



Smart-Réno

Fiabiliser, Professionnaliser, Valoriser la Rénovation Energétique

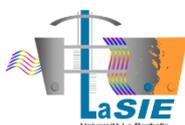
**Deuxième volet : Industrialiser les processus de rénovation et
professionnaliser les acteurs**

Livrable T2.2 - Diagnostic automatisé des bâtiments existants

Auteurs : Mathias Bouquerel, Alexandre Girard, Guillaume Terrasse, Jérôme Trailin

Contributeurs : Thierry Duforestel, Jean-Michel Dumay

Décembre 2020



Sigles et abréviations

3CL : Méthode réglementaire pour le Calcul de la Consommation Conventiionnelle des Logements, utilisé pour réaliser les DPE

API : Application Public Interface (Méthode d'accès à un applicatif)

DPE : Diagnostic de Performance Energétique

ECS : Eau Chaude Sanitaire

IA : Intelligence Artificielle

IAD : Intelligence Artificielle pour la Détection

ODEAL : Outil de Diagnostic Energétique Automatisé du Logement

SIG : Système d'Information Géographique

Sommaire

1	Introduction.....	5
1.1	Objectifs du livrable.....	5
1.2	Méthodologie appliquée.....	5
1.3	Architecture de l’outil complet.....	6
1.4	Plan du rapport.....	7
2	Module SIG : interrogation des bases SIG en open data.....	8
2.1	Traitement des informations de type cadastral.....	8
2.2	Récupération des photos de façade depuis Google Street View.....	10
3	Module IAD : détermination des caractéristiques de l’enveloppe par intelligence artificielle.....	13
3.1	Reconnaissance des éléments visibles de l’enveloppe.....	13
3.2	Extraction des caractéristiques utiles pour le diagnostic énergétique.....	15
3.3	Perspectives.....	17
4	SolRen : diagnostic énergétique et évaluation des solutions de rénovation.....	18
4.1	Cœur de calcul pour la simulation thermique dynamique: DiagSérie.....	18
4.2	Solveur dynamique Simfast.....	18
4.3	Evaluation des solutions de rénovation : SolRen.....	19
5	Traitement IA pour la période de construction.....	20
5.1	Présentation de la démarche.....	20
5.2	Exemples d’application.....	24
5.3	Généralisation à la détermination de la période de construction d’une maison parmi un ensemble de périodes de construction données.....	28
5.4	Synthèse et pistes de poursuite.....	28
6	Audit Numérique par analyse de la courbe de charge électrique.....	30
6.1	Concept de l’Audit Numérique.....	30
6.2	Méthode.....	30
6.2.1	Construction d’une base de données de caractéristiques de courbes de charges.....	30
6.2.2	Traitement de la courbe de charge client.....	33
6.2.3	Identification des signatures énergétiques les plus proches.....	33
6.2.4	Estimation des caractéristiques du logement du client.....	35
6.3	Perspectives.....	37
7	Conclusion.....	38
7.1	Etat actuel de l’outil.....	38
7.2	Limites actuelles de l’outil et perspectives.....	38
7.2.1	Démarche générale d’ODEAL.....	38
7.2.2	Module SIG.....	39

7.2.3	Module IAD.....	39
7.2.4	Module SolRen	41
7.2.5	Identification de la période de construction.....	41
7.2.6	Audit numérique par analyse de la courbe de charge	42
8	Annexe 1 – Configuration logicielle requise pour exécuter ODEAL	43

1 Introduction

1.1 Objectifs du livrable

L'ambition majeure du volet 2 est de d'imaginer des services permettant de professionnaliser les filières de rénovations, en se limitant à la maison individuelle puisque c'est le cadre défini par Smart-Réno. Ces services sont dans un premier temps destinés à produire le diagnostic personnalisé d'un bâtiment existant dans son état initial et après applications de différents gestes de rénovation. Ces gestes de rénovation de l'enveloppe du bâtiment étant définis par les études menées dans les volets 1 et 3 du Programme Smart Réno afin d'écarter tout risque de pathologie et de contre référence que pourrait engendrer la mise en œuvre de solutions non adaptées au cas traité. Ce diagnostic initial et les préconisations de travaux appropriés doivent intégrer des critères d'efficacité énergétique, mais aussi des critères concernant le confort, les aspects économiques et financiers, et les émissions de CO2. Ces services doivent ensuite pouvoir produire des devis et une planification des travaux, mais aussi éventuellement développer une offre de contrôle de la qualité. Ces services sont en priorité destinés aux particuliers et aux propriétaires-bailleurs, les clients de la rénovation, mais ils pourraient être à terme adaptés et mis à la disposition des artisans, des industriels, ou de tout acteur du secteur de la rénovation du bâtiment.

En améliorant le professionnalisme des filières de rénovation, ces services permettront d'avoir un impact favorable sur la confiance chez les clients. Ces services doivent apporter une justification technique et pratique aux solutions proposées, et éventuellement évoquer des solutions alternatives et les raisons qui ont amené à les écarter. Le développement du scan 3D, du BIM, de l'analyse d'images, la multitude sans cesse croissante des données disponibles sur Internet, issues de base de données ouvertes au public, mais aussi des objets connectés ou des compteurs communicants, sont des sources à exploiter pour objectiver les travaux de rénovation, pour en industrialiser les processus, pour produire quasi automatiquement du diagnostic, des métrés, des devis.

La première tâche consiste à se pencher sur la question du diagnostic énergétique automatisé. Il s'agit d'être capable de mener un diagnostic énergétique d'un bâtiment existant sans aucun autre apport par l'utilisateur que l'adresse de la maison concernée, en collectant et en traitant de manière automatisée et transparente les données disponibles qui concernent ce bâtiment. Ce livrable décrit le travail réalisé à EDF R&D sur le développement d'un tel service de diagnostic automatisé. L'objectif poursuivi par EDF R&D dans le cadre de ce volet 2, est de démontrer la faisabilité de ce service en s'appuyant sur ses acquis en matière de modélisation énergétique des bâtiments et au-delà, en explorant de nouvelles approches (IA, traitement de photos, traitement de courbes de charge,...) susceptibles d'améliorer les méthodes déjà éprouvées et de tester la potentialité de ses nouvelles approches au travers d'un chaînage logiciel prototype.

1.2 Méthodologie appliquée

EDF R&D disposait avant le début de Smart-Réno d'un certain nombre de briques logicielles offrant de manière plus ou moins complète certaines fonctionnalités nécessaires pour la mise en place d'un outil de diagnostic complet : récupération des données géométriques de type cadastral, réseaux de neurones adaptés à l'analyse et l'exploitation de photos, outils de traitement de courbe de charge, outils de simulation thermique dynamique pour le bâtiment... Le travail réalisé dans le cadre de la tâche 2.2 peut se résumer aux étapes suivantes :

- Inventaire des briques logicielles disponibles et intéressantes pour la construction de l'outil
- Adaptation de ces briques au problème considéré
- Construction d'un outil complet par assemblage des différentes briques logicielles

Dans la suite de ce rapport, le terme « module » sera utilisé pour désigner une brique logicielle plus ou moins indépendante offrant une ou plusieurs fonctionnalités, et le terme « outil » pour désigner l'outil complet issu de l'assemblage de certains modules. Cet outil sera aussi désigné sous l'acronyme ODEAL : Outil de Diagnostic Energétique Automatisé du Logement.

Pour la majorité des modules, le code source est du python. Par soucis de simplicité, c'est le code qui a également été utilisé pour la construction de l'outil complet ODEAL, ce qui n'empêche pas de faire appel à des briques logicielles qui ne sont pas en python. L'objectif suivi par EDF R&D a été de produire un démonstrateur d'outil de diagnostic automatisé, qui s'exécute en local sur un PC préalablement configuré par simple appel à une unique fonction python prenant en entrée une adresse et un numéro de département, sans autre intervention manuelle. La priorité a été mise sur l'automatisation du diagnostic et sur la mesure de la qualité des résultats obtenus. Par conséquent, ni la mise en place d'un service web pour accéder à cet outil, ni la création d'une interface graphique n'ont pour l'instant été intégrés au développement de l'outil.

1.3 Architecture de l'outil complet

Le processus de diagnostic réalisé par l'outil complet ODEAL se fait selon trois étapes successives, qui sont assurées par les trois modules principaux suivants :

- Module SIG (Système d'Information Géographique) pour la récupération automatique des données disponibles sur les bases de données publiques et ouvertes de type SIG : données géométriques issues du cadastre, photographie de la façade issue de Google Street View
- Module IAD (Intelligence Artificielle pour la Détection) pour la détection d'éléments de l'enveloppe et d'éléments architecturaux utiles à la mise en données du diagnostic énergétique (détection réalisée par application d'un réseau de neurones à la photographie de la façade issue de Google Street View)
- Module SolRen pour réaliser le diagnostic énergétique de l'état initial et par la suite évaluer différentes solutions de rénovations.

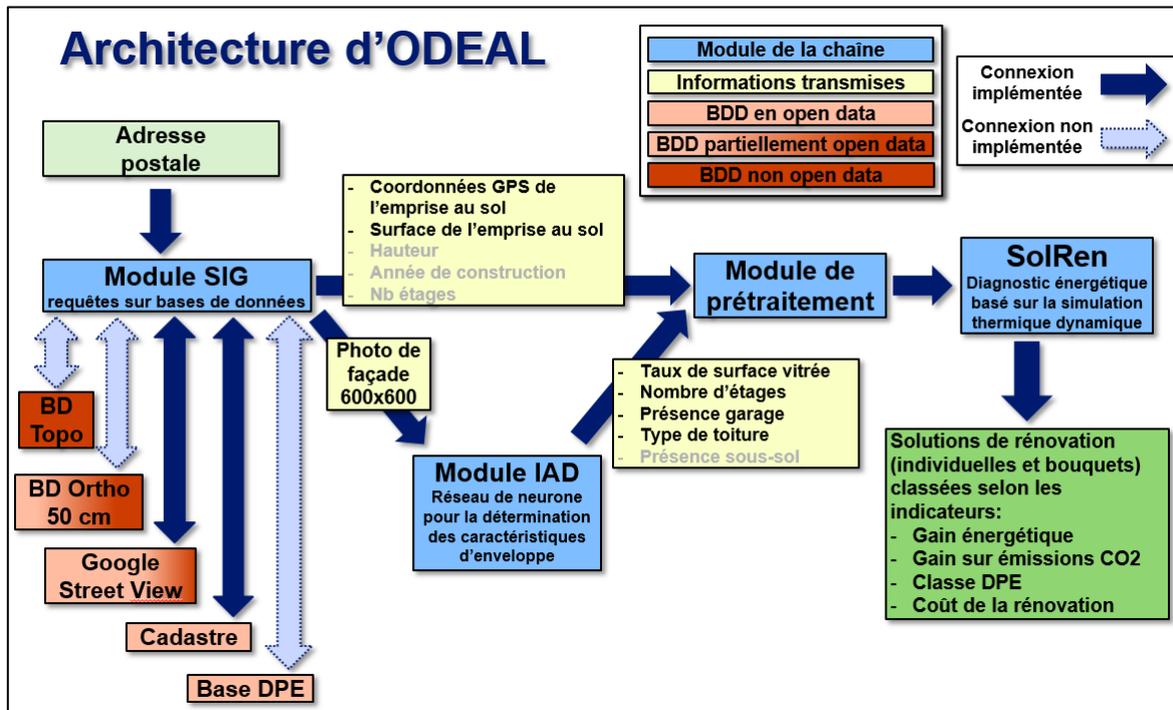


Figure 1: Architecture simplifiée de l'outil complet ODEAL

Le cœur de calcul pour la simulation thermique dynamique est le logiciel DiagSérie, un outil métier développé par EDF R&D qui sera détaillé dans les paragraphes suivant. Ce logiciel est encapsulé dans le module SolRen. On peut voir ODEAL comme une succession d'étapes de prétraitement et de post-traitement autour du cœur de calcul DiagSérie.

L'architecture simplifiée de l'outil est représentée sur la Figure 1. On peut y voir les différents modules et leurs interactions. Certaines interactions ou fonctionnalités ont été prévues mais non implémentées faute de temps, elles sont matérialisées par les flèches bleu clair. L'architecture fait également apparaître les informations transmises en sortie des modules (avec en gris les caractéristiques qui ne sont pas encore extraites).

Deux autres travaux ont été menés dans le cadre de la tâche 2.2 mais non intégrés à ODEAL compte tenu de leur maturité à date de diffusion de ce document (Décembre 2020), ils seront néanmoins présentés ici :

- Identification de la période de construction par application d'un réseau de neurone à la photo de façade
- Identification des caractéristiques du logement (enveloppe et usages) par analyse d'une courbe de charge électrique au pas horaire pour les maisons 100 % électrique (chauffage et eau chaude sanitaire)

1.4 Plan du rapport

Ce rapport présente tout d'abord les modules intégrés dans ODEAL dans les trois chapitres qui suivent cette introduction : le module SIG, le module IAD, et SolRen. Le fonctionnement des différents modules sera illustré par l'application de l'outil sur une adresse réelle : le 5 rue du Docteur Pouillot, à Melun (77).

Deux chapitres sont ensuite consacrés aux deux travaux qui n'ont pas été intégrés à ODEAL : l'identification de la période de construction par réseau de neurone, et l'audit numérique par analyse de courbe de charge électrique.

Enfin, un chapitre de conclusion reprendra les principales avancées et présentera les limites actuelles du démonstrateur ODEAL.

2 Module SIG : interrogation des bases SIG en open data

Le module SIG actuellement implémenté permet d'interroger les bases données ouvertes suivantes :

- Base des adresses normalisées par département au format csv,
- Données cadastrales des parcelles par commune au format json,
- Données cadastrales des bâtiments par commune au format json,

ainsi que de l'API Google StreetView, qui autorise un accès gratuit jusqu'à un certain plafond de consultation non contraignant l'utilisation d'ODEAL comme démonstrateur (de l'ordre de 10 000 à 20 000 consultations par mois).

Le module SIG prend en entrée une adresse sous la forme de deux paramètres, une chaîne de caractère « Numéro de rue + Nom de rue + Ville » ainsi que le numéro de département, et va interroger les bases ci-dessus pour récupérer toutes les informations nécessaires.

D'un point de vue informatique, les différents algorithmes développés et détaillés ci-dessous ont fait l'objet de codes Python intégrables dans un environnement logiciel complet.

2.1 Traitement des informations de type cadastral

Les opérations suivantes sont réalisées :

- Le numéro de département permet de télécharger la base d'adresses correspondante et de créer des fichiers de données d'adresses par commune, de façon à ce qu'ils puissent être réutilisés pour une adresse dans le même département.
- Un traitement du texte de l'adresse est réalisé pour identifier dans un premier temps la ville, puis, à partir du fichier d'adresses correspondant à cette ville, la rue et le numéro de rue. Une fois l'adresse identifiée, les coordonnées longitude/latitude de référence correspondantes vont être utilisées dans les étapes suivantes.
- Le fichier des données cadastrales des parcelles correspondant à la ville est alors téléchargé. L'objectif suivant est de relier la parcelle à l'adresse par une association géométrique entre les longitude/latitude de l'adresse et celles du polygone correspondant à l'adresse. Une attention particulière a été portée au cas de parcelles fortement non convexes comme des chemins privés via des critères géométriques.
 - o Les fichiers json ne contiennent pas l'information de référencement cadastral classique (par exemple AM450), c'est ce qui oblige à faire des recherches exclusivement géométriques
- Une fois la parcelle déterminée, il faut ensuite rechercher les bâtiments associés à la parcelle. Ceci passe à nouveau par des considérations géométriques : dans un premier temps, un test de distance est réalisé entre les centres d'un bâtiment donné et de la parcelle pour éliminer les bâtiments trop loin de la parcelle, puis un test d'appartenance plus fin est réalisé : des désajustements entre les fichiers de parcelles et des bâtiments sur les frontières impliquent qu'il faut prendre une marge sur l'appartenance d'un bâtiment à la parcelle.
 - o Il est à noter que les formats json ne sont pas complètement unifiés entre départements : il faut ajouter un test systématique

Une fois que l'on dispose de la parcelle et des bâtiments associés, il faut convertir les coordonnées en longitude/latitude en coordonnées métriques pour obtenir des dimensions exploitables pour la modélisation de la maison, en particulier la surface de l'emprise au sol.

Ci-dessous les informations issues du Cadastre pour l'adresse 5 rue du Docteur Pouillot à Melun (77).

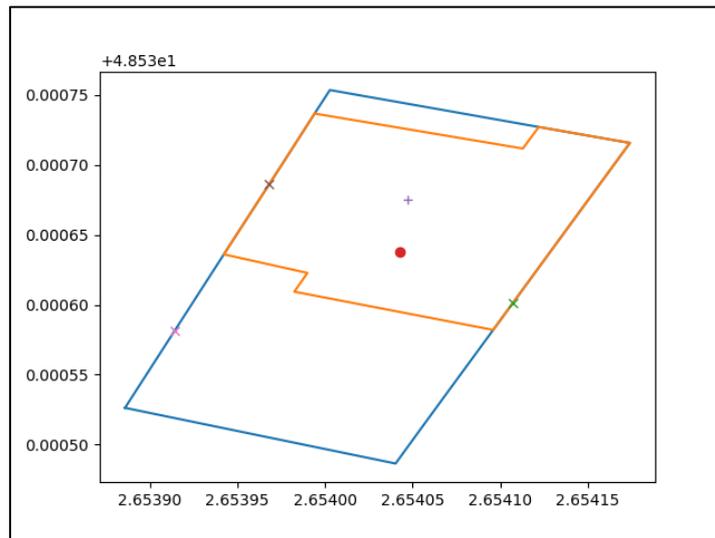


Figure 2 : Contour de la parcelle cadastrale (bleu) et du bâtiment situé sur la parcelle (orange) correspondant au 5 rue du Docteur Pouillot, à Melun

Tableau 1 : Extrait des données cadastrales

Coordonnées angulaires de l'adresse (longitude, latitude)	
2,654042954	48,53063795
Surface(s) du/des bâtiments	
Bâtiment n°1 [m ²]	
181,80	
Coordonnées métriques du/des bâtiment(s) dans un repère local	
Bâtiment n°1	
-9,30	0,00
-9,30	-3,97
-9,30	-11,77
-5,43	-11,77
-5,43	-13,35
3,59	-13,35
3,59	-5,97
3,59	2,49
-0,50	2,49
-0,50	0,66
-9,80	0,66

Le point rouge sur la Figure 2 correspond aux coordonnées angulaires du point de référence correspondant à la parcelle qui est spécifié dans le cadastre. Le « + » correspond au centre du bâtiment défini comme la moyenne des coordonnées angulaires définissant le bâtiment. Les « x » correspondent aux milieux des segments des parcelles qui sont les plus proches du point rouge, c'est un moyen de définir les points de vue pour les photos en définissant une direction allant du point de référence cadastre à ces milieux.

Dans le cas présent, comme le point rouge est à l'intérieur de la parcelle, ce n'est pas le bon moyen de définir le point de vue pour l'image Google : on considère une alternative utilisant la direction donnée par la ligne passant par le centre du bâtiment et le centre de la parcelle. Des tests permettent d'éliminer les points de vue qui ne seront pas bons, mais ça ne marche pas toujours.

Concernant les coordonnées métriques du Tableau 1, elles correspondent à une double conversion : d'abord le passage de coordonnées angulaires en coordonnées métrique Lambert 93, puis la rectification des coordonnées pour récupérer des angles droits.

2.2 Récupération des photos de façade depuis Google Street View

L'objectif est d'obtenir un point de vue ciblé vers la maison, ce qui revient à chercher un point de prise de vue, une orientation et un angle d'ouverture, sous la contrainte des points de prise de vue disponibles sous Google. Deux approches sont possibles, sachant que selon les cas l'une ou l'autre pourra ne pas donner de résultat :

- Une approche « géométrique » : partant du polygone décrivant la parcelle, on utilise le côté le plus proche du point associé à l'adresse et le point de prise de vue est alors donné par le milieu de ce côté et la direction par la normale intérieure au côté.
- Une approche « cadastre » : partant des coordonnées du point défini dans la base d'adresses, la direction est donnée par le vecteur reliant ce point au centre du bâtiment principal.

Des tests de validité complémentaires géométriques sont faits pour ne pas rechercher d'images qui ne correspondraient pas à la maison.

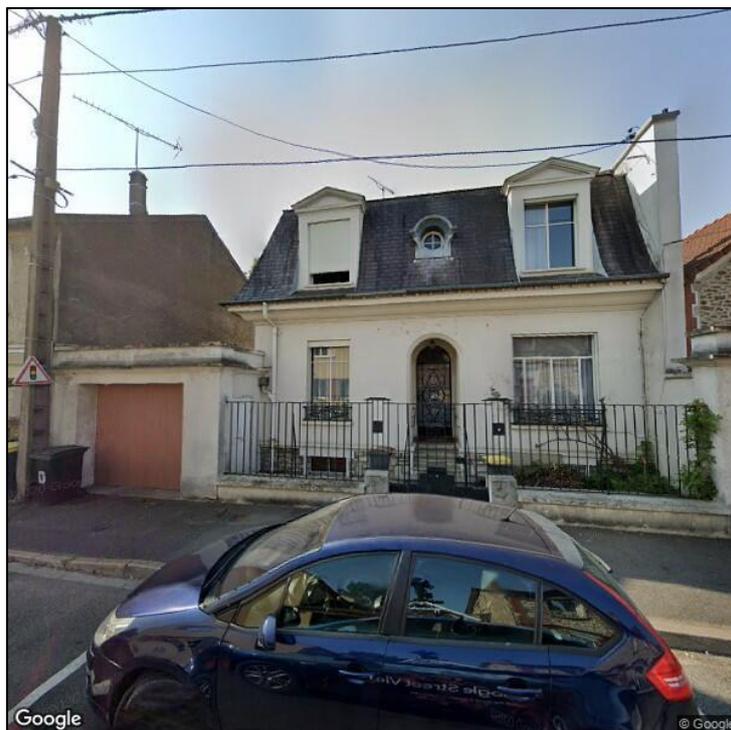


Figure 3 : Vue Google Street View du 5 rue du Docteur Pouillot à Melun (approche « géométrique » ou approche « cadastre »)

La Figure 3 donne la photo Google Street View pour l'adresse 5 rue du Docteur Pouillot à Melun (77). Les deux approches ci-dessus donnent exactement la même vue pour cette adresse, et cette vue est pleinement exploitable. Ce n'est pas systématiquement le cas pour toutes les adresses testées : selon les cas on peut avoir des vues identiques ou non, et toutes les vues ne sont pas exploitables.

Les figures ci-dessous illustrent plusieurs cas rencontrés lors de tests systématiques sur une rue de Melun :

- Figure 4 (75 rue du Docteur Pouillot) : les deux vues sont exploitables, de qualité comparable
- Figure 5 (67 rue du Docteur Pouillot) : les deux vues sont potentiellement exploitables, mais l'une est bien meilleure
- Figure 6 (27 rue du Docteur Pouillot) : une seule vue est exploitable
- Figure 7 (42 rue du Docteur Pouillot) : aucune vue n'est exploitable

Dans l'état actuel, on ne sait pas prévoir a priori quelles seront les vues exploitables, et aucun test n'a été implémenté pour vérifier si la vue est exploitable ou non. Les deux vues sont systématiquement exportées et traitées, et il faut ensuite vérifier manuellement quelle vue est la plus intéressante. Une automatisation de ce traitement serait à envisager dans le futur.



Figure 4 : Vue Google Street View du 75 rue du Docteur Pouillot à Melun : approche « géométrique » (gauche) ou approche « cadastrale » (droite)



Figure 5 : Vue Google Street View du 67 rue du Docteur Pouillot à Melun : approche « géométrique » (gauche) ou approche « cadastrale » (droite)



Figure 6 : Vue Google Street View du 27 rue du Docteur Pouillot à Melun : approche « géométrique » (gauche) ou approche « cadastre » (droite)



Figure 7 : Vue Google Street View du 42 rue du Docteur Pouillot à Melun : approche « géométrique » (gauche) ou approche « cadastre » (droite)

3 Module IAD : détermination des caractéristiques de l'enveloppe par intelligence artificielle

Le module IAD permet, à partir d'une photographie d'extérieur de maison individuelle transmise par le module SIG, d'en extraire ses caractéristiques visuelles. Les caractéristiques sont alors transmises au module SolRen pour mettre en données le diagnostic de la maison individuelle et réaliser des simulations numériques de consommation énergétique.

Le module IAD fonctionne en 2 étapes :

1. La reconnaissance des éléments de façade visibles dans l'image : celle-ci s'appuie sur trois réseaux de neurones indépendants spécialisés dans la reconnaissance de certains éléments.
2. L'extraction de caractéristiques utiles pour le diagnostic énergétique : cette étape recoupe l'ensemble des éléments détectés par les réseaux de neurones pour en déduire des caractéristiques de plus haut niveau.

3.1 Reconnaissance des éléments visibles de l'enveloppe

La première étape avant de se lancer dans la mise-en-place de réseau de neurones est de constituer une base d'exemples d'images associées à leur vérité terrain. Pour cela, EDF a réalisé la labélisation d'une base de données de 12666 images extraites par l'API Google Street View avec le module SIG.

Pour extraire les informations nécessaires en entrée du module SolRen, nous proposons de passer par une étape intermédiaire exploitant une reconnaissance de chaque élément composant les maisons individuelles à partir de réseaux de neurones pour la **segmentation par instance**. Dans cette catégorie de réseau de neurones, visant la reconnaissance d'objets dans les images, on distingue généralement quatre tâches (voir Figure 8) :

- **la détection d'objets** consiste à localiser une ou plusieurs classes d'objets et leurs positions dans une image,
- **la segmentation par instance** consiste, pour chaque objet, à avoir une classification au pixel près. Ainsi pour chaque objet détecté un contour de celui-ci est estimé.
- **la segmentation sémantique** consiste à classer suivant une¹ ou plusieurs classes d'objets chaque pixel d'une image. Toutefois ici, il n'y a pas de distinction entre deux objets d'une même classe.
- **la segmentation panoptique** est l'association de la segmentation sémantique et de la segmentation par instance où chaque pixel de l'image est classifié tout en distinguant des objets différents (en leur attribuant un identifiant).

Lors de nos investigations, nous avons identifié de meilleures performances avec une décomposition de la problématique en deux sous-problèmes indépendants pour la segmentation par instance sur deux ensembles disjoints de classe. Ainsi un premier réseau de segmentation par instance vise à détecter les classes de façade et de toiture alors que le second réseau a pour objectif de détecter les classes de fenêtres, volets, portes d'entrée, portes de garages et cheminées. Les deux réseaux de neurones pour la segmentation par instance reposent sur la même architecture : **Mask-RCNN**² développée par Facebook AI Research.

¹ En général, on ne tient pas compte de la classe « arrière-plan » qui regroupe l'ensemble des classes non inclus dans la liste des classes que l'on cherche à détecter.

² K. He et al, «Mask R-CNN» Proceedings of the IEEE international conference on computer vision, pp. 2961-2969, 2017.

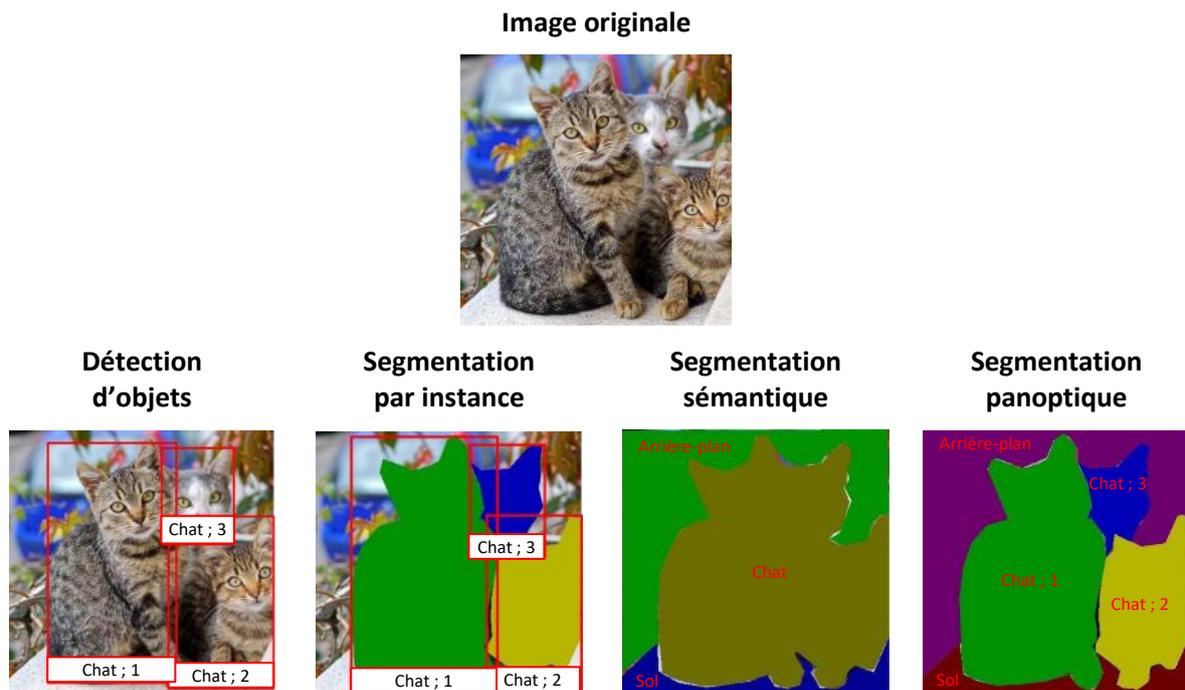


Figure 8 : Illustrations des quatre tâches de labélisation des images – Source : Medium3

Dans le cas où plusieurs maisons individuelles sont visibles sur une image, nous avons également mis en place un réseau de neurones de détection de maison individuelle (cf l'exemple de détection d'objets en Figure 8) afin d'associer chaque élément détecté par les réseaux de segmentation à une maison individuelle seulement. Ici, le réseau de neurones utilisé est **Faster-RCNN**⁴. Il repose sur une architecture identique à celle de Mask-RCNN à l'exception que Mask-RCNN intègre, en plus, une partie qui réalise une régression pour raffiner la classification au niveau pixelique.

Pour l'entraînement et l'évaluation des performances des trois réseaux de neurones, la base de données est découpée en trois ensembles :

- la base d'entraînement (80 % de la base de données) utilisée pour l'entraînement du réseau, c'est-à-dire l'ajustement de ses poids à partir d'une optimisation mathématique,
- la base de validation (10 % de la base de données) utilisée pour évaluer les performances du réseau de neurones en fonction d'un jeu de paramètres. Dans la pratique, on réalise plusieurs entraînements en fonction de différents paramètres (paramètres de la fonction d'optimisation, paramètres de l'architecture du réseau, ...) ajustés « manuellement ». Au final, on conserve le jeu de paramètres qui obtient les meilleures performances sur la base de validation,
- la base de test (10 % de la base de données) utilisée pour évaluer les performances du réseau non biaisées sur des données qui n'ont pas été impliquées dans l'entraînement (c'est-à-dire l'ajustement des poids du réseau) ou dans le choix des paramètres.

La Figure 9 présente les résultats qualitatifs des trois réseaux de neurones.

³ <https://medium.com/@danielmechea/what-is-panoptic-segmentation-and-why-you-should-care-7f6c953d2a6a>

⁴ Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2016): 1137-1149.



Figure 9 : Résultats obtenus pour les trois réseaux de neurones. A gauche : le résultat de la détection de maison individuelle. Au centre : le résultat de la segmentation des catégories façades et toitures. A droite : le résultat de la segmentation des catégories fenêtres et autres éléments

Ainsi pour une image, nous obtenons à la sortie des trois réseaux de neurones :

- le nombre de maisons individuelles, leur position et leur emprise dans l'image,
- une segmentation au pixel près des caractéristiques de la maison individuelle. Cette segmentation nous permet alors d'estimer des ratios de surface (par exemple le ratio de surface de fenêtres par rapport à celui de la façade ; cf §3.2)

3.2 Extraction des caractéristiques utiles pour le diagnostic énergétique

A partir des résultats des trois réseaux de neurones présentés précédemment, nous souhaitons en déduire des caractéristiques de haut-niveau qui seront utilisées en entrée du module de simulation énergétique SolRen.

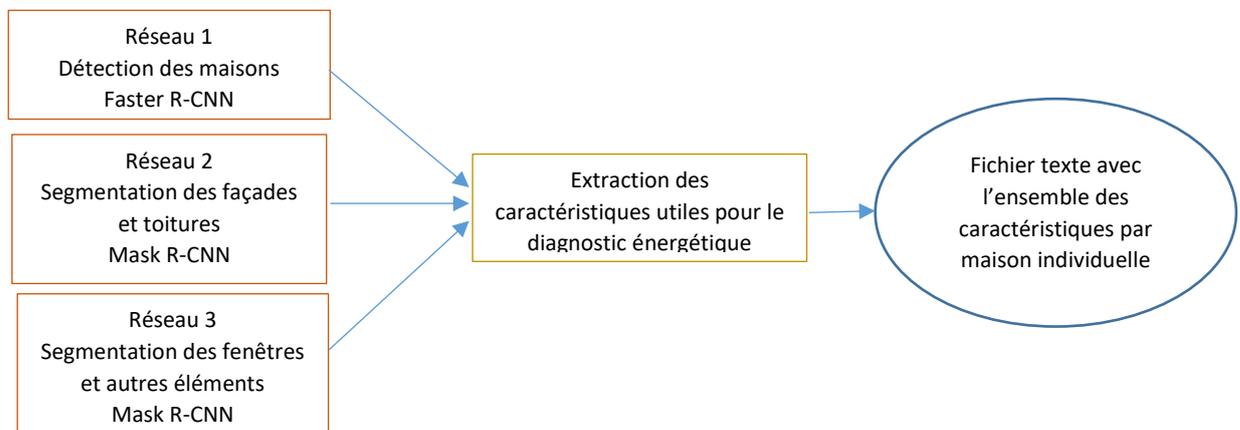


Figure 10 : Schéma du processus de traitement des images par le module IAD

Pour cela, on attribue à chaque prédiction des deux réseaux de neurones de segmentation par instance (fenêtres, façades, ...) une maison individuelle détectée par le réseau de détection de maisons. L'association se fait par intersection entre la boîte englobante issue du réseau de détection et les sorties des réseaux de segmentation.

Il reste ensuite à déduire pour chaque maison individuelle :

- **Taux de vitrage** : pour chaque façade, on calcule l'intersection avec les fenêtres détectées. On calcule ensuite un ratio en pixel entre les fenêtres et une façade.
- **Nombre d'étages** est estimé pour chaque maison à partir de la répartition en hauteur des fenêtres et portes.

- **Type de toiture** : de la même manière que pour le calcul du taux de vitrage, on calcule, pour chaque toiture, l'intersection avec des fenêtres. Si au moins une fenêtre est détectée alors on considère la toiture comme aménagée.
- **Présence d'un garage** : pour chaque maison, on vérifie si une porte de garage a été détectée.

Au final, à la sortie du post-traitement des réseaux de neurones, nous générons un fichier JSON avec les caractéristiques des maisons détectées dans l'image. Un exemple de fichier JSON est présenté en Figure 11 à partir des résultats des réseaux de neurones présentés précédemment (Figure 9).

<pre>{ "house": [{ "ID": 1, "Nombre facades": 3, "Facades": [{ "ID": 1, "Taux vitrage": 0.15081144880495723, "Nombre fenetres": 3 }, { "ID": 2, "Taux vitrage": 0.0, "Nombre fenetres": 0 }, { "ID": 3, "Taux vitrage": 0.0, "Nombre fenetres": 0 }], "Garages": { "Nombre garages": 1, "Differentes niveaux": false }, "Sous-sol": "Bool pas implemente", "Presence cheminee": true, "Nombre etages": 1, "Toiture amenee": true }, </pre>	<pre>{ "ID": 2, "Nombre facades": 0, "Facades": [], "Garages": {}, "Sous-sol": "Bool pas implemente", "Presence cheminee": false, "Nombre etages": 0, "Toiture amenee": false }]</pre>
---	---

Figure 11 : Fichier JSON produit en sortie du post-traitement des réseaux de neurones pour l'extraction de caractéristique à partir des résultats présentés en Figure 9.

Le module IAD permet donc de déterminer, à partir d'une photo de maison individuelle :

- La surface apparente (calculée en pixel sans correction des effets de perspectives), le nombre et la localisation par façade de tous les éléments de l'enveloppe visibles sur la photo : façades, toitures, fenêtres, volets, portes, portes de garage et cheminée.
- Des caractéristiques architecturales utiles pour le diagnostic énergétique : Taux de vitrage par façade (calculé par ratio de surface), type de toiture, nombre d'étage, présence d'un garage.

Le module ne permet néanmoins pas de récupérer toutes les caractéristiques de l'enveloppe du bâtiment à partir d'une seule photo, en particulier les caractéristiques des façades qui ne sont pas visibles sur cette photo ne sont pas accessibles.

3.3 Perspectives

Nous identifions principalement deux pistes d'amélioration pour ces travaux :

1. **Prolonger et améliorer l'étape de post-traitement des réseaux de neurones en vue de l'extraction des caractéristiques.** Actuellement, nous n'avons pas encore implémenté la présence ou non de sous-sol. De plus, cette partie de travail peut être améliorée. En effet, certaines particularités ne sont pas encore prises en compte comme par exemple la présence de volet dans le calcul du taux de vitrage. Ainsi en fonction que les volets soient ouverts ou fermés, cela aura une incidence sur le taux de vitrage.
2. **Conversion des surfaces en pixel en une surface en mètre.** Les surfaces en sortie de réseau sont calculées en comptant le nombre de pixels. En intégrant les informations de dimension de l'emprise au sol de la maison individuelle, extraites du module SIG, nous pourrions convertir la surface en pixel en une surface en mètre.
3. **Le redressement des images.** Aujourd'hui, les surfaces sont estimées à partir des images directement extraites de l'API Google Street View. Ces dernières sont donc biaisées du fait de l'effet de perspective dans certains cas. Ainsi, nous souhaitons redresser les images afin de corriger les effets de perspectives.

4 SolRen : diagnostic énergétique et évaluation des solutions de rénovation

SolRen est le module utilisé en dernière étape de l'outil ODEAL pour réaliser le diagnostic énergétique du bâtiment dans son état initial, et évaluer l'impact de différentes solutions de rénovation, seules ou en bouquet. Il est écrit en python, et fait appel à DiagSérie pour la simulation thermique dynamique. DiagSérie est un exécutable qui utilise comme interface d'entrée et de sortie des tableaux Excel.

La fonctionnalité d'évaluation de différentes solutions de rénovation énergétique est plutôt du ressort de la tâche 2.3, néanmoins cette fonctionnalité étant intégrée dans le même module que le diagnostic de l'état actuel qui ressort de la tâche 2.2, l'ensemble du fonctionnement de ce module est présenté dans ce chapitre.

4.1 Cœur de calcul pour la simulation thermique dynamique: DiagSérie

Le cœur de calcul utilisé pour faire le diagnostic énergétique est DiagSérie. C'est un outil de simulation thermique dynamique simplifié, développé par EDF R&D depuis plusieurs années.

DiagSérie effectue l'estimation des consommations d'un logement (maison ou appartement) par usage (chauffage, ECS, climatisation, autres usages) au pas de temps horaire à partir de données simples dites « données client » et produit des données complémentaires (coûts, émissions de CO₂, pistes de gains) exploitées par les outils supports de services développés autour de DiagSérie.

Les principaux composants du cœur de calcul DiagSérie sont les suivants :

- un solveur dynamique SimFast qui calcule au pas de temps horaire les consommations par usage (chauffage, climatisation, eau chaude, autres usages,...) ;
- des modules de prétraitement (appelés façades) qui transforment les données client en données natives du solveur ;
- des modules de post-traitement qui produisent des données complémentaires (émissions CO₂, facture énergétique, coût des solutions...);
- un module 3CL qui calcule les étiquettes énergie et climat selon la méthode réglementaire ;
- un ensemble de données intermédiaires utilisées par les différents solveurs et modules.

4.2 Solveur dynamique Simfast

Le solveur dynamique SimFast a été créé par EDF R&D. Il s'appuie sur des modèles réduits paramétrables de logements (aussi appelés « boîtes grises » ou « boîtes semblables »). Ces modèles réduits permettent avec un nombre limité de paramètres de représenter le comportement thermique d'un bâtiment, et d'obtenir des résultats de simulation avec un temps de calcul très limité (inférieur à la seconde).

Le solveur comporte 5 modèles réduits différents : maison sur terre-plein, maison sur vide-sanitaire, appartement en rez-de-chaussée, appartement intermédiaire, appartement sous toiture. Ils sont le résultat de l'application de deux méthodes successives sur des modèles détaillés de logements (construits sous CLIM2000) : méthode des plans d'expérience pour générer des modèles linéaires correspondants à chaque logement type, puis méthode de Moore pour la réduction de ces modèles.

Le solveur Simfast effectue des simulations en régime dynamique, en monozone (tout le logement constitue une seule zone thermique à température homogène), et avec un pas de temps d'une heure. La simulation tient compte de la configuration du logement (géométrie et performance thermique de

l'enveloppe), de la météo (température extérieures et apports solaires), des apports internes dus à l'occupation et aux équipements énergétique (électro-ménager, éclairage, électronique...)

4.3 Evaluation des solutions de rénovation : SolRen

Comme indiqué sur le schéma de la Figure 1, un petit module de prétraitement sert à préparer l'exécution de SolRen. Ce module sert simplement à récupérer les données issues des modules SIG et IAD et à les mettre en forme pour SolRen, qui prend en entrée un tableau Excel formaté pour être lu par DiagSérie. Une fois cette étape de prétraitement réalisée, SolRen effectue les actions suivantes :

- Simulation de la performance énergétique du logement dans l'état initial
- Constitution des configurations correspondant à une chaque solution de rénovation appliquée individuellement parmi une liste pré-définie
- Simulation de la performance énergétique correspondant à ces configurations de solutions individuelles
- Sélection des meilleures solutions (selon leur ordre de mérite évalué par un critère choisi par l'utilisateur, par exemple coût du point d'amélioration d'un critère donné) et constitution de configurations correspondant à des bouquets associant 2, 3 ou 4 parmi les meilleures solutions individuelles
- Simulation de la performance énergétique correspondant à ces nouvelles configurations de bouquets de solution
- Génération des fichiers de résultats correspondants à toutes les simulations effectuées
- Export de la liste des configurations évaluées, avec les solutions individuelles et les bouquets de solution

5 Traitement IA pour la période de construction

5.1 Présentation de la démarche

L'objectif est, à partir d'une image de maison issue de Google Street View, de pouvoir lui associer une période de construction. Trois périodes de construction sont recherchées :

- Avant 1948,
- Entre 1948 et 1974,
- Après 1974.

Il s'agit donc de relier une information de type image à une information « abstraite ». On est dans une démarche de type apprentissage machine pour laquelle il faut une base de données permettant d'associer des images et des périodes de construction.

Un premier essai de création de base de données avait été fait en associant les fichiers de demandes de valeurs foncières avec les fichiers de ventes des notaires pour créer des couples (adresse, année) via un croisement des fichiers par des informations communes (prix, nombre de pièces, surface), mais les résultats n'ont pas été concluants, au sens où les prix ne peuvent pas être comparés exactement mais avec une marge (prise en compte ou non d'honoraires d'agence, achat simultané de terrains avec une maison) et parfois les nombres de pièces ne sont pas cohérents. L'apport a été de réaliser une base de photos Google qui a été utilisée dans la suite du processus.

Une avancée importante a pu être faite récemment avec la publication d'une base de données des Diagnostics de Performance Énergétique (DPE) fournissant des couples (adresse, année). Un recouplement a été fait entre la banque d'images réalisée précédemment et les DPE disponibles. Il existe environ 6400 images auxquelles une période de construction a pu être associée.

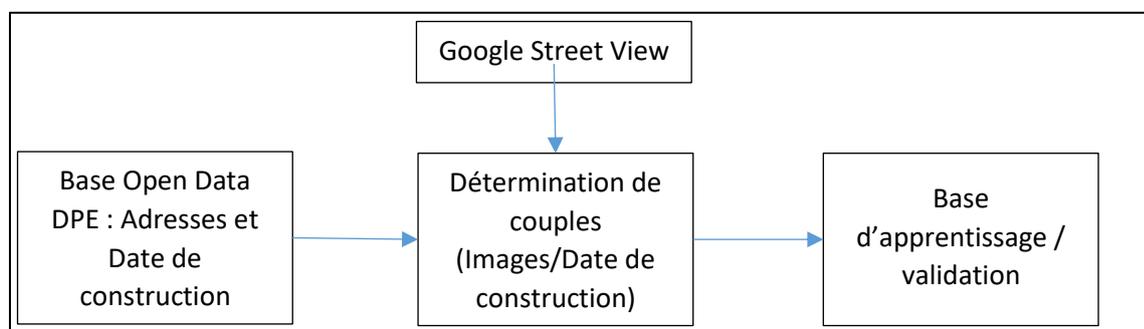


Figure 12 : Schéma de création des couples (images, année) avec les différentes sources de données

La répartition des images de notre échantillon en périodes de construction est la suivante :

- Avant 1948 : 42.6%
- Entre 1948 et 1974 : 19.2%
- Après 1974 : 38.2%

Il est intéressant de faire la comparaison avec la statistique réelle des maisons. Pour cela, les données fournies par IRIS agrégées à la France métropolitaine sont utilisées.

La répartition réelle du parc existant est la suivante :

- Avant 1945 : 25%
- Entre 1946 et 1970 : 22%
- Entre 1971 et 1990 : 29%
- Après 1990 : 27%



Pour une image donnée, pour chacun des 3 canaux de couleur, une intercorrélacion 2D normalisée va être réalisée entre l'image formée par la norme du gradient de l'image sur un canal donné et la norme du gradient de l'imagette sur le même canal. Le maximum d'intercorrélacion en valeur absolue va alors être stocké ainsi que la moyenne de l'intercorrélacion sur l'image. Une imagette va donc fournir 6 informations.

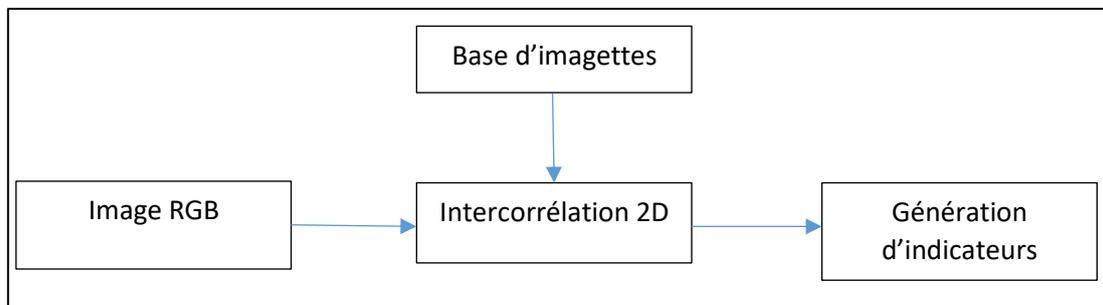


Figure 14 : Schéma de création des indicateurs synthétisant l'information des images RGB

Nous allons donc disposer d'une matrice de données avec pour chaque ligne les différentes informations de corrélation avec les imagettes. On est arrivé ainsi à un tableau de données qui va pouvoir être utilisé pour définir une démarche d'apprentissage. Les sorties pour l'apprentissage vont être les classes de périodes de construction (1 pour « avant 1948 », 2 pour « entre 1948 et 1974 » et 3 pour « après 1974 »). Une première sélection des colonnes est effectuée en calculant les corrélations avec le vecteur des classes : les corrélations supérieures à un seuil en valeur absolue (autour de 0.25, à affiner) sont les seules retenues. Ce seuillage conduit à ne conserver qu'environ 200 entrées sur 9361 calculées, ce qui implique qu'il faut une base d'images conséquente pour pouvoir avoir un ensemble d'imagettes de référence conséquent. Cette matrice est divisée aléatoirement entre la partie « apprentissage » (84% des données en moyenne) et la partie « validation » (16%). Cette division est répétée 100 fois pour voir la sensibilité à cette division.

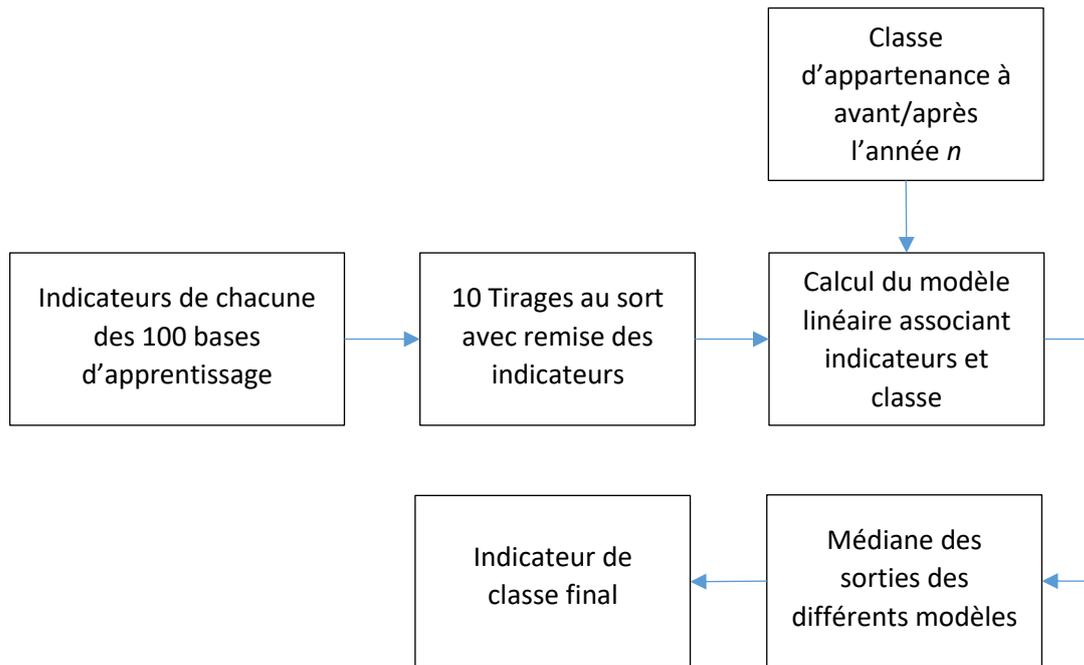


Figure 15 : Schéma global de création des modèles de période de construction

L'approche choisie est de créer des classifieurs binaires (avant/après 1948, avant/après 1974). Pour chaque classifieur binaire, un tirage aléatoire des colonnes est fait, puis un algorithme de moindres carrés est appliqué entre la matrice des coefficients et le vecteur des classes (pour le classifieur « avant/après 1948 », les éléments de classe 3 sont pris égaux à 2 et, réciproquement, « avant/après 1974 », les éléments de classe 1 sont pris égaux à 2). Dix tirages aléatoires sont appliqués, et la médiane est prise parmi les 10 sorties de classifieur et comparée par rapport à un seuil (1.5 pour le classifieur « avant/après 1948 » et 2.5 pour le classifieur « avant/après 1974 ») pour déterminer la classe. La démarche est illustrée sur la Figure 15.

Les Tableau 2 et Tableau 3 donnent les résultats en termes des critères de précision et rappel usuels en machine learning sur les ensembles de validation et d'apprentissage respectivement : la précision est définie comme le rapport entre le nombre d'éléments correctement attribués à une classe donnée et le nombre d'éléments attribués à une classe donnée et le rappel est défini comme le rapport entre le nombre d'éléments correctement attribués à une classe donnée et le nombre d'éléments total d'une classe donnée. Le critère F1 qui correspond à la moyenne harmonique des critères de précision et de rappel est aussi fourni par classement (« Avant/après »).

Les résultats entre les bases d'apprentissage et de validation sont proches, donc il n'y a pas de surapprentissage, mais les performances sont trop faibles.

Par définition, le hasard total donnerait une performance de 50% : il faut donc voir les pourcentages par rapport à ce seuil. Un très bon score serait d'atteindre 90% sur le critère F1 et un score correct 80% sur l'ensemble des classes en validation. Les résultats actuels montrent des performances de l'ordre de 73% sur les critères en validation : comme il existe un effet d'asymptote sur les performances, au sens où il faut de plus en plus d'images pour gagner des pourcents. Il faudrait aussi voir ce que ça donne lorsque des parties d'âge différent sont présents, ou même des rénovations qui conduisent à mélanger les styles.

Tableau 2 : Résultats en Précision/Rappel sur les blocs de validation

	Précision	Rappel	Critère F1
Avant 1948	69.6 +/- 5.2%	66.5 +/- 4.7%	72.2 +/- 3.3%
Après 1948	75.2 +/- 3.6%	77.7 +/- 3.9%	
Avant 1974	79.2 +/- 3.3%	82.5 +/- 4.0%	73.8 +/- 3.1%
Après 1974	69.1 +/- 5.6%	64.4 +/- 5.4%	

Tableau 3 : Résultats en Précision/Rappel sur les blocs d'apprentissage

	Précision	Rappel	Critère F1
Avant 1948	74.1 +/- 1.1%	71.6 +/- 1.3%	76.7 +/- 0.7%
Après 1948	79.6 +/- 0.7%	81.5 +/- 1.0%	
Avant 1974	82.3 +/- 0.7%	85.2 +/- 0.9%	78.2 +/- 0.6%
Après 1974	74.7 +/- 0.7%	70.6 +/- 1.5%	

5.2 Exemples d'application

Les images suivantes présentent des exemples de résultats, la vraie période étant donnée par la partie gauche du titre : quand, dans le titre, les périodes indiquées sont identiques, l'algorithme a bien fonctionné.



Avant 1948 - Avant 1948



Avant 1948 - Après 1948



Après 1948 - Avant 1948



Après 1948 - Après 1948



Avant 1948 - Avant 1948



Avant 1948 - Après 1948



Après 1948 - Avant 1948



Après 1948 - Après 1948



Avant 1974 - Avant 1974



Avant 1974 - Après 1974



Après 1974 - Avant 1974



Après 1974 - Après 1974



Avant 1974 - Avant 1974



Avant 1974 - Après 1974



Après 1974 - Avant 1974



Après 1974 - Après 1974



Les deux figures suivantes présentent des exemples d'images sur lesquelles la période de construction a été bien identifiée.



Figure 16 : Exemples d'images identifiées avant 1948



Figure 17 : Exemples d'images identifiées après 1948

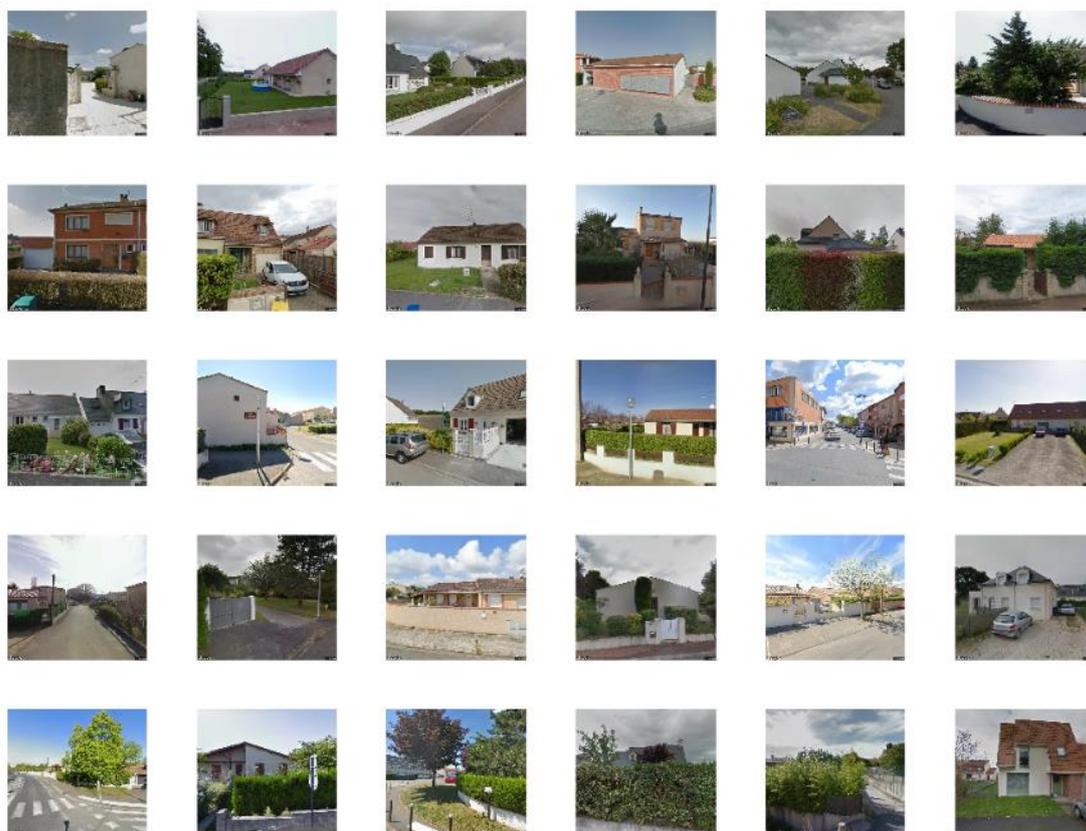


Figure 18 : Exemples d'images identifiées après 1974

5.3 Généralisation à la détermination de la période de construction d'une maison parmi un ensemble de périodes de construction données

Comme la méthode présentée ci-dessus repose sur la superposition d'indicateurs binaires sur le fait qu'une maison donnée soit construite avant ou après une date donnée, il est facile de généraliser à un ensemble quelconque de périodes en construisant autant d'indicateurs binaires qu'il existe d'années de séparation des périodes de construction. Il faudra vérifier quelle est la longueur minimale de période de construction en fonction de la base d'apprentissage qui permette d'avoir un nombre suffisant d'images permettant de contribuer à l'élaboration de chacun des indicateurs.

Cette généralisation n'aura naturellement de sens que si la méthode peut être améliorée du point de vue des performances.

5.4 Synthèse et pistes de poursuite

La démarche présentée présente des résultats qui doivent être améliorés. Pour cela, deux pistes existent :

- D'une part, poursuivre les intercorrélations avec les imagerie : il existe une base de 250 images avant 1980 et 250 images après 1980 sur laquelle des imagerie ont été extraites et dont on n'a exploité que 47 et 15 images respectivement, donc le calcul peut être poursuivi, mais cela suppose la constitution d'une base d'images correctement étiquetées plus étendue.
- D'autre part, la base d'images de maisons a été obtenue avant la création du module SIG qui doit fournir des images plus centrées sur les maisons et comprendre moins de végétation par exemple. Il faut donc que la base d'images plus étendue mentionnée au point précédent soit aussi plus adaptée aux exigences de l'apprentissage. La mise à disposition en open data de la

BD Topo donne également la possibilité d'améliorer la base d'apprentissage, puisque la période de construction est renseignée dans cette base de données pour plus de 90% des constructions, ce qui donne beaucoup plus de liberté pour la sélection des vues à intégrer dans cette base

- Ces deux pistes d'amélioration peuvent nécessiter la création d'un nouveau réseau de neurones, opérant en amont de la démarche, destiné à sélectionner automatiquement sur une image la partie utile (hors végétation) et à exclure automatiquement de l'échantillon les images inexploitable.

6 Audit Numérique par analyse de la courbe de charge électrique

6.1 Concept de l'Audit Numérique

L'Audit Numérique consiste à inverser la problématique habituelle de réalisation d'un audit énergétique du logement basé sur une collecte de données spécifiques et une analyse a posteriori du cas client.

Un objectif particulier de l'audit numérique est donc de réaliser un diagnostic en analysant la courbe de charge électrique du client et en lui posant un minimum de questions (dans cette première phase de l'étude, aucune question ne lui est posée). Cette courbe de charge serait alors comparée à des courbes de charges générées en amont à l'aide de jeux de données générées aléatoirement selon une approche statistique (afin d'envisager tous les cas possibles rencontrés sur le terrain) et d'un outil de simulation dynamique du bâtiment.

L'identification des courbes de charges ayant les caractéristiques les plus proches de celle du client permettrait alors d'estimer les caractéristiques du logement du client, ou à minima, d'avoir des jeux de données permettant de simuler un comportement thermique proche.

Dans un premier temps, ce sont les maisons individuelles thermosensibles au froid, c'est à dire avec un chauffage électrique uniquement (pas d'appoint avec une autre énergie) et sans système de climatisation, qui ont été ciblées. Selon les résultats, le champ d'étude pourra être élargi à d'autres maisons individuelles (avec chauffage non électrique, avec climatisation...) voire aux appartements collectifs dès lors qu'une courbe de charge issue d'un compteur « intelligent » puisse être récupérée.

6.2 Méthode

6.2.1 Construction d'une base de données de caractéristiques de courbes de charges

L'objectif est de créer une base de données de courbes de charge représentatives du parc immobilier en France. A priori, il est impossible de d'accéder précisément à toutes les caractéristiques des logements (surfaces, performances du bâti...) et habitudes des occupants (absence, températures de consignes...) qui conditionnent sa dynamique de consommation de chauffage.

Afin de créer une base de données suffisamment représentative et contenant le maximum de typologies de maisons individuelles qu'il est possible de rencontrer, on va générer près de 500.000 combinaisons aléatoires de jeux de données en propageant des incertitudes sur les paramètres d'entrée (surface, niveau d'isolation, températures de consigne...) à l'aide de lois uniformes. Ces lois s'appliqueront autant sur les valeurs quantitatives/continues (entre des bornes minimum et maximum déterminées à dire d'expert) que sur les valeurs qualitatives/catégorielles.

Paramètre	Choix possibles
G	Entre 0.5 et 1.8 (W/m ³ /°C)
Performance des systèmes de chauffage	Standard / Effet joule Norme NF Cat. C
Surface habitable	Entre 70 et 160 (m ²)
Surface chauffée	= Surface habitable
Eclairage	Classique / Fluo Compact / Mixte
Nombre d'occupants en journée	Entre 0 et 4
Durée des congés en hiver	Aucun / 1 Semaine / 2 Semaines / 3 Semaines
Durée des congés en été	Aucun / 1 Semaine / 2 Semaines / 3 Semaines

Ouverture quotidienne des fenêtres	Oui / Non
fermeture des volets la nuit	Oui / Non
type de volets	Persienne / Roulant / Bois
Volume du ballon de l'ECS	50 / 75 / 100 / 150 / 175 / 200 / 250 (L)
Habitudes ECS des occupants	Plutôt des douches / Des douches et quelques bains
Emplacement du ballon ECS	En volume chauffé / En local accolé
Jaquette Ballon ECS	Si emplacement = volume chauffé alors NON sinon OUI
pavillon mitoyen	Oui / Non
Etage	Oui / Non
Type de toiture	En pente avec combles perdus / En pente avec combles aménagés
Garage	Oui / Non
Forme du logement	En rectangle / en U / en L
Type de plancher	Sur sous-sol / Sur vide-sanitaire / Sur terre-plein
Ventilation	VMC autoréglable / VMR / Naturelle
Type de montant des ouvrants	Bois / Aluminium avec rupteur de ponts thermique / Aluminium / PVC
Surface Nord des ouvrants	Entre 2 et 20 (m ²)
Surface Sud des ouvrants	Entre 2 et 20 (m ²)
Surface Est des ouvrants	Entre 2 et 20 (m ²)
Surface Ouest des ouvrants	Entre 2 et 20 (m ²)
Année de simulation	Entre 2009 et 2018
Température de confort	Loi normale centrée autour de 21°C avec un écart type de 2°C
Température de réduit	Loi normale centrée autour de 18°C avec un écart type de 2°C
Jour type 1	Scenario de Température avec Tconfort et Tredit et horaires variables
Jour type 2	Scenario de Température avec Tconfort et Tredit et horaires variables
Puissance des équipements spécifiques	Entre 1 et 2000 (W)
Nombre heures de fonctionnement des équipements spécifiques	Entre 1 et 120 (heures/semaines)

La génération des 500.000 courbes de charges s'effectue en batch avec un outil de simulation thermique des consommations du bâtiment. Ici, l'outil utilisé est DiagSérie mais un autre outil équivalent pourrait convenir.

En parallèle des courbes de charges horaires, il est indispensable de récupérer aussi les données de températures extérieures correspondantes au pas de temps horaire.

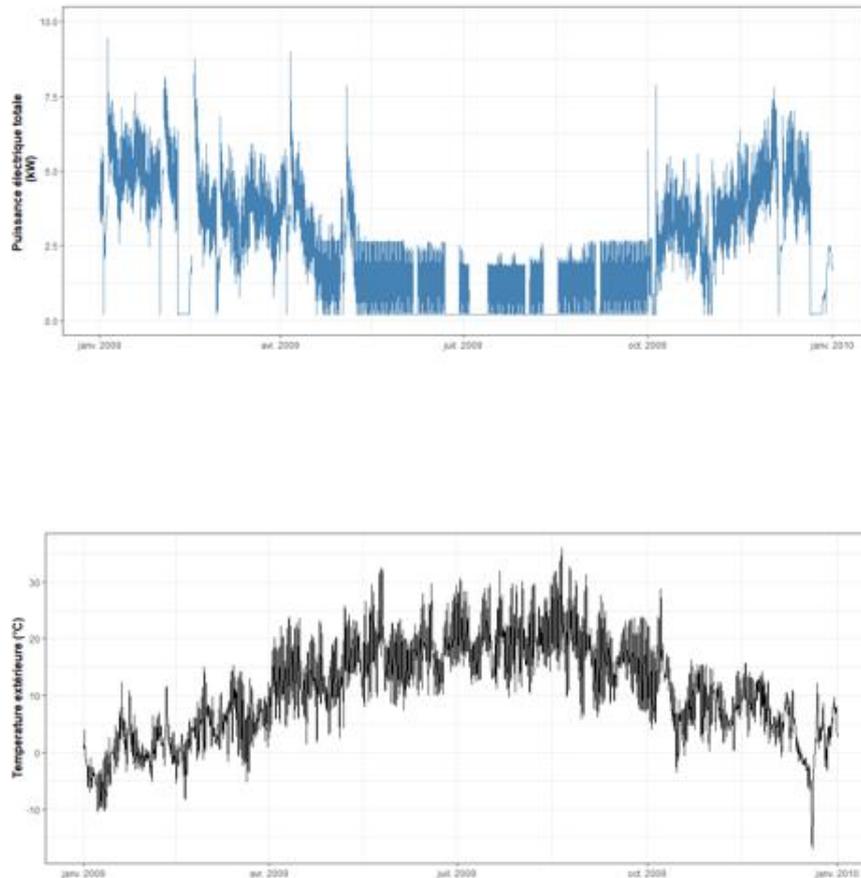


Figure 19: Exemple de sortie dynamique d'une simulation - Puissance électrique (haut) et température extérieure (bas)

Pour chaque courbe de charge, on va regrouper toutes les puissances horaires par valeur de température extérieure, avec un pas de 0.1 °C pour cette température extérieure, et calculer la moyenne de l'appel en puissance pour chacune des températures.

Ensuite, une régression linéaire par morceaux permettra de récupérer la signature énergétique du logement qui résume les caractéristiques de la courbe de charge à l'aide de seulement 3 points (les points en vert sur la Figure 20).

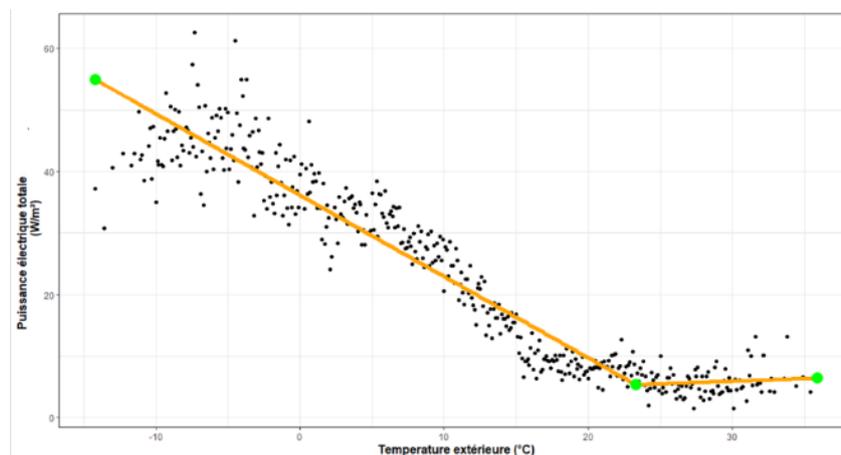


Figure 20 : Régression linéaire par morceaux sur la distribution des puissances électriques appelées en fonction de la température extérieure

6.2.2 Traitement de la courbe de charge client

L'objectif est de récupérer la courbe de charge client ainsi que les données de températures extérieures correspondantes au pas de temps horaire.

Ensuite, la signature énergétique du logement du client est calculée de la même façon qu'expliqué ci-dessus.

Dans le cadre de Smart-Réno, les courbes de charge client utilisées proviennent du panel Particuliers d'ICAME (Département *Innovation Commerciale Analyse des Marchés et de leur Environnement* d'EDF R&D).

6.2.3 Identification des signatures énergétiques les plus proches

L'objectif est de filtrer et de classer les signatures énergétiques de la base de données par ordre décroissant de proximité avec celle du client. On cherche donc à ne conserver que les signatures énergétiques simulées dont l'écart par rapport à celle du client est minimum.

Afin de ne pas extrapoler des valeurs en dehors des intervalles de calcul, on va restreindre l'intervalle des températures à l'intersection des données disponibles. Par exemple si la caractéristique du client est calculée sur l'intervalle $[-1^{\circ}\text{C}; +28^{\circ}\text{C}]$ et que les caractéristiques de la base de données sont calculés sur l'intervalle $[0^{\circ}\text{C}; +30^{\circ}\text{C}]$, on retiendra pour le calcul l'intervalle $[0^{\circ}\text{C}; +28^{\circ}\text{C}]$.

Ensuite, afin d'optimiser les temps de calcul, on va filtrer les caractéristiques trop éloignées, en conservant uniquement les caractéristiques dont les points sont compris dans un intervalle autour des points de la caractéristique client. Si on a par exemple pour le client $P(\text{Text} = 0^{\circ}\text{C}) = 3000 \text{ W}$, on ne va conserver que les caractéristiques de la base de données telles que $2100 \text{ W} < P(\text{Text} = 0^{\circ}\text{C}) < 3900 \text{ W}$, c'est-à-dire avec un écart de 30% - valeur modifiable)

Enfin, pour toutes les caractéristiques restantes, on va interpoler tous les points entre les valeurs bornes des Text, avec un pas de 0.1°C , et calculer l'erreur cumulée (RMSE) entre chaque point de la caractéristique client et celle de la base de données.

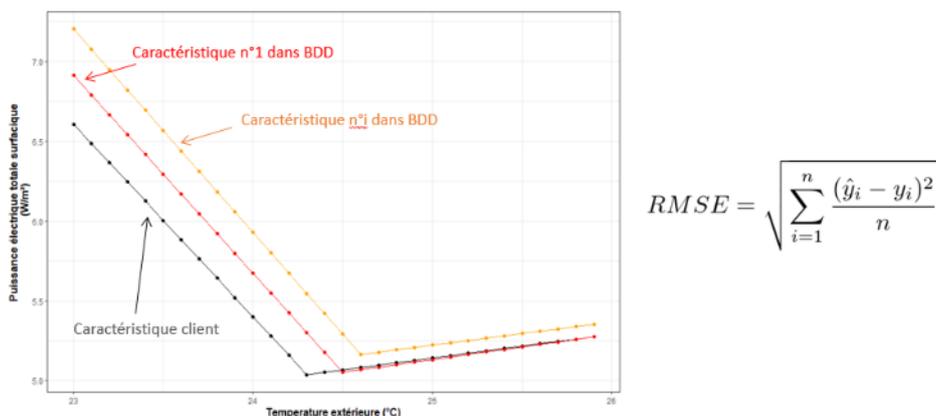


Figure 21 : Illustration de la comparaison entre la caractéristique client et celles de la base de données

Ci-dessous, quelques résultats graphiques présentant une signature énergétique client et les 5 simulations les plus proches.

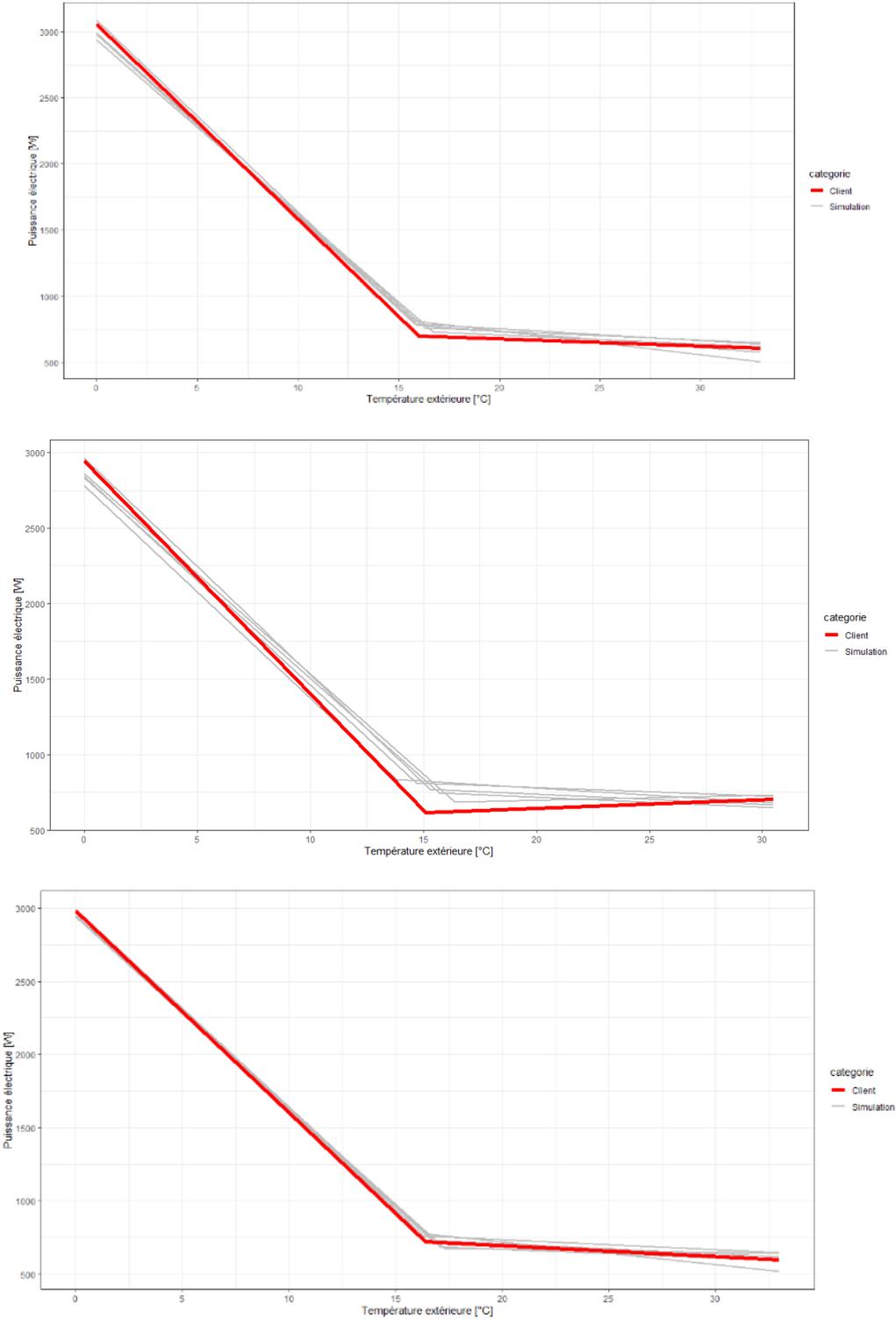


Figure 22: Exemple de caractéristiques de courbe client comparée à des caractéristiques de courbe de charges issues de simulation

6.2.4 Estimation des caractéristiques du logement du client

L'objectif est de pouvoir identifier les caractéristiques du logement et du comportement client à partir des caractéristiques des modèles présentant les signatures énergétiques les plus proches.

Sur les graphiques ci-dessous, l'estimation des surfaces a été essayée avec plus ou moins de modèles proches retenus : Les carrés bleus représentent les surfaces réelles déclarées par chaque client, les boîtes à moustache orange représentent la répartition des surfaces des simulations les plus proches

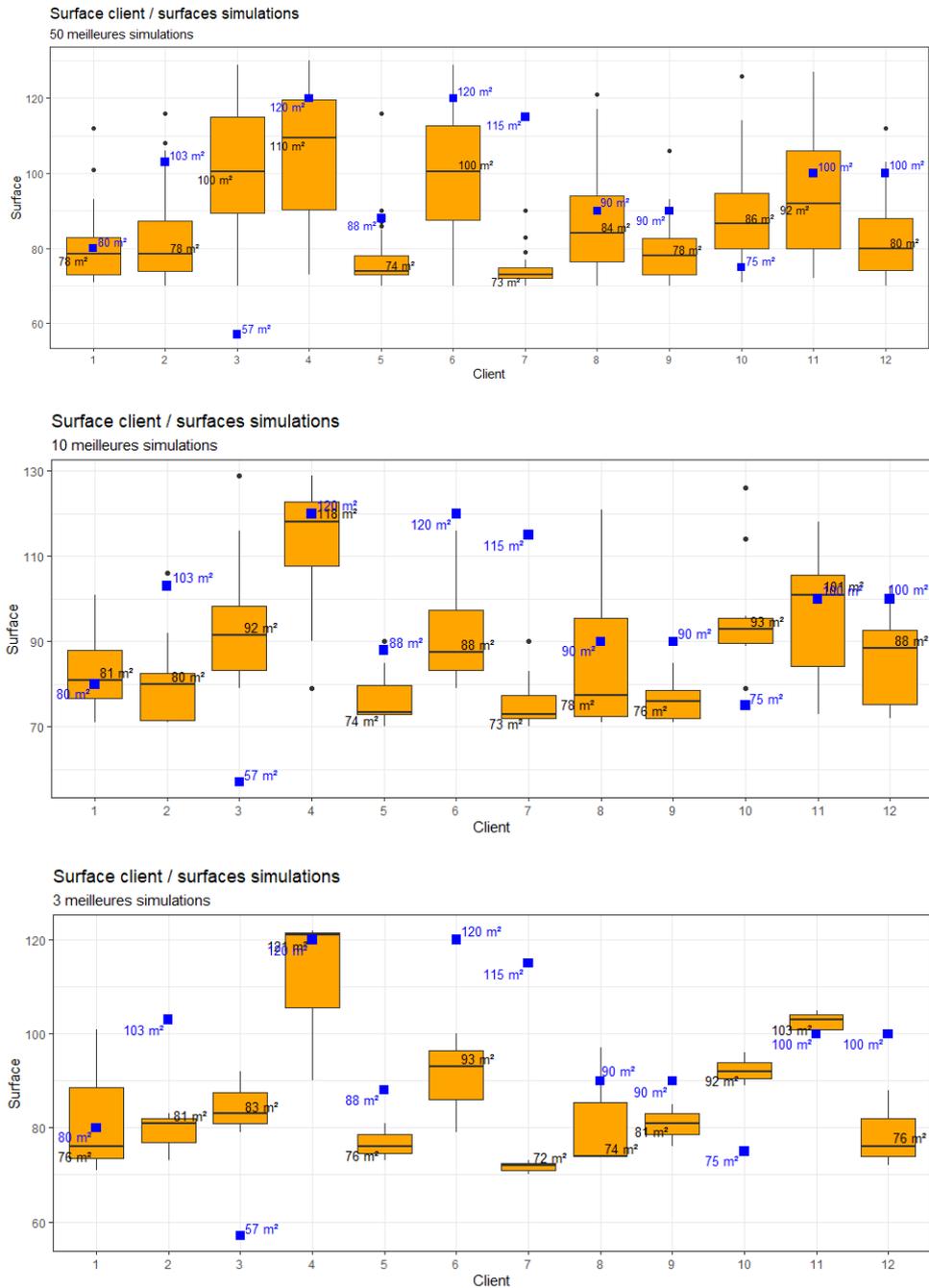


Figure 23 : Exemple de comparaison entre la surface réelle du logement du client et la distribution des surfaces simulées

Les graphes ne montrent pas de corrélation satisfaisante entre la surface réelle et la surface estimée via les courbes caractéristiques les plus proches. Les points 3 et 7 par exemple posent question, une analyse plus fine est nécessaire pour évaluer la sensibilité des algorithmes aux paramètres déterminants.

Il en est de même pour les températures de confort :

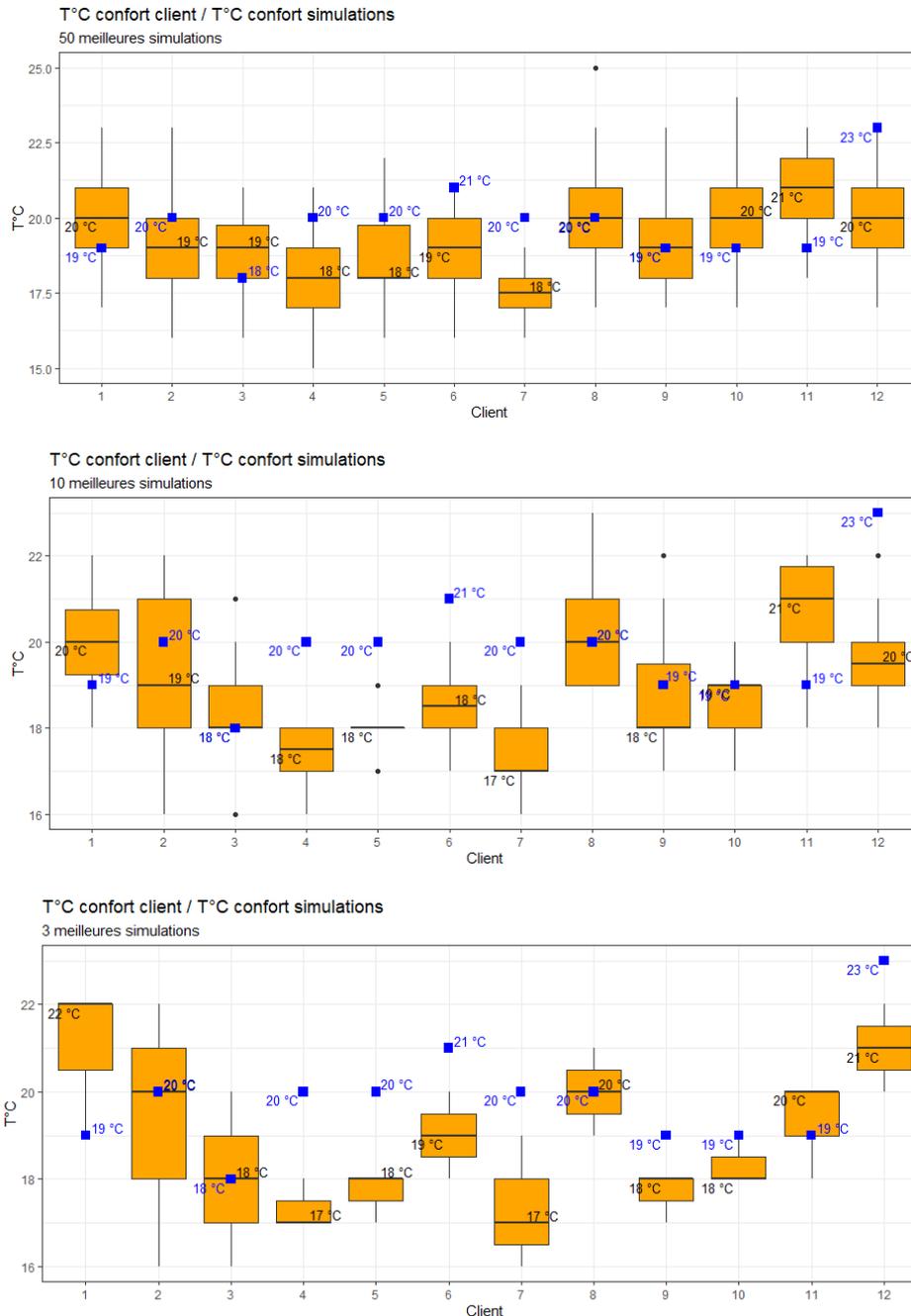


Figure 24 : Exemple de comparaison entre la température de confort du client et la distribution des température de confort simulées

A ce stade, il semble encore très aléatoire d'estimer les caractéristiques d'un logement et/ou des habitudes des occupants seulement à partir des modèles les plus proches. En effet, il est possible que des logements aient des signatures énergétiques proches tout en ayant des caractéristiques

différentes. En conclusion, les résultats obtenus démontrent la potentialité de l'approche, mais soulignent la nécessité d'en améliorer certains maillons.

Ces travaux complémentaires doivent porter sur l'analyse plus fine qui est nécessaire pour évaluer la sensibilité des algorithmes aux paramètres déterminants d'une part, et d'autre part, sur l'amélioration du score d'identification des courbes de charge simulées correspondant à celle du client. Pour ce dernier point, il conviendrait par exemple, d'encadrer plus précisément les caractéristiques du logement du client, en le questionnant sur des points clés comme la surface habitable de son bien, et/ou en positionnant ce bien par rapport à des typologies de maisons connues, et/ou en allant chercher les caractéristiques de la maison du client à partir de la collectes d'informations obtenues via le croisement de bases de données géo localisées (BD Topo, base cadastrale, ...) à l'instar des approches explorées par ailleurs par EDF R&D et pour partie développées dans le cadre du volet 2 de Smart Réno (voir les paragraphes précédents traitant d'ODEAL). Ce couplage avec les travaux précédemment présentés n'a pu être réalisé faute de temps, mais apparaît comme une piste prometteuse à instruire.

L'évaluation des gains associés à chaque geste de MDE est statistiquement robuste et peut d'ores et déjà être exploité pour qualifier l'impact des gestes de rénovation appliqués sur des maisons individuelles.

6.3 Perspectives

- Certains travaux restant à mener pourraient viser à **consolider les algorithmes et la chaîne de traitement** :
 - **Améliorer le modèle de régression** pour améliorer la précision. Une régression affine par morceaux (avec davantage de partitions), voire une régression polynomiale, pourrait être plus pertinente.
En outre, il pourrait être intéressant de regarder l'impact d'autres variables que la température extérieure (jour type, heure...)
 - **Améliorer le processus d'identification** des courbes les plus proches. Actuellement, le calcul se base sur des caractéristiques simplifiées de la courbe de charge mais il pourrait être intéressant d'analyser directement les courbes de charge brutes entre elles (ex : méthode « Dynamic Time Warping »). Attention, dans ce cas, il pourrait y avoir des difficultés en termes de stockage des données et en temps de calcul. Un travail sur le filtrage de la base de données de caractéristiques à partir de caractéristiques déclarées par le client (ex : surface) pourrait aussi être envisagée afin d'améliorer la précision des résultats.
 - Utiliser un **moteur de calcul de simulation thermique permettant plus de finesse** dans la définition des paramètres d'entrées, afin d'augmenter l'éventail des solutions de MDE proposées (localisation des rénovations sur une partie de l'enveloppe seulement par exemple)
- D'autres travaux pourraient concerner la limitation des informations à identifier via cette méthode en limitant le champ de l'identification aux seules caractéristiques qui ne peuvent pas être automatiquement obtenues autrement.
- Si les résultats sont concluants, il faudra dans un deuxième temps chercher à **étendre le champ d'application** à d'autres maisons individuelles (avec chauffage non électrique, avec climatisation...) voire aux appartements collectifs.

7 Conclusion

7.1 Etat actuel de l'outil

A fin 2020, l'outil ODEAL se présente sous la forme d'un démonstrateur d'outil de diagnostic automatisé pour maison individuelle, qui s'exécute en local sur un PC préalablement configuré. La configuration requise est synthétisée en Annexe 1. L'utilisation de ce démonstrateur se fait par simple appel à une unique fonction python prenant en entrée deux variables : une adresse (numéro de rue, nom de rue et ville) et un numéro de département. La sortie de ce démonstrateur est un diagnostic de la performance énergétique de la maison en l'état actuel, et une liste contenant des solutions de rénovation individuelles et des bouquets de solutions avec le gain apporté par chaque solution selon plusieurs indicateurs :

- Consommation énergétique
- Emissions CO2
- Etiquette énergie et CO2 selon la méthode réglementaire 3CL
- Coût de la rénovation

Ces données de sortie se présentent sous la forme d'un onglet dans un fichier Excel, et une partie de ces informations sont exportées dans un PDF.

7.2 Limites actuelles de l'outil et perspectives

Dans ce chapitre de conclusion, les limites observées dans la version actuelle seront listées, et pour chacune de ces limites, on identifiera l'impact sur la qualité de l'outil, et les perspectives associées en mentionnant les améliorations envisagées et la confiance que l'on a dans leur mise en place. Un tableau récapitulatif regroupe ces analyses pour chaque module actuellement implémentés dans ODEAL, puis pour les fonctionnalités sur lesquels du travail a été mené mais qui n'ont pas pu être intégrées à ODEAL.

7.2.1 Démarche générale d'ODEAL

Limite actuelle	Impact sur la qualité de l'outil	Amélioration à apporter	Confiance dans la mise en place de l'amélioration
Impossible de trouver la maison pour certaines adresses : lieux dits, copropriété...	Faible : pourcentage limité des adresses concernées	-	
Aucune vue exploitable dans Google Street View (maison cachée par de la végétation, une clôture, un autre bâtiment...)	Modéré : pourcentage limité des adresses concernées (à confirmer)	-	

7.2.2 Module SIG

Limite actuelle	Impact sur la qualité de l'outil	Amélioration à apporter	Confiance dans la mise en place de l'amélioration
Difficulté d'identifier la bonne maison pour les parcelles qui contiennent plusieurs constructions (garages, ateliers...)	Modéré : pourcentage limité des adresses concernées (à confirmer)	- Tri des empreintes par surface décroissante - Utiliser la BD TOPO pour trouver	- Facile mais fiabilité non garantie - Fiabilité de la BD TOPO à tester
Point de vue parfois moins bon que celui créé par Google à partir de l'adresse	Photo moins centrée sur la maison, présence de bâtiments voisins	- Définir un point de vue plus ciblé sur le bâtiment principal	- Assez facile à faire, performance à confirmer

7.2.3 Module IAD

Limite actuelle	Impact sur la qualité de l'outil	Amélioration à apporter	Confiance dans la mise en place de l'amélioration
Plusieurs maisons	Les caractéristiques extraites ne correspondent pas à l'adresse	Méthode simple par sélection de la maison de plus grande taille dans l'image. Dans des cas non triviaux avec des maisons mitoyennes dont les dimensions dans l'image sont identiques, il faudrait alors intégrer une alerte au niveau du module SIG	Facile dans les cas non triviaux Difficile dans le cas contraire (mais marginal ?)
Détection de volets ouverts ou fermés pour la prise en compte du taux de vitrage	Taux de vitrage erroné	Intégration dans l'outil d'extraction de caractéristiques d'une détection de volets ouverts ou fermés	Moyennement difficile. Nécessite de mettre en place une nouvelle méthode plus robuste.
Présence d'un sous-sol	Impact sur la simulation énergétique dans SolRen	Evolution de l'outil d'extraction	Moyennement difficile. Nécessite de mettre en place une nouvelle méthode plus robuste.

Toiture terrasse	Impact sur la simulation énergétique dans SolRen	<ul style="list-style-type: none"> - Evolution des réseaux de neurones intégrant la notion de toiture terrasse - Prise en compte des contours de toitures 	<ul style="list-style-type: none"> - Difficile : cela demande de refaire une labélisation la base d'apprentissage sans garanties de performances des réseaux de neurones (potentiellement peu d'exemples dans la base) - Difficile : potentiellement inefficace.
Effet de perspective	Taux de vitrage erroné	Méthode de redressement d'image	<ul style="list-style-type: none"> - Facile dans le cas manuel - Moyennement difficile dans le cas tout automatique
Façade partiellement obstruée	Extraction de caractéristiques incomplète	Détection d'obstacle puis extrapolation (ou recherche d'un autre point de vue avec module SIG)	<ul style="list-style-type: none"> - Difficile : les premiers tests de détection d'obstacles par réseaux de neurones n'ont pas été concluants (classe mal définie avec une grande diversité dans la nature des obstacles).
Identification des blocs	Impact sur la simulation énergétique dans SolRen	Evolution de l'outil d'extraction à mettre en interface l'emprise du bâtiment (module SIG) et les façades détectées	Difficile
Association d'une façade avec une orientation géographique	Impact sur la simulation énergétique dans SolRen	Evolution de l'outil d'extraction à mettre en interface la position de prise de vue par rapport à la maison (module SIG) et les façades détectées	Moyennement difficile (certains cas non triviaux seront difficile à gérer)

7.2.4 Module SolRen

Limite actuelle	Impact sur la qualité de l'outil	Amélioration à apporter	Confiance dans la mise en place de l'amélioration
Outil de modélisation énergétique monozone	Tendance à surestimer les consommations avec ce genre de modélisation, et donc surestimer le gain espéré par une solution de rénovation	Utilisation d'un outil de simulation multizone	Pas de difficultés techniques majeures, mais des ressources sont nécessaires
Modélisation limitée à des logements type	Modérée sur la partie diagnostic énergétique. Majeure quand il s'agira de produire métrés et devis	Exploiter BD Topo pour récupérer des informations géométriques plus précises	A tester
Peu de finesse sur la définition des caractéristiques thermiques du bâti, uniquement choisies selon la période de construction	Majeure : la performance réelle sera différente en cas de rénovation thermique partielle ou globale depuis la construction	Coupler avec la démarche de l'Audit Numérique pour affiner les caractéristiques thermiques	Pas de difficultés techniques majeures

7.2.5 Identification de la période de construction

Limite actuelle	Impact sur la qualité de l'outil	Amélioration à apporter	Confiance dans la mise en place de l'amélioration
Base de photos Google de test avec des points de vue sur les maisons pas optimaux	Perte de performance notable	Créer une base à partir du module SIG développé dans SmartReno	Bonne
Manque d'images corrélées avec la base de photos-test	Résultats de classement faibles	Poursuivre la corrélation avec les images	Assez bonne
Fiabilité et nombre de données sur les années de construction limité	Limites sur les performances	Utilisation de la BD Topo (à voir avec l'augmentation des open data au 01/01/2021) pour pouvoir créer une base d'images importante	Assez bonne

7.2.6 Audit numérique par analyse de la courbe de charge

Limite actuelle	Impact sur la qualité de l'outil	Amélioration à apporter	Confiance dans la mise en place de l'amélioration
Manque de fiabilité dans la détection des caractéristiques du logement	Majeur : les caractéristiques sont mal identifiées, les préconisations ne seront pas pertinentes	<ul style="list-style-type: none"> - Coupler ODEAL et Audit numérique - Améliorer les traitements mathématiques pour identifier les - Utiliser un cœur de calcul pour la simulation thermique plus détaillé 	Améliorations prometteuses, à tester
Méthode uniquement applicable aux maisons individuelles 100% électrique (chauffage et ECS)	Majeur : ce n'est pas le cas de la majorité des maisons	Tester et adapter la méthode pour d'autres types de maison	Inconnue

8 Annexe 1 – Configuration logicielle requise pour exécuter ODEAL

Pour exécuter l'ensemble des modules contenu dans ODEAL, les logiciels suivants doivent être installés

- Distribution Python 3.8
- Package python externes (les versions sont détaillées dans les différents packages python) :
 - o Numpy
 - o Scipy
 - o Matplotlib
 - o opencv-python
 - o scikit-image
 - o PyTorch
 - o xlutils
- Excel
- DiagSerieSolveur 10.2.0 (propriété d'EDF)
- LaTeX et MiTeX