Réfractaires	-	*3REFDEPO*

Les fichiers ont été architecturés façon à avoir toutes les informations nécessaires ainsi qu'à permettre l'automatisation des calculs statistiques. Les données de dispersion sont donc classées dans des dossiers journaliers. Par exemple, pour un jour donné, les données d'aérosols pour la centrale de Beznau sont nommées ainsi (le préfixe *hys* indique que les données ont été produites avec le programme *HYSPLIT*) :

hys_BEZ_2AER_concplot.kml

Les indicatifs pour les différentes centrales sont les suivant :

- Beznau : BEZ
- Bugey : BUG
- Goesgen : GOE
- Leibstadt : LEI
- Mühleberg : MUH

Ceux-ci seront utilisés dans les scripts *Python* et *R* afin de définir les données de quelle centrale sont calculées.

2.5 But du stage

Ce stage, sous la supervision de Dr. Grégory Giuliani, a pour but d'appliquer les connaissances théoriques obtenues lors des cours du Certificat Complémentaire en Géomatique à l'UNIGE.

La mission principale de ce stage est d'utiliser les données de dispersion de la matière radioactive lors d'un hypothétique accident nucléaire en Suisse (ou proche) afin d'établir une statistique sur la population touchée ainsi que sur les différents types de sols à l'aide d'outils géomatiques. Comme mentionné précédemment l'étude est faite pour plusieurs centrales nucléaires avec plusieurs types de rejets et ce, sur une période de plusieurs années, ainsi il est nécessaire d'automatiser les processus au vu du très grand nombre de données et de traitements. Il faut aussi choisir quelles couches (population et couverture du sol) seront utilisées pour l'étude statistique. Et enfin, une fois que les résultats ont été produits, des solutions quant à leur organisation doivent être trouvées afin de permettre leur bonne lecture.

3. Méthode

3.1 Programmes utilisés

Puisque *Sortir du nucléaire* est une association qui n'est pas directement liée à la géomatique, *QGIS*⁶ a été choisi afin de mener à bien cette étude car c'est un outil GIS (Geographic Information System) robuste et open source. Dans ce cas il s'agit de *QGIS 2.18 'Las Palmas'* avec une console utilisant *Python 2.7.10*⁷. Je trouvais personnellement intéressant de compléter mes connaissances en utilisant un outil différent de celui utilisé pendant le cursus du Certificat Complémentaire en Géomatique. *R*⁸ quant à lui a permis d'organiser les résultats des statistiques une fois tous les traitements effectués.

Le choix de la version 2.18 de *QGIS* est une conséquence au fait que *QGIS 3* a été publié début 2018. J'ai préféré utiliser la version précédente qui a été jugée stable et aboutie (publiée en 2016).

3.2 Stratégie

Comme mentionné précédemment, les fichiers produits par le programme de dispersion *HYSPLIT* ont une extension *.kml* basée sur le langage *XML* qui permet l'affichage de données spatiales dans de nombreuses applications, ainsi que leur partage⁹. Cependant, ces fichiers ne permettent pas une utilisation optimale des données qu'ils contiennent. Une première étape du projet est alors de convertir ces fichiers en données vectorielles utilisables par *QGIS* : *.shp* (*ESRI shapefile*).

Les données de nuages et déposition (AER et REF) devant être étudiées séparément mais aussi cumulées (AER + REF et AERDEPO + REFDEPO) il faut convertir les *.shp* en fichiers raster (ici : *TIFF (.tif)*). De plus, les fichiers raster permettent une visualisation plus esthétique des données de dispersion.

Pour produire les statistiques de population et de types de sols touchés il faut appliquer un outil de statistiques de zones afin de croiser les différentes isolignes des fichiers de dispersion avec des couches de densité de population et de couverture du sol. Pour permettre une étude à l'échelle européenne mais aussi par pays (Suisse et pays limitrophes), une couche de frontières doit aussi être utilisée dans cette partie. Les statistiques de zones sont effectuées en croisant des couches et en utilisant une extension de *QGIS* (*SAGA*) contient un outil de statistique de zone.

Pour finir, quand les résultats sont récoltés, ceux-ci sont sous forme de fichiers journaliers séparés. Une agrégation de ces résultats journaliers par mois/année et par pays doit être effectuée afin de permettre une lecture efficace.

⁶ <https://www.qgis.org/fr/site/>

⁷ <https://www.python.org/>

⁸ <https://www.r-project.org/>

⁹ <https://developers.google.com/kml/>

En résumé :

1. Conversion KML - SHP
2. Rastérisation
3. Cumule des données (nuage et déposition)
4. Statistique de zone
5. Organisation des résultats

3.3 Les données statistiques

Le choix des couches de densité de population et de couverture du sol est très important car les résultats en sont dépendants. L'objectif est de trouver des cartes en libre accès, actuelles et dont la résolution est suffisante pour permettre une étude précise.

3.3.1 Population

Les données de densité de population proviennent du dataset *Global Human Settlement* de l'*EUROPEAN COMMISSION*. La couche choisie est la plus récente (2015) et à une résolution de 250m :

*GHS_POP_GPW42015_GLOBE_R2015A_54009_250*¹⁰

La projection utilisée est la projection équivalente (qui conserve les surfaces) *Mollweide* (*EPSG:54009*).

Ces données ont été choisies car elle sont récentes (2015) et leur résolution (250m) permet une étude très précise.

3.3.2 Couverture du sol

Le programme *Copernicus*¹¹ offre des données obtenues par imagerie satellite en libre accès. La couche de couverture du sol utilisée qui provient de ce programme est la *Corine Land Cover (CLC) 2012, Version 18.5.1 (g250_clc12_V18_5)*. Avec une résolution de 250m similaire à celle obtenue pour la densité de population et représentant la couverture du sol de l'année 2012 cette couche est le candidat parfait pour l'étude.

Le système de référence de la couche est l'*European Terrestrial Reference System 1989 (ETRS 1989)*, correspondant au code *EPSG:3035*.

Afin d'avoir un jeu de données cohérent, cette couche a été re-projetée à la projection équivalente *EPSG:54009* qui deviendra la projection standard du projet.

¹⁰ European Commission, Joint Research Centre (JRC); Columbia University, Center for International Earth Science Information Network - CIESIN (2015): GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a

¹¹ Copernicus programme - <https://www.copernicus.eu/en>

Le nombre des classes de la *CLC* est de 44 au total, ce qui est bien trop détaillé pour l'étude. Nous avons décidé de diminuer ce nombre de classes en regroupant certaines pour avoir au final 10 classes :

1. Impermeable urban areas (zones urbaines denses à >80%)
2. Urban areas (zones urbaines à <80%)
3. Non vegetal exploitations (zones de constructions, décharges, exploitations minérales)
4. Recreational areas (parcs, infrastructures sportives)
5. Agricultural areas (zones agraires)
6. Grasslands (pâturages, prairies)
7. Forests (forêts)
8. Other natural areas (autres zones naturelles)
9. Unproductive areas (zones rocheuses, déserts)
10. Water bodies (lacs, cours d'eau)

3.4 Les scripts

Comme mentionné plus tôt les fichiers d'origine sont des fichiers *.kml* lesquels sont ensuite convertis en *.shp* puis en *.tif*. Pour éviter toute confusion ou modifications des fichiers originaux, des dossiers différents sont créés au préalable afin d'organiser les fichiers. L'automatisation de la production de ces dossiers varie selon le système d'exploitation de la machine hôte (la Figure 3-1 contient le code que j'ai utilisé sur *Mac*). La Figure 3-2 quant à elle présente l'architecture de l'environnement de travail.

```
1 startdate=2016-06-01
2 enddate=2018-12-31
3
4 sDateTs=`date -j -f "%Y-%m-%d" $startdate "+%s"`
5 eDateTs=`date -j -f "%Y-%m-%d" $enddate "+%s"`
6 dateTs=$sDateTs
7 offset=86400
8
9 while [ "$dateTs" -le "$eDateTs" ]
10 do
11     date=`date -j -f "%s" $dateTs "+%Y%m%d"`
12     printf '%s\n' $date
13     mkdir "$date"
14     dateTs=$(( $dateTs+$offset ))
15 done
```

Figure 3-1 : Script d'automatisation de la production de dossier de dates (*Mac*)

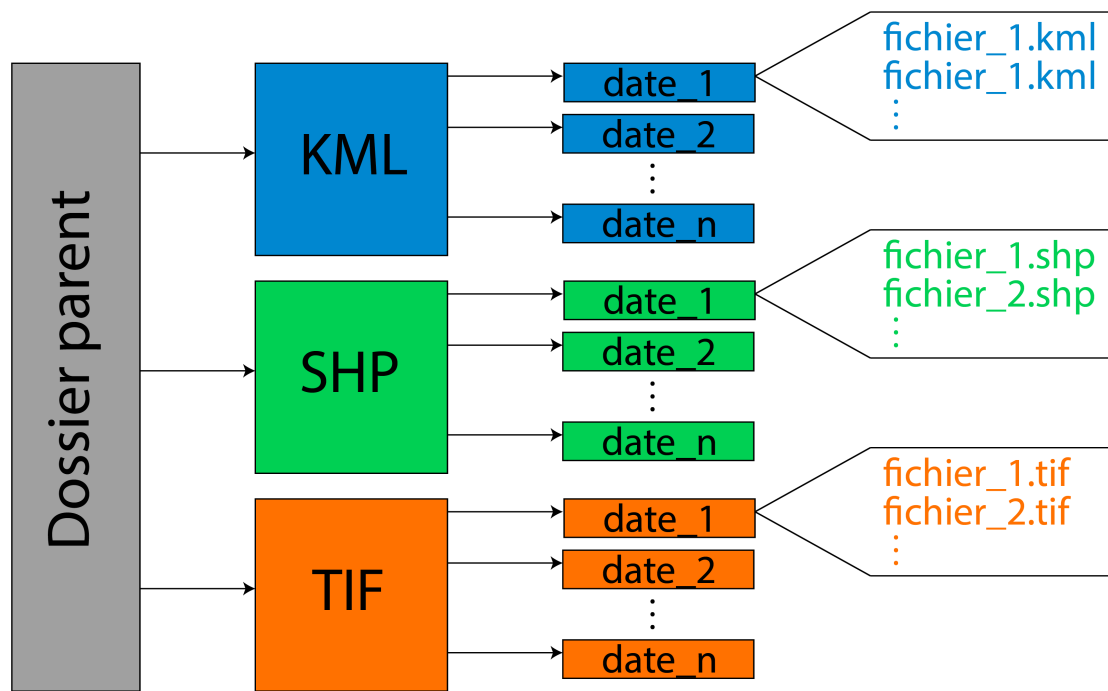


Figure 3-2 : Architecture des dossiers

Les dossiers de dates sont au format *aaaammjj*, par exemple le dossier du 21 mars 2017 est nommé 20170321.

Les scripts *Python* utilisés dans *QGIS* commencent tous par le même code. Les deux premières lignes (Figure 3-3) commandent l'importation du temps système de début lorsque l'utilisateur lance le script, ce qui va permettre, avec les dernières lignes de chaque script (Figure 3-4) d'estimer le temps de calcul de chaque étape. Les lignes 3 et 4 permettent d'importer les bibliothèques d'outils nécessaires aux traitements. Les lignes suivantes attribuent les chemins d'accès pour l'importation ainsi que pour l'exportation des fichiers, cependant ceux-ci ne sont pas complets. Les dossiers de dates sont ajoutés aux chemins d'accès grâce à la variable « *newdate* » qui est une adaptation au format texte de la variable de date utilisée pour les itérations. Les lignes 11 à 14 sont ajoutées afin de permettre l'itération par date à travers les différents fichiers. Ces lignes contiennent les informations quant aux dates de début et de fin entre lesquelles les traitements vont être effectués.

L'itération des dates se fait en utilisant la boucle *Python* « *while* » et une condition, ici « *mydate<end* ». La boucle prendra donc fin quand l'objet « *mydate* » qui commence à la valeur « *start* » (équivalent à la date de début) sera égal à la date de fin « *end* ». Pour que la boucle fonctionne et ne rentre pas dans une itération infinie, il faut faire varier « *mydate* » en lui réattribuant une valeur de la façon suivante, à la fin des traitements souhaités pour une date :

mydate=mydate+day » où « *day* » a la valeur un jour (voir ligne 14, Figure 3-3). Ces notions de base ont été utilisées dans tous les scripts *Python* sur *QGIS* afin d'importer les outils, définir les chemins d'accès aux données, déterminer le temps d'exécution du programme et d'itérer sur les dates.

```

1  import time
2  tmps1=time.time()
3  import os, processing, gdal
4  from qgis.PyQt.QtCore import QVariant
5
6
7  path1="/Users/Bastien/Documents/STAGE/GIS/ALL/ALL_KML/"
8  path="/Users/Bastien/Documents/STAGE/GIS/ALL/ALL_SHP/"
9
10 # setting start and end dates to iterate through the data by date
11 from datetime import date, timedelta
12 start=date(2017,02,28)
13 end=date(2017,03,1)
14 day=timedelta(days=1)
15
16 # starting date
17 mydate=start
18
19 # loop through dates to import kml and produce shp
20 while mydate<end:
21     #project CRS
22     crs=QgsCoordinateReferenceSystem("EPSG:4326")
23     newdate=str(mydate).replace("-", "")
24     end_date=mydate+3*day
25     newenddate=str(end_date).replace("-", "")

```

Figure 3-3 : Premières lignes de code des scripts Python

```

318
319
320     print newdate
321
322     QgsMapLayerRegistry.instance().removeAllMapLayers
323
324     mydate=mydate+day
325
326 tmps2=time.time()-tmps1
327 print "Exec time = %f" %tmps2

```

Figure 3-4 : Dernières lignes des codes Python

3.4.1 Conversion des KML en SHP

La conversion des fichiers *KML* en *SHP* est nécessaire pour la production de données de qualité ainsi que pour la manipulation des données dans un GIS. L'observation des fichiers *KML* permet de se familiariser avec les données qu'ils contiennent, nécessaires ou non.

Tout d'abord, à l'importation des *KML*, *QGIS* indique que plusieurs données sont disponibles et il faut choisir laquelle ou lesquelles doivent être ouvertes. Celles-ci sont définies par un identifiant et dans notre cas, les données importantes (données contenant les couches vectorielles) sont regroupées sous l'identifiant « 4 » (Figure 3-5). Cet identifiant doit donc figurer dans le code d'importation :

« QgsVectorLayer(path+newdate+"/hys_BEZ_2AER_concplot.kml|layerid=4","", "ogr") »

Identifiant de la couche	Nom de la couche	▲ Nombre d'entités	Type de géométrie
4	<pre>Concentration (Valid:20170120 0000UTC)</pre>	12	Polygon
5	<pre>Deposition (Valid:20170120 0000UTC)</pre>	0	Unknown
3	EPA AIRNow Air Quality Index (AQI)	0	Unknown
2	NOAA NESDIS kml Smoke/Fire Data	0	Unknown
1	NOAA NWS kml Weather Data	0	Unknown
0	Source Locations	1	Point

Figure 3-5 : Boîte de dialogue à l'importation d'un KML

Où le premier argument appelle l'outil d'importation de vecteur dans QGIS. Le chemin d'accès au dossier complet contient « path » et la date « newdate ». Le nom du fichier est ensuite donné entre guillemets avec l'argument d'identifiant de couche « layerid=4 ». Les guillemets suivants sont utilisés pour donner un nom à la couche dans QGIS. Attribuer un nom n'est pas nécessaire dans notre cas. Enfin, « ogr » définit la librairie à utiliser pour importer les différents formats.

Les KML contiennent le nom des différentes isolignes en format texte dans la colonne « Name », les dates de début et de fin de l'intégration (évolution sur 72h) dans les colonnes « begin » et « end ». Les autres colonnes sont vides ou présentent des données inutilisables. Aussi, le dernier attribut de la première colonne, « Maximum Value Grid Cell », ne correspond pas à une isoligne (voir Figure 3-6).

Une première étape est de « nettoyer » les fichiers en enlevant les données qui ne sont pas nécessaires. Cependant, les KML ne sont pas modifiables dans QGIS. Il faut donc les convertir en SHP en utilisant l'outil « QgsVectorFileWriter »

Les fichiers sont parfois corrompus, ils peuvent contenir des erreurs de topologie. Deux opérations de base vont permettre de corriger les erreurs :

- ajouter un identifiant unique aux éléments (avec « qgis:fieldcalculator »)
- appliquer un buffer de distance fixe, de distance 0 afin de ne pas modifier l'emprise des isolignes (avec « qgis:fixeddistancebuffer »).

	Name	description	timestamp	begin	end	altitudeMode	tessellate	extrude	visibility	drawOrder	icon	snippet
1	Contour Level: 1.2E+08 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
2	Contour Level: 2.3E+08 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
3	Contour Level: 3.5E+08 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
4	Contour Level: 1.7E+09 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
5	Contour Level: 2.1E+09 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
6	Contour Level: 6.6E+09 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
7	Contour Level: 6.9E+09 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
8	Contour Level: 1.7E+10 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
9	Contour Level: 1.7E+10 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
10	Contour Level: 3.4E+10 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
11	Contour Level: 3.5E+10 Bq-s/m3			2017-01-17...	2017-01-20...		-1	1	-1			
12	Maximum Value Grid Cell	The s... of max... size of		2017-01-17...	2017-01-20...		-1	1	0			

Figure 3-6 : Table d'attribut d'un fichier KML

La ligne d'attribut non nécessaire nommée « Maximum Value Grid Cell » peut être supprimé à l'aide d'une recherche « request » et de l'opération « deleteFeature ». À noter que dans le cas où la couche n'existe pas (nous verrons des cas spécifiques plus tard), cette partie du code provoquerait une erreur. Pour éviter cela ces quelques lignes ont été placées sous la condition « if » que la couche ait bien été importée. La Figure 3-7 présente un exemple des lignes commandant ces opérations.

```

aer=QgsVectorLayer(path+newdate+"/hys_BEZ_2AER_concplot.kml|layerid=4","", "ogr")
writer=QgsVectorFileWriter.writeAsVectorFormat(aer,path+newdate+"/BEZ_aer1_shp.shp","utf-8",None,"ESRI shapefile")
processing.runalg("qgis:fieldcalculator",path+newdate+"/BEZ_aer1_shp.shp","id",1,13,1,True,"$Id",path+newdate+"/BEZ_aer2_shp.shp")
processing.runalg("qgis:fixeddistancebuffer",path+newdate+"/BEZ_aer2_shp.shp",0.000,5,0,path+newdate+"/BEZ_aer_shp.shp")

aer=QgsVectorLayer(path+newdate+"/BEZ_aer_shp.shp","", "ogr")

if aer:
    aer.startEditing()
    request=QgsFeatureRequest().setFilterExpression( u'"Name" = \'Maximum Value Grid Cell\' )
    for f in aer.getFeatures(request):
        aer.deleteFeature(f.id())
    aer.commitChanges()

```

Figure 3-7 : Exemple des lignes permettant l'importation des KML, la conversion en SHP, l'ajout d'un identifiant unique, la correction par l'outil buffer, ainsi que la suppression de la ligne d'attribut inutile

Certaines données de dispersion ont été produites en plusieurs fichiers séparés car certaines des erreurs apparaissaient lors des calculs sur *HYSPLIT*. Par exemple, les fichiers GAS sont séparés en *1GAS_a*, *1GAS_b* et *1GAS_c*. Les traitements décrits plus haut doivent être appliqués à chaque fichier avant de les regrouper en un avec l'outil « qgis:mergevectorlayers ». Parfois, des isolignes n'apparaissent pas et certains fichiers sont donc vides ce qui produit des erreurs. Pour les éviter, il a fallu ajouter des lignes de code conditionnelles pour chaque cas possible (voir Table 3).

Table 3 : Table des différentes possibilités de validité des couches. La couleur verte indique que la couche dont le nom apparaît en début de ligne est valide, la couleur rouge indique que celle-ci n'est pas valide.

	Toutes valides	2 valides	2 valides	2 valides	1 valide	1 valide	1 valide	0 valide
GAS_a								
GAS_b								
GAS_c								

Lorsqu'une couche n'est pas valide, son type de géométrie est égal à 0. L'argument « wkbType() » permet d'obtenir le type de géométrie, lorsque celui-ci n'est pas égal à 0 (« !=0 »), la couche est valide. La Figure 3-8 montre un exemple des lignes permettant l'agrégation des couches (*1GAS_a*, *1GAS_b*, *1GAS_c*) sans erreur. À noter que dans le cas où aucune couche n'est valide aucun traitement n'est prévu puisque le résultat sera nul dans tous les cas.

```

if GAS_A.wkbType()!=0 and GAS_B.wkbType()!=0 and GAS_C.wkbType()!=0:
    processing.runalg("qgis:mergevectorlayers",[GAS_A,GAS_B,GAS_C],path+newdate+"/BEZ_GAS_shp.shp")
if GAS_C.wkbType()==0 and GAS_A.wkbType()!=0 and GAS_B.wkbType()!=0:
    processing.runalg("qgis:mergevectorlayers",[GAS_A,GAS_B],path+newdate+"/BEZ_GAS_shp.shp")
if GAS_B.wkbType()==0 and GAS_A.wkbType()!=0 and GAS_C.wkbType()!=0:
    processing.runalg("qgis:mergevectorlayers",[GAS_A,GAS_C],path+newdate+"/BEZ_GAS_shp.shp")
if GAS_A.wkbType()==0 and GAS_B.wkbType()!=0 and GAS_C.wkbType()!=0:
    processing.runalg("qgis:mergevectorlayers",[GAS_B,GAS_C],path+newdate+"/BEZ_GAS_shp.shp")
if GAS_C.wkbType()==0 and GAS_B.wkbType()==0 and GAS_A.wkbType()!=0:
    writer=QgsVectorFileWriter.writeAsVectorFormat(GAS_A,path+newdate+"/BEZ_GAS_shp.shp","utf-8",None,"ESRI shapefile")
if GAS_C.wkbType()==0 and GAS_A.wkbType()==0 and GAS_B.wkbType()!=0:
    writer=QgsVectorFileWriter.writeAsVectorFormat(GAS_B,path+newdate+"/BEZ_GAS_shp.shp","utf-8",None,"ESRI shapefile")
if GAS_A.wkbType()==0 and GAS_B.wkbType()==0 and GAS_C.wkbType()!=0:
    writer=QgsVectorFileWriter.writeAsVectorFormat(GAS_C,path+newdate+"/BEZ_GAS_shp.shp","utf-8",None,"ESRI shapefile")

```

Figure 3-8 : Exemple de lignes conditionnelles permettant la fusion des couches *_a*, *_b* et *_c* sans erreur

3.4.2 Production de fichiers raster

Avant de passer à la conversion en fichiers *TIFF*, quelques opérations sont nécessaires. Un outil de simplification de géométrie « *qgis:simplifygeometries* » permet une dernière vérification/correction des couches. Puisque les couches de statistique ont été projetées suivant la projection équivalente Mollweide, les données de nuage et déposition doivent avoir la même projection. Or, celle-ci sont projetées suivant la projection standard *EPSG : 4326*. L'outil « *qgis:reprojectlayer* » permet d'effectuer la conversion. À noter que la reprojection d'une couche vecteur est plus rapide que la reprojection d'une couche raster, c'est pour cette raison que ce traitement est effectué avant la conversion. Les champs d'attributs (colonnes) non nécessaire sont ensuite supprimés (Figure 3-9) avec le traitement « *deleteAttributes* ». Ces champs étant toujours les mêmes dans tous les fichiers, ils sont supprimés en sélectionnant leur numéro de colonne (à noter que l'indexation dans *Python* commence à 0). De nouveaux champs sont ensuite ajoutés aux fichiers. Ces nouveaux champs comprennent les valeurs (en format numérique afin de produire une base de donnée réutilisable) en $Bq \cdot s/m^3$ et mSv pour les nuages ou Bq/m^2 pour la déposition. Les fichiers raster contenant des nombres entiers étant bien plus légers et le fait que la statistique de zone utilisée n'arrivait pas à utiliser des nombres décimaux, des catégories correspondant aux diverses isolignes ont été produites et ajoutées au *SHP*. Ces catégories doivent permettre de reconnaître les valeurs d'isolignes même pour les données qui seront cumulées (2AER + 3REF et 2AERDEPO + 3REFDEPO). Pour ce faire des catégories différentes ont été données aux différents types de rejet. Par exemple, les données AER (nuage ou déposition) ont reçu les catégories suivantes :

1*100 à $n(\text{isolignes}) \cdot 100$

alors que les données REF (nuage ou déposition) ont reçu les catégories :

1 à $n(\text{isolignes})$.


```

gas=QgsVectorLayer(path+newdate+"/BEZ_gas"+newdate+".shp", "", "ogr")
processing.runalg("qgis:simplifygeometries", gas, 0.001, path+newdate+"/BEZ_gas"+newdate+"-simp.shp")
gas=QgsVectorLayer(path+newdate+"/BEZ_gas"+newdate+"-simp.shp", "", "ogr")
processing.runalg("qgis:reprojectlayer", gas, 'EPSG:54009', path+newdate+"/BEZ_gas"+newdate+"-reproj.shp")

# 1) import .shp layers by dates.
gas=QgsVectorLayer(path+newdate+"/BEZ_gas"+newdate+"-reproj.shp", "", "ogr")
# 2) delete unnecessary attributes (coming with kml files) and keeping only starting date,
# ending date and name
res_gas=gas.dataProvider().deleteAttributes([1,2,5,6,7,8,9,10,11])

```

Figure 3-9 : Simplification des géométries, reprojection et suppression des champs d'attributs inutiles

Ainsi, si une partie d'un nuage cumulé correspond au cumule de la deuxième isoligne AER avec la septième isoligne REF la catégorie cumulée est 207 ce qui permet de facilement reconnaître les isolignes en jeu. Puisque certaines isolignes peuvent ne pas apparaître, l'ajout des valeurs d'isolignes et des catégories correspondantes est fait de façon conditionnelle (Figure 3-10).

```

for feat in features:
    attrs = feat.attributes()
    fl=attrs[0]
    if str(fl)=="Contour Level: 9.0E+09 Bq-s/m3":
        gas.changeAttributeValue(feat.id(), idx_cat, 1)
        gas.changeAttributeValue(feat.id(), idx_bq, 8.973E+09)
        gas.changeAttributeValue(feat.id(), idx_msv, 1.24E-02)
    elif str(fl)=="Contour Level: 2.7E+09 Bq-s/m3":
        gas.changeAttributeValue(feat.id(), idx_cat, 2)
        gas.changeAttributeValue(feat.id(), idx_bq, 2.688E+10)
        gas.changeAttributeValue(feat.id(), idx_msv, 3.72E-02)
    elif str(fl)=="Contour Level: 8.1E+10 Bq-s/m3":
        gas.changeAttributeValue(feat.id(), idx_cat, 3)
        gas.changeAttributeValue(feat.id(), idx_bq, 8.052E+10)
        gas.changeAttributeValue(feat.id(), idx_msv, 1.11E-01)
    elif str(fl)=="Contour Level: 2.4E+11 Bq-s/m3":
        gas.changeAttributeValue(feat.id(), idx_cat, 4)
        gas.changeAttributeValue(feat.id(), idx_bq, 2.412E+11)
        gas.changeAttributeValue(feat.id(), idx_msv, 3.34E-01)
    elif str(fl)=="Contour Level: 4.8E+11 Bq-s/m3":
        gas.changeAttributeValue(feat.id(), idx_cat, 5)
        gas.changeAttributeValue(feat.id(), idx_bq, 4.817E+11)
        gas.changeAttributeValue(feat.id(), idx_msv, 6.67E-01)

```

Figure 3-10 : Ajout conditionnel des valeurs d'isolignes

Une fois que les SHP contiennent toutes les informations utiles, la conversion en TIFF peut être effectuée en utilisant l'outil « gdal_rasterize ». Des limites géographiques sont données afin que tous les fichiers aient la même extension sans quoi, le cumule et les opérations statistiques ne sont pas possibles. La résolution est fixée à 250m pour correspondre à celle des couches de population et de couverture du sol et les images produites sont aussi compressées afin de ne pas utiliser trop d'espace disque.

3.4.3 Cumule des données nuage et déposition

Une fois les fichiers TIFF produits les données AER et REF ainsi que AERDEPO et REFDEPO peuvent être agrégées pour former des nouvelles données cumulées. Les fichiers ayant une forme facilement utilisable, l'automatisation de ce traitement est aisée et est présenté à la Figure 3-11.

```

# Loop through dates to
while mydate<end:
    newdate=str(mydate).replace("-", "")

    #####
    #####                               SUM                               #####
    #####

    # 1) import .tif layers by dates to sum
    aer=QgsRasterLayer(path1+newdate+"/BEZ_aer"+newdate+".tif", "")
    ref=QgsRasterLayer(path1+newdate+"/BEZ_ref"+newdate+".tif", "")
    os.system('gdal_calc.py --calc "A+B" --format GTiff --type Float32 --NoDataValue 0 --co="COMPRESS=DEFLATE" -A '

    aerDEP0=QgsRasterLayer(path1+newdate+"/BEZ_aerDEP0"+newdate+".tif", "")
    refDEP0=QgsRasterLayer(path1+newdate+"/BEZ_refDEP0"+newdate+".tif", "")
    os.system('gdal_calc.py --calc "A+B" --format GTiff --type Float32 --NoDataValue 0 --co="COMPRESS=DEFLATE" -A '

    print newdate

    mydate=mydate+day

tmps2=time.time()-tmps1
print "Exec time = %f" %tmps2

```

Figure 3-11 : Lignes d'automatisation du cumule des couches. À noter que les chemins d'accès d'importation et exportation des fichiers n'apparaissent pas sur la figure ci-dessus mais se trouvent à la fin des lignes commençant par « os.system ».

Présent à la fin de chaque itération, et ce pour chaque script, le code « print newdate » permet d'afficher dans la console la date traitée afin que l'utilisateur ait une idée de l'avancée du calcul.

```

while mydate<end:
    newdate=str(mydate).replace("-", "")

    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_gas"+newdate+".tif", [POPULATION,BORDERS],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_aer"+newdate+".tif", [POPULATION,BORDERS],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_ref"+newdate+".tif", [POPULATION,BORDERS],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_aerECRR"+newdate+".tif", [POPULATION,BORDER],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_aerHPRESS"+newdate+".tif", [POPULATION,BORDER],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_refHPRESS"+newdate+".tif", [POPULATION,BORDER],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_CLOUD"+newdate+".tif", [POPULATION,BORDERS],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_aerDEP0"+newdate+".tif", [LANDCOVER,BORDERS],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_aerDEP0"+newdate+".tif", [POPULATION,BORDER],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_refDEP0"+newdate+".tif", [LANDCOVER,BORDERS],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_refDEP0"+newdate+".tif", [POPULATION,BORDER],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_DEP0"+newdate+".tif", [LANDCOVER,BORDERS],
    processing.runalg('saga:zonalgridstatistics', path1+newdate+"/BEZ_DEP0"+newdate+".tif", [POPULATION,BORDERS],

    print newdate

    #QgsMapLayerRegistry.instance().removeAllMapLayers()

    mydate=mydate+day

tmps2=time.time()-tmps1
print "Exec time = %f" %tmps2

```

Figure 3-12 : Exemple de script permettant l'automatisation des calculs statistiques. À noter que les chemins d'accès pour l'exportation des fichiers traités n'apparaissent pas sur l'image.

3.4.4 Statistiques

Tous les fichiers étant prêts il s'agit maintenant de faire la statistique de population et couverture du sol touchées. L'outil *zonal grid statistics* de la librairie SAGA (« saga:zonalgridstatistics ») a été choisi car il permet une statistique de zone raster - raster. Cet outil prend en entrée les couches de nuage et déposition et les croise avec les couches de densité de population, de couverture du sol ainsi qu'un couche

de frontières. Ces dernières sont importées au début et stockées dans la mémoire tampon du programme afin d'éviter des allers - retours dans les dossiers. La Figure 3-12 présente en exemple un extrait du script d'automatisation des calculs statistiques pour la centrale de Beznau.

3.4.5 Organisation des résultats

Les fichiers en sortie de la statistique sont des fichiers CSV organisés par centrale, et par jour, et même s'ils contiennent les informations, leur lecture et leur étude sont compliquées, voir Figure 3-13. Le but est ici d'organiser les résultats de façon à faciliter leur utilisation. Pour ce faire j'ai choisi d'utiliser le programme de statistique *R* avec l'extension « xlsx » permettant la création de fichiers .xlsx. La stratégie est la suivante :

Tout d'abord, il faut distinguer dans cette étude deux aspects :

1. les impacts sur la population
2. les impacts sur l'environnement

J'ai donc choisi de produire deux fichiers *EXCEL* pour l'Europe, la Suisse, la France, l'Allemagne, l'Italie et l'Autriche, regroupant :

1. les données statistiques concernant la population
2. les données statistiques concernant la couverture du sol.

Le premier fichier *EXCEL* est organisé par page et contient les statistiques concernant la population vis à vis de :

- GAS, AER, REF, CLOUD (AER + REF), AERECRR, AERHPRESS, REFHPRESS, AERDEPO, REFDEPO et DEPO (AERDEPO + REFDEPO)

Le deuxième fichier *EXCEL* est organisé par page et contient les statistiques concernant la couverture du sol vis à vis de :

- AERDEPO, REFDEPO et DEPO (AERDEPO + REFDEPO).

L'architecture générale du script *R* est similaire à celle des scripts *Python* (Figure 3-14). La toute première ligne active l'extension *xlsx* qui va permettre l'exportation des fichiers dans ce format. (À noter qu'il est nécessaire d'avoir installé au préalable cette extension avec un code du type « `install.packages("xlsx")` »). Comme pour les autres scripts il est possible de déterminer le temps de calcul en important le temps système au début du code (voir ligne 3). Les chemins d'accès d'importation et d'exportation sont à définir. L'automatisation a été faite en itérant les dossiers de dates afin de récupérer les résultats de la statistique de zone au format CSV. Pour ce faire, il est nécessaire de définir une date de début ainsi qu'une date de fin (voir lignes 9 et 10). Sans entrer dans trop de détails, il s'est avéré plus aisé de créer une « dataframe » (tableau de données) vide et de le remplir avec les données au fur et à mesure. La création de cette « dataframe » nécessite des dimensions (nombre de lignes et nombre de colonnes). Le nombre de colonne est inconnu à l'avance car il représente le nombre d'isolignes ou de catégories (et celles-ci peuvent varier surtout dans le cas des nuages et dépôts cumulés) par contre le nombre de lignes est connu puisqu'il représente le nombre de jours calculés. Pour faciliter l'utilisation du programme sans avoir besoin de se soucier de ce genre de détails, « *tdelta* » est donc défini comme l'intervalle de temps.

	BEZaer20170101.tif	GHSEUPOPT.tif	gadm36EU.tif	Count.UC
column 0: rownames	0	0	-2147483648	2870000
2	0	0	0	52703399
3	0	0	8	582624
4	0	0	12	1302827
5	0	0	20	6571
6	0	0	40	1148247
7	0	0	56	282163
8	0	0	70	783756
9	0	0	100	882522
10	0	0	112	2080366
11	0	0	191	835957
12	0	0	203	700485
13	0	0	208	574637
14	0	0	233	566454
15	0	0	250	7456167
16	0	0	276	4152476
17	0	0	292	24

Showing 1 to 18 of 39,997 entries

Figure 3-13 : Exemple d'un fichier CSV de résultat pour une centrale, pour un jour et pour un type de rejet

« mydate » est l'équivalent de « newdate » dans les scripts *Python* et permet l'itération au travers des dates en gardant la valeur de date de début stockée dans une autre variable. Les lignes suivantes stockent les valeurs des isolignes en mSv ou Bq/m^2 puisque les valeurs sortant de la statistique de zone sont des catégories.

De façon similaire à ce que l'on a pu voir dans *Python*, l'itération est effectuée à l'aide de la fonction « while » à laquelle est ajoutée une condition (ici : « while(mydate<=enddate) »). Chaque fichier est « nettoyé ». Effectivement, la statistique de zone prend en compte toute l'étendue géographique des *TIFF*. Il faut donc enlever les données où il n'y a pas de valeurs de radioactivité (où il n'y a pas d'isoligne). Pour ce faire il faut tout d'abord trouver les duplicatas de catégories car on ne peut pas savoir si toutes les isolignes sont présentes, le cas échéant, lesquelles sont effectivement présentes. Ceci est tout particulièrement important pour les données cumulées (AER + REF et AERDEPO + REFDEPO). Une table de données est alors créée en agrégeant sans distinction les données de tous les jours calculés et une fonction permettant de trouver les duplicatas (« duplicated() ») est appliquée à la colonne correspondant aux catégories d'isolignes. Une fois ces valeurs redondantes trouvées le nombre de colonnes à créer dans la « dataframe » peut être déterminé et les isolignes présentes sont connues. Il est alors possible de cumuler les données pour chaque catégorie (voir Chapitre 3.4.2). La statistique de population étant une couche de densité de population il faut multiplier le nombre de

personnes par pixel (colonne 2 de la Figure 3-13) avec le décompte de ces mêmes pixels.

```

1 library(xlsx)
2
3 start_time <- Sys.time()
4
5
6 path<-"/Users/Bastien/Documents/STAGE/GIS/ALL/ALL_CSV/"
7 path1<-"/Users/Bastien/Documents/STAGE/GIS/ALL/ALL_R/OUTPUT_BEZ/"
8
9 startdate<-as.Date("2017-01-01")
10 enddate<-as.Date("2017-01-01")
11 tdelta<-as.numeric(enddate-startdate)
12 mydate<-startdate
13
14 aermSv<-c("mSv", 3.34E-01, 6.67E-01, 1.00E+00, 5.00E+00, 6.00E+00, 1.90E+01, 2.00E+01,
15             4.90E+01, 5.00E+01, 9.90E+01, 1.00E+02)
16 aerECRRmSv<-c("mSv", 1.00E+00, 6.00E+00, 1.90E+01, 2.00E+01, 4.90E+01, 5.00E+01, 9.80E+01,
17                1.00E+02)
18 refmSv<-c("mSv", 7.305E-03, 3.340E-01, 6.670E-01, 7.016E-01, 7.381E-01, 7.764E-01, 8.167E-01,
19             8.591E-01, 9.038E-01, 9.507E-01, 1.000E+00)
20 gasmSv<-c("mSv", 1.24E-02, 3.72E-02, 1.11E-01, 3.34E-01, 6.67E-01, 1.00E+00, 1.19E+00,
21            1.41E+00, 1.68E+00, 2.00E+00)
22 aerDEPObqs<-c("bqs", 3.525E+05, 4.749E+05, 7.560E+05, 1.203E+06, 1.763E+06, 2.375E+06, 3.544E+06,
23                    5.288E+06, 7.124E+06, 1.410E+07, 1.900E+07, 5.000E+07, 1.000E+08)
24 refDEPObqs<-c("bqs", 1.000E+03, 1.000E+04, 3.036E+04, 9.216E+04, 1.606E+05, 2.798E+05,
25                4.875E+05, 8.494E+05, 1.480E+06, 3.000E+06, 6.000E+06)
26 aerHPRESSmSv<-c("mSv", 1.250E-01, 8.350E-01, 1.668E+00, 2.500E+00, 5.000E+00, 1.250E+01, 4.000E+01,
27                  1.000E+02, 1.625E+02, 2.500E+02, 3.000E+02)
28 refHPRESSmSv<-c("mSv", 1.000E-01, 3.340E-01, 6.670E-01, 1.000E+00, 1.170E+00, 1.369E+00, 1.601E+00,
29                  1.873E+00, 2.192E+00, 2.564E+00, 3.000E+00)
30
31

```

Figure 3-14 : Premières lignes du script R pour Beznau

```

92 seq1<-seq(1,11, by=1)
93 ca<-length(seq1)
94 cats<-seq(1,11, by=1)
95
96 gas_df<-cbind(cats, gasmSv)
97
98 cl<-ncol(EU_POP_gas)
99 a=1
100 lendf<-nrow(gas_df)
101 b=1
102 while(a<=cl){
103   while(b<=lendf){
104     if(EU_POP_gas[1,a]==gas_df[b,1]){
105       EU_POP_gas[1,a]<-gas_df[b,2]
106     }
107     b=b+1
108   }
109   a=a+1
110 }

```

Figure 3-15 : Exemple d'attribution conditionnelle pour les valeurs d'isolignes

Il est alors possible d'attribuer de façon conditionnelle (Figure 3-15) aux catégories la valeur correspondante en mSv pour les nuages et en Bq/m^2 pour la déposition avec l'opérateur « if ». Cette opération est relativement complexe pour les données cumulées car le nombre d'isolignes potentielles est très élevé ($n_{AER(DEPO)} \times n_{REF(DEPO)}$). Aussi, pour les statistiques de couverture du sol, le nombre de colonnes correspond à $n_{(AER/REF)DEPO} \times n_{catégories}$. Il a alors été décidé que les valeurs d'isolignes cumulées (nuage et déposition) prendraient la forme d'intervalles afin de réduire le nombre de colonnes et faciliter la lecture des résultats. Les statistiques de couvertures du sol présentent le nombre de pixels de chaque catégorie touché par chaque isoligne (Figure 3-16). Il a été choisi de présenter ces résultats en km^2 ainsi, les surfaces sont calculées en multipliant le nombre de pixel de chaque catégorie par $0.0625 km^2$ (chaque pixel représente une surface de $250m$ par $250m$).

	BEZaerDEPO20170101.tif	g250clc12EU54009.tif	gadm36EU.tif	Count.UC
1	0	1	0	914
2	0	1	8	127
3	0	1	40	1594
4	0	1	56	867
5	0	1	70	32
6	0	1	100	107
7	0	1	191	81
8	0	1	203	89
9	0	1	208	955
10	0	1	233	26

Figure 3-16 : Exemple de résultats de statistiques de couverture du sol non organisés

Une fois que ces traitements ont été faits, le jour calculé est entré comme nom de ligne dans le fichier *EXCEL* et les isolignes représentent les noms de colonne. Les données statistiques du jour sont donc organisées en lignes. Le prochain jour peut alors être calculé et ajouté au fichier à la ligne suivante.

La séparation des résultats en fonction des différents pays est obtenue en sélectionnant au début du code *R* les données dont le code *UN* (troisième colonne dans la Figure 3-13) équivaut à une certaine valeur (756 pour la Suisse, 250 pour la France, 276 pour l'Allemagne, 40 pour l'Autriche et 380 pour l'Italie).

4. Résultats

Les résultats de ces différents traitements sont d'une part la production d'un grand nombre de couches de nuages et déposition radioactifs aux formats *SHP*, *TIFF*. Ce dernier permet une visualisation esthétique des données (Figure 4-1) qui est un atout pour l'objectif de sensibilisation de la population à la question du nucléaire en Suisse.

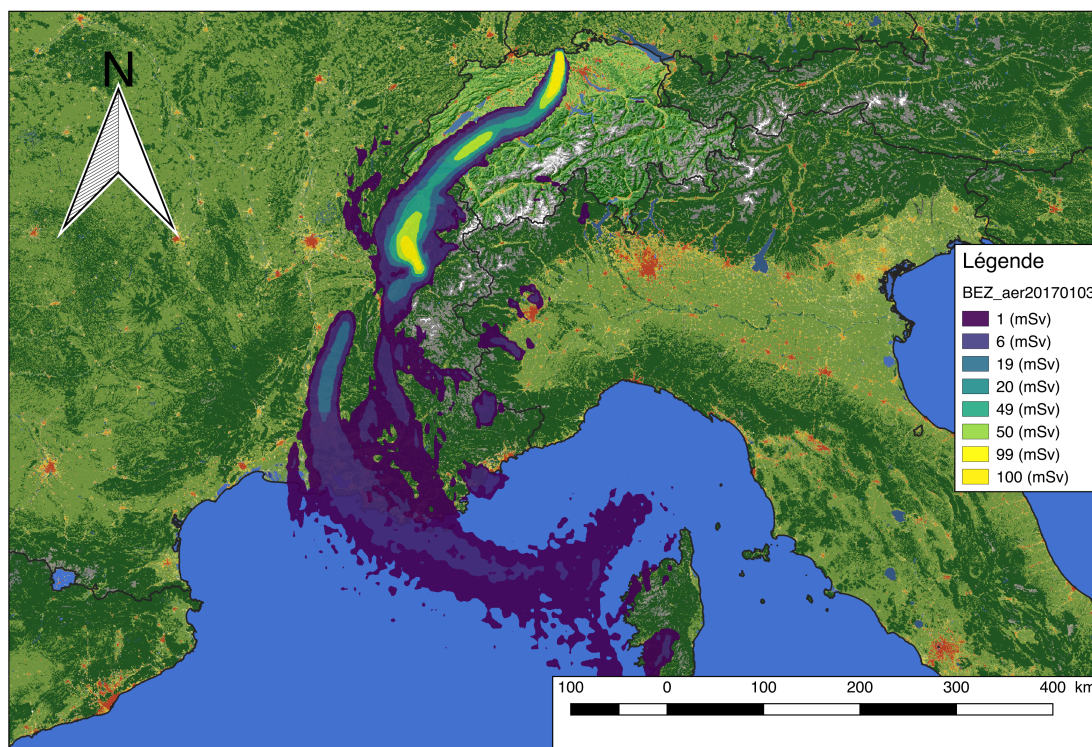


Figure 4-1 : Exemple de visualisation des données (ici, la dispersion d'un nuage d'aérosols sur 72h lors d'un hypothétique accident de la centrale nucléaire de Beznau le 03.01.2017)

Ce travail a aussi permis d'automatiser les processus permettant d'estimer la population et les types d'environnements touchés par un tel accident. L'organisation des résultats favorise la lecture et l'interprétation des résultats de l'étude. Les Figure 4-2a-b présentent des exemples de ces résultats statistiques organisés.

Ces résultats organisés vont permettre une bonne analyse. Il est possible d'en tirer des graphiques et autres statistiques de base. Des exemples de graphiques tirés de ces résultats sont présentés dans les Figure 4-3 et Figure 4-4. Ces simples graphiques de nuage et déposition d'aérosols permettent déjà de se rendre compte de l'impact qu'un accident pourrait avoir sur l'Europe. Il était attendu que les isolignes basses (par exemple 1mSv), qui indiquent que les rejets ont été dispersés de façon extensive, impactent un plus grand nombre de personnes que les isolignes hautes (par exemple 100mSv). Cela dit, le nombre de personnes affectées à plus de 100 mSv est loin d'être anodin (voir Figure 4-3). Il semblerait aussi que de début janvier à fin juin il y aurait eu une augmentation des situations météorologiques défavorables et le nombre relatif de personnes potentiellement touchées par mois aurait augmenté. Il est intéressant d'observer le fait que les personnes affectées à 1 mSv est très supérieur à celui des personnes affectées à 20, 50 et 100 mSv. Par contre, ces derniers taux ne présentent pas d'aussi grandes différences entre eux.

	A	B	C	D	E	F	G	H	I	J	K	L
1	mSv	0.334	0.667	1	5	6	19	20	49	50	99	100
2	BEZ_aer20180101	7126301	4432485	8431281	51786	138198	1200	21185	87	15088	14	
3	BEZ_aer20180102	6903782	2440074	2061837	62967	119774	2298	15496	479	2119	923	
4	BEZ_aer20180103	1591343	715714	1221564	19087	128969	769	27036	868	22314	84	a
5	BEZ_aer20180104	7324680	3328074	6031668	62428	143793	1273	21314	195	3098	923	
6	BEZ_aer20180105	7984043	2303140	3142122	90246	182494	1026	20483	304	14921	0	47
7	BEZ_aer20180106	2445801	347767	844800	57823	353957	11393	405182	7851	220671	2189	202149
8	BEZ_aer20180107	3933803	1245797	4911080	263122	1973070	24353	372616	10283	197045	2940	203614
9	BEZ_aer20180108	3041937	512956	2516173	256993	1393890	35482	290657	5800	140523	2377	179652
10	BEZ_aer20180109	13560662	6025601	12151524	374592	1112755	22159	700410	26847	291524	1505	114520
11	BEZ_aer20180110	6123693	1832364	4234545	527605	4108505	7948	93540	1414	40775	2387	129706
12	BEZ_aer20180111	1604222	1180678	2083564	69219	323605	13450	219931	5283	190600	908	215306
13	BEZ_aer20180112	8421125	2957348	4901052	112258	537895	14607	217212	3676	157192	1107	106036
14	BEZ_aer20180113	15619098	9622718	3583765	44781	426518	21262	285555	1875	150526	7481	141537
15	BEZ_aer20180114	6836480	1494794	1990741	98400	548566	17731	292292	5719	278850	4774	511183
16	BEZ_aer20180115	3687483	1565789	3350688	198764	583614	5353	133369	2600	40842	242	84625
17	BEZ_aer20180116	1217663	501113	748920	14198	61102	1564	30796	571	366	9233	0
18	BEZ_aer20180117	640932	288265	665425	46619	75530	828	16659	177	999	9233	0
19	BEZ_aer20180118	5198722	2350577	1646467	37627	113043	1735	18860	174	1516	9233	0
20	BEZ_aer20180119	519517	235806	688973	40089	109704	521	16978	226	724	9233	0

	Impermeable urban areas (>80 Bq/m2)	Impermeable	Impermeable	Impermeable	Impermeable	Urban areas (30-80%)	Urban area	Urban area	Urban area	Urban area	Non-vegetal exp
2	Bq/m2	<3.7E+04	>=3.7E+04	<=5.55E+05	>=1.48E+06	Bq/m2	<3.7E+04	>=3.7E+04	<=5.55E+05	>=1.48E+06	Bq/m2
3	BEZ_DEPO_LC20180101	379.5	31.125	19.4375	0.6875	BEZ_DEPO_LC20180101	6775.688	3389.938	1486.938	136.5625	BEZ_DEPO_LC20180101
4	BEZ_DEPO_LC20180102	613.25	27.8125	4.1875	0.25	BEZ_DEPO_LC20180102	9509.875	1277.063	278.4375	103.5	BEZ_DEPO_LC20180102
5	BEZ_DEPO_LC20180103	55.9375	2.625	0.625	0.3125	BEZ_DEPO_LC20180103	9500	543.8125	255.3125	100	BEZ_DEPO_LC20180103
6	BEZ_DEPO_LC20180104	68.6875	56.0625	20.875	0.6875	BEZ_DEPO_LC20180104	14613.56	2296.438	913.375	126.5625	BEZ_DEPO_LC20180104
7	BEZ_DEPO_LC20180105	83.25	14.5	8.75	0.625	BEZ_DEPO_LC20180105	10624.13	1627.75	714.5625	167.75	BEZ_DEPO_LC20180105
8	BEZ_DEPO_LC20180106	57.375	0.625	3.5	8.4375	BEZ_DEPO_LC20180106	4806.25	359.875	259.25	587.5	BEZ_DEPO_LC20180106
9	BEZ_DEPO_LC20180107	30.875	15.1875	26.625	21.375	BEZ_DEPO_LC20180107	3652.75	1348.688	1390.063	1252.188	BEZ_DEPO_LC20180107
10	BEZ_DEPO_LC20180108	100.8125	10.5625	7.25	17.25	BEZ_DEPO_LC20180108	6994.375	470.4375	745.3125	1199.438	BEZ_DEPO_LC20180108
11	BEZ_DEPO_LC20180109	49.4375	23.1875	15.75	11.1875	BEZ_DEPO_LC20180109	6259.938	4386.25	2271.875	1254.938	BEZ_DEPO_LC20180109
12	BEZ_DEPO_LC20180110	41.5625	5.25	8.875	17.75	BEZ_DEPO_LC20180110	3617.563	1106.375	643.5625	1725.188	BEZ_DEPO_LC20180110
13	BEZ_DEPO_LC20180111	14.125	18.625	4.8125	2.375	BEZ_DEPO_LC20180111	1141.188	680.125	302.5	457.75	BEZ_DEPO_LC20180111
14	BEZ_DEPO_LC20180112	117.6875	23.625	15.1875	5.75	BEZ_DEPO_LC20180112	7481.75	1890.563	862	692.375	BEZ_DEPO_LC20180112
15	BEZ_DEPO_LC20180113	179.5625	93.9375	6.125	8.8125	BEZ_DEPO_LC20180113	9773.063	2768.125	490.75	548.625	BEZ_DEPO_LC20180113
16	BEZ_DEPO_LC20180114	33.9375	6.1875	1.8125	13.3125	BEZ_DEPO_LC20180114	7749.5	1099.625	424.0625	780.125	BEZ_DEPO_LC20180114
17	BEZ_DEPO_LC20180115	22.1875	13.875	2.6875	4.9375	BEZ_DEPO_LC20180115	5768.813	880.3125	458.75	665	BEZ_DEPO_LC20180115
18	BEZ_DEPO_LC20180116	99.375	1.5625	0.4375	0	BEZ_DEPO_LC20180116	11189.81	365.625	150.375	58.4375	BEZ_DEPO_LC20180116
19	BEZ_DEPO_LC20180117	53.5	0.9375	0.875	0.625	BEZ_DEPO_LC20180117	1388.625	152.6875	94.5	80.625	BEZ_DEPO_LC20180117
20	BEZ_DEPO_LC20180118	110.4375	22.3125	0.6875	0.3125	BEZ_DEPO_LC20180118	14802.44	862.4375	171.0625	95.0625	BEZ_DEPO_LC20180118

Figure 4-2 : a) Exemple de résultats organisés de population touchée, pour les nuages d'aérosol de la centrale Beznau en 2018; b) Exemple de résultats organisés de types de sols touchés pour la déposition cumulée AERDEPO + REFDEPO de la centrale Beznau en 2018.

Du point de vue des types de sols touchés, la Figure 4-4 présente 4 catégories :

- Zones urbaines
- Zones agraires
- Pâturages
- Forêts

Ces catégories font parties des 10 catégories agrégées depuis les 44 catégories du CLC (Corine Land Cover). Elles ont été choisies pour illustrer le type de résultats obtenus et l'utilisation qui peut en être faite une fois que ceux-ci sont organisés. Il a été choisi ici de regrouper les statistiques de chaque jour pendant la période allant du 1^{er} janvier 2018 au 30 juin 2018 pour une déposition équivalent à 354'000 Bq/m² et d'en faire des box plots. Ce type de graphique donne des informations sur la dispersion des données et permet de comparer des jeux de données de tailles différentes. Ici, on se rend compte qu'il y a une grande variation au sein des données et que les impacts sur l'environnement vont dépendre des situations météorologiques. Cependant, ces graphiques permettent de déterminer rapidement quel type de couverture du sol sera le plus affecté en cas d'accident majeur à Beznau. Pour le premier semestre de 2018 il semblerait que les surfaces les plus

affectées soient les zones agricoles. Ce résultat provient probablement des proportions relatives des surfaces aux alentours de la centrale.

Ces deux graphiques ne sont pas représentatifs de l'étude en général mais permettent de se rendre compte des possibilités offertes par les divers traitements et surtout l'importance d'organiser les résultats de façon cohérente. Comme mentionné plus haut, avant l'exécution du script *R* permettant leur organisation les résultats sont sous forme de fichiers *CSV* journaliers et suivant le type de rejet. Sous cette forme il serait très fastidieux et long de les analyser.

De plus, quand tous les résultats auront été obtenus, il sera possible de produire une base de donnée les contenant ce qui permettra à d'autres personnes de les utiliser pour compléter l'étude ou les utiliser de façon différente.

4.1 Points forts

L'utilisation de logiciels open source comme *QGIS* et *R* permet la reproductibilité de cette étude dans des environnements n'ayant pas forcément accès à des logiciels coûteux. Les scripts produits peuvent servir de base de travail pour d'autres études car ceux-ci ne sont pas spécifiques et s'occupent uniquement de l'automatisation de certains traitements. Bien évidemment, certaines variables devraient être modifiées mais le principe de base de chaque script reste le même :

- Conversion *KML* - *SHP*
- Rasterisation
- Addition de raster
- Statistique de zones
- Organisation des résultats

De plus, ces scripts sont dissociés et peuvent être utilisés séparément.

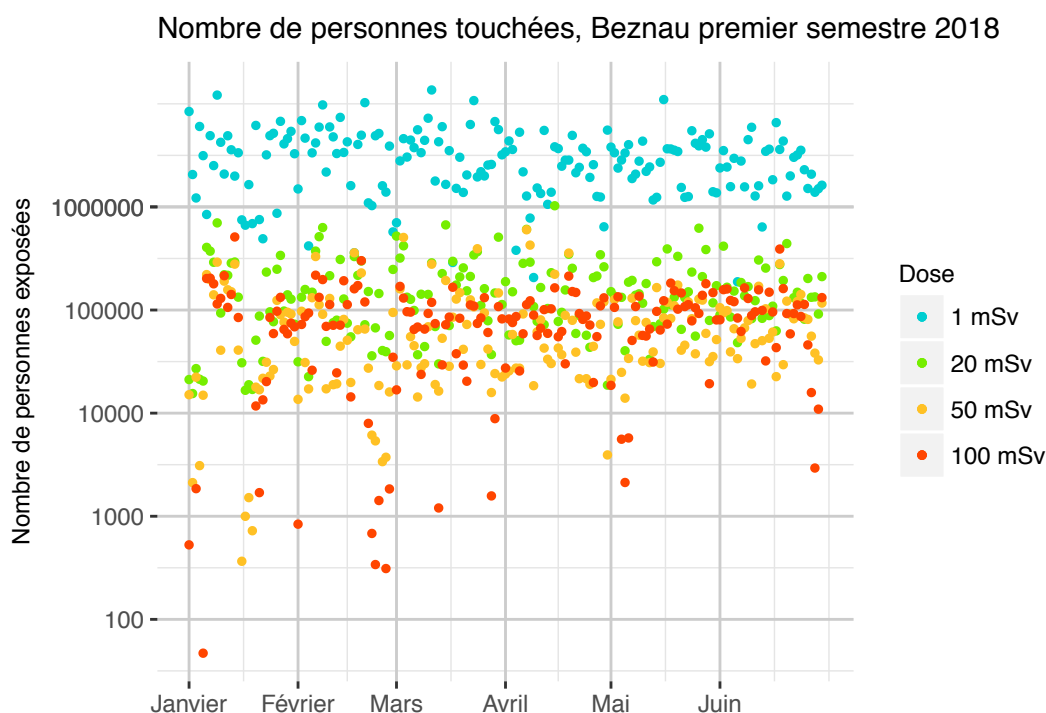


Figure 4-3 : Nombre de personnes touchées en cas d'accident à Beznau durant le premier semestre 2018

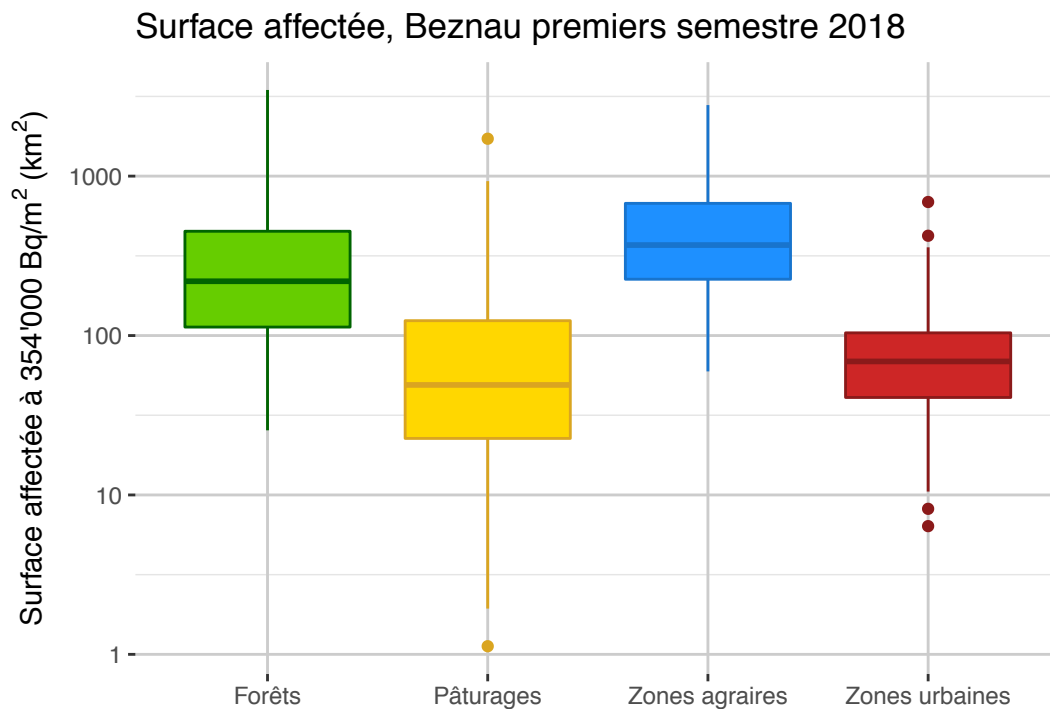


Figure 4-4 : Surfaces affectée à 354'000 Bq/m² en cas d'accident à Beznau durant le premier semestre 2018

4.2 Points faibles

Malgré le fait que l'automatisation des processus permet le traitement et la production d'une très grande quantité de données alors qu'effectuer les mêmes traitements manuellement ne serait pas envisageable, le temps de calcul des différentes étapes n'est pas anodin. La durée est dépendante de la machine utilisée (dans mon cas : *MacBook Pro* de mi-2012 avec un processeur 2.9 GHz et 8 Go de RAM). En prenant cela en compte, il faut environ 10 minutes par mois et par centrale pour le premier script *Python* (conversion des *KML* en *SHP*, voir Chapitre 3.4.1), troisième script *Python* (cumule des données de nuage et déposition, voir Chapitre 3.4.3) et le script *R* (organisation des résultats, voir Chapitre 3.4.5) ; 15 minutes par mois et par centrale pour le deuxième script *Python* (production des fichiers raster, voir Chapitre 3.4.2) et 3h par mois et par centrale pour le quatrième script *Python* (production des statistiques, voir Chapitre 3.4.4). Au total, il faut compter deux jours non-stop de calcul pour produire la statistique d'une centrale pour une année.

D'autres points faibles liés au programme *QGIS* ont été observés et aucune solution satisfaisante n'a été trouvée à ce jour. Effectivement, il semblerait que certains outils utilisés remplissent une mémoire tampon qui n'est jamais vidée. Par exemple, le premier script *Python* permettant la conversion des *KML* en *SHP* peut être utilisé pour un maximum de trois mois consécutifs avant que les fichiers produits soient corrompus et que des messages d'erreur apparaissent dans la console. La solution permettant de contourner ce problème est de fermer et rouvrir *QGIS* pour calculer les mois suivants. Un souci similaire est observé pour le quatrième script qui s'occupe de produire les statistiques de population et de types de sols touchés. Dans ce cas,

l'outil (« *saga:zonalgridstatistics* ») accumule des données dans une mémoire tampon sur le disque et remplit ainsi le disque dur hôte. Fermer et rouvrir *QGIS* permet de réinitialiser le disque si cette opération est effectuée avant la saturation dudit disque. Si le périphérique de stockage s'avère être totalement saturée, *QGIS* ne répond plus et doit être forcé à quitter. À ce moment, la seule solution pour réinitialiser la mémoire est de redémarrer la machine. Par exemple, sur mon ordinateur qui comporte 250 Go de mémoire disque libre, il m'est possible de calculer le script de statistiques pour une centrale pour une durée d'un mois et demi avant la saturation de la mémoire. Mes recherches n'ont pas abouti à une solution. Il semblerait que ces soucis soient liés directement à la version de *QGIS* (2.18) et sont possiblement corrigés dans la dernière version de *QGIS* (>3). Malheureusement, les scripts utilisés pour ce travail ne sont pas aisément adaptables à *QGIS3* qui utilise une version différente de *Python* (3.6).

5. Bilan du stage

Ce stage m'a permis de compléter mes connaissances en géomatique, de par l'utilisation du programme open source *QGIS*, de sa console de programmation sous *Python* et la recherche de données statistiques en libre accès correspondant le mieux possible aux demandes du mandataire. Le fait que l'association *Sortir du Nucléaire* n'est pas en lien direct avec le domaine de la géomatique m'a obligé à remplir ce mandat de façon indépendante et efficace. J'ai effectivement dû appliquer mes connaissances acquises lors des cours du certificat mais aussi faire preuve d'initiatives afin de trouver une stratégie applicable à la problématique donnée. La programmation des traitements appliqués dans cette étude a dû être conçue depuis le début, ce qui a permis une grande liberté d'action mais a aussi engendré une grande responsabilité quant aux problèmes rencontrés ainsi qu'aux points faibles de ce travail.

À ce jour, les traitements statistiques n'ont pas encore été produits dans leur totalité au vu du temps de calcul nécessaire, mais les scripts sont quant à eux établis. Le mandat a été plus long que prévu car cette étude innovante a demandé de nombreux changements et re-calibrations à divers niveaux (calculs de dispersion, choix des isolignes etc.).

Les perspectives pour la suite des opérations sont la complétion des calculs statistiques pour toutes les centrales pour l'année 2017 et leur organisation. Les données de dispersion de 2016 et 2018 doivent être recalculées lorsque le « shutdown » des activités gouvernementale U.S sera terminé, car les données météo proviennent de la NOAA et ne sont plus accessibles à ce jour (15.01.2019). Il sera alors possible de produire les données statistiques pour 2016 et 2018.

Lors de discussions avec le groupe de travail, il a été décidé d'ajouter aux statistiques de population et de types de sols touchés leur regroupement selon les situations météorologiques typiques.

L'étude se poursuit mais les objectifs du stage, étant principalement le choix de données de densité de population et de couverture du sol à une bonne résolution ainsi que l'automatisation des traitements permettant d'obtenir une statistique journalière (sur de longues périodes de temps) sur les impacts d'un accident nucléaire majeur en Suisse, ont été atteints.

5.1 Perspectives

Grâce à ce travail, les données produites vont pouvoir être analysées et les impacts évalués. La suite de l'étude va se focaliser sur une potentielle corrélation entre les différentes situations météorologiques et la gravité des impacts. Aussi, un petit clip vidéo intégrant une situation choisie va être produit afin de montrer l'évolution heure par heure d'un nuage de rejets radioactifs dans l'atmosphère. Ce film va permettre de sensibiliser la population.

6. Conclusion

Le travail effectué durant ces quelques mois a abouti au développement des divers scripts *Python* et *R* permettant l'automatisation de la conversion de fichiers *KML* en fichiers *SHP*, la consolidation et la complétion de ces derniers, la conversion en fichiers de type raster afin de les cumuler, la production d'une statistique de zones et enfin, l'organisation des résultats dans un format facilitant leur lecture. Des données statistiques en libre accès ont été récoltées et l'étude a été faite en utilisant des programmes open source.

Ce stage m'a appris à travailler de façon indépendante et responsable. Il m'a permis d'appliquer les connaissances acquises lors des cours du Certificat Complémentaire en Géomatique mais aussi de développer une méthode de travail personnelle. La collaboration pluridisciplinaire d'une telle étude nécessite une vision d'ensemble afin de travailler de façon efficace. Cette complémentarité au sein de l'équipe a révélé l'atout que représente la maîtrise des systèmes d'information géographique.

Cette première expérience professionnelle dans le domaine de la géomatique a été très enrichissante et a permis de tester avec succès les compétences développées durant le cursus du Certificat. Les diverses notions enseignées m'ont permis de rapidement mettre en place une stratégie pour répondre aux besoins de l'étude et aussi à faire face aux imprévus.

7. Références

Corine Land Cover (Copernicus program) :

<https://www.copernicus.eu/en> (dernière consultation 16.01.2019)

GHS (Global Human Settlement, European Commission) :

https://ghsl.jrc.ec.europa.eu/ghs_pop.php (dernière consultation 16.01.2019)

IAEA (International Atomic Energy Agency) :

<https://www.iaea.org/> (dernière consultation 16.01.2019)

IFSN (Inspection Fédéral de la Sureté Nucléaire) :

<https://www.ensi.ch/fr> (dernière consultation 16.01.2019)

NOAA (National Oceanic Atmospheric Administration) :

<https://www.noaa.gov/> (dernière consultation 16.01.2019)

Office Fédérale de l'Énergie :

<http://www.bfe.admin.ch/> (dernière consultation 16.01.2019)

Python : <https://www.python.org/> (dernière consultation 16.01.2019)

QGIS : <https://www.qgis.org/fr/site/> (dernière consultation 16.01.2019)

R : <https://www.r-project.org/> (dernière consultation 16.01.2019)

Sortir du nucléaire :

<http://sortirdunucleaire.ch/> (dernière consultation 16.01.2019)

Report of the Chernobyl Forum Expert Group 'Environnement', *Environmental Consequences of the Chernobyl Accident and their Remediation : Twenty Years of Experience*, (2006), IAEA

8. Annexes

Les scripts produits pour ce travail sont disponibles en annexes pour la centrale de Beznau et sont représentatifs des scripts utilisés pour les autres centrales. Les scripts *Python* sont sous un format *.py* et le script *R* sous le format *.R*.

1. Table de reclassification pour CLC
2. 01BEZ_imp_attr2.py (Numérique)
3. 02BEZ_rasterize2.py (Numérique)
4. 03BEZ_raster_sum.py (Numérique)
5. 04BEZ_table.py (Numérique)
6. BEZNAU_4.0.R (Numérique)

Table 4 : Table de correspondance des classes de la couche CLC et des nouvelles catégories (voir Chapitre 3.3.2 pour les nouveaux labels).

# QGIS Generated Color Map Export File		
Nouvelles catégories	Catégories originales	
1	1	111 - Continuous urban fabric
2	2	112 - Discontinuous urban fabric
2	3	121 - Industrial or commercial units
2	4	122 - Road and rail networks and associated land
2	5	123 - Port areas
2	6	124 - Airports
3	7	131 - Mineral extraction sites
3	8	132 - Dump sites
3	9	133 - Construction sites
4	10	141 - Green urban areas
4	11	142 - Sport and leisure facilities
5	12	211 - Non-irrigated arable land
5	13	212 - Permanently irrigated land
5	14	213 - Rice fields
5	15	221 - Vineyards
5	16	222 - Fruit trees and berry plantations
5	17	223 - Olive groves
6	18	231 - Pastures
5	19	241 - Annual crops associated with permanent crops
5	20	242 - Complex cultivation patterns
5	21	243 - Land principally occupied by agriculture with significant areas of natural vegetation
5	22	244 - Agro-forestry areas
7	23	311 - Broad-leaved forest
7	24	312 - Coniferous forest
7	25	313 - Mixed forest
6	26	321 - Natural grasslands
8	27	322 - Moors and heathland
8	28	323 - Sclerophyllous vegetation
8	29	324 - Transitional woodland-shrub
9	30	331 - Beaches - dunes - sands
9	31	332 - Bare rocks
9	32	333 - Sparsely vegetated areas
9	33	334 - Burnt areas
10	34	335 - Glaciers and perpetual snow
8	35	411 - Inland marshes
8	36	412 - Peat bogs
8	37	421 - Salt marshes
3	38	422 - Salines
0	39	423 - Intertidal flats
10	40	511 - Water courses
10	41	512 - Water bodies

10	42	521 - Coastal lagoons
0	43	522 - Estuaries
0	44	523 - Sea and ocean
0	48	999 - NODATA
0	49	990 - UNCLASSIFIED LAND SURFACE
0	50	995 - UNCLASSIFIED WATER BODIES
0	255	990 - UNCLASSIFIED