



# **Rapport de stage**

**Lydia SI MOUSSI**

Troisième année d'Ecole d'Ingénieurs SUP Galilée  
Mathématiques Appliquées et Calcul Scientifique



De Avril 2019 à Septembre 2019

**Tuteur du stage :** Sophie BERCU, Ingénieur Chercheur  
Département MIRE, EDF LAB Saclay

**Superviseur académique :** Bastien MALLEIN, Maître de conférences  
Laboratoire LAGA, Université de Paris 13

## Table des matières

<b>1</b>	<b><i>Introduction .....</i></b>	<b>5</b>
<b>2</b>	<b><i>Présentation de l'entreprise .....</i></b>	<b>6</b>
2.1	<b>Le Groupe EDF.....</b>	<b>6</b>
2.1.1	L'histoire d'EDF.....	6
2.1.2	EDF en chiffres.....	7
2.1.3	EDF en bourse .....	8
2.1.4	La concurrence d'EDF .....	8
2.2	<b>La R&amp;D.....</b>	<b>8</b>
2.2.1	Les propriétés stratégiques de la R&D .....	8
2.2.2	La R&D en France .....	9
	<b>Département MIRE.....</b>	<b>10</b>
2.3	<b>Le groupe R48 .....</b>	<b>11</b>
<b>3</b>	<b><i>Présentation du sujet contexte et problématique .....</i></b>	<b>12</b>
3.1	<b>Système d'information .....</b>	<b>12</b>
3.1.1	Communication au sein de l'équipe projet .....	12
3.1.2	Ressources fournies .....	12
3.1.3	Choix du langage de programmation .....	13
3.2	<b>Problématique et Objectifs du Stage .....</b>	<b>13</b>
	<b>Missions et tâches .....</b>	<b>14</b>
<b>4</b>	<b><i>Encrassement des transformateurs électrique.....</i></b>	<b>15</b>
4.1	<b>Le réseau de transport, de distribution et de poste de transformation.....</b>	<b>15</b>
4.1.1	Le réseau de transport .....	15
4.1.2	Le réseau de distribution.....	16
	Les postes de transformation .....	17
4.2	<b>Transformateurs électriques.....</b>	<b>17</b>
4.2.1	Comprendre le fonctionnement du transformateur.....	18
4.2.2	Système de refroidissement.....	20
4.3	<b>Encrassement du système aéroréfrigérants.....</b>	<b>22</b>
4.3.1	Intérêt du suivi de l'encrassement .....	23
4.3.2	Modes d'encrassement .....	23
4.3.3	Méthodes de suivi de l'encrassement.....	25
4.4	<b>Répertorier les données .....</b>	<b>27</b>
4.4.1	Poste source .....	27
4.4.2	Mesure d'encrassement.....	27
4.4.3	Données de pollinisation .....	29
<b>5</b>	<b><i>Deep Learning pour segmentation sémantique d'images .....</i></b>	<b>35</b>
5.1	<b>État de l'art .....</b>	<b>35</b>
5.2	<b>Jeux de données.....</b>	<b>36</b>
5.3	<b>Approche CNN pour la segmentation sémantique .....</b>	<b>37</b>
5.4	<b>Base de données d'entraînement.....</b>	<b>39</b>

5.5	Mise en œuvre .....	40
5.6	Résolution de l'overffiting .....	41
5.7	Résultats .....	43
<b>6</b>	<b><i>Segmentation sémantique par intensité de pixel.....</i></b>	<b>44</b>
6.1	OpenStreetMap .....	44
6.2	Le webservice Overpass.....	45
6.3	Segmentation par classification des cartes OSM statiques .....	47
6.3.1	La méthode des k-means .....	47
6.3.2	Scraper les cartes OSM.....	48
6.3.3	La détection et la description de Features .....	49
6.3.4	Ségmentation de cartes .....	53
6.3.5	Validation du Model .....	54
6.3.6	Reconstitution de données pour l'entrainement du model U-net.....	55
<b>7</b>	<b><i>CONCLUSION.....</i></b>	<b>57</b>
7.1	Mission accomplies et perpectives .....	57
7.2	Apport du stage,.....	57
<b>8</b>	<b><i>Bibliographie.....</i></b>	<b>58</b>

## REMERCIEMENTS

Je tiens à remercier toutes les personnes qui ont contribué au bon déroulement de mon stage et qui m'ont aidé lors de la rédaction de ce rapport.

Tout d'abord, j'adresse mes remerciements à ma tutrice, **Mme Sophie Bercu, ingénieur chercheur au sein du groupe R48 du département MIRE**, pour son accueil, pour sa patience, le temps passé ensemble et le partage de son expertise au quotidien. Elle fut d'une aide précieuse dans les moments les plus délicats.

Je remercie également mon encadrant académique, **Mr Bastien MALLEIN, Maître de conférences au laboratoire LAGA, Université de Paris 13**. Qui m'a rendu visite à l'entreprise pour s'assurer du bon déroulement de mon stage, et pour sa contribution dans la rédaction de ce rapport. Ainsi que pour toute l'équipe pédagogique de l'université de Paris 13 en particulier le département de mathématique.

Je remercie également toute l'équipe **R48** pour leur accueil, leur esprit d'équipe et en particulier tous les stagiaires, qui ont contribué à la bonne ambiance de travail au long de toute la période du stage.

Enfin, je tiens à remercier toutes les personnes qui m'ont soutenu : ma famille, mon fiancé et mon camarade de promotion qui m'a toujours aidé **Laydine**.

Figure 1 : Répartition de la production et de la distribution d'électricité en France.....	6
Figure 2 : Production d'électricité en TWH .....	7
Figure 3 : EDF LAB France .....	9
Figure 4 : Vue Aérienne du centre EDF de recherche et développement de Paris Saclay.....	9
Figure 5 : Réseau Très Haute Tension (THT) en France .....	16
Figure 6 : Transformateur très haute tension, .....	18
Figure 7 : Transformateur monophasé idéal .....	18
Figure 8 : Liste des composants constituant un transformateur .....	19
Figure 9 : le banc de radiateur avec des ventilateurs dessous .....	20
Figure 10 : Refroidissement force (OF) .....	21
Figure 11 : Refroidissement dirigé (OD) .....	21
Figure 12 : Transformateur ONAN Source .....	22
Figure 13 : Transformateur ONFWF .....	22
Figure 14 : Filtres des aéroréfrigérants .....	24
Figure 15 : Encrassement des filtres .....	24
Figure 16 : Faisceaux des aéroréfrigérants.....	25
Figure 17 : Encrassement des Faisceaux .....	25
Figure 18 : Méthode de suivi de l'encrassement.....	26
Figure 19 : Deux configurations de l'aéroréfrigérant .....	26
Figure 20 : Poste source en France.....	27
Figure 21 : Relevés de la différence de pression sur le poste de Pailler .....	28
Figure 22 : Carte de Vigilance Septembre 2019 .....	29
Figure 23 : Position des capteur de pollinisation .....	30
Figure 24 : la variation du taux du pollen (Platane, Peuplier, saule) .....	32
Figure 25 : Variation des graminées fourragères et céréalières .....	32
Figure 26 : Les station de mesure de la qualité de l'aire dans la région nouvelle aquitain .....	33
Figure 27 : Niveau de Pm10 mesuré par la station de bordeaux en 2016 .....	33
Figure 28 : Scraping web .....	36
Figure 29 : Image Satellite (à droite), la carte d'occupation correspondante (à droite) - Source (Postadjian, et al. s.d.) .....	37
Figure 30 : ARCHITECTURE STANDARD D'UN RESEAU A CONVOLUTIONS SOURCE : WIKIPEDIA.....	38
Figure 31 : Architecture Réseaux U-net.....	38
Figure 32 : Donnée référence pour entraînement .....	40
Figure 33 : La fonction loss.....	41
Figure 34 : Parmi les transformations possibles pour augmenter .....	42
Figure 35 : La fonction Loss après Data-Augmentation .....	43
Figure 36 : Image à 8 bandes et sa prédiction par le réseau Unet .....	43
Figure 37 : Utilisation des bases de données d'OSM .....	44
Figure 38 : Extraction fond de carte.....	45
Figure 39 : Algorithme du k-means .....	48
Figure 40 : Exemple Carte statistique OSM résumant .....	49
Figure 41 : répartition de la couleur feature foret .....	51
Figure 42 : carte statique .....	53
Figure 43 : occupation de sol .....	54
Figure 44 : carte donnée par le modèle .....	54

## 1 Introduction

Leader mondial des énergies bas carbone, le groupe EDF rassemble tous les métiers de la production, du commerce et des réseaux d'électricité. En s'appuyant sur l'expertise de ses équipes, sa R&D et son ingénierie, son expérience d'exploitant industriel et l'accompagnement attentif de ses clients, EDF apporte des solutions compétitives qui concilient développement économique et préservation du climat.

J'ai effectué mon stage de troisième année d'école d'ingénieurs SUP Galilée au sein du département MIRE (Mesure et système d'Information des Réseaux Electriques) de la R&D d'EDF (Campus de Paris Saclay) du 01 Avril 2019 au 27 Septembre 2019 sur le thème de maintenance des transformateurs du réseau électrique.

Le réseau de transport et de distribution d'électricité transporte l'énergie électrique d'un site de production vers des consommateurs. Au sein de ce réseau se trouvent d'importants transformateurs de puissance à l'interface entre différents niveaux de tension. Le système aéroréfrigérants de ces transformateurs subissent un encrassement, ce qui cause par la suite une augmentation de la vitesse de leur vieillissement. La diminution des capacités d'échange de l'aéroréfrigérants réduit également les capacités de surcharge du transformateur, ce qui peut être préjudiciable à l'exploitation lors des manœuvres suite à un incident.

Enedis a décidé de procéder au déploiement du système de monitoring AGATH sur son parc de transformateurs du réseau de distribution. Le système AGATH est constitué d'une centrale d'acquisition et de différents capteurs permettant le suivi en temps réel du transformateur. A partir de ces mesures, Enedis souhaite que soient mis en place des indicateurs permettant une optimisation de la maintenance et des conditions d'exploitation.

A la suite de la réunion du 24 octobre 2017 entre R&D d'EDF et Enedis, les indicateurs suivants ont été définis comme prioritaires :

- Encrassement des aéroréfrigérants ;
- Perte d'efficacité de l'assécheur d'air ;
- Risques spécifiques au Changeur de Prises En Charge

Dans le cadre de mon stage de fin d'étude que j'ai effectué dans le groupe R48 du département MIRE, il s'agit de déterminer précisément et localement les causes du phénomène d'encrassement, afin de permettre une classification des transformateurs selon leur risque d'encrassement et de réaliser les opérations de nettoyage au meilleur moment afin de définir les cycles de maintenance pour chaque transformateur en fonction de son environnement géographique.

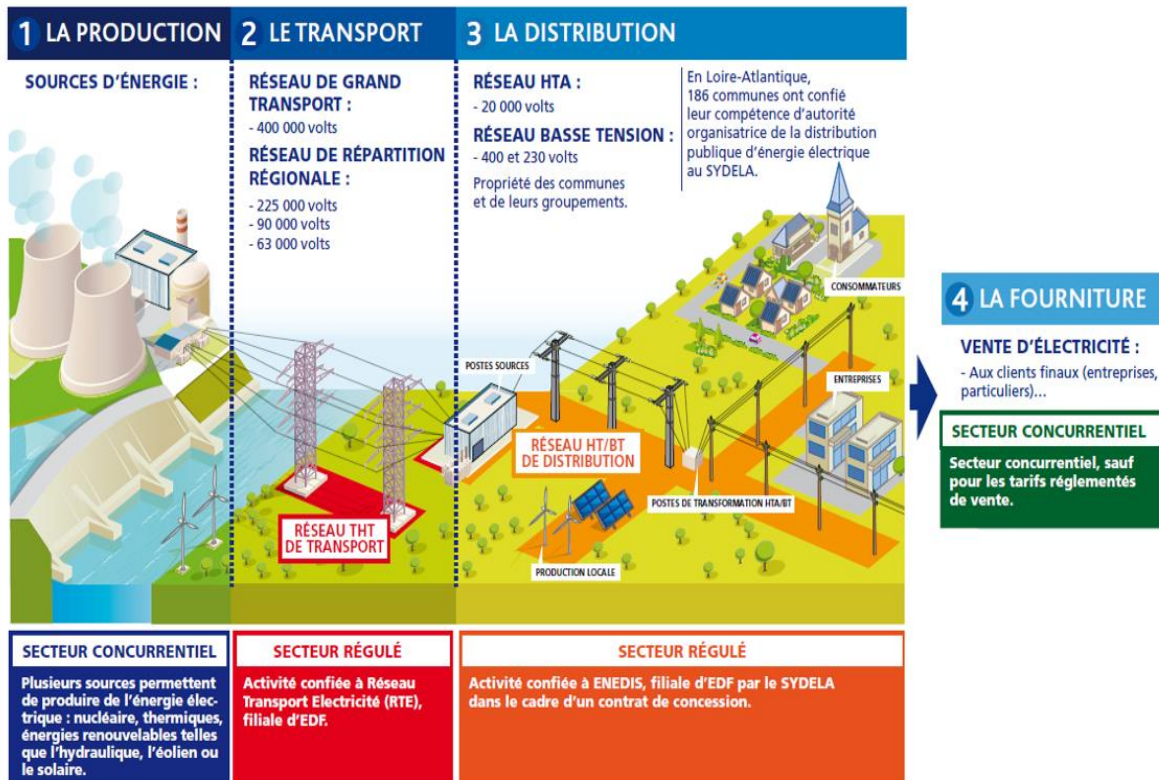
## 2 Présentation de l'entreprise

### 2.1 Le Groupe EDF

Le groupe EDF (Électricité de France) est le principal producteur et fournisseur d'électricité en France et à l'étranger. Le groupe international a longtemps été l'unique fournisseur français d'électricité, jusqu'en 2007 où le marché de l'électricité s'est ouvert à la concurrence. Face à ces changements de paramètres le groupe a opéré un revirement dans sa stratégie de croissance en nommant Jean-Bernard Lévy PDG d'EDF en 2014. Celui-ci a succédé à Henri Proglio dans le but de sécuriser les parts de marché acquises par EDF.

Le groupe EDF intervient tout le long du parcours de l'électricité, de sa conception jusqu'à sa distribution. La principale source d'électricité exploitée par EDF est le nucléaire : de janvier à novembre 2018, la production nucléaire s'élevait à 354,3 TWh, soit 3,7% de plus que l'année 2016. Le groupe EDF reste le premier producteur mondial de l'électricité nucléaire.

Le groupe EDF est différent d'EDF fournisseur d'électricité : en effet, le groupe EDF était à l'origine chargé de la production, de la gestion, de la distribution et de l'acheminement de l'électricité. Les différentes fonctions autour de l'électricité ont été divisées en plusieurs filiales du même groupe (voir la figure ci-après). EDF désigne donc à la fois le groupe comprenant plusieurs filiales et la filiale qui s'occupe de la distribution de l'électricité, aujourd'hui Enedis.



**FIGURE 1 : REPARTITION DE LA PRODUCTION ET DE LA DISTRIBUTION D'ELECTRICITE EN FRANCE**  
SOURCE : SITE EDF

#### 2.1.1 L'histoire d'EDF

EDF été créé sous l'égide de Maurice Thorez et Marcel Paul le 8 avril 1946. Réunissant toutes les infrastructures existantes pour la production, la gestion et la distribution d'électricité. En effet un grand nombre de ces structures avaient été créées pendant l'entre-deux guerre. Toutes les structures qui produisaient et distribuaient l'électricité, depuis son apparition en France un siècle plus tôt, furent réunies en une seule grande entreprise.

Au cours de son histoire EDF a participé à la construction et à l'entretien de centrales nucléaires afin de produire de l'énergie renouvelable qui remplace les énergies fossiles comme le charbon et le pétrole.

Avec la conférence du climat COP21 de 2015, EDF a montré sa volonté de s'engager dans la transition énergétique, en investissant notamment dans des infrastructures permettant de passer sur le long terme à un plus grand pourcentage d'énergies vertes. EDF se soucie des énergies nouvelles, et notamment de la mise en place d'un réseau électrique intelligent.

Les *smart grids* sont toutes les interconnexions formées par le réseau électrique, qui forme comme une grille qui couvre tout un territoire urbain. Ces grids sont dites "smart" c'est à dire intelligentes : Les compteurs Linky sont en cours de déploiement sur tout le territoire français, ils seront automatisés et reliés au réseau national

### 2.1.2 EDF en chiffres

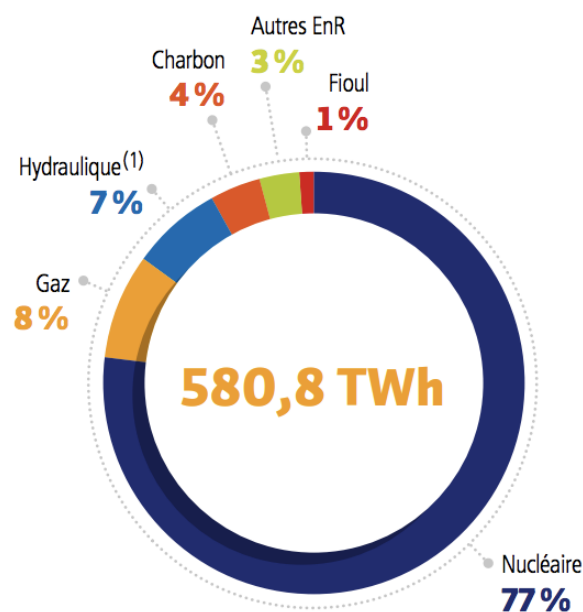
En 2017, le groupe Électricité de France a produit plus de 580 TWh d'électricité à travers le monde. On estime que le groupe EDF a plus de 35 millions de clients à travers le monde. En 2017 EDF comptait plus de 152 000 employés dans le monde.

Le nucléaire, première source d'électricité d'EDF, représente pour le groupe 77% de sa production d'électricité. EDF a conçu avec le groupe AREVA, le premier réacteur nucléaire : un EPR (Réacteur Pressurisé Européen) est un réacteur nucléaire de troisième génération, qui propose plus de sécurité et de rentabilité économique que ses prédécesseurs.

La production thermique représente pour EDF 7% de sa production en électricité. EDF produit 4% de son électricité à base de charbon et 1% à base de fioul.

La production hydraulique représente la première des sources d'énergies renouvelables exploitées par EDF, avec 640 barrages et 435 centrales hydroélectriques installées en France.

La part d'énergie renouvelable produite par EDF est de 24% en 2017. Durant cette même année, le groupe a renforcé ses capacités solaires et éoliennes de 1,5 GW. Le groupe souhaite continuer à développer ce type d'énergie à l'avenir.



**FIGURE 2 : PRODUCTION D'ELECTRICITE EN TWH**  
SOURCE : SITE EDF



### 2.1.3 EDF en bourse

Comme de nombreuses entreprises, EDF est cotée en bourse. La valeur EDF fait partie de l'indice CAC Next 20 depuis le 21 décembre 2015. Au deuxième trimestre 2018, son principal actionnaire était l'État français, à hauteur de 83,7%. Le cours de l'action et de la valeur EDF est disponible en temps réel sur son site et sur le site Boursorama où on peut également directement effectuer des transactions.

### 2.1.4 La concurrence d'EDF

Les fournisseurs alternatifs comme Direct Énergie, Planète Oui ou ENI sont producteurs et fournisseurs ou seulement fournisseurs d'électricité. Leur rôle est de produire de l'électricité et de la commercialiser aux particuliers. Ils ne gèrent ni le transport (rôle de RTE) de l'électricité, ni l'entretien des lignes et des compteurs (rôle d'ENEDIS).

Le nombre de fournisseurs d'énergie a considérablement augmenté depuis 2007. De plus en plus de fournisseurs proposent notamment des offres d'énergie renouvelable, qui produit lui-même 100% d'énergies issues de modes de productions renouvelables. La loi NOME (Nouvelle Organisation du Marché de l'Electricité), promulguée en 2010, oblige Electricité de France à revendre un quart de sa production d'électricité nucléaire aux fournisseurs d'électricité présents sur le marché, dits "alternatifs". Le but de la loi NOME est de stimuler la concurrence sur le marché de l'électricité.

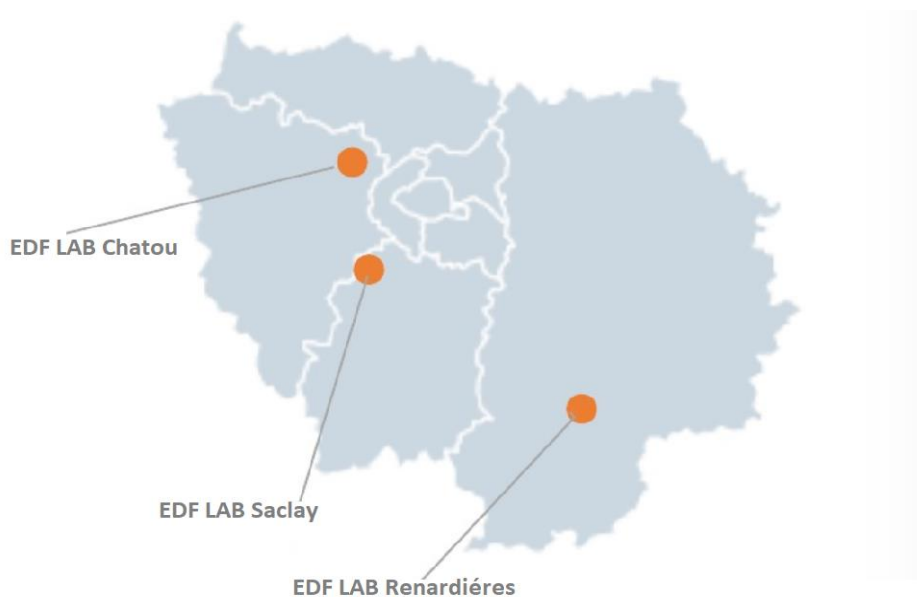
## 2.2 La R&D

La R&D mobilise ses experts de l'énergie pour améliorer les performances du groupe, et explorer les technologies de ruptures qui construiront le mode de demain. Elle travaille pour les différents métiers et filial d'EDF et aussi pour des acteurs extérieurs dans tous les domaines du secteur énergétique. Pour relever ses défis, près de 2000 collaborateurs conduisent des activités dans le monde entier, 3 centres en France ainsi que 6 centres à l'internationales (Etats-Unis, Angleterre, Chine, Allemagne, Italie, Singapour, Pologne). Ces collaborateurs qui forment les équipes, qui sont des chercheurs au profil varié et aux compétences multiples, dont 117 doctorants et 1900 salariés.

### 2.2.1 Les propriétés stratégiques de la R&D

Les priorités stratégiques de la R&D s'organisent autour de trois axes. Le premier consiste à consolider et développer des mix de production compétitifs et décarbonnés comme des réacteurs nucléaires de faible puissance. Le deuxième est destiné à développer et expérimenter de nouveaux services énergétiques pour les clients. Enfin, le troisième axe est chargé de préparer les systèmes électriques de demain.

## 2.2.2 La R&D en France



**FIGURE 3 : EDF LAB FRANCE**  
SOURCE SITE EDF

### 2.2.2.1 EDF LAB Paris Saclay



**FIGURE 4 : VUE AERIEENNE DU CENTRE EDF DE RECHERCHE ET DEVELOPPEMENT DE PARIS SACLAY**  
SOURCE : SITE EDF

Achevé en mars 2016, EDF Lab Paris-Saclay est l'un des 10 sites de la R&D. Il concentre en un même lieu des compétences d'excellence et des moyens d'essais sur des sujets stratégiques

pour l'avenir du Groupe EDF : la production d'électricité décarbonée, les systèmes électriques du futur et les nouveaux services énergétiques.

Ses centres d'expertise sont variés : analyses en mécanique avancée, développement des systèmes d'information pour les réseaux électriques, fonctionnement et étude des systèmes énergétiques, technologies et modélisation des infrastructures du système électrique, innovation commerciale, analyse des marchés et de leur environnement, management des risques industriels, simulation neutronique, technologies de l'information, calcul scientifique, mathématiques appliquées.

#### *2.2.2.2 EDF LAB Les Renardières*

Ses recherches s'étendent de l'étude du vieillissement des matériaux des centrales nucléaires aux recherches sur les technologies de rupture pour les énergies renouvelables et le stockage, en passant par les smart grids. Il conçoit et teste de nouveaux services d'efficacité énergétique ainsi que d'offres pour l'habitat et l'industrie.

#### *2.2.2.3 EDF LAB Chatou*

Ses études se concentrent essentiellement dans le domaine de la production hydraulique, nucléaire et énergies renouvelables, entre autres, en simulation numérique, campagnes de mesures en nature, chimie, écotoxicologie et microbiologie, outils de modélisation, mais aussi Thermo-hydraulique, Sûreté incendie, Evaluation de la maturité des nouvelles technologies de l'énergie, et développe des méthodes d'accompagnement dans la conduite, la surveillance et le maintien des installations existantes et à venir. A cela, s'ajoute l'équipe du Programme Production qui définit et arbitre les programmes de recherche dans le domaine de la production d'électricité, et celle de l'Innovation Hub qui accompagne la transformation des métiers et de la R&D par les tendances, l'open innovation et le design.

### Département MIRE

Le département exerce ses compétences sur l'ensemble du cycle de vie des réseaux, de la conception jusqu'à la gestion (mesure des performances, pilotage de la qualité, comptage), en passant par l'exploitation (protections, contrôle-commande, supervision) et la sûreté de fonctionnement (maintenabilité, disponibilité, fiabilité). Il apporte également à ses différents clients un éclairage sur les évolutions potentielles de leur système d'information et des échanges sous-jacents. Il réalise des études techniques aussi bien que des spécifications, développements ou tests (logiciels et matériels).

Bien entendu, les activités du Département s'inscrivent pleinement dans les réflexions visant à développer le SmartGrids. Les compétences qui sont cultivées au sein du département, ainsi que les partenariats avec des organismes de recherche en France et à l'étranger, sont au cœur des thématiques qui préfigurent les "SmartGrids" : gestion optimisée des actifs, solutions de comptage intelligent, reconfiguration automatique des réseaux de distribution intégrant le développement des énergies renouvelables, urbanisme des systèmes d'information et normalisation, qualité de l'onde et impact des nouveaux usages.

Ce département compte une centaine de personnes, réparties en cinq groupes d'activité :

- Le groupe R42 « Qualité de l'Electricité et Développement des Réseaux »
- Le groupe R43 « Systèmes de Comptage »
- Le groupe R44 « Automatisation et Protection des Réseaux »
- Le groupe R48 « Fiabilité et Disponibilité des Réseaux Electriques »
- Le groupe R49 « Interopérabilité et Applications Métiers »

### 2.3 Le groupe R48

Le groupe dans lequel j'ai effectué mon stage, est en charge de la fiabilité et de la disponibilité des réseaux électriques. Le groupe dispose de deux pôles de compétences. Le premier concerne la sûreté de fonctionnement et le second concerne la gestion optimisée des actifs. Il regroupe des compétences statistiques, fiabilistes et électrotechniques afin de couvrir les domaines suivants :

- Maîtriser la fiabilité des réseaux électriques et améliorer la connaissance des facteurs influents de la fiabilité des réseaux électriques.
- Analyse du retour d'expérience et des informations contenues dans les bases de données disponibles, extraction des données pertinentes, traitements statistiques de ces données en vue d'établir des modèles de fiabilité et de vieillissement des composants et des réseaux, prise en compte de la fiabilité dans les spécifications de matériels. Prise en compte des facteurs influents de la fiabilité.
- Contribuer à l'optimisation des politiques de gestion des actifs des réseaux.
- Intégration des modèles de fiabilité dans une vision système intégrant des arbitrages financiers. Simulation probabiliste. Quantification des actions de maintenance.
- Assister les Directions opérationnelles dans l'identification et la gestion des risques qui pèsent sur leurs réseaux.
- Application et mise au point de méthodes d'analyse et de modélisation des risques, élaboration de méthodes ou d'outils d'aide à la décision. Etudes, modélisations et simulations de structures de réseaux pour en évaluer la sûreté de fonctionnement.
- Anticiper les modes de défaillance des « Smart Grids ».

## 3 Présentation du sujet contexte et problématique

### 3.1 Système d'information

#### 3.1.1 Communication au sein de l'équipe projet

Les membres du groupe R48 se trouvant au même étage du bâtiment OPALE, la communication verbale est le moyen de communication privilégié. L'équipe étant très disponible, il est aisé d'obtenir des réponses rapides à des questions ponctuelles ou d'aide informatique auprès des gens spécialisés. Autrement, il m'a été possible de demander des informations et des réunions par mail à l'ensemble des agents dans d'autres départements de la R&D, notamment au département de PRISME d'EDF R&D sur le Campus de Chatou. Pour des questions simples et nécessitant une réponse rapide, notamment les questions concernant certains détails et l'installation de modules soit, dans l'environnement informatique locale soit, sur l'espace informatique Data Innovation Lab (DIL).

#### 3.1.2 Ressources fournies

Plusieurs ressources m'ont été fournies dans le cadre de mon stage. Dans un premier temps, un ordinateur fixe a été mis à ma disposition, avec le système d'exploitation Windows et connecté au réseau Gardian. Ensuite un ordinateur portable m'a été attribué pour mes jours de déplacement dans le cadre du stage au campus d'EDF R&D Chatou.

J'ai également accès en DRP et en SSH (un protocole pour la connexion à distance sécurisée et d'autres services réseau sécurisés sur un réseau non sécurisé) au serveur DIL sur laquelle j'ai effectué les opérations de Deep Learning sur le traitement d'images. Enfin, un centre de logiciel EDF Store était accessible depuis mon poste et contenant différentes applications très utiles.

##### 3.1.2.1 Le Data Innovation Lab (DIL)

Le Data Innovation Lab (DIL) basé à EDF Lab Paris-Saclay, a pour objectif de valoriser l'ensemble des données par des techniques de Data Analytics et par une approche agile, en forte interaction avec les métiers d'EDF. Il s'agit d'une nouvelle manière de travailler pour pratiquer et développer la Data Science au sein de la R&D et du Groupe.

Les enjeux de l'analyse des données pour le groupe EDF couvrent l'éventail de ses métiers de la production et l'optimisation amont-aval, à la distribution et la commercialisation. Avec son Data Innovation Lab et son expertise, la R&D d'EDF souhaite accélérer la valorisation des données avec des techniques de Data Analytics pour répondre à l'ensemble des besoins métier.

Le DIL propose 3 niveaux d'offres adaptées au degré de maturité :

1. Le « **Data Shaker** » pour les utilisateurs qui souhaitent faire émerger des sujets de valorisation de leurs données ;
2. Le « **Use-Case** » pour les utilisateurs qui ont déjà une problématique bien définie pour laquelle on souhaite des apports techniques ;
3. Le « **Socle** », qui englobe la composante recherche scientifique approfondie de la R&D pour ceux qui souhaitent lever des verrous techniques sur une problématique bien précise.

Aujourd'hui, plus d'une quarantaine de chercheurs de la R&D collaborent et déclinent au sein du DIL des thèmes variés tels que la prévision, l'optimisation et le machine Learning, le marketing et la relation client, l'open data, le traitement d'image ou la maintenance.



### 3.1.3 Choix du langage de programmation

Le choix de Python comme langage de programmation n'est pas seulement parce que c'est le langage avec lequel je me sens la plus à l'aise, mais aussi du fait de son interactivité avec internet et avec l'environnement du DIL. Du plus, j'ai programmé dans l'environnement Jupiter (notebook), qui permet de rassembler, dans un même document, du texte, des images, des formules mathématiques et du code informatique exécutable. Les notebooks sont manipulables interactivement dans un navigateur web. Cela facilite du web scraping pour l'extraction de données, notamment les données images que j'ai largement utilisées au cours de mon stage.

Enfin peut-être le point le plus important, est l'ensemble des outils, des librairies, la communauté, bref, tout ce qui gravite autour du langage lui-même. Python possède des librairies pour à peu près tout ce dont j'ai eu besoin durant mon projet, et notamment :

- Numpy et scipy pour les calculs numériques sur des tableaux de données
- Matplotlib et Seaborn pour la visualisation
- Scikit-learn pour les algorithmes d'apprentissage automatique.
- OpenCV pour le traitement d'image
- Pandas pour les gérer les données (les charger, appliquer des opérations d'algèbre relationnelle, etc.)
- Tensorflow et PyTorch pour le Deep Learning

Pour son installation, la façon la plus rapide et efficace est d'installer la distribution anaconda qui permet non seulement d'avoir la dernière version du Python, mais aussi le Jupiter notebook et presque toutes les librairies citées précédemment. Sans oublier que cette méthode d'installation convient à la politique de sécurité du réseau informatique de l'entreprise gérée entre autres par un « proxy ».

## 3.2 Problématique et Objectifs du Stage

Le long des lignes dans le réseau de transport et de distribution, l'énergie électrique est guidée, répartie et sa tension est abaissée successivement dans des postes de transformation pour être livrée en quantité et en tension adaptées aux besoins des différents consommateurs et pour alimenter les postes sources du réseau de distribution.

Ces transformateurs sont des éléments essentiels jouent un rôle décisif et doivent être soigneusement entretenus. L'huile du transformateur doit être transportée et refroidie de manière sûre, durable et efficace. Les aéroréfrigérants jouent le rôle d'échangeur de chaleur dont l'efficacité diminue avec son niveau d'encrassement. L'encrassement d'un échangeur de chaleur peut être défini comme la dégradation avec le temps de ses capacités d'échange entre les fluides le traversant, par rapport aux capacités nominales. Les causes de cette évolution peuvent être réversibles (dépôts dans les filtres et faisceaux ; dégradation de l'huile) ou irréversibles (détérioration des filtres, faisceaux et ventilateurs ; corrosion ; recirculation de l'air chaud dû à une mauvaise installation).

L'encrassement des aéroréfrigérants augmente la vitesse de vieillissement du transformateur. La diminution des capacités d'échange de l'aéroréfrigérants réduit également les capacités de surcharge du transformateur, ce qui peut être préjudiciable à l'exploitation lors des manœuvres suite à un incident.

La pollution et les particules d'origines végétales notamment le pollen et les moisissures dans l'air sont à l'origine des dépôts dans les filtres des aéroréfrigérants.

Déterminer précisément et localement les causes du phénomène d'encrassement permettrait de classer les transformateurs selon leur risque d'encrassement et de réaliser les opérations de nettoyage au meilleur moment et définir les cycles de maintenance pour chaque transformateur en fonction de son environnement géographique.

L'objet de la proposition est de définir pour chaque transformateur son cycle de maintenance préventive systématique en fonction de son environnement géographique (zone à fort pollen, a fort niveau de pollution ...etc.) et en temps réel de surveiller en continue les prévisions des indicateurs environnementaux (pic de pollution, ...) et l'état de l'encrassement des aéroréfrigérants pour ajuster ce cycle systématique de nettoyage.

### Missions et tâches

Les étapes pour mettre en œuvre l'objectif sont les suivantes :

- I. Comprendre le fonctionnement du réseau électrique.
- II. Identifier toutes les causes externes environnementales à l'origine de l'encrassement des aéroréfrigérants.
- III. Répertoire où trouver les données identifiées.
- IV. Définir pour chaque transformateur son environnement : zone boisée avec des espèces à fort taux de pollen, zone urbaine à fort taux de pollution.
- V. Construire un modèle d'encrassement à partir des données déjà disponibles dans AGATH.

## 4 Encrassement des transformateurs électrique

### 4.1 Le réseau de transport, de distribution et de poste de transformation

L'électricité circule depuis le lieu où elle est fabriquée jusqu'à l'endroit où elle est consommée, par l'intermédiaire d'un réseau de lignes électriques aériennes ou souterraines. Il permet de transporter et de distribuer l'énergie électrique sur l'ensemble du territoire français et même vers d'autres pays d'Europe.

Le réseau de transport et de distribution de l'électricité est organisé à la manière d'un réseau routier avec ses grands axes, ses axes secondaires et ses échangeurs :

#### 4.1.1 Le réseau de transport

Joue le rôle du réseau des autoroutes et des routes nationales, il transporte l'énergie électrique des centres de production, les centrales électriques, aux zones de consommation.

Il est présent à deux échelles :

À l'échelle nationale, via le réseau de grand transport et d'interconnexion, vers :

1. Les grandes zones de consommation ;
2. Les pays frontaliers : Italie, Espagne, Allemagne, Belgique, Suisse et Royaume-Uni (par le câble sous-marin IFA 2000) afin d'assurer la stabilité du réseau, la sécurité d'approvisionnement, les échanges commerciaux.

À l'échelle régionale et départementale, via le réseau de répartition, vers :

1. Les agglomérations ;
2. Les entreprises fortement consommatrices comme la SNCF, la RATP, ou les industries (chimiques, sidérurgiques et métallurgiques).

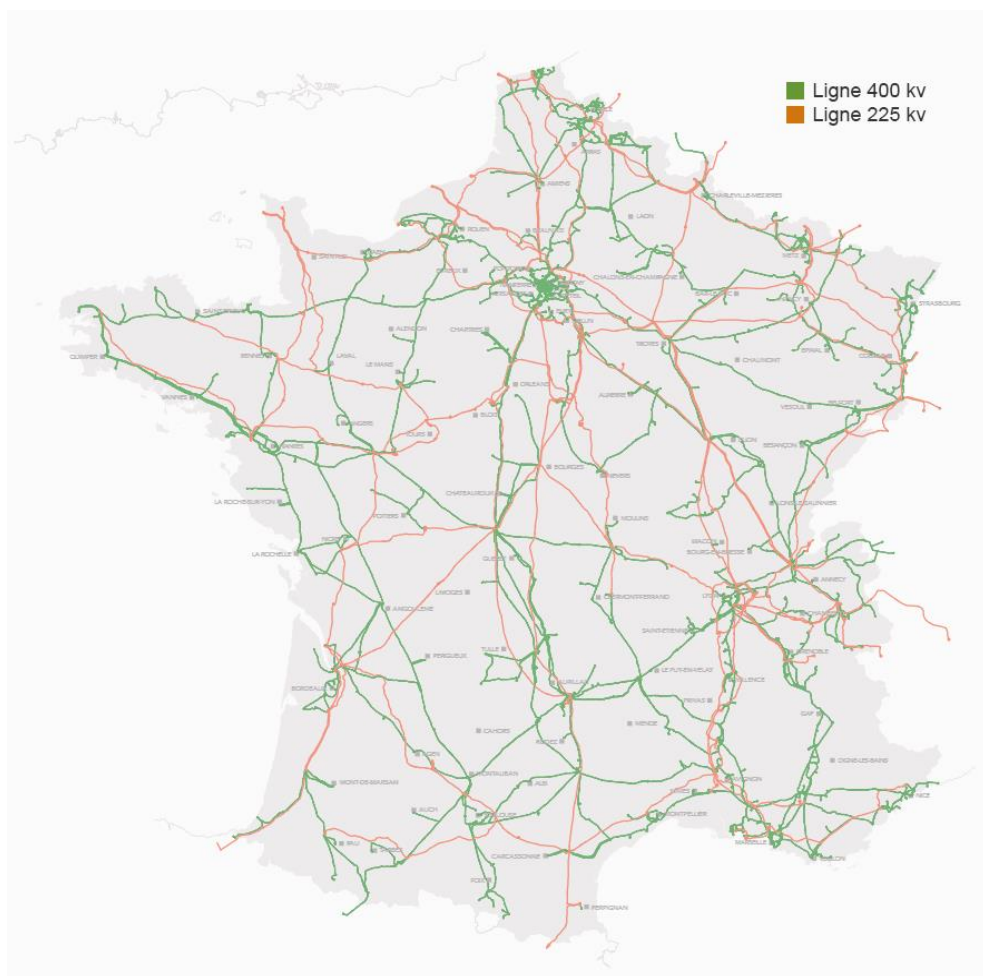
Cela représente 105 000 km de lignes Très Haute Tension (THT) et Haute Tension (HT) et 46 lignes transfrontalières exploitées, entretenues et développées par Réseau Transport Électricité (RTE).

Le réseau THT, est composé de lignes à :

- 400 000 volts (pour le réseau de grand transport),
- 225 000 volts (pour le réseau de répartition).

À 400 000 volts, les lignes THT permettent de limiter les pertes d'énergie pour le transport de quantités très importantes d'électricité sur de longues distances. Grâce à des postes de transformation, la tension est ensuite abaissée à 225 000 (THT), 90 000 (HT) ou 63 000 volts (HT) pour acheminer l'électricité en quantité moindre et sur de plus courtes distances.





**FIGURE 5 : RESEAU TRES HAUTE TENSION (THT) EN FRANCE**  
**SOURCE : SITE RTE**

#### 4.1.2 Le réseau de distribution

Joue celui du réseau des routes départementales, et il est exploité, entretenu et développé par ENEDIS, le réseau de distribution permet de transporter l'énergie électrique à l'échelle locale, des centres de distribution vers le client final : les petites et moyennes entreprises, les villes, les grandes surfaces, les commerces, les artisans, les particuliers.

Il peut exister localement des sources de production qui injectent de l'électricité sur le réseau (éolien, microcentrales hydrauliques, photovoltaïques...).

Il est constitué de deux types de réseaux électriques

- Le réseau Moyenne Tension (MT)
- Le réseau Basse Tension (BT)

#### 4.1.2.1 Le réseau Moyenne Tension (MT)

Long de 622 187 kilomètres, le réseau MT comporte une part rurale très importante. Ainsi, le réseau aérien, bien que décroissant progressivement, reste à ce jour plus long que le réseau souterrain. Il comporte des automatismes destinés à en faciliter l'exploitation et, en particulier, à en assurer la remise sous tension sans intervention humaine après une défaillance temporaire.

Le réseau MT alimente plus 760 000 postes de distribution MT/BT.

#### 4.1.2.2 Le réseau Basse Tension (BT)

Issues des postes de transformation MT/BT, les lignes BT sont construites, exclusivement depuis le décret d'avril 1991, en faisceaux de conducteurs isolés sur poteaux ou sur façades ou en câbles souterrains : 701 858 kilomètres de réseau dont plus de 260 000 km sous terre.

Chaque circuit BT est protégé par un jeu de fusibles placé en sortie de transformateur. Le branchement se situe entre le réseau BT et le point de départ de l'installation intérieure de l'utilisateur. Sa protection est assurée par des fusibles côté distributeur et un disjoncteur côté utilisateur.

### Les postes de transformation

Jouent le rôle d'échangeurs de tension. Pour rendre l'électricité transportable sur de longues distances dans le réseau d'interconnexion avec des pertes minimales d'énergie, un transformateur placé à la sortie des principales centrales élève la tension à 400 000 volts.

L'énergie électrique n'a pas la même tension quand elle entre dans le poste et quand elle en sort. Exemple : un poste 225 kV/ 90 kV signifie que du courant 225 000 volts entre dans le poste et qu'il en sort du courant à 90 000 volts.

Ces installations électriques sont composées de 3 types d'appareils :

- Des transformateurs qui abaissent la tension
- Des disjoncteurs, capables d'interrompre automatiquement en cas de nécessité
- Des sectionneurs, qui établissent ou interrompent un circuit à la demande, permettant l'aiguillage de l'énergie électrique.

Ce sont des lieux fermés et commandés à distance à partir de postes principaux, appelés Pupitres de Commandes Groupées.

## 4.2 Transformateurs électriques

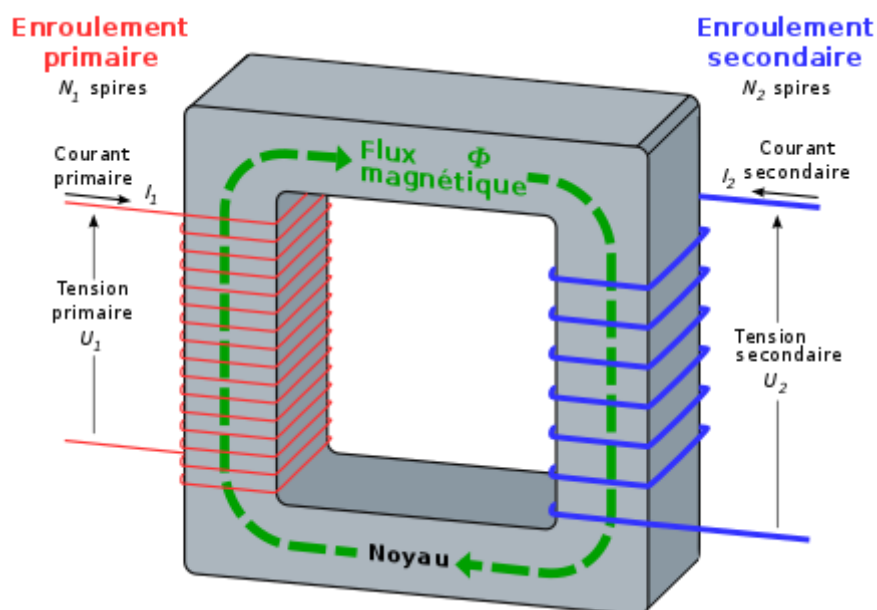
Le transformateur électrique a plusieurs fonctions : il permet d'abaisser ou augmenter la tension du courant électrique qui traverse le réseau. Dès la centrale de production d'électricité jusqu'au domicile du consommateur, le transport du courant électrique se fait par des câbles de très haute, haute, moyenne puis basse tension. Pour permettre à l'électricité de passer d'une ligne à l'autre, il est nécessaire d'abaisser sa tension. Le transformateur permet au courant électrique d'être acheminé chez le client sans aucun danger.



**FIGURE 6 : TRANSFORMATEUR TRES HAUTE TENSION,**  
SOURCE : WIKIPEDIA

#### 4.2.1 Comprendre le fonctionnement du transformateur

Le transformateur, intégré dans un poste électrique, est composé d'un noyau de fer et de deux bobines de cuivre. Le courant passe dans la bobine primaire puis dans la bobine secondaire, qui contient moins de spires, ce qui permet d'en diminuer la tension. Le transformateur monophasé est composé d'un noyau et deux bobines, tandis que le transformateur triphasé comporte 3 noyaux et 6 bobines. Comme dans l'image ci-dessous.



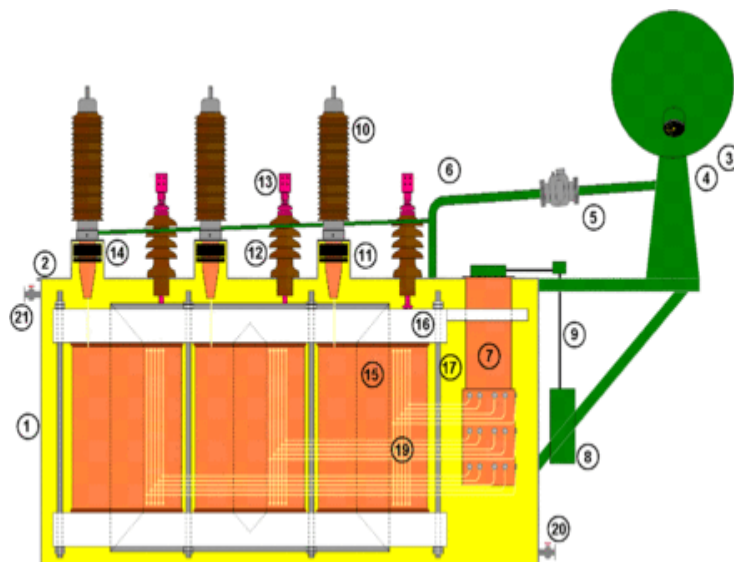
**FIGURE 7 : TRANSFORMATEUR MONOPHASE IDEAL**  
SOURCE : WIKIPEDIA

En plus de son noyau magnétique, de ses enroulements et de son isolation, un transformateur dispose de nombreux composants secondaires. Certains sont indispensables comme un

dispositif permettant la dilatation de l'huile, d'autres ne le sont pas comme le changeur de prises. Ceux présentés ici correspondent à un transformateur de puissance « classique ».

Sur le schéma ci-dessous, sont représentés :

1. Cuve
2. Couvercle
3. Conservateur
4. Indicateur de niveau d'huile
5. Relais Buchholz
6. Tuyau d'huile
7. Changeur de prises
8. Moteur électrique du changeur de prises
9. Transmission mécanique du changeur de prises
10. Traversée du primaire, avec connexion à son extrémité
11. Dôme avec transformateurs de courant à l'intérieur
12. Traversée du secondaire
13. Connexion du secondaire avec l'extérieur
14. Dôme avec transformateurs de courant à l'intérieur
15. Enroulements
16. Noyau magnétique
17. Élément mécanique maintenant le noyau magnétique et les enroulements ensemble exerçant une force de compression
18. (Non représenté)
19. Connexion du changeur de prises aux enroulements
20. Robinet d'huile
21. Robinet d'air



**FIGURE 8 : LISTE DES COMPOSANTS CONSTITUANT UN TRANSFORMATEUR DE PUISSANCE**  
**SOURCE : WIKIPEDIA**

On peut y ajouter l'huile, l'isolation papier et le système de refroidissement qui ne sont pas représentés. Des parafoudres sont présents au niveau des connexions avec les lignes.

#### 4.2.2 Système de refroidissement



FIGURE 9 : LE BANC DE RADIATEUR AVEC DES VENTILATEURS DESSOUS  
SOURCE : RTE

Les pertes des enroulements, principalement par effet Joule, mais également par courant de Foucault, oblige à les refroidir. L'évacuation de la chaleur par conduction, convection et radiation. Idéalement l'huile doit être en mouvement par rapport à tous les conducteurs afin que la chaleur soit transmise par convection depuis ces derniers vers les premiers.

Le système de refroidissement d'un transformateur a pour but de maintenir les températures des enroulements et de l'huile à des niveaux acceptables. Le choix du mode de refroidissement se fait en particulier en fonction des paramètres suivants :

- Pertes à évacuer
- Température ambiante extérieure
- Contraintes de bruit
- Contraintes dimensionnelles limitant la taille du transformateur
- Le coût.

Un système de refroidissement est caractérisé par un code de quatre lettres.

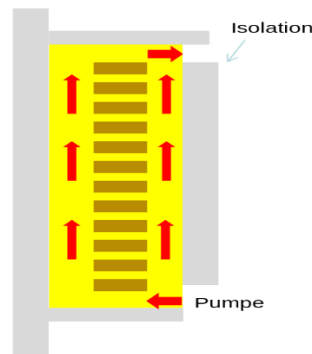
1. La première lettre désigne le fluide de refroidissement interne :

- O pour l'huile minérale (99 % des cas),
- K pour les liquides isolant ayant un point de feu  $>300\text{ }^{\circ}\text{C}$ ,
- L pour les liquides à point de feu non mesurable.

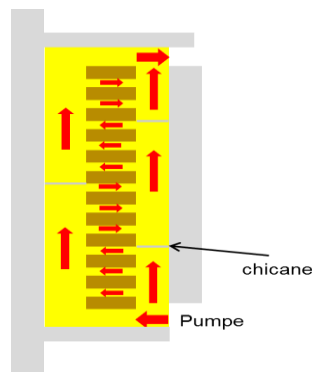
2. La seconde lettre désigne le mode de circulation du fluide de refroidissement interne :

- N pour naturel,
- F pour forcer (présence d'une pompe, mais l'huile circule librement),
- D pour forcer et dirigé (pompe et l'huile est forcée et dirigée à travers les enroulements).

Les figures ci-après montrent dans un premier temps les canaux de refroidissement forcés, où les conducteurs sont peu refroidis en leur centre. La seconde est le cas du refroidissement dirigé (OD), les chicane dirigent l'huile dans les canaux radiaux ce qui refroidit de manière plus uniforme les conducteurs



**FIGURE 10 : REFROIDISSEMENT FORCÉ (OF)**  
SOURCE: WIKIPEDIA

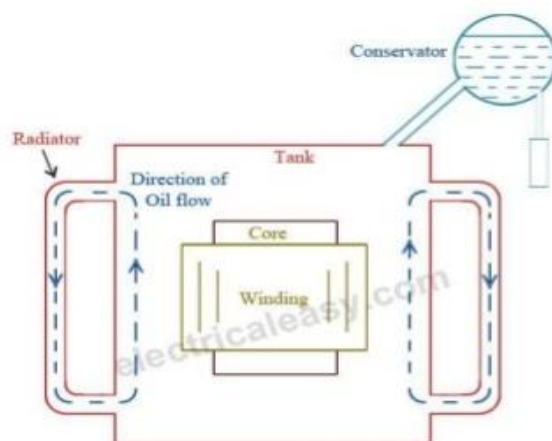


**FIGURE 11 : REFROIDISSEMENT DIRIGÉ (OD)**  
SOURCE : WIKIPEDIA

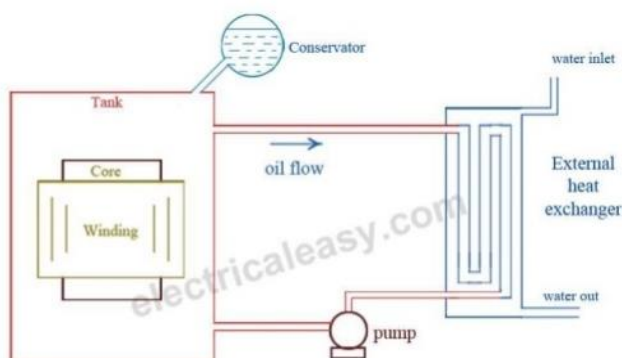
3. La troisième lettre indique le fluide de refroidissement externe :
  - A pour air,
  - W pour eau.
4. Enfin la quatrième lettre désigne le mode de circulation du fluide de refroidissement externe :
  - N pour naturel,
  - F pour forcer (ventilateurs).

Ainsi le mode de refroidissement le plus simple est l'ONAN. Les figures suivantes sont des illustrations de deux exemples de fonctionnement ONAN et ONAF.





**FIGURE 12 : TRANSFORMATEUR ONAN SOURCE**  
**SITE : ELEC13 WORDPRESS**



**FIGURE 13 : TRANSFORMATEUR ONFWF**  
**SOURCE : ELEC13 WORDPRESS**

Concrètement l'huile colporte la chaleur vers l'extérieur du transformateur et est refroidie à l'intérieur du radiateur, qui évacue la chaleur par convection. Des ventilateurs permettent d'augmenter cette dernière (voir **Figure 9**).

Les transformateurs utilisés à RTE sont de type immergé dans l'huile et de technique ODAF sur le réseau de transport, et c'est encore plus varié pour le réseau de distribution disposé par Enedis qui sont de type ODAF, OFAF, ONAF ou ONAN.

### 4.3 Encrassement du système aéroréfrigérants

En raison de leur conception, les aéroréfrigérants s'encrassent avec les polluants présents dans l'air (poussières, pollens, pollution, etc.) qui peuvent se coller si présence d'huile. A partir d'un certain seuil, ces polluants empêchent la bonne circulation de l'air et donc le refroidissement de l'huile. Le dégrassage périodique ou palliatif semble être la pratique la plus répandue.

Un encrassement important du transformateur se caractérise par :

- Une élévation de la température d'huile.
- Une incapacité à supporter les surcharges.

L'élévation de la température de fonctionnement du transformateur conduit à la réduction de la durée de vie de ce dernier, par une usure prématurée des isolants celluloseux des câbles constituant les bobines. Sous l'action d'une température de l'ordre de 120°C et d'une forte humidité, des bulles peuvent se former conduisant à un défaut diélectrique.

L'objectif du suivi en continu de l'encrassement des aéroréfrigérants est d'agir avant que les signes d'un encrassement important n'apparaissent.

L'encrassement, dans son expression la plus large, a les conséquences suivantes :

- Investissements initiaux supplémentaires ;
- Coût énergétiques supplémentaires et impact environnemental induit ;
- Pertes de production pendant les périodes d'arrêt pour nettoyage ;
- Coûts de maintenance.

#### 4.3.1 Intérêt du suivi de l'encrassement

Le suivi en continu du niveau d'encrassement des aéroréfrigérants doit permettre :

- D'augmenter la durée de vie : un fonctionnement du transformateur à une température trop élevée conduit à une accélération de la dégradation des isolants celluloseux de la partie active. De même, limiter les changements de stade permet de diminuer le nombre de démarrage des moto-ventilateurs et de prolonger leur durée de vie.
- De préserver les capacités de surcharge de l'appareil : en fonction des besoins d'exploitation du réseau de distribution, des reports de charge peuvent être effectués entre les transformateurs. Un entretien insuffisant des aéroréfrigérants peut ne pas avoir de conséquence thermique en régime normal mais, en régime exceptionnel, il pourrait être la cause de défauts en cascade.
- D'éviter des arrêts d'exploitation : un encrassement faible, s'il n'est pas traité à temps, peut dégénérer en bouchage important de l'aéroréfrigérants. Le traitement de ce dernier peut nécessiter une consignation du transformateur, préjudiciable à l'exploitation.
- De passer d'une maintenance périodique ou palliative à une maintenance prévisionnelle : le suivi de l'encrassement doit permettre de n'effectuer les opérations de nettoyage qu'au besoin. En effet, un entretien périodique n'est pas optimal car les conditions environnantes varient au cours de l'année et sont différentes pour chacun des transformateurs.

#### 4.3.2 Modes d'encrassement

Les transformateurs de type immergé dans l'huile et de technique ODAF (Circulation d'huile forcée et dirigée, avec système de refroidissement forcé, externe, à l'air). L'huile assure l'extinction des éventuels arcs électriques internes au transformateur ainsi que l'isolement entre les parties actives et la cuve. Ici on va présenter une analyse des éléments pouvant conduire à son encrassement en mode réversible (dans ce cas des composants peuvent être corrigés par des opérations simples de nettoyage).



- Filtres (ou grilles) : ils permettent d'intercepter un maximum de poussières et d'éviter le colmatage des parties intérieures de l'aéroréfrigérant. Ils sont démontables pour permettre leur nettoyage.



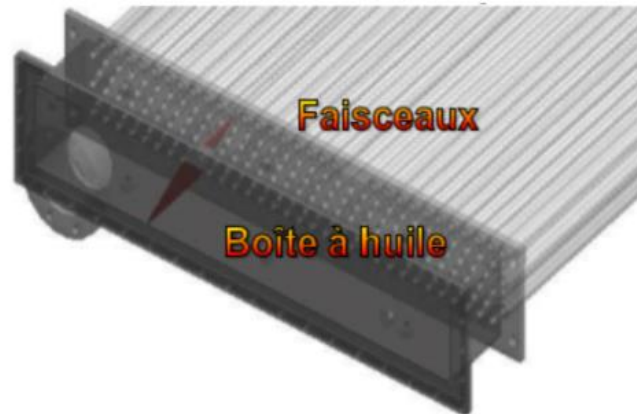
**FIGURE 14 : FILTRES DES AÉRORÉFRIGÉRANTS**  
SOURCE : (REMI DESQUIENS, EDF 2018)

Une appréciation visuelle du niveau d'encrassement des filtres peut être faite. Le niveau 3 dans la figure d'après doit déclencher un nettoyage.



**FIGURE 15 : ENCRASSEMENT DES FILTRES**  
SOURCE : (REMI DESQUIENS, EDF 2018)

- Faisceaux : ils constituent la surface d'échange entre l'huile du transformateur et l'air environnant.



**FIGURE 16 : FAISCEAUX DES AÉRORÉFRIGÉRANTS**  
SOURCE (REMI DESQUIENS, EDF 2018)

La poussière non captée par le filtre vient obstruer les faisceaux de l'échangeur réduisant les capacités thermiques de l'aéroréfrigérant. Visuellement, le niveau 3 doit déclencher un nettoyage.



**FIGURE 17 : ENCRASSEMENT DES FAISCEAUX**  
(REMI DESQUIENS, EDF 2018)

- Dégradation de l'huile : le vieillissement de l'huile du transformateur conduit à la formation de particules qui peuvent se déposer sur la surface d'échange (à l'intérieur des faisceaux). Un filtrage ou un renouvellement de celle-ci peut alors être fait. Les huiles actuelles ont permis de réduire le risque de formation de boues dans le transformateur.

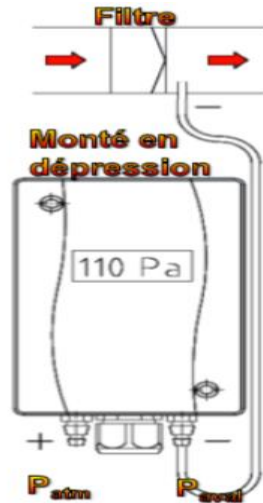
#### 4.3.3 Méthodes de suivi de l'encrassement

Différents indicateurs peuvent être mis en place pour suivre en temps réel l'encrassement des aéroréfrigérants, on les classe en 2 catégories :

- Méthodes directes : Les mesurent une grandeur directement liée à la présence d'un dépôt.
  - Suivi du poids des filtres ;
  - Suivi de la différence de pression entre l'entrée et la sortie du filtre ;
  - Mesure de la section de passage entre les mailles du filtre ou entre les ailettes.
- Méthodes indirectes : elles consistent à suivre une grandeur affectée par l'encrassement de l'échangeur.
  - Suivi du débit d'air ;
  - Suivi de la résistance thermique de l'aéroréfrigérants ;
  - Suivi de la température d'huile du transformateur avec l'utilisation d'un modèle d'échauffement.

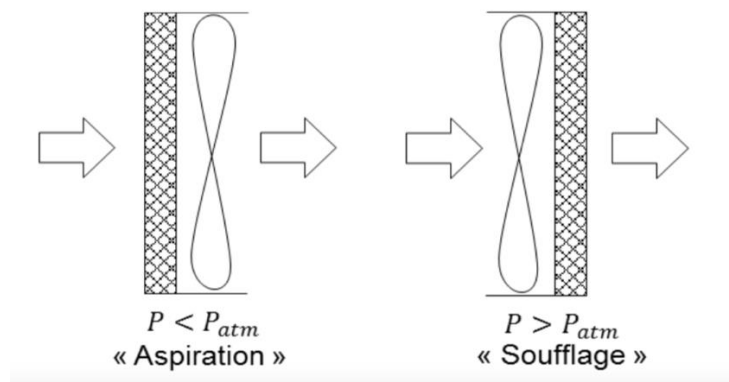
#### 4.3.3.1 Différence de pression

Dans ce projet il s'agit de conduire une étude basée sur la différence de pression entre l'entrée et la sortie du filtre. Les unités de mesure de pression sont placées entre e l'air ambiant et la partie intérieure de l'aéroréfrigérant. L'image ci-dessous montre une méthode de suivi de l'encrassement.



**FIGURE 18 : METHODE DE SUIVI DE L'ENCRASSEMENT**  
SOURCE : (REMI DESQUIENS, EDF 2018)

Dans la majorité des cas, les aéroréfrigérants sont en configuration « aspiration » (voir la figure ci-après) mais, pour certaines configurations en bâtiment, les aéroréfrigérants peuvent être en configuration « soufflage ».



**FIGURE 19 : DEUX CONFIGURATIONS DE L'AÉRORÉFRIGÉRANT**  
SOURCE : (REMI DESQUIENS, EDF 2018)

Le suivi en temps réel de la pression différentielle du filtre de l'aéroréfrigérant,  $\Delta P = |P - P_{atm}|$ , permet de voir 2 phénomènes :

- L'encrassement du filtre à un stade fixe
  - « Aspiration » :  $P \searrow$  d'où  $\Delta P = P_{atm} - P \nearrow$ .
  - « Soufflage » :  $P \searrow$  d'où  $\Delta P = P - P_{atm} \searrow$ .

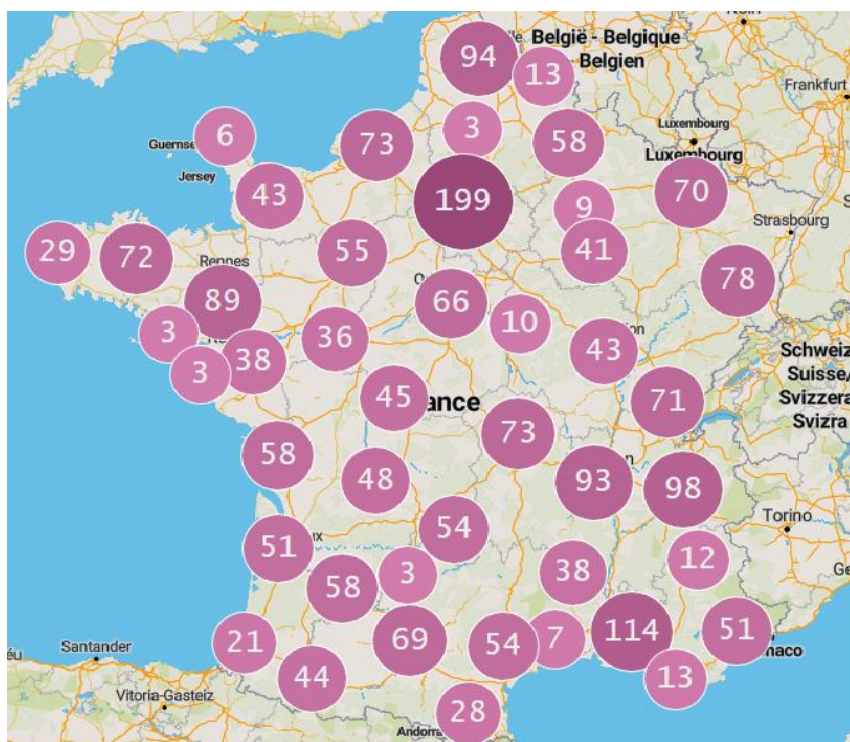
➤ Le changement de stade :

- « Aspiration » :  $P \searrow$  d'où  $\Delta P = P - P_{atm} \nearrow$ .
- « Soufflage » :  $P \nearrow$  d'où  $\Delta P = P - P_{atm} \nearrow$ .

## 4.4 Répertoirer les données

### 4.4.1 Poste source

On s'intéresse dans ce projet aux transformateurs appartenant aux postes sources gérés par Enedis. Les coordonnées de ces postes sont disposées sur le site d'Enedis Open Data. La figure suivante montre la répartition de ces postes en France.



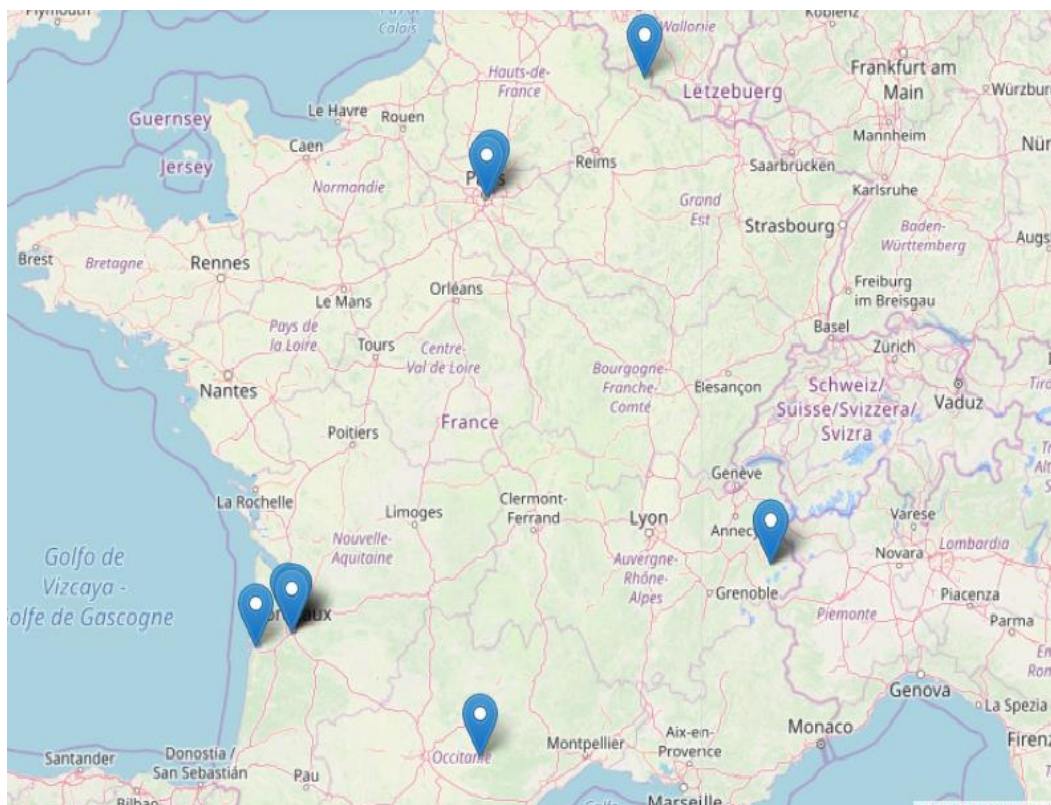
**FIGURE 20 : POSTE SOURCE EN FRANCE**  
**SOURCE SITE ENEDIS**

### 4.4.2 Mesure d'encrassement

Enedis dispose d'un système de monitoring appelé AGATH constitué d'une centrale d'acquisition et de différents capteurs permettant le suivi en temps réel du transformateur. A partir de ces mesures, Enedis souhaite que soient mis en place des indicateurs permettant une optimisation de la maintenance et des conditions d'exploitation.

Les mesures dont nous disposons pour ces études sont réduites aux postes montrés sur la carte suivante



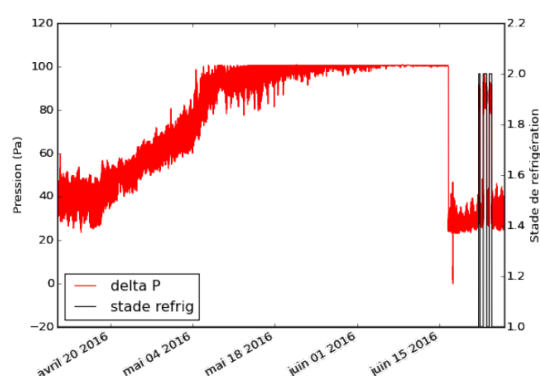


Pour chacun de ces postes sources, les données disponibles contiennent des mesures par seconde prises sur un de ses transformateurs sur une période donnée. Parmi ces mesures on trouve : la température en haut de la cuve, la température de l'air, la résistance thermique et la pression différentielle entre l'air ambiant et la partie intérieure de l'aéroréfrigérants.

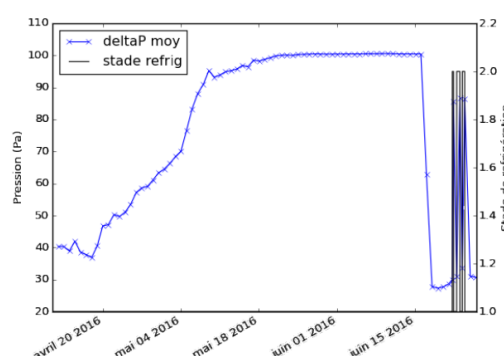
Le suivi de la différence de pression est une méthode de suivi des capacités d'échange thermique de l'échangeur. Son augmentation est caractéristique d'un encrassement.

Prenons l'exemple des relevés effectués sur le poste de Paillère entre le 11 avril 2016 et le 26 juin 2016.

#### Valeurs instantanées :



#### Valeurs moyennées par stade et par périodes de 24 h :



**FIGURE 21 : RELEVES DE LA DIFFERENCE DE PRESSION SUR LE POSTE DE PAILLER**

On note une augmentation continue de la différence de pression jusqu'au 16 juin 2016, ce qui correspond à une phase d'encrassement de l'aéroréfrigérants. Un nettoyage du filtre a alors dû être réalisé puisqu'on observe une réduction importante de la différence de pression.

Les valeurs instantanées de différence de pression peuvent fluctuer de manière importante (instrumentation, conditions météorologiques...). C'est pourquoi on réalise un traitement par moyenne par stade et par période de 24 heures.

#### 4.4.3 Données de pollinisation

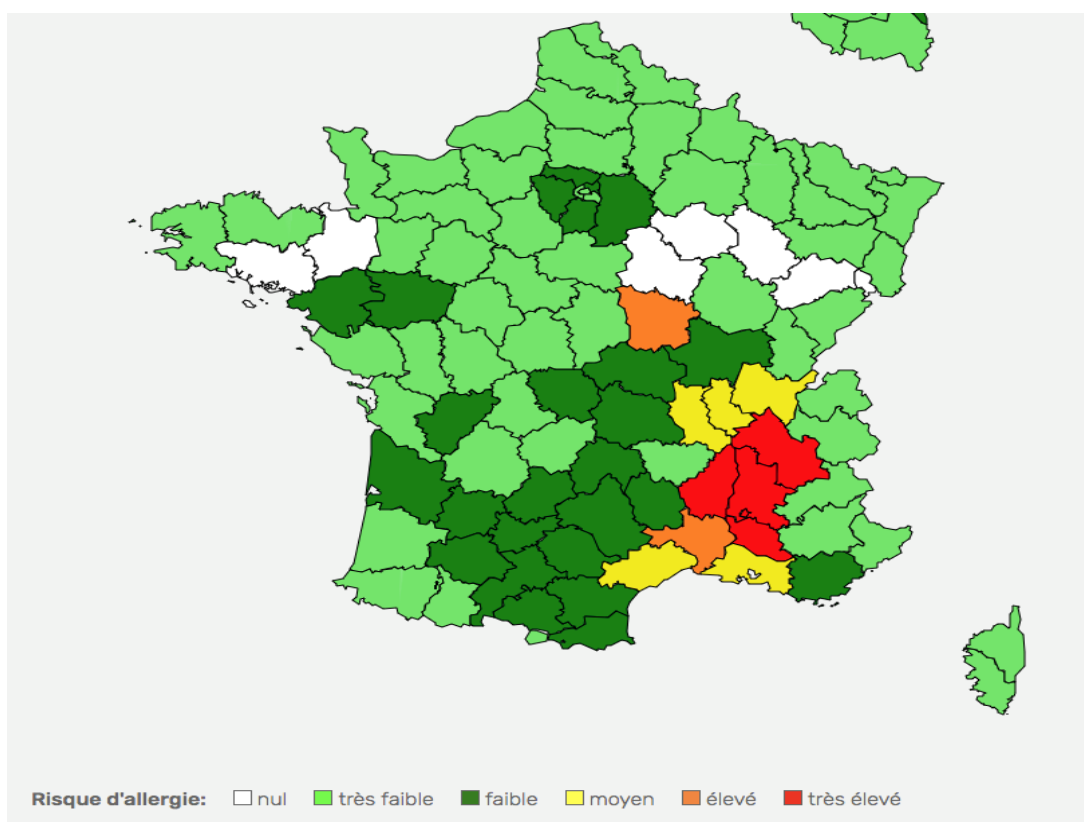
Les pollens sont des grains microscopiques (entre 20 et 30 microns), libérés par les plantes à fleurs, qui vont leur permettre de se reproduire. La variété des pollens est un indice de la biodiversité.

Les végétaux sont fixés au sol. Pour permettre le transport de leurs pollens, ils possèdent des stratégies de reproduction. Il existe deux principaux modes de fécondation croisée chez les plantes à fleurs, certaines utilisent le vent (plantes anémophiles), d'autres les insectes (plantes entomophiles).

##### 4.4.3.1 Réseau nationale de surveillance agrobiologique

Le réseau national de surveillance agrobiologique (R.N.S.A) est une association créée en 1985 pour poursuivre les travaux réalisés depuis 1985 par le laboratoire de l'aérobiologie à l'institut Pasteur de Paris.

Le site internet de l'RNSA fournit des bulletins allegro-pollinique, bulletin de moisissures, et bulletin phréologique. Ci-après une prévision de la pollinisation pour la France métropolitaine en mois de septembre 2019.



**FIGURE 22 : CARTE DE VIGILANCE SEPTEMBRE 2019**  
SOURCE : SITE RNSA

Le site internet dispose aussi d'une base de données ouverte. Les données exprimées dans les fichiers disposés par le site sont les concentrations journalières en nombre de grains par mètre cube d'air pour les pollens et en nombre de spores par mètre cube pour les moisissures. Ces données sont à usage privé ou scientifique, ne pouvant être utilisées pour des fins commerciales.

#### 4.4.3.2 Les capteurs de pollens et moisissures atmosphériques

Le principe de fonctionnement du RNSA repose en premier lieu sur la constitution des sites de captures qui sont des sites de recueil de fond. Chaque site a été choisi par rapport à des critères climatiques, botaniques et de densité de populations. Chaque site est constitué d'un capteur, d'un responsable du capteur, d'un analyste et d'un médecin responsable clinique du site.

Pour une identification de pollinisation optimale autour des transformateurs, on s'intéresse à répertorier les positions de ces capteurs de pollens. On trouve 69 capteurs en 2019.

Une cartographie des données rassemblées nous donne la carte suivante, qui montre la répartition de ces capteurs sur la France métropole.



**FIGURE 23 : POSITION DES CAPTEUR DE POLLINISATION**  
SOURCE REALISE SOUS PYTHON

#### 4.4.3.3 Les données pollinisation

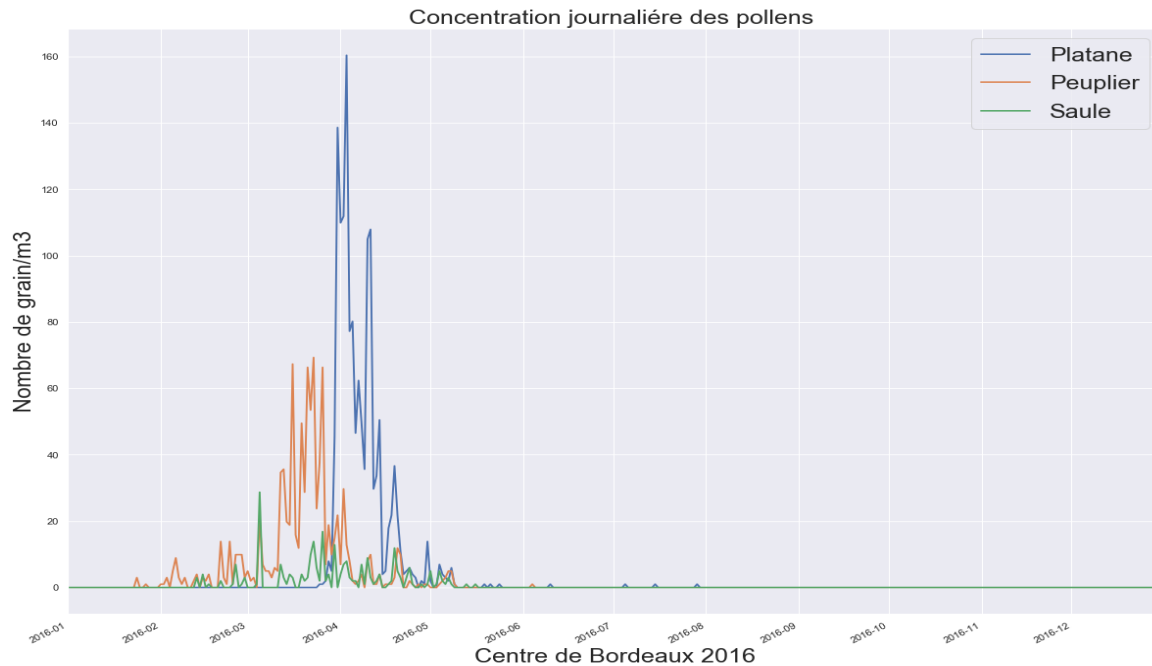
Les données sur les concentrations journalières en nombre de grains par m<sup>3</sup> d'air pour les pollens, et en nombre de spores par mètre cube d'air pour les moisissures, sont disponibles dans la rubrique base de données du site de l'RNSA. Pour des données complémentaires, notamment celles des dernières années, elles ne sont pas disponibles sur le site, il fallait s'approcher auprès de l'RNSA afin de les demander.

Dans le tableau suivant on trouve les différentes espèces de pollens et de moisissure, dont on a pu avoir les données depuis 2008 jusqu'à 2019.

Pollen	Moississures
Aulne	Taxa
Ambroisie	Alternaria
Armoise	Cladosporium
Bouleau	
Charme	
Châtaignier	
Noisetier	
Cupressacées (cyprès, thuya, genévrier)	
Hêtre	
Frêne	
Poacées (graminées fourragères et céréalières)	
Mûrier	
Olivier	
Plantain	
Platane	
Peuplier	
Chêne	
Oseille	
Saule	
Tilleul	
Urticacées (pariétaire, ortie)	

Prenons à titre d'exemple l'évolution des pollens (Platane, Peuplier et Saul) à Bordeaux en 2016.

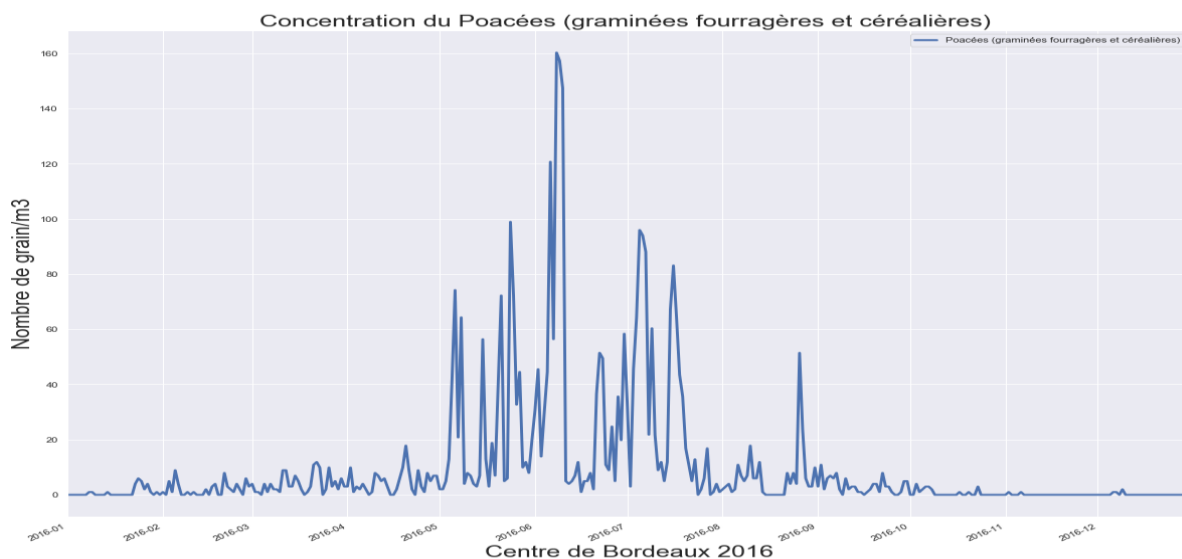




**FIGURE 24 : LA VARIATION DU TAUX DU POLLEN (PLATANE, PEUPLIER, SAULE)  
ANNEE 2016**

On note une augmentation dans l'air des Pollens dans la période entre fin février et début juin. Le pic pour les platanes étant en fin mars et début avril pour les peupliers. C'est la période qui précède la saison d'encrassement pour le transformateur de Pailler qui appartient à la zone du capteur des pollens à Bordeaux.

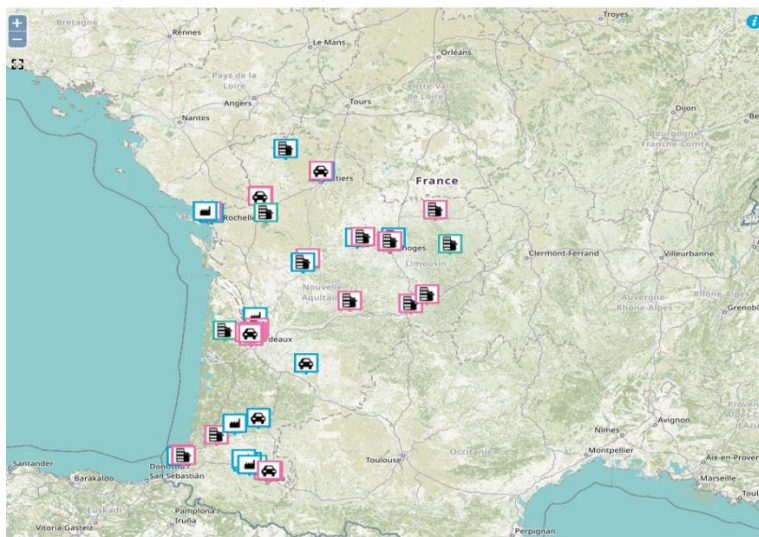
La figure ci-dessous représente la variation du nombre de grain des graminée Fourragères et céréalières sur un mètre cube d'air en 2016 à Bordeaux. On note une augmentation importante dans la période entre début mai et fin juillet. Cette période couvre celle d'augmentation du niveau d'encrassement du transformateur électrique à Pailler. Donc, si des champs agricoles font partie de l'environnement du transformateur, on peut faire l'hypothèse qu'en plus des pollens, la graminée peut se révéler comme une raison d'encrassement du système de refroidissement.



**FIGURE 25 : VARIATION DES GRAMINEES FOURRAGERES ET CEREALIERES**

#### 4.4.3.4 Les données pollution

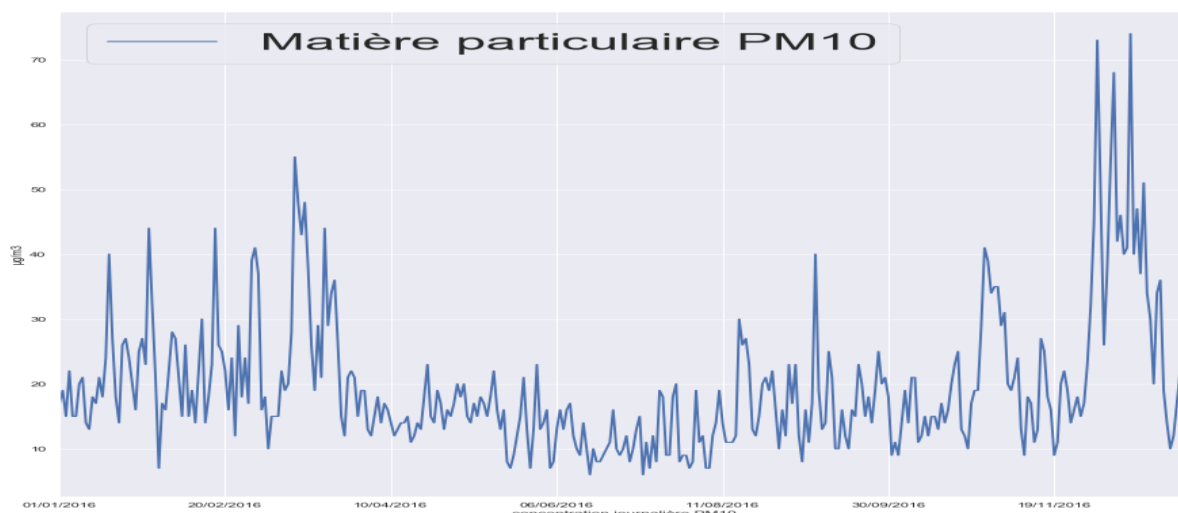
L'encrassement causé par le dépôt de polluants se présente très souvent par noircissement. Ce phénomène particulièrement observé dans les grandes villes et zones où le trafic est important. En France il existe plusieurs organismes agréés par le ministère de l'Environnement pour la surveillance de la qualité de l'air dans les différentes régions, comme Airparif pour la région ile de France, ATMO-Nouvelle aquitaine, etc.



**FIGURE 26 : LES STATION DE MESURE DE LA QUALITE DE L'AIRE DANS LA REGION NOUVELLE AQUITAINE**

Comme pour les autres régions, le site ATMO- Nouvelle-Aquitaine met à disposition en temps réel, les données de qualité d'air issues de son réseau de mesures, ainsi que les différentes mesures par polluant et station durant les 10 dernières années.

La figure ci-après représente la concertation journalière du PM10 durant l'année 2016.



**FIGURE 27 : NIVEAU DE PM10 MESURE PAR LA STATION DE BORDEAUX EN 2016**

PM10 est la matière particulaire de diamètre inférieur à 10 µm qui a tendance à se déposer sur les matériaux (Varney-Caronn s.d.). En ville la concentration de cette particule dépasse très souvent la norme annuelle moyenne de 20 µm par mètre cube d'air ambiant dans presque tous

les environnements suburbain (défini par une distance par rapport aux larges sources de pollution de 3 à 10 kilomètres), urbain de fond (Urbain de fond défini moins de 2500 véhicules par jour dans un rayon de 50 mètres), et en bordure de territoire (signifie à proximité des voies de trafic).

## 5 Deep Learning pour segmentation sémantique d'images

Dans le chapitre précédent nous avons effectué une analyse du comportement de certains facteurs environnementaux, tels que la pollution, les pollens et les graminées. Cette analyse a mis en évidence la possible corrélation entre ces facteurs et l'encrassement des systèmes de refroidissement. La connaissance de la couverture du sol entourant les postes sources est dans ce cas un enjeu majeur pour le suivi d'encrassement. Dans ce contexte, les cartes d'Occupation du Sol (OCS, Land Cover en anglais) caractérisant la couverture biophysique des terres émergées, sont un puissant outil scientifique et décisionnel, notamment dans les travaux de recherche. Dans notre cas, nous allons nous intéresser à la variable qui va permettre de déterminer les zones boisées, urbaines et agricoles, afin de classer les transformateurs selon leur risque d'encrassement et de réaliser les opérations de nettoyage au meilleur moment, en d'autres termes définir les cycles de maintenance pour chaque transformateur en fonction de son environnement géographique.

### 5.1 État de l'art

Dans le domaine de la télédétection sur des images satellites ou aériennes, il y a deux grandes familles de méthodes pour procéder à une segmentation :

- Les méthodes denses (initiées par Long, J., Shelhamer, E. et Darrell, T., 2015.) utilisant les architectures de réseau de neurones type « encodeur-décodeur » telles que le réseau Alex net.
- Les méthodes « par Patch » (décrites dans Grishick et al., 2014), qui consistent à faire glisser un patch d'une dimension bien choisie sur une image. Ce patch jouera le rôle d'un filtre et on attribuera à chaque pixel sur lequel le patch est centré une probabilité d'appartenance à l'une des classes prédéfinies.

Pour le calcul d'une carte d'occupation des sols à partir d'une image aérienne avec un nombre de classes réduit (bâtiment, végétation, champs, routes, eau), des travaux (introduits par Postijan, T., Le Bris, A., Sahbi, H., Mallet, C., Univ. Paris Est, LASTIG MATIS, IGN, CNRS, LIP6 UPMC) utilisent l'approche « par Patch » sur des réseaux de neurones convolutifs pour détecter une carte d'occupation sur des images satellites SPOT à 4 bandes.

#### *Deux architectures testées pour mettre en évidence l'impact du nombre de couches*

Entre les réseaux de neurones à 3 et 4 couches, les résultats ont montré que le réseau à 4 couches est plus performant, c'est donc le seul réseau considéré dans la suite des travaux. Chaque couche contient un filtre de convolution ( $2 \times 2$ ) suivi d'un filtre d'activation Relu et un filtre de maxpooling ( $2 \times 2$ ) qui sépare deux couches consécutives. À la fin de chacune des structures se trouve un filtre de fully connected qui permet de combiner les filtres issus de la dernière couche pour avoir un attribut pour le patch d'entrée, celui-là choisit de dimensions  $65 \times 65 \times 65 \times 46$  (4 étant le nombre de bandes de l'image R, V, B et IR) afin de prendre en compte des petits objets dans l'image.

#### *La création du jeu de données d'entraînement :*

Le jeu de données d'entraînement utilisé est extrait des bases de données géographiques nationales (IGN), constituées d'images géoréférencées appartenant aux classes d'OCS que l'on cherche à classer.

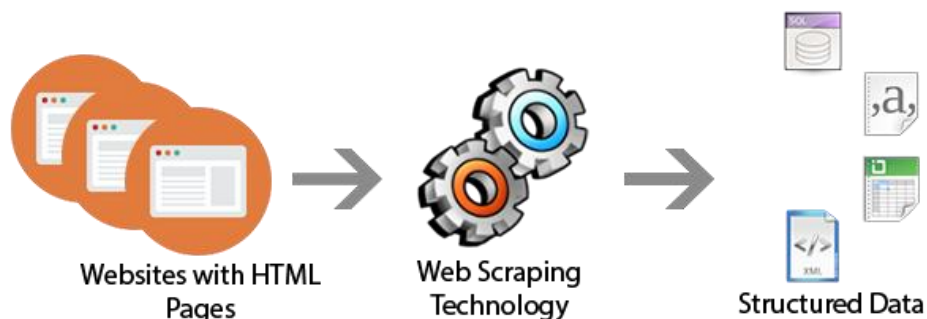
#### *Deux stratégies d'apprentissage du CNN*

Deux stratégies (complémentaires) ont été considérées lors de la construction du réseau :

- Initialisation aléatoire du modèle (RWI), pour laquelle les paramètres des CNNs (en anglais CNN ou ConvNet pour Convolutional Neural Networks) initialisés aléatoirement. Dans ce cas le modèle n'étant pas pré-entraîné à l'avance sur des images à 4 bandes, la convergence était très lente (1-2 jours), et par ailleurs se faisait ressentir un besoin d'augmenter l'échantillon d'apprentissage.
- Fine tune, pour laquelle on garde la même architecture que dans le cas du modèle RWI, mais ici le modèle est réutilisé avec les paramètres récupérés à la fin d'entraînement du réseau précède auparavant entraînée et ayant convergé sur un autre jeu de données qui représente d'autres caractéristiques absentes dans le nouveau jeu de données. Cela va permettre :
  - Une convergence très rapide
  - Un besoin plus faible en échantillons
  - Une adaptation de domaine (bonne généralisation)

## 5.2 Jeux de données

Chaque poste source est géolocalisé et ses coordonnées GPS sont disponibles dans la rubrique OPEN Data du site d'Enedis. Afin de pouvoir définir des cartes d'occupation du sol par méthode de réseau de neurones, on utilisera des images qu'on va extraire de Google Maps. Les postes source gérés par Enedis sont au nombre de 2235, ce qui nous oblige à extraire ces données par une méthode automatique de « Scraping ». Scraper des données, c'est « gratter » des pages web pour stocker l'information voulue. Cela revient à copier-coller du contenu d'une page web de manière à pouvoir récupérer les données dont on a besoin depuis un site internet (voir la figure ci-après). La technologie utilisée dans notre cas est Python.



**FIGURE 28 : SCRAPING WEB**  
SOURCE: [HTTP://SAYONETECH.COM](http://sayonetech.com)

Le Scraping de données type images sur Google MAP permet de récupérer des images à trois bandes (RGB) centrées sur le point de coordonnées (latitude, longitude) avec un niveau de zoom égale à 17 (le zoom étant compris entre 1 et 18, le niveau 18 correspond à la carte du monde entier).

La distance  $S$  représentée par pixel est donnée par  $S = C \frac{\cos(y)}{2^{z+8}}$ , où

- $C$ 'est la circonférence de la Terre à l'Équateur
- $z$  est le niveau de zoom
- $y$  est la latitude

Ci-dessous un échantillon des images restituées centrées sur des postes sources.



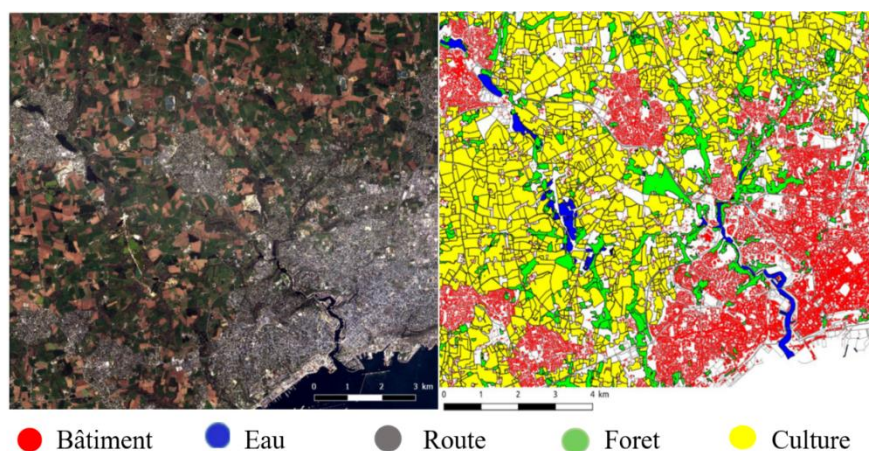


**TABLEAU 1 : IMAGES SATELLITES CENTREE SUR DES POSTES SOURCE DANS DIFFERENTES TYPE ENVIRONNEMENT**

Les postes sources ont des environnements très diversifiés les uns des autres. On note par exemple : forêt, urbain, agricole ... ou très souvent un mélange de plusieurs classes (voir la ligne deux du tableau).

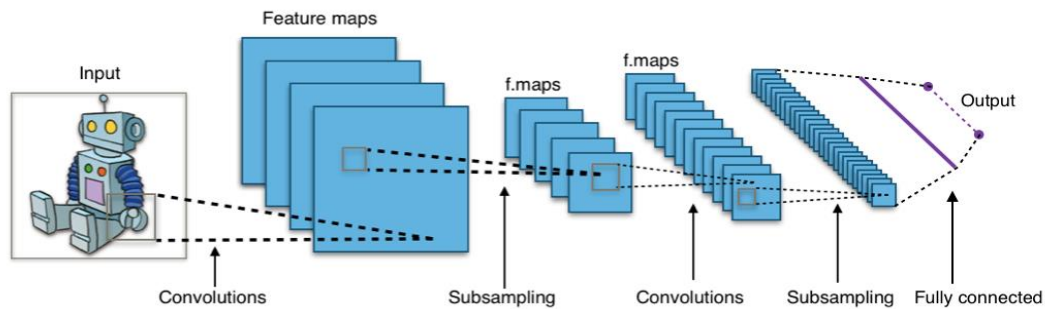
### 5.3 Approche CNN pour la segmentation sémantique

La classification d'une image acquise par télédétection géospatiale (aérienne ou satellite), dans un contexte de cartographie d'occupation des sols (OCS), consiste à attribuer à chaque pixel de cette image une classe. L'image ci-dessous est une synthèse de notre objectif.



**FIGURE 29 : IMAGE SATELLITE (A DROITE), LA CARTE D'OCCUPATION CORRESPONDANTE (A DROITE) - SOURCE (POSTADJIAN, ET AL. S.D.)**

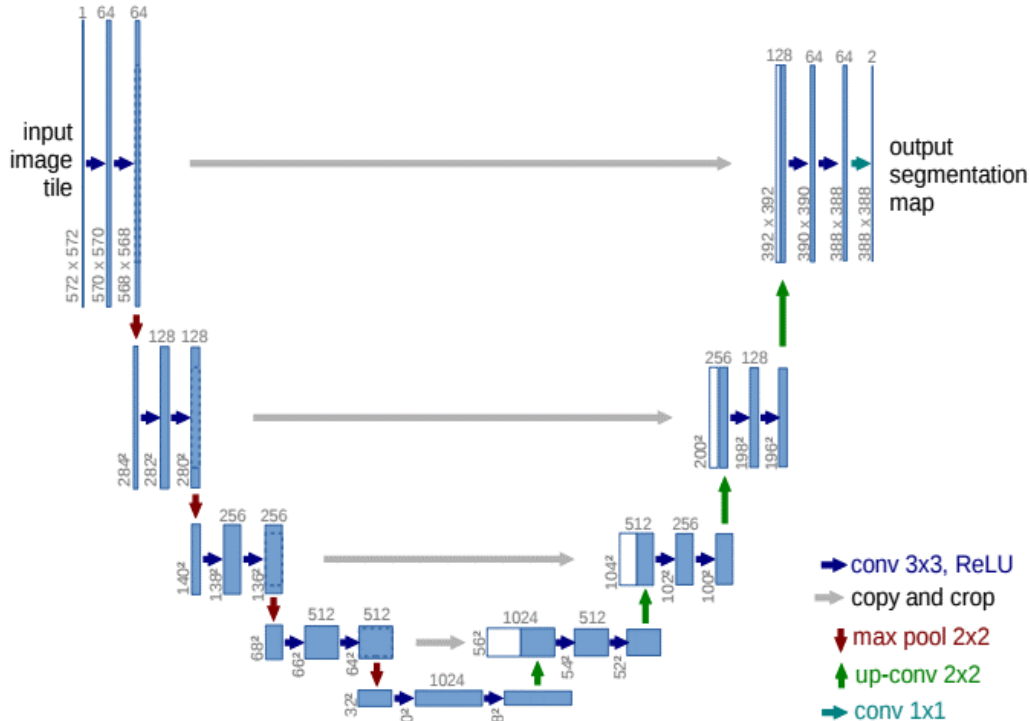
On utilisera dans ce projet les techniques d'apprentissage profond (Deep learning), et plus particulièrement les réseaux de neurones convolutifs (CNNs pour Convolutional Neural Networks). Ci-dessous une architecture standard d'un réseau de neurones convolutifs.



**FIGURE 30 : ARCHITECTURE STANDARD D'UN RESEAU A CONVOLUTIONS**  
SOURCE : WIKIPEDIA

Cette méthode nous affranchit du travail fastidieux de l'extraction des features en amont de la classification. Comme dans ce cas, le réseau de neurones s'occupe de la tâche fastidieuse d'extraction de features requérant des connaissances expertes, subjectives, et qui varient selon les tâches à effectuer.

On s'intéresse particulièrement au réseau de neurone UNET, introduit pour la première fois dans (Ronneberger, Fischer et Brox s.d.) pour la segmentation des images biomédicales, décrit comme suit (Figure ci-après) :



**FIGURE 31 : ARCHITECTURE RESEAUX U-NET**

L'architecture « Unet » utilisée dans ce projet comprend 10 couches (la capacité de l'environnement de calcul DIL nous le permet).

L'architecture ressemble à un «U» qui justifie son nom. Cette architecture comprend trois sections : la **contraction**, le **goulet d'étranglement** et la section d'**expansion**.

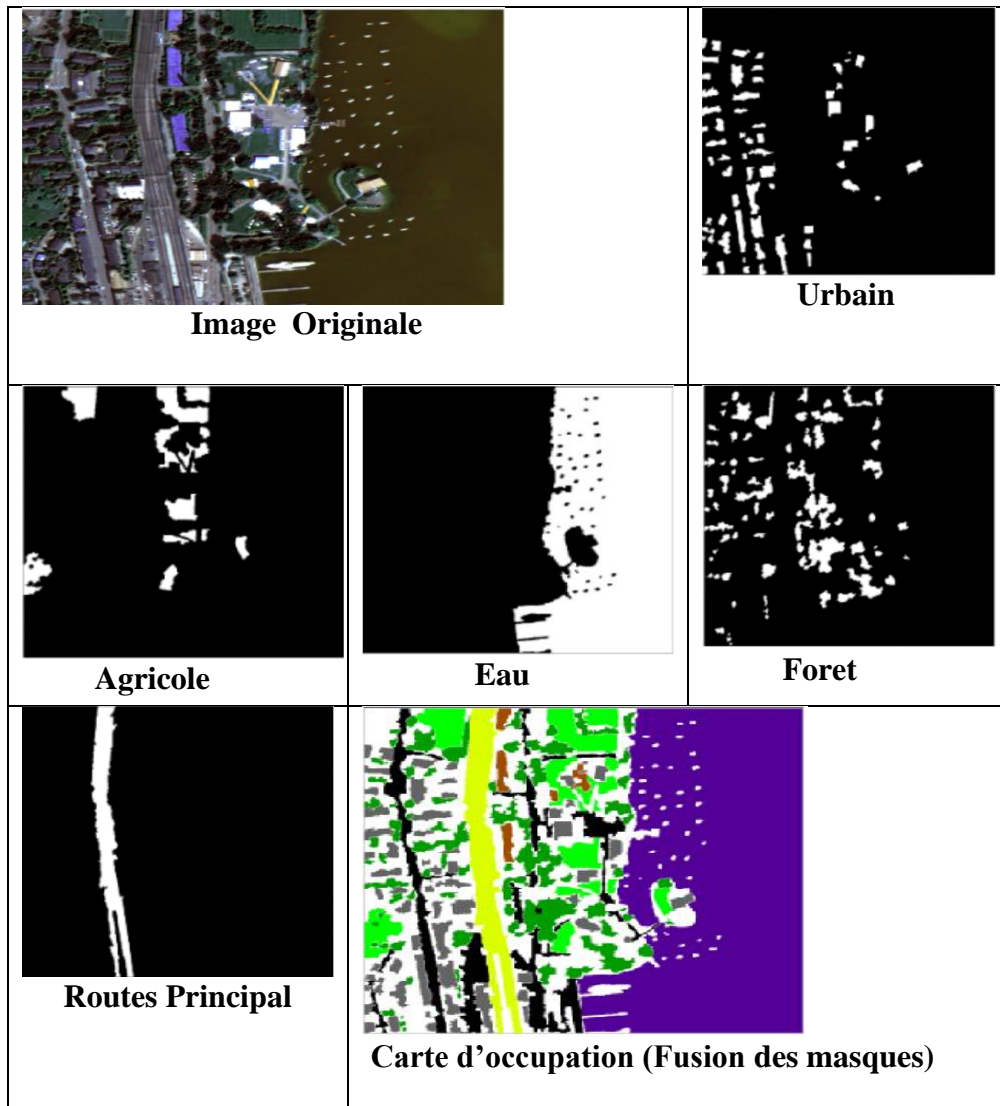
La partie qui rend intéressante cette structure de réseau de neurones est la partie d'**expansion**. Semblable à la couche de **contraction**, elle se compose également de plusieurs blocs d'expansion. Chaque bloc transmet l'entrée à deux couches de convolution de dimension  $(3 \times 3)$  suivies d'une couche de max-pooling de  $(2 \times 2)$ . De plus, après chaque bloc, le nombre de map-of-features utilisé par la couche de convolution est réduit à la moitié afin de maintenir la symétrie. Cependant, à chaque fois que l'entrée est également ajoutée, elle est accompagnée d'une map-of-features de la couche de contraction correspondante. Cette action garantit que les fonctionnalités apprises lors de la contraction de l'image seront utilisées pour la reconstruire. Après cela, les features-maps résultants traversent une autre couche  $(3 \times 3)$  de réseaux de neurones convolutifs avec un nombre de features-maps égal au nombre de classes souhaitées.

#### 5.4 Base de données d'entraînement

Nos données référentes sont prises de Kaggle Data-Set. Kaggle est une plateforme web organisant des compétitions où les entreprises proposent des problèmes en data science et offrent un prix aux meilleures performances. Ces données correspondent à 24 locations sur des environnements différents. Chaque location correspond à une image sur 8 bandes spectrales (rouge, bords rouge, côtier, bleu, vert, jaune, infrarouge à ondes courtes, infrarouge à ondes courtes  $> 2 \mu\text{m}$ ), ainsi que 10 masques qui correspondent aux 10 classes (forêt, route, eau, piscine, voie ferrée, non classifié, ...).

Or dans notre base de données les images que nous devons prédire sont sur 3 bandes (RGB), ce qui n'est pas le cas pour les images de référence qui sont sur 8 bandes. Nous allons donc procéder à une extraction de bandes dans l'ordre R G B à partir des 8 bandes. On fera la même chose pour les masques, car pour chaque position, les 10 masques dans le fichier du concours correspondent à une image mask.TIF qui n'est rien d'autre que la superposition des 10 images binaires. Donc de la même manière, nous allons transformer ces images mask.TIF à 10 superpositions de masques en des images mask.TIF qui ne contiennent cette fois-ci que les 5 classes qui nous intéressent. Tel que par exemple : `mask.TIF[:, :, 0]=masque_foret`, `mask.TIF[:, :, 1]=masque_agricole`. La figure ci-après montre l'image originale sur 3 bandes, ainsi que les masques qui nous intéressent.





**FIGURE 32 : DONNEE REFERENCE POUR ENTRAINEMENT**

### 5.5 Mise en œuvre

Afin de mettre en œuvre une méthode d'apprentissage pour notre réseau de neurone, pour chaque localisation on va procéder comme tel,

- 75% des images et masques seront utilisés pour l'entraînement
- 25% seront utilisés dans la phase de validation.

Les fichiers les plus importants pour la méthode sont :

- *Generation\_patch* : Prend en argument une image, le masque correspondant et la taille du patch « patch\_size ». Retourne la liste des patches générés par l'image et la liste des patches générés par le masque. Chaque élément de la première liste est de taille (patch size × patch\_size×3) et de (patch size × patch\_size×5) pour la liste des masques générés à partir de la même image.
- *Gen\_rand\_patch* : On va l'utiliser pour essayer de résoudre le problème du sur-apprentissage. Elle prend les mêmes arguments que *Generation\_patch* et retourne différentes transformations aléatoires, géométriques sur les images et les masques.
- *U\_net* qui contient l'implémentation du modèle U\_net.

- Le fichier *main* où on entraîne le modèle U\_net sur les patchs générés à partir des données d'entraînement prérépérées (Adaptation aux types de nos données, et au nombre de classes qui nous intéressent). On fixe le nombre d'époques d'entraînement à 200.
- Fonction de prédiction qui prend en argument le modèle entraîné, la taille du patch et l'image à prédire, et retourne une image décodée selon un code couleur choisit

## 5.6 Résolution de l'overfitting

Une visualisation de la fonction Loss durant l'entraînement (voir la figure ci-après) permet de dire que le modèle apprend trop bien mais trop vite sur les données d'entraînement au bout de la 5ème époque, ensuite il stagne. Ce qui détecte un problème de sur-apprentissage pour ce modèle. L'Overfitting (sur-apprentissage) désigne le fait que le modèle prédictif produit par le modèle s'adapte "trop" au Training Set. C'est-à-dire que le modèle prédictif pourra donner de très bonnes prédictions sur les données du Training Set (les données qu'il a déjà "vues" et auxquelles il s'est adapté), mais il prédira mal à partir de données qu'il n'a pas encore vues lors de sa phase d'apprentissage. On dit que la fonction prédictive se généralise mal.



FIGURE 33 : LA FONCTION LOSS

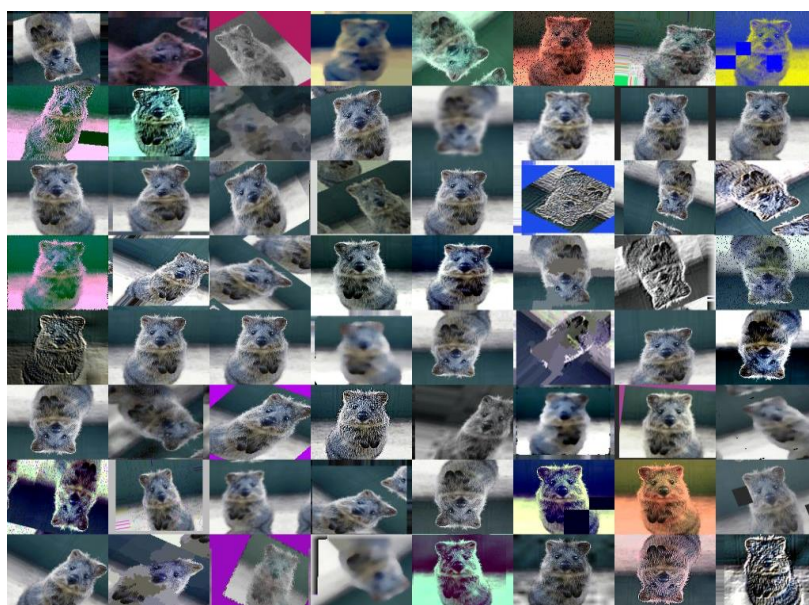
### Augmenter les Training-set ?

Il existe plusieurs manières de réduire l'overfitting dans les modèles de deep-learning. La meilleure option est d'obtenir plus de données d'entraînement. Malheureusement, dans notre situation cela n'est pas possible vu la source de nos données d'entraînement.

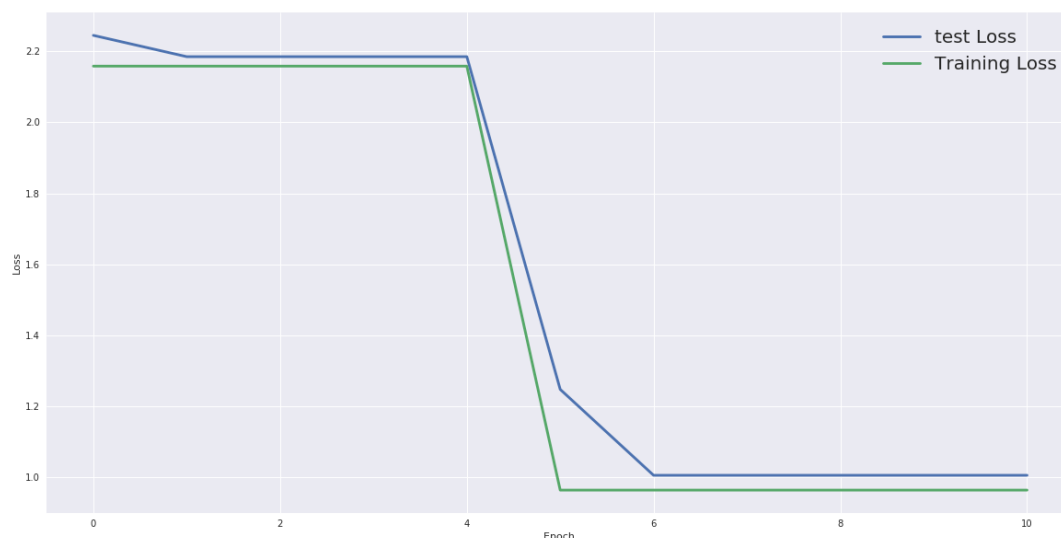
### Data-Augmentation pour la résolution du sur-apprentissage

Par cette expression, on ne veut pas dire que nous allons augmenter le training-set, mais plutôt appliquer des transformations sur nos images d'entraînement. Car lorsque le modèle U\_Net apprend il doit réellement ajuster ses paramètres pour qu'il mappe une entrée particulière sur une sortie (une étiquette). Notre objectif est d'enlever cet endroit idéal où la loss du model est faible, ce qui se produit lorsque les paramètres sont réglés correctement.

La data-augmentation consiste à faire des opérations telles que la rotation aléatoire de l'image, le zoom avant, l'ajout d'un filtre de couleur, etc. L'augmentation des données ne concerne que l'ensemble d'entraînement et non pas l'ensemble de validation. Dans notre cas c'est la fonction *Gen\_rand\_patch* qui prend en argument une image avec le masque correspondant, et nous retourne une liste de patches auxquels on aurait appliqué d'une façon aléatoire une transposition, une permutation des axes ou des rotations. Ci-dessous on montre les différentes transformations qui peuvent être appliquées à une image pour augmenter le nombre de patches.



**FIGURE 34 : PARMI LES TRANSFORMATIONS POSSIBLES POUR AUGMENTER LES DONNEES**



**FIGURE 35 : LA FONCTION LOSS APRES DATA-AUGMENTATION**

## 5.7 Résultats

La prédiction sur nos images a montré une confusion importante entre les classes, notamment entre surface d'eau et forêt.

On a noté une légère amélioration du modèle après la data-augmentation, même si on est toujours dans un problème de sur-apprentissage.

Le teste de prédiction sur une image à 8 bandes (Une image de plus dans le data set récupéré) entraîné par le même modèle montre le résultat suivant :



**FIGURE 36 : IMAGE A 8 BANDES ET SA PREDICTION PAR LE RESEAU UNET**

Malgré une qualité de prédiction qui n'est pas la bonne. Ce résultat est relativement meilleur que la prédiction sur nos images à 3 bandes. C'est-à-dire que l'adaptation faite au départ (extraire les 3 bandes RGB à partir de l'image d'entraînement à 3 bandes) n'est pas favorable pour l'apprentissage du réseau, ceci d'un part. D'un autre par, les images d'entraînement utilisées venant de Kaggl Data-set étaient prises sur des paysages d'Amérique Latines, dont la couverture du sol ne correspond pas vraiment à celle en France (Le réseaux qui a vite très bien appris des paysages n'arrivent plus à reconnaître les nouveaux).

## 6 Segmentation sémantique par intensité de pixel

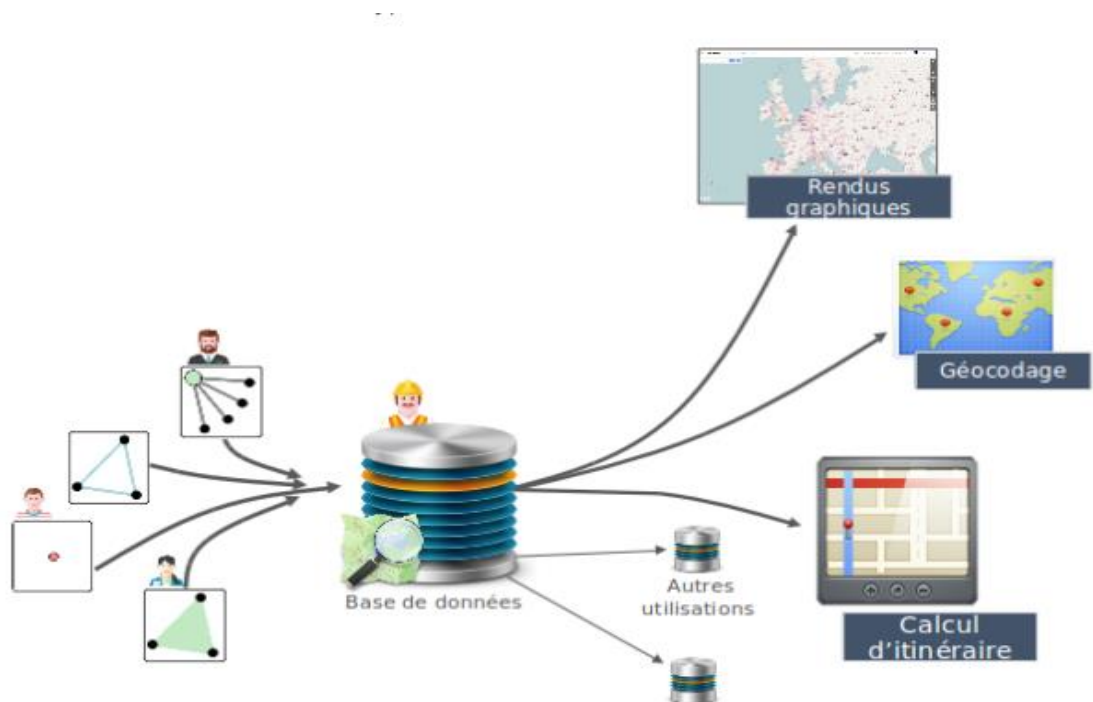
Dans les chapitres précédents, on a vu que notre problème d'identification de l'occupation du sol entourant les postes de transformation électrique, se ramène à un problème de segmentation sémantique d'image satellite (extraite de Google MAP dans notre cas). Ce dernier est un problème totalement différent de toute tâche de classification ou de détection d'objet sur une image. Contrairement à la classification où nous prédisons la classe de l'image d'entrée dans l'une des classes disponibles ou les coordonnées de la boîte en cas de détection d'objet, la segmentation nécessite de générer l'étiquette de la classe de chaque pixel de l'image. Dans cette partie, on présentera une méthode de segmentation sémantique fondée sur la classification ou le seuillage des pixels en fonction de leur intensité sur des cartes statiques d'OpenStreetMAP.

### 6.1 OpenStreetMap

OpenStreetMap est une carte basée sur des informations géographiques, elle est donc une base de données géographique avant d'être une carte : il est donc possible de télécharger directement les données brutes et d'en faire la réutilisation de son choix, cartographique ou non. Ces bases de données contiennent, grâce à chaque contribution, des informations très diverses relatives à l'espace public : des routes, des bâtiments, des adresses.

En fonction de ce que vous souhaitez réaliser avec OpenStreetMap, vous utiliserez plutôt

- Un fond de carte ou une carte interactive réalisée à partir des données.
- Le webservice Overpass, qui permet de réaliser des extractions thématiques.
- Les données brutes de la base de données, disponibles dans plusieurs formats.



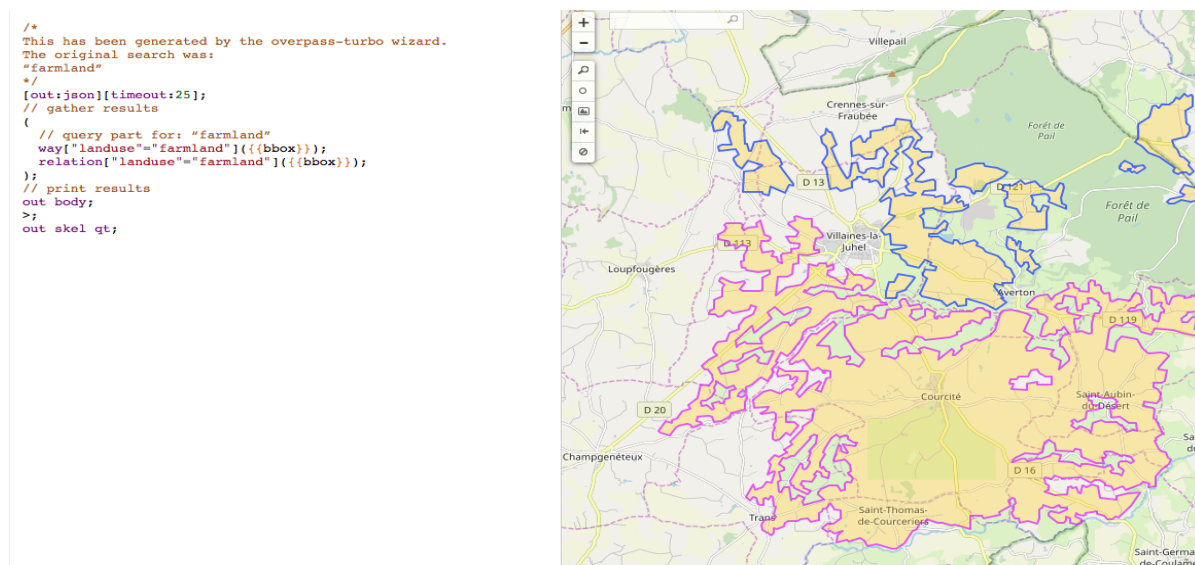
**FIGURE 37 : UTILISATION DES BASES DE DONNEES D'OSM**  
**SOURCE : SITE OPENSTREETMAP FRANCE**



## 6.2 Le webservice Overpass

Overpass est le bon outil pour télécharger des données thématiques : par exemple extraire la liste des écoles d'une ville avec leur nom, ou encore effectuer une carte des voies de bus d'une région.

Dans un premier temps on va utiliser overpass-Turbo qui est un outil d'exploration et de filtrage des données OpenStreetMap. Il permet de formuler sa requête, et de visualiser le résultat sur une carte interactive. Dans la figure ci-dessous, à droite c'est la requête qui permet de détecter le feature «farmland» ce qui correspond aux terres agricoles qui se trouve sur la zone géographique sélectionnée en coordonnées GPS.



**FIGURE 38 : EXTRACTION FOND DE CARTE**  
SOURCE : OVER PASS TURBO

Il propose aussi des téléchargements dans de nombreux formats géographiques sur l'API : les coordonnées GPS en format GeoJSON, JPX, KML, et en carte soit en forme carte interactif ou bien statique sous format PNG. Vu que nous sommes devant un problème de cartographie d'occupation du sol pour les 2235 postes de transformations, on doit procéder au téléchargement d'une façon automatique en utilisant Python pour interroger les données OSM avec l'API overpass. Cette méthode permet de récupérer les données sous les différents formats, mais non pas les cartes que ça soit statique ou dynamique. La chose qui aurait pu résoudre notre problème de segmentations via la détection du fond de carte qui nous intéresse (forêts, agriculture, eau et habitation) en chaque position centrée sur nos zones d'intérêt. Les données qui nous intéressent à chaque position sont stockées sous format json est un dictionnaire de points. Chaque point a un identifiant unique donné par une chaîne de caractères (par exemple : '1265651281'). Cet identifiant sert de clé pour identifier le point dans le dictionnaire. Pour chaque clé, le dictionnaire point contient la longitude et la latitude du point en question. Par exemple, le dictionnaire commence par

```

points = {
    '1265651281': (5.870229, 45.641839),
    '1265650638': (5.870642, 45.643893),
    '341949246': (5.8722016, 45.6391549),
    ...

```



Comme l'extraction de fond de carte résultant d'une requête overpass API n'est pas possible sous format de carte, on va essayer à partir des coordonnées dessiner une carte contenant que la classe qui nous intéresse (soit forêt, zone habitation, etc.).

Pour pouvoir afficher les points, il faut traduire leurs coordonnées longitude/latitude en pixels. Pour simplifier, nous allons supposer que toutes les cartes que nous allons dessiner sont carrées, autrement dit, le morceau du globe terrestre est que l'on regarde est spécifié par un centre (donné par sa longitude et sa latitude), et un rayon (donnée en degrés). Le morceau du globe terrestre considéré est donc délimité par les quatre coins de coordonnées

Si on suppose que le globe est localement plat, on peut transformer les coordonnées longitude / latitude d'un point en coordonnées x/y en pixels avec une règle de trois. On trouve :

- $x = (\text{longitude} - (\text{longitude\_centre} - \text{rayon})) * \text{taille} / (2 * \text{rayon})$ ,
- $y = \text{taille} - (\text{latitude} - (\text{latitude\_centre} - \text{rayon})) * \text{taille} / (2 * \text{rayon})$ .

Où taille est la taille de la fenêtre qui contient la carte que l'on dessine (en pixels).

Les résultats obtenus dans le cas où les requêtes se font sur des points « nodes » sont identiques. Dans le cas où les éléments sont de type « chemin » les résultats sont moins satisfaisants. Un exemple d'un dessin d'une image binaire de « forêt » est dans la figure suivante :



Ce qui ne correspond pas vraiment à la carte résultante de l'API. Car à priori les déformations sont à cause des éléments dont les coordonnées ont été choisies d'être retournées par l'API, quand l'objet identifié est sur une surface et non pas un point. Ainsi, théoriquement il est impossible de plaquer un morceau de sphère sur un carré sans déformation. Même si ces déformations peuvent être minimisées. Par exemple on peut considérer des rayons différents pour la longitude et la latitude : on peut par exemple calculer la latitude minimale et maximale avec

latitude-rayon/cos(longitudePi/180)) et latitude+rayon/cos (longitude Pi/180)) au lieu de latitude-rayon et latitude+rayon.

Mais nous avons quand même décidé de passer encore à une méthode plus efficace. Cette dernière va permettre de profiter de la diversité du code couleur utilisé par la carte.

### 6.3 Segmentation par classification des cartes OSM statiques

On part ici d'un rapport qu'entretient chaque pixel individuellement avec des informations calculées sur toute l'image, comme la moyenne des niveaux de gris de l'ensemble des pixels, ou la médiane, permettant de construire  $n$  classes d'intensité. Lorsque les classes sont déterminées par le choix d'un seuil, on parle de seuillage. Les pixels appartenant à une même classe et étant connexes forment des régions *de l'image*.

Dans la méthode d'Otsu, le seuil qui minimise la variance intra-classe est recherché à partir de tous les seuillages possibles :

$$\sigma_{\omega}^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t)$$

Otsu montre que minimiser la variance intra-classe revient à maximiser la variance inter-classe :

$$\sigma_b^2(t) = \sigma^2 - \sigma_{\omega}^2(t) = \omega_1(t)\omega_2(t)(\mu_1(t) - \mu_2(t))^2$$

qui est exprimée en termes des probabilités de classe  $\omega_i$  et des moyennes de classes  $\mu_i$  qui à leur tour peuvent être mises à jour itérativement. Cette idée conduit à un algorithme efficace mais elle n'est utilisable que sur des images en niveau de gris.

#### 6.3.1 La méthode des k-means

La méthode des k-means a été introduite par J. McQueen en 1971 et mise en œuvre sous sa forme actuelle par E. Forgy. De nombreuses variantes se sont succédées depuis afin d'étendre ses capacités de classification (séparations non linéaires) : *kernel k-means* (*k-means* basée sur des méthodes à noyaux), améliorer ses performances : *global-k-means*, *k*-Harmonic means, automatiser le choix du nombre de clusters : Gaussian-means, *X*-means.

Dans le cadre de la classification non supervisée, on cherche généralement à partitionner l'espace en classes concentrées et isolées les unes des autres. Dans cette optique, l'algorithme des *k-means* vise à minimiser la variance intra-classe, qui se traduit par la minimisation de l'énergie suivante

$$\begin{aligned} E &= \frac{1}{2} \sum_{c \in C} \sum_{x \in C} \|x - m_c\| \\ &= \frac{1}{2} \sum_{x \in C} \#c \times V(c) \\ &= \frac{1}{2} \sum_{x \in D} \min_{c \in D} \|x - m_c\|^2 \end{aligned}$$

Avec  $C$  l'ensemble des clusters et pour chaque cluster  $c$ ,  $m_c$  son centre (appelé noyau),  $V(c)$  sa variance,  $\#c$  le nombre de ses éléments et  $D = U_{c \in C}$  l'ensemble des données que l'on cherche à classer.

##### 6.3.1.1 Algorithme

La minimisation de l'équation de l'énergie peut se réaliser par une descente de gradient sur les noyaux dont les propriétés de convergence ont été vérifiées. Elle peut se traduire par les étapes suivantes :

- initialisation des noyaux.

- b) mise à jour des clusters.
- c) réévaluation des noyaux.
- d) itérer les étapes a, b et c) jusqu'à stabilisation des noyaux.

Il se présente plus formellement sur la figure suivante :

**Algorithme 1 - k-means**

On note  $\mathbb{E}(c)$  la fonction qui associe à un cluster  $c$  son centre de gravité défini par  $\frac{1}{\#c} \sum_{x \in c} x$ .

**Début**

// Initialisation

$\forall c \in \mathcal{C}$

$m_c \leftarrow x \mid x \in \mathcal{D}$  // Tirage uniforme dans la base de données

$c \leftarrow \{x \in \mathcal{D} \mid \min_{q \in \mathcal{C}} \|x - m_q\|^2 = \|x - m_c\|^2\}$

**Fin**  $\forall$

**Tant que**  $\exists c \in \mathcal{C} \mid m_c \neq \mathbb{E}(c)$

$\forall c \in \mathcal{C}$

$m_c \leftarrow \mathbb{E}(c)$

$c \leftarrow \emptyset$

**Fin**  $\forall$

$\forall x \in \mathcal{D}$

$\hat{c} \leftarrow \arg \min_{c \in \mathcal{C}} \|x - m_c\|^2$

$\hat{c} \leftarrow \hat{c} \cup x$

**Fin**  $\forall$

**Fin tant que**

**Fin**

**FIGURE 39 : ALGORITHME DU K-MEANS**  
**SOURCE : SITE TELECOM PARIS TECH**

### 6.3.1.2 Choix du nombre $k$ de clusters

L'algorithme classique des  $k$ -means laisse un paramètre libre : le nombre de clusters, ce qui dans le cas de la segmentation d'images correspond au nombre d'intensités utilisées pour représenter l'image. Généralement le choix de  $k$  est fait empiriquement en sélectionnant la valeur de  $k$  qui minimise la fonction de l'énergie.

Différents critères permettent d'estimer le nombre de clusters en minimisant la distance intra-classes et à maximiser la distance inter-classes (par exemple en segmentation d'images couleurs).

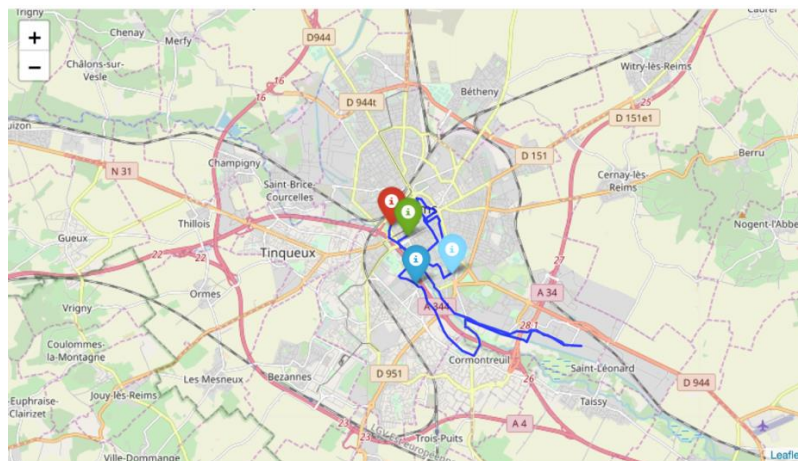
De nombreuses stratégies permettent de déterminer le nombre de clusters pendant le déroulement de l'algorithme. Ces méthodes sont basées sur un processus itératif au cours duquel on choisit de subdiviser les clusters précédemment établis en se basant sur un critère statistique (par exemple grâce aux BIC : *Bayesian Information Criterion*) : Gaussian-means et X-means.

### 6.3.2 Scraper les cartes OSM

Le module « Folium » est un module Python qui facilite l'utilisation des données OpenStreetMap, en permettant de créer une carte OSM et d'y dessiner des marqueurs, des lignes brisées, des polygones, etc.

- La méthode `folium.Map` (`location=[latitude, longitude]`, `tiles="OpenStreetMap"`, `zoom=zoom_start`).
- La méthode `Marker` () place un marqueur sur la carte au point de coordonnées (latitude,longitude) avec une couleur (paramètre `color`) choisie dans une palette de 147 nuances et un titre (paramètre `popup`).
- La méthode `PolyLine`() dessine une ligne brisée à partir d'une liste de couples de coordonnées.
- La méthode `save`() génère un fichier HTML qui embarque la carte, ainsi que tous les éléments ajoutés par le script (marqueurs, lignes, polygones, etc).
- Enfin jusqu'à là, on put générer une carte interactive qui se centre sur une position donnée, ensuite on convertit la carte HTML en carte statique avec la méthode `asPng` (), ce qui donnera une image en forme binaire qu'on convertit à nouveau en format PNG qui se sauvegarde en RGBA (Rouge Vert Bleu Alpha(l'opacité)) qui une extension du codage ne couleurs RVB.

Ci-dessous un exemple de carte où un parcours est représenté sous la forme d'une ligne brisée bleue. On placera aussi quatre marqueurs : un sur le départ, un sur l'arrivée et deux autres correspondant respectivement aux points de plus faible et plus grande altitude.



**FIGURE 40 : EXEMPLE CARTE STATISTIQUE OSM RESUMANT**

### 6.3.3 La détection et la description de Features




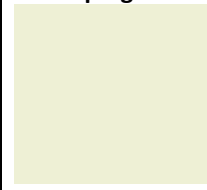


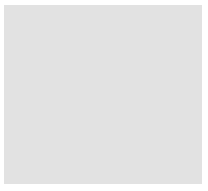
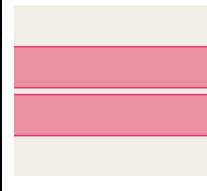
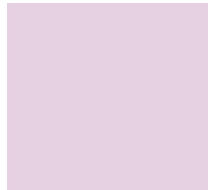

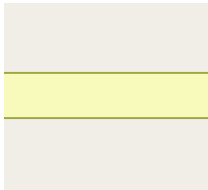
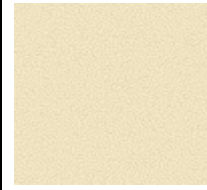
Malgré une visualisations à l'œil nu qui montre que le nombre de couleurs dans l'image ci-dessus n'est relativement pas très élevé. Une façon de décompter via python a donné plus de 4000 couleurs sur cette carte. D'où l'intérêt d'une méthode de clustering afin de réduire le nombre de couleurs à un nombre assez proche de celui qu'on identifie par visualisation, car ça nous permet une vraie distinction entre les objets de la carte.

#### 6.3.3.1 Quesque c'est un features dans notre cas ?

La détection et la description de *features* constituent la première étape indispensable pour de nombreuses autres tâches en vision par ordinateur, comme la classification d'images.

Dans ce domaine, le terme de features désigne des zones intéressantes de l'image numérique. Ces zones peuvent correspondre à des contours, des points ou des régions d'intérêt. OpenStreetMap Wiki fournit un aperçu de la définition et des différentes clés (tags) concernant l'occupation des sols (landuse) pour l'extraction de donnée via Overpass API. Ce qui est de plus intéressant, est non pas la clé qui identifient chaque tag dans notre situation (Même si c'est le cas dans overpass turbo API), mais plutôt la représentation en image pour chacun des features

utilisés par la carte. C'est ce qui représentera pour nous « feature-map ». Ci-dessous quelques exemples parmi les features-map utilisés.

<b>Forêt</b> 	<b>Parcs</b> 	<b>Non cultivé</b> 	<b>Champ Agricole</b> 
<b>Eau</b> 	<b>Construction</b> 	<b>Résidence</b> 	<b>Autouroute</b> 
<b>Industrie</b> 	<b>Vois ferré</b> 	<b>Route secondaire</b> 	<b>Sable</b> 

**TABLEAU 2 FEATURE MAP**  
**SOURCE : OPENSTREETMAP WIKI**

Plus généralement, à chaque feature détectée est associé un vecteur, appelé descripteur (feature descriptor ou feature vector), qui, comme son nom l'indique, décrit la zone concernée. Vu la nature de nos features-maps, le features-descripteur sera un vecteur (R, G, B) ce qui correspond au stockage du taux de saturation en rouge, en vert et en bleu.

### 6.3.3.2 Nos features-descripteurs, sont-ils un bon choix ?

Une zone constitue un bon choix de *features* si elle est :

- Répétable : une *feature* doit se retrouver dans les images représentant la même scène malgré les différences géométriques et photométriques. Une *feature* doit donc présenter des propriétés d'invariance à ces transformations. Or, dans notre cas le même features-map est utilisé sur le territoire Français pour des données extraites la même année (voir OpenStreetMap Wiki France). La contrainte de différence géométrique ne s'impose pas, comme on garde le même niveau de zoom pour toutes les cartes statiques, ainsi on procède par intensité de pixel et non pas par forme géométrique des entités à identifier. Pareillement pour les contraintes photométriques, on garde le même contraste et features-descripteurs pour toutes nos régions d'étude.
- Distinctive : une *feature* doit être suffisamment unique et non-ambiguë au sein d'une image pour faciliter le *matching*. Ce sont les informations contenues dans son descripteur qui doivent mettre en valeur sa particularité. En effet dans le tableau ci-dessus, on voit que chaque couleur du fond (sans les motifs) caractérise chaque tag et ne peut être confondu avec une autre couleur d'un autre tag. Ainsi, que chaque catégorie exemple verdure (foret, parc, non cultivé) est caractérisée par une plage d'intensité de pixel assez éloignée pour ne pas être confondu lors d'une méthode clustering, si



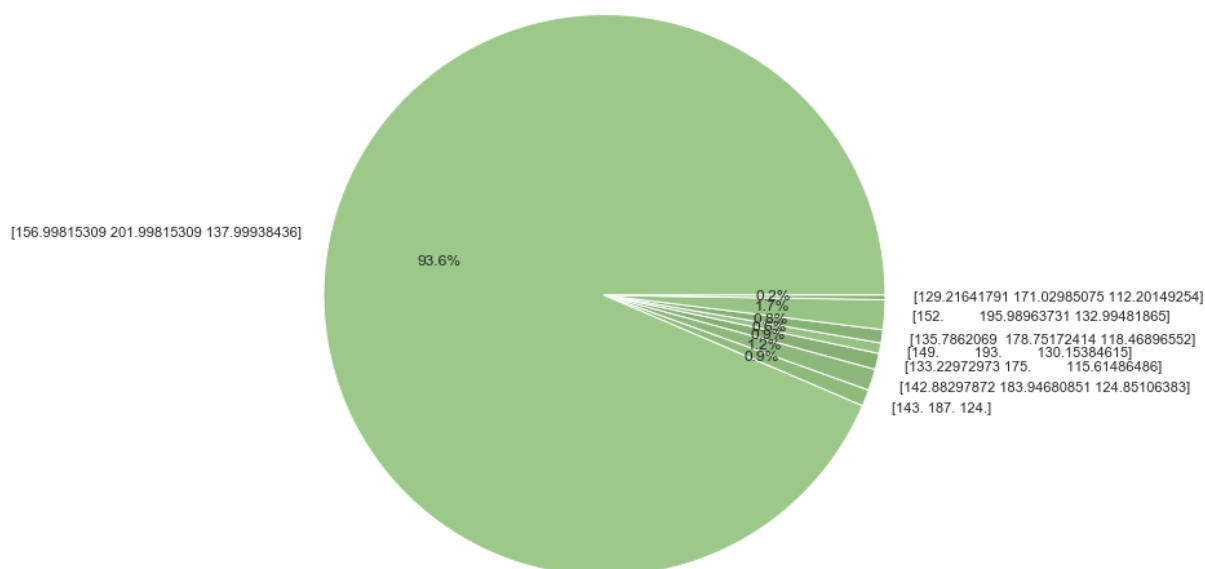
l'émersion des tags de même catégorie est appliquée (c'est-à-dire, si on veut confondre par exemple le feature-map forêt avec, celui du tag parc).

### 6.3.3.3 Extraction des features –descripteur

Dans ce chapitre on s'intéresse à un plus que le nombre de classe fixé au début du projet. Cela est dû au nombre d'information disposé par les cartes OSM, qui permet une classification plus raffinée d'un part. D'un autre part, dans une méthode de classification par clustring le nombre de classe est fixé avant l'opération donc, plus on a de classes moins on confondra la classe qui nous intéresse à une autre classe qui nous concerne moins dans le cadre de ce projet.

On présentera dans la suite des méthodes simples qui vont permettre l'extraction d'un « bon » features-descripteur, pour plusieurs features-map. Ceux qui nécessitent la même méthode ne sont mentionnés que par une seule méthode.

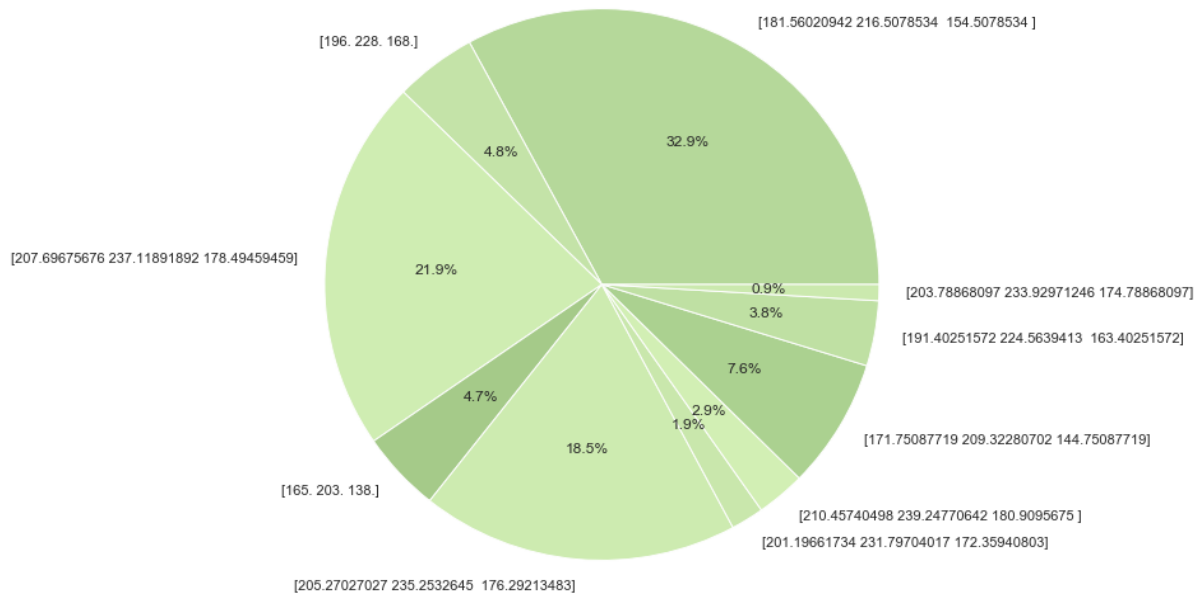
- L'application de la méthode des k-means, avec initialisation aléatoire sur le feature-map forêt nous donne la répartition suivante :



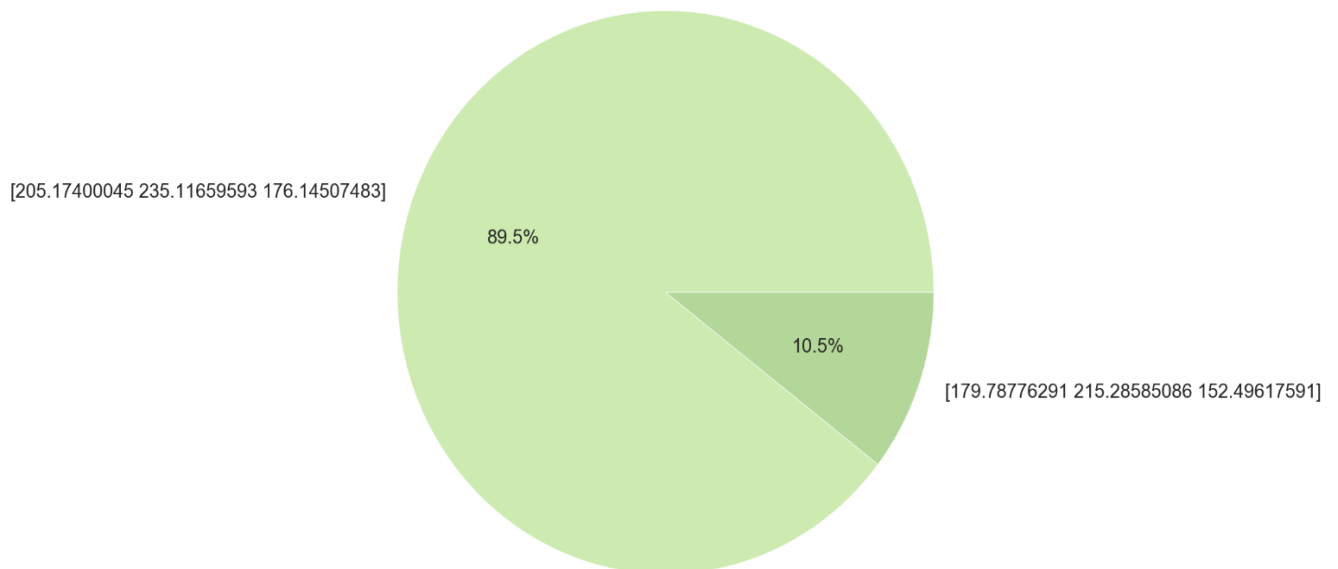
**FIGURE 41 : REPARTITION DE LA COULEUR FEATURE FORET**

On voit clairement qu'il y a une couleur dominante à plus de 93%, même avec une initialisation aléatoire et plusieurs itérations de l'algorithme. Sachant que les autres couleurs sont des dégradés de la couleur principale. On choisit dans ce cas le feateurs-descripteur (156.99815309, 201.99815309, 137.99938436) pour la classe forêt.

- En procédant de la même façon qu'avec la classe zone non cultivée, avec initialisation aléatoire. On trouve le résultat suivant



Malgré que dans le feature-map on voit la présence d'une couleur dominante. Mais à chaque application on obtient une nouvelle répartition. Ceci est à cause de l'initialisation aléatoire. Nous allons supposer que le feature-map ne dispose que des deux couleurs que nous visualisons, et on initialisera la méthode sur ces deux couleurs. On obtiendra alors la répartition sur la figure suivante où, on remarque qu'il y a une couleur dominante à un pourcentage important, qui sera par conséquent notre features-descripteur pour cette classe. Sachant que on obtient la même répartition pour d'autre applications à un différent nombre d'itération de l'algorithme. D'où l'importance de l'initialisation qui est un facteur déterminant dans la qualité des résultats (minimum local).



On procédera de cette manière pour les features-map qui ont cette façon d'être représenté. C'est-à-dire on commencera par une initialisation aléatoire, et on verra s'il existe une classe dominante avec un pourcentage assez important qui soit relativement constant pour plusieurs applications de l'algorithme. Si non, on procédera à une utilisation de la méthode, et on augmentera de plus en plus le nombre de points d'itinérisation de telle façon à avoir un features-map qui soit vraiment caractéristique de la classe.

Pour les classes unicouleurs (comme la classe des champs agricoles, résidence, eau, industrie, etc.) le travail est moins fastidieux comme le features-descripteur sera n'importe quelle valeur du pixel dans le features-map.

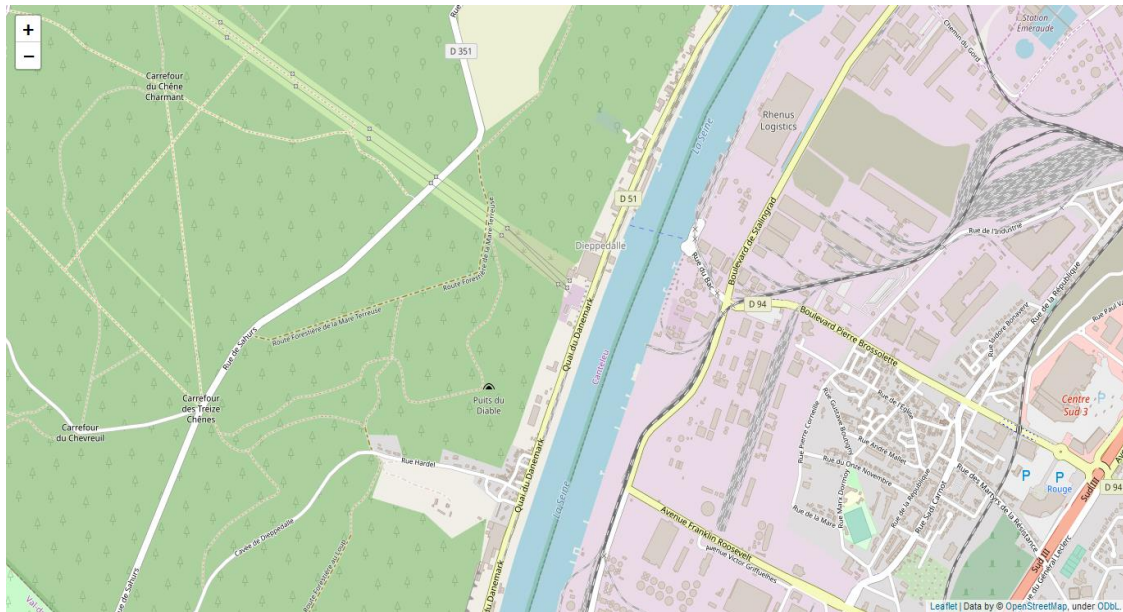
Dans la catégorie « routes » on procédera aussi de la même façon, il y a une couleur caractéristique pour chaque classe et, c'est cette couleur qui sera notre features-descripteur.

#### 6.3.4 Ségmentation de cartes

La méthode des k-means est un outil de classification classique qui permet de répartir un ensemble de données en k-classes homogènes. La plupart des images (photos, dessins vectoriels 2D, synthèses 3D, ...) vérifient localement des propriétés d'homogénéité, notamment en termes d'intensité lumineuse. L'algorithme des k-means permet donc d'apporter une solution à la segmentation à nos cartes statiques.

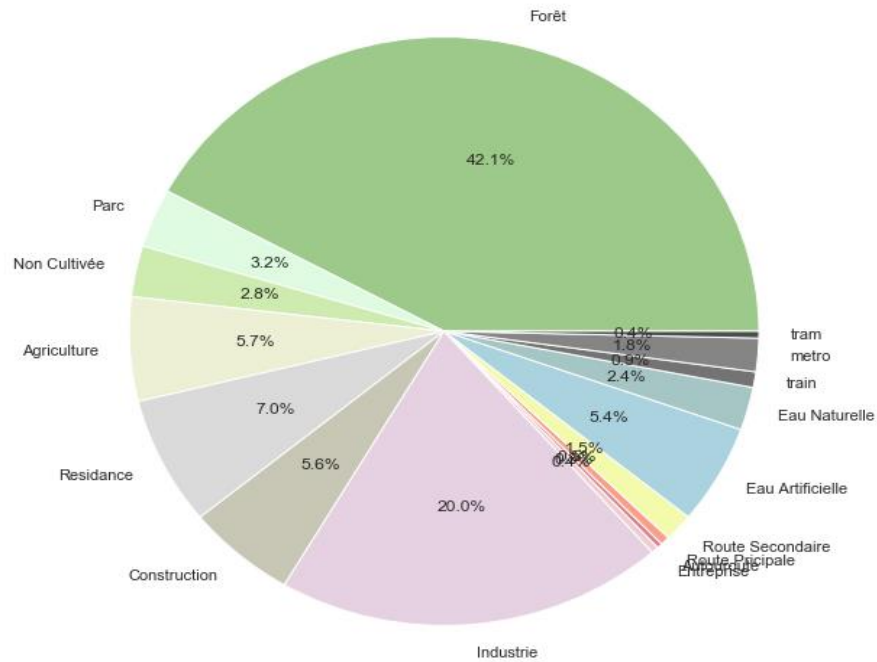
Pour une initialisation optimale on choisit comme noyaux d'initialisation les features-descripteurs choisis avec soin.

On choisit pour cela une carte statique centrée sur un poste source électrique avec 15 comme niveau de zoom (voir la carte ci-dessous). On choisit un environnement diversifié.



**FIGURE 42 : CARTE STATIQUE**  
**SOURCE : OPENSTREETMAP**

Une application de la méthode avec 16 classes sur la carte ci-dessus donne la répartition suivante :



**FIGURE 43 : OCCUPATION DE SOL**

### 6.3.5 Validation du Model

A partir des labels récupérés pour chaque pixel, on reconstituera une nouvelle carte afin de valider le model, on obtiendra la carte suivante :



**FIGURE 44 : CARTE DONNEE PAR LE MODEL**

Les résultats sont très satisfaisants. A noter une confusion entre la classe construction classifiée comme résidence (uniquement au niveau de zoom 14), car cette classe se représente souvent entourée de résidence donc classifiée comme résidence par le model de clustering comme les

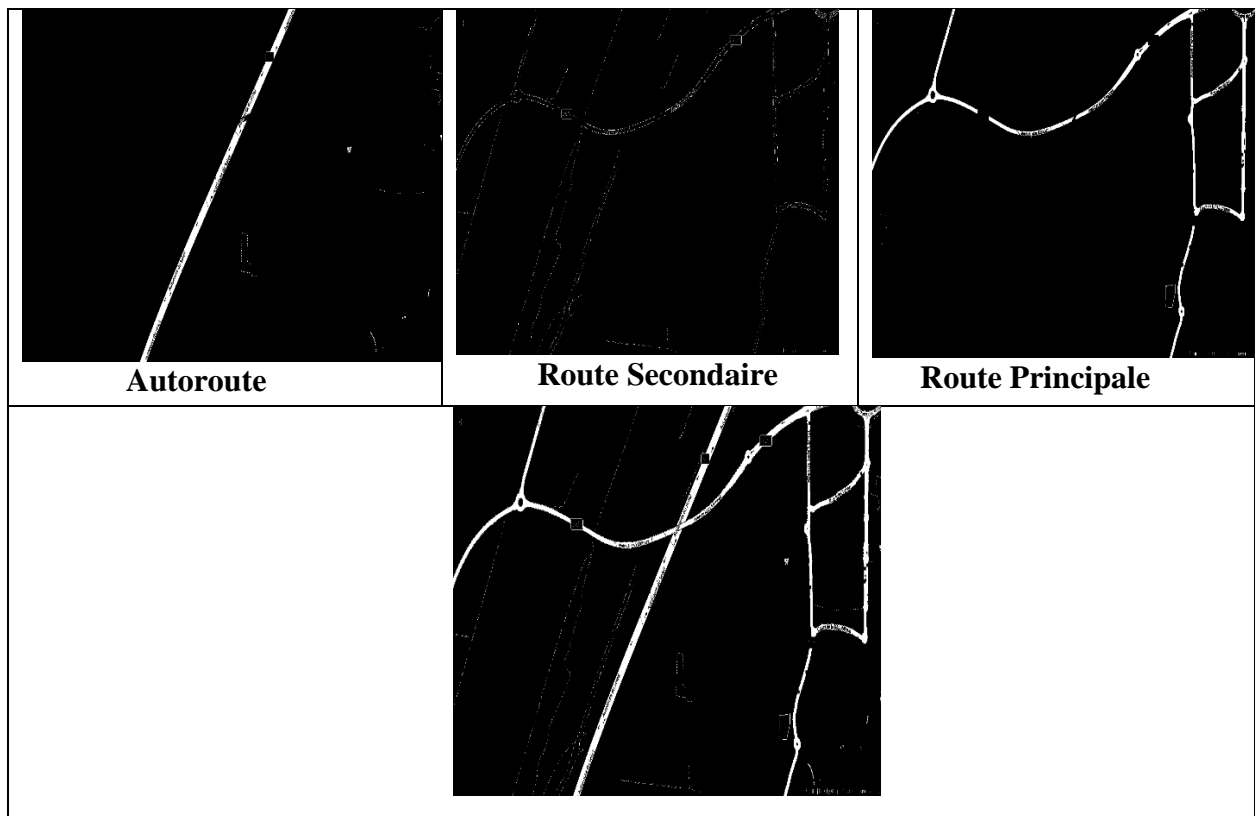
features-descripteur. Cela ne déminuera pas de la qualité satisfaisante de l’algorithme comme dans le cadre de notre projet les deux classes sont confondues. Donc on valide cette méthode de détection du sol.

### 6.3.6 Reconstitution de données pour l’entrainement du model U-net

Dans le chapitre précédent, on était affronté à un problème d’entrainement dû au manque de données référentes. Ainsi, par une connaissance à priori du sujet le type de paysages sur lequel le model était entrainé aurait pu ne pas s’adapter à nos données.

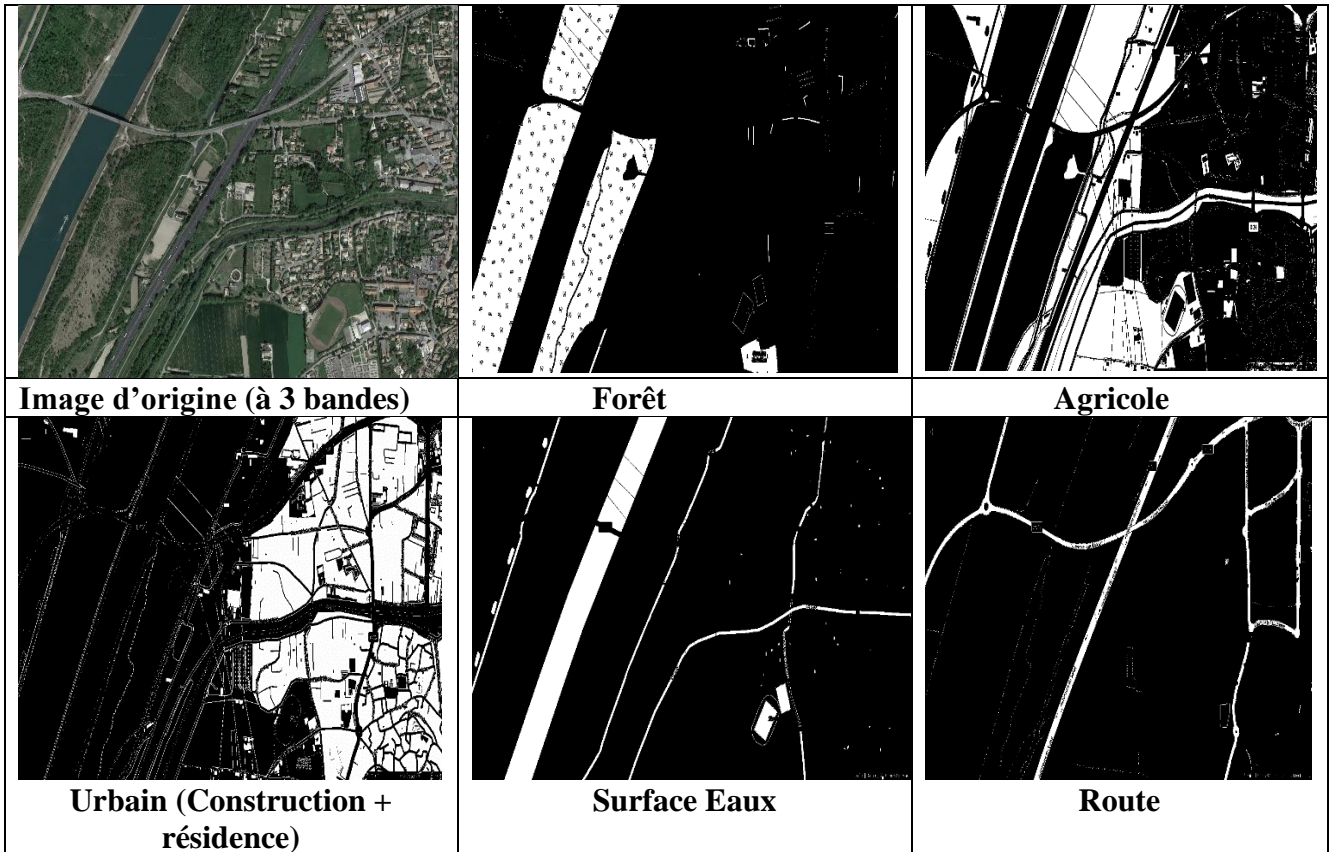
L’occupation du sol détectée par la méthode de clustring précédente, nous permet de créer une nouvelle base de données propre à nous.

Pour notre problème d’occupation on se ramène dans fusionner quelques classes entre elles, comme dans la figure suivante. On met la classe autoroute, route secondaire et route principale dans la même classe « route ».



On fait pareil pour la classe résidence et construction. La figure suivante représente, notre donnée d’entrainement pour notre réseau.





## 7 CONCLUSION

### 7.1 Mission accomplies et perspectives

Pour la mise en œuvre de mon projet, j'ai été amené dans un premier temps à comprendre le fonctionnement du réseau électrique, la conception des transformateurs électriques et notamment son système aéroréfrigérant, pour pouvoir mettre en œuvre la pression différentielle entre l'entrée et la sortie des filtres des aéroréfrigérants comme indicateur d'une situation d'encrassement.

Dans un deuxième temps, j'ai identifié toutes les causes externes et environnementales qui sont à l'origine de l'encrassement, comme les pollens (Platans, Pauplier et Saule), les graminées et la polluants de l'air (PM10). Ensuite, j'ai répertorié toutes les données possibles depuis les organismes concernés, et localisé tous leurs capteurs de mesure afin de pouvoir associer chaque poste source à un capteur mesurant un « polluant » parmi ceux qui sont identifiés.

La concentration de l'air autour du transformateur est directement lié à son environnement. Donc connaître l'environnement du transformateur est primordiale.

Pour accomplir cette tâche, j'ai procédé par la segmentation sémantique sur des images satellites centrées sur ces postes sources, et des méthodes de clustering sur les cartes d'Openstreetmap. Dans la perspective de croiser toutes ces données et afin de définir pour chacun transformateur son système pré-dectif d'encrassement à l'aide des indicateurs environnementaux.

### 7.2 Apport du stage,

Ce stage a été l'opportunité pour moi d'appréhender mes connaissances sur le fonctionnement du réseau électrique de la conception des centrales électriques à la distribution aux particuliers. Ainsi, que d'apprendre la démarche d'une équipe R&D dans son travail tel que le management des équipes R&D, les processus de développement et les activités de conception des produits. Ce stage m'a permis ainsi d'améliorer mes compétences informatiques, notamment en data scraping et manipulation de données images par méthode de réseau de neurone.

Au-delà d'enrichir mes connaissances, ce stage m'a permis de comprendre mon intérêt pour le domaine de l'industrie, et aux entreprises répandant aux enjeux du future en matière d'énergie, toute en me formant aux métiers d'ingénieurs en mathématiques appliquées, capable de s'adapter aux différentes problématiques nécessitant à la fois des connaissances mathématiques variées et des compétences informatiques.

## 8 Bibliographie

- A, Verney Carron. «Impact de la pollution de l'aire et du climat sur l'encrassement des vitrages.» 2017.
- Girshick, Ross , Jeff Donahue, Trevor Darrell, et Jitendra Malik. «Rich feature hierarchies for accurate object detection and semantic segmentation.» UC Berkeley, s.d.
- Long, Jonathan, Evan Shelhamer, et Trevor Darrell. «Fully Convolutional Networks for Semantic Segmentation.» s.d.
- Postadjian, Tristan, Arnaud Le Bris, Hichem Sahbi, et Clément Mallet. «Classification à très large échelle d'images satellites à très haute résolution spatiale par réseaux de neurones convolutifs.» Univ. Paris Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mandé, CNRS, LIP6 UPMC Sorbonne Universités, Paris, France, s.d.
- Remi DESQUIENS, EDF. «Etude d'un indicateur d'encrassement des aéroréfrigérants pour le système de monitoring AGATH d'Enedis.» Palaiseau , 2018.
- Ronneberger, Olaf, Philipp Fischer, et Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Germany: University of Freiburg, s.d.