

Un prédicteur
énergétique interactif
robuste pour
révolutionner
l'adaptation des
bâtiments à l'ère du
changement climatique

Mostafa Saad

UNIVERSITÉ DE CONCORDIA Montréal, Québec

Table des matières

1	RÉSUMÉ	3
2	INTRODUCTION	6
2.1	LE CONTEXTE	6
2.2	PERSPECTIVE	11
3	MÉTHODOLOGIE	14
3.1	INFORMATIONS SUR L'ÉTUDE DE CAS	18
3.2	MESURES DE RÉNOVATION DES BÂTIMENTS (MRB)	19
3.3	DÉVELOPPEMENT DU MODÈLE D'ÉMULATION	21
4	RÉSULTATS	28
4.1	ANALYSE DES DONNÉES	29
4.2	RÉSULTATS DU RLMV	31
4.3	TEMPS DE CALCUL	41
4.4	COMPARAISON AVEC D'AUTRES MODÈLES D'APPRENTISSAGE AUTOMATIQUE	43
4.5	PRÉDICTION SUR D'AUTRES ENSEMBLES DE DONNÉES NON VUS	44
5	SOUTIEN À LA PRISE DE DÉCISION	45
6	DISCUSSION ET RECOMMANDATIONS	47
6.1	RECOMMANDATIONS	47
6.2	LIMITATIONS DE L'ÉTUDES	49
7	REFERENCES	50

Figure 1 prise de décision pour l'adoption de la rénovation des bâtiments.....	7
Figure 2 Méthodologie du projet	16
Figure 3 Sélection de caractéristiques et modèles basés sur la régularisation et les objectifs utilisés pour le développement.....	25
Figure 4 Analyse et distribution des histogrammes des variables de sortie	30
Figure 5 Performances des modèles de base de la RLMV	33
Figure 6 Processus de sélection pas à pas à rebours des caractéristiques.....	35
Figure 7 BSFS RLMV modèle performance	35
Figure 8 Processus de sélection récursive des caractéristiques - R^2 en fonction du nombre de caractéristiques.....	37
Figure 9 modèle RFE Performances	38
Figure 10 Coefficients de régression élastique Processus de rétrécissement	40
Figure 11 Performance du Elastic Net.....	41
Figure 12 Comparaisons entre les modèles développés et le modèle de simulation conventionnel en termes de temps de calcul	43

1 Résumé exécutif

Les changements climatiques et les émissions de gaz à effet de serre (GES) sont au cœur des préoccupations internationales, car ils causent des problèmes partout dans le monde. Le gouvernement du Québec souhaite atteindre la neutralité carbone d'ici 2050, ce qui nécessitera des mesures et des politiques rigoureuses. En plus de poursuivre les Objectifs de développement durable (ODD) des Nations Unies, les mesures de décarbonisation sont une priorité pour les différents secteurs qui touchent la vie humaine. Le secteur du bâtiment joue un rôle clé, en raison de l'immense consommation d'énergie et des émissions de carbone qu'il génère (41 % à l'échelle mondiale). Montréal, avec son parc immobilier vieillissant, est la quatrième ville la plus ancienne du Canada. Les structures post-industrielles de l'emblématique canal de Lachine en font partie, avec une surface totale de plus de 5 400 000 pieds carrés. Pendant longtemps, la demande énergétique totale a été le seul paramètre pris en compte lors de la rénovation des bâtiments. Or, ce paramètre peut être trompeur lorsqu'il s'agit des émissions de carbone. Alors que les rénovations de bâtiments visant à éliminer les émissions de carbone gagnent en popularité, les propriétaires et les gestionnaires de bâtiments sont limités par un manque d'expertise et des processus de rénovation compliqués, qui reposent sur des calculs exhaustifs qui limitent l'accessibilité générale. Une méthode robuste, abordable sur le plan informatique et évolutive est nécessaire de toute urgence pour combler les lacunes et remplacer les méthodes conventionnelles tout en incorporant davantage de variables.

Un modèle hybride, conçu pour combler les lacunes en matière de calcul, prédit l'effet des mesures d'économie d'énergie sur le bâtiment concerné. Il combine les capacités prédictives de

l'apprentissage automatique avec les modèles énergétiques des bâtiments. Sur la base de l'ensemble des données relatives à la rénovation, le modèle reproduit les performances basées sur la physique à des fins de formation, de test et de réglage. Une combinaison de stratégies sera donc attestée pour faire des prédictions assez précises. Il peut s'agir de méthodes d'ingénierie et de divers algorithmes. Le modèle vise à atteindre une grande précision pour faciliter la prise de décision préliminaire.

La modélisation par émulation utilise une technique d'approximation pour surmonter les inefficacités de calcul des méthodes conventionnelles, tout en donnant accès à un plus grand nombre de scénarios d'amélioration énergétique créés. Cette étude a créé trois classes de modèles d'émulation pour prédire l'énergie, les émissions de carbone et les coûts de construction associés à la mise en œuvre de différents MRB. Les modèles ont utilisé les algorithmes interprétables de RLMV pour approximer le comportement des modèles conventionnels et fournir des prédictions en temps quasi réel. En outre, une analyse approfondie de l'effet des techniques de sélection des caractéristiques et de la prise en compte de la multicollinéarité a été réalisée afin de mieux comprendre le comportement du modèle et d'optimiser la structure du modèle et le nombre de caractéristiques requises.

Les modèles les plus précis développés ont fourni une précision acceptable et réduit le temps de calcul de 1000 fois par rapport aux modèles conventionnels. Les résultats de l'étude ouvrent la voie à des modèles plus accessibles et à des analyses complètes des méthodes de transformation des bâtiments durables, permettant à terme d'identifier des mesures de décarbonisation à l'échelle.

Le projet vise à fournir un modèle possible qui peut être utilisé pour combler le fossé entre les méthodes d'ingénierie et les parties prenantes qui cherchent à évaluer le potentiel de transformation de leur bâtiment. Étant donné que les principes de l'ODD 16.7 appellent à la démocratisation de la prise de décision, les modèles développés peuvent être utilisés pour développer des interfaces d'engagement citoyen en temps réel afin d'assurer l'accès à la connaissance et d'accroître la sensibilisation.

2 Introduction

2.1 Le contexte

Compte tenu de l'aggravation des effets du changement climatique, il est impératif de souligner l'urgence de mettre en œuvre des politiques efficaces pour atténuer les émissions de gaz à effet de serre (GES) [1]. L'évolution rapide de l'urbanisation a créé un scénario dans lequel les centres urbains sont devenus l'épicentre prédominant de ce problème mondial. Les projections suggèrent que d'ici 2050, près de 67 % de la population mondiale vivra dans ces agglomérations urbaines [2]. De nos jours, les centres urbains revêtent le rôle de principaux acteurs dans la consommation d'énergie à l'échelle planétaire, contribuant à hauteur de 70 % de la consommation totale. De plus, ils constituent la source de plus de 70 % des émissions mondiales de dioxyde de carbone [3]. Les bâtiments sont l'une des entités les plus importantes qui influencent les performances d'une ville et, en tant que tels, ils jouent un rôle essentiel dans la métamorphose durable des villes [4,5]. Les secteurs commercial et résidentiel représentent environ un quart de la consommation nationale d'énergie dans le contexte canadien. Le secteur des bâtiments représente à lui seul 13,06 % des émissions de dioxyde de carbone du Canada [6]. Au Québec, les bâtiments consomment 35 % de toute l'énergie produite, alors qu'à Montréal, cette proportion atteint 88 % de la consommation globale d'énergie électrique. Les bâtiments commerciaux accaparent 47 % des ressources énergétiques émettrices de carbone, ce qui fait de cette catégorie la plus importante en émissions de gaz à effet de serre au Québec [7]. De plus, la plupart des bâtiments de Montréal datent d'avant 1970, ce qui les rend obsolètes en matière de conformité au code du bâtiment et d'efficacité énergétique. C'est pourquoi une multitude d'initiatives et de programmes incitatifs ont été conçus par les décideurs à la disposition des

propriétaires de bâtiments, dans le but de stimuler les études préliminaires et les études de faisabilité afin d'accélérer le processus de rénovation des bâtiments (RB) [8,9]. La plupart de ces mesures incitatives visent à donner accès à des possibilités d'évaluation par des consultants qui sont subventionnés pour permettre la prise de conscience comme la montre la Figure 1.

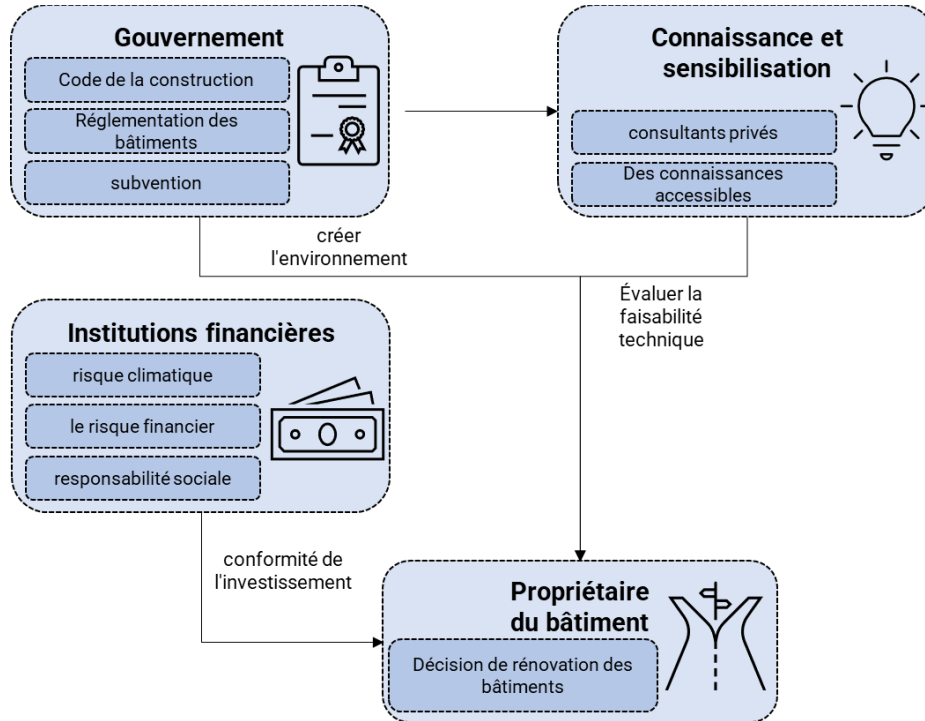


Figure 1 prise de décision pour l'adoption de la rénovation des bâtiments

L'insuffisance des connaissances et l'absence d'outils accessibles constituent un frein considérable à l'adoption du processus d'amélioration énergétique des bâtiments. Cette carence peut exercer une influence marquée à différents niveaux, entravant non seulement l'application des technologies appropriées, mais influençant également la manière dont les politiques et les réglementations sont mises au point et mises en pratique. Le rôle des municipalités et des

gouvernements se révèle primordial dans cette dynamique, et ces entités peuvent intervenir de manière décisive pour pallier ces déficits et encourager une adoption plus étendue et accélérée de la rénovation énergétique des bâtiments.

Le manque de connaissances peut constituer un obstacle à l'adoption de mesures d'amélioration de l'efficacité énergétique des bâtiments à différents niveaux. Sans une compréhension claire des avantages et des méthodes de rénovation énergétique, les propriétaires et les gestionnaires de bâtiments peuvent être réticents à investir dans de telles technologies. La carence en connaissances spécialisées peut également aboutir à un choix inapproprié de technologies ou à une mise en œuvre incorrecte, diminuant de ce fait l'efficacité des mesures adoptées.

Par ailleurs, la disponibilité d'outils accessibles pour évaluer et mettre en œuvre la rénovation énergétique des bâtiments s'avère également essentielle. Sans les instruments adéquats pour évaluer les besoins en énergie, sélectionner les technologies appropriées et suivre les progrès, le processus peut s'avérer complexe et coûteux. Cette situation peut dissuader les propriétaires de bâtiments et autres intervenants concernés d'opter pour des mesures de rénovation énergétique, même si ces dernières se révèlent économiquement rentables à long terme. La présente analyse souligne donc la nécessité d'une collaboration accrue entre les différents niveaux de gouvernance.

Dans cette démarche, les modèles énergétiques des bâtiments (MEB) se révèlent indispensables. Ces MEB jouent un rôle capital dans l'évaluation des stratégies de décarbonation et des approches de rénovation énergétique, offrant aux décideurs une aide pour mettre en œuvre les mesures les plus efficaces en vue de réduire la consommation énergétique et les émissions de carbone.

Et les parties prenantes pour surmonter ces obstacles et réaliser les objectifs de durabilité énergétique dans le secteur du bâtiment.

D'une manière générale, la modélisation énergétique des bâtiments appliquée au domaine de la rénovation peut être classée en deux approches principales : ascendante et descendante. Les approches ascendantes sont généralement privilégiées pour leur capacité à représenter les bâtiments individuels à un niveau granulaire, ce qui permet d'éviter les analyses statistiques cumulées [10].

Différents logiciels de simulation, tels que EnergyPlus et TRNSYS, sont couramment utilisés pour mettre en œuvre ces modèles et effectuer des analyses de performance. EnergyPlus et TRNSYS sont largement reconnus comme des outils de modélisation fiables, et leur précision éprouvée leur a valu une place de choix dans les milieux universitaires [11]. Toutefois, cette précision est diminuée par la complexité des données d'entrée et les fortes charges de calcul substantielles associées aux simulations de haute-fidélité. En outre, la construction d'un modèle énergétique peut être une entreprise longue et laborieuse, nécessitant un investissement substantiel en heures de travail. En outre, des modèles informatiques inefficaces peuvent entraver l'analyse des options d'efficacité énergétique [12]. Une analyse exhaustive de la rénovation énergétique d'un édifice intègre une multiplicité de variables, ce qui requiert des durées de calcul substantielles et des ressources énergétiques.

Afin de surmonter les lacunes inhérentes aux approches susmentionnées, des modèles d'apprentissage automatique (MAA) sont déployés pour remplacer les méthodes de modélisation en boîte blanche et fournir une approximation rapide. Le MAA est entraîné sur des bases de données échantillonnées dérivées des résultats de la modélisation en boîte blanche, dans le but

d'accroître l'efficacité en réduisant le nombre de simulations nécessaires. Ce concept, connu sous le nom de modélisation par émulation, permet à la MAA de remplacer les modèles à boîte blanche [18]. La configuration de référence des modèles est principalement exploitée pour surmonter le défi inhérent au manque de données sur les bâtiments avant et après la rénovation [10]. De multiples études ont examiné l'utilisation des modèles d'émulation pour fournir des prévisions de conception à un stade précoce. Les modèles d'émulation utilisent divers algorithmes d'apprentissage, tels que les réseaux neuronaux artificiels (RAN), les machines à vecteurs de support (MVS), les forêts aléatoires (FA), la régression linéaire multivariée (RLMV) et les splines de régression adaptative multiple (RSAM) [19]. Ces substituts offrent un équilibre important entre le temps de calcul et la précision.

Le choix du modèle d'émulation est essentiellement influencé par la recherche d'une précision optimale. Toutefois, dans certaines circonstances, un compromis peut être recherché entre l'optimisation de la précision et l'adoption d'une structure de modèle permettant une interprétation optimale [19].

La rénovation énergétique des bâtiments englobe une variété considérable de mesures de réduction d'émissions, ce qui engendre une complexité accrue et élargit l'espace des solutions, donnant lieu à une tâche sophistiquée et complexe souvent confiée à des experts-conseils[20]. La diversité des consultants se fonde sur des suggestions émanant de l'expérience ou d'un nombre limité d'études, conséquence du temps restreint et de l'expansion exponentielle du temps nécessaire pour enquêter sur de nombreuses MRB au sein d'un modèle conventionnel. Ainsi, l'élaboration d'un outil de faible coût computationnel, apte à fournir une analyse plus exhaustive et couvrant de multiples mesures, pourrait représenter une avancée considérable et apporter un

soutien inestimable dans le processus de prise de décision, susceptible de catalyser de façon significative les efforts en faveur de la décarbonisation des bâtiments.

L'étude se concentre sur le développement d'un modèle d'émulation hybride, combinant une simulation de haute fidélité avec un algorithme basé sur des données. Dans cette optique, la recherche vise à identifier les méthodes d'ingénierie des caractéristiques pour l'application envisagée. Plus précisément, l'objectif est de tester plusieurs méthodes sur un ensemble de données synthétiques représentatif d'un bâtiment préexistant dans la région du canal de Lachine, en vue de concevoir un prédicteur optimal, précis et robuste de la performance du bâtiment. Ce prédicteur prend en compte les paramètres de rénovation du bâtiment et leurs implications en matière d'émissions de carbone et de coûts associés. La contribution majeure de cette étude réside dans la traduction de ce processus complexe en une interface interactive, rendant les concepts complexes accessibles au grand public et aux parties prenantes.

2.2 Perspective

Les rénovations de bâtiments représentent un enjeu d'envergure mondiale et nécessitent des analyses approfondies de l'impact des mesures de rénovation énergétique. Toutefois, la nature combinatoire de ce problème peut entraîner des coûts de calcul élevés, ce qui limite la portée de l'analyse pour les décideurs, en particulier lorsque l'évaluation doit prendre en compte la consommation d'énergie, les émissions de carbone et les coûts associés. Cette limitation conduit souvent à des enquêtes restreintes, étant donné les contraintes informatiques. De plus, les modèles de substitution, s'ils peuvent être une solution, ont leurs propres limites, en particulier l'absence de prise en considération simultanée de l'énergie, des émissions de carbone et du coût,

qui sont des aspects cruciaux de la prise de décision en matière d'amélioration de l'efficacité énergétique.

Un examen de la littérature existante a mis en évidence une lacune importante dans la recherche sur l'impact des méthodes d'ingénierie et de sélection des caractéristiques des performances des modèles d'émulation, en particulier dans le domaine de la recherche sur les bâtiments. Cette lacune empêche une compréhension globale et laisse des possibilités inexploitées pour configurer de manière optimale les modèles en utilisant un nombre réduit de caractéristiques d'entrée. Bien que des études antérieures aient exploré le développement de modèles de substitution et de modèles conventionnels pour évaluer la performance des bâtiments, il est important de noter que l'analyse de l'effet des paramètres du modèle sur la précision des modèles de substitution reste limitée dans le domaine de la recherche sur les bâtiments. L'étude se concentre sur le développement d'un modèle hybride de substitution combinant une simulation haute fidélité et un algorithme piloté par les données. Dans cette optique, la recherche vise à identifier les méthodes optimales d'ingénierie des données et des caractéristiques pour l'application visée. La recherche vise à tester plusieurs méthodes sur un ensemble de données synthétiques d'un bâtiment existant dans la région du canal de Lachine afin de développer un prédicteur optimal, précis et robuste de la performance du bâtiment en tenant compte des éléments de rénovation du bâtiment et de leurs émissions de carbone et coûts correspondants. La recherche entreprise dans cette étude vise donc à combler ces lacunes en proposant une méthodologie complète qui intègre des modèles d'émulation basés sur différentes facettes de la performance du bâtiment, tout en explorant les méthodes de sélection des caractéristiques pour améliorer la précision et la compréhension des modèles développés.

L'objectif de cette recherche est d'examiner l'applicabilité de la création d'un modèle de substitution pour émuler le modèle conventionnel de performance des bâtiments, et d'étendre sa capacité prédictive pour inclure les coûts ainsi que les émissions de carbone. Le projet vise à étudier les performances des modèles de substitution interprétables, tels que l'algorithme RLMV. À cette fin, il s'efforce de fournir une étude complète de l'impact des différentes méthodes d'ingénierie et de sélection des caractéristiques, dans le but d'établir une compréhension claire et précise qui pourra être reproduite et enrichie dans le cadre de recherches futures.

Dans cette étude, un modèle ascendant de performance des bâtiments traditionnels est développé pour un bâtiment de référence à Montréal, au Canada. Un modèle de simulation robuste et de haute fidélité est conçu pour tenir compte de la performance du bâtiment lors de l'analyse d'un large éventail de MRB. Parallèlement, les données générées par le modèle physique sont utilisées pour développer un modèle de performance du bâtiment de substitution qui prend en compte de multiples objectifs, tels que la consommation d'énergie, les émissions de carbone et les coûts.

La contribution scientifique de la recherche peut être articulée de la manière suivante :

1. L'étude fusionne de nombreux systèmes de MRB et de CVC pour examiner leur influence sur la performance des bâtiments, jetant ainsi les bases d'un modèle de substitution complet. L'effet combinatoire du problème complexe examiné est déterminé, illustrant les déficiences manifestes des modèles de simulation basés sur la physique en ce qui a trait aux exigences computationnelles.
2. La recherche propose une méthodologie élaborée pour développer des modèles de substitution interprétables afin de prédire divers aspects de la performance du bâtiment,

tout en étudiant les configurations optimales à travers l'ingénierie des caractéristiques et le couplage de sélection.

3. L'étude offre une analyse détaillée des impacts des différents paramètres des modèles de substitution, tels que la sélection des caractéristiques, sur la précision et la maturation des modèles.
4. l'étude suggère de nouvelles méthodes de sélection de caractéristiques ajustées, modifiées dans le but d'optimiser les performances des modèles, de stabiliser leurs capacités prédictives et de pallier leurs insuffisances.
5. Les modèles développés dans cette étude ouvrent la voie à leur intégration dans de futures plateformes interactives qui pourront être utilisées pour l'aide à la prise de décision.

3 Méthodologie

La méthodologie de recherche développée dans cette étude s'articule autour de trois étapes clés comme le montre Figure 2, formant une approche systématique et méthodique. La première étape englobe le développement d'un modèle de performance du bâtiment pour un bâtiment de référence à Montréal, impliquant la collecte de données sur site, la configuration du modèle, et son étalonnage en accord avec les normes couramment utilisées. La seconde étape comprend l'élaboration des mesures de rénovation des bâtiments (MRB) et des fourchettes incluses, où les solutions alternatives sont classées en différentes catégories, telles que les murs extérieurs ou les systèmes de Chauffage, Ventilation et Climatisation (CVC), et où les combinaisons variées des diverses matrices de rénovation des bâtiments (MRB) définissent les scénarios possibles de

rénovation des bâtiments, soumis à une évaluation approfondie. L'étape finale est le développement de modèles de substitution, un processus complexe impliquant l'ingénierie des caractéristiques, la sélection des caractéristiques, l'ajustement des modèles, l'optimisation des hyper paramètres et l'évaluation rigoureuse des modèles. Cette structuration permet non seulement une compréhension claire et précise de la recherche, mais aussi une mise en application concrète de la méthodologie dans le domaine de la performance des bâtiments, en offrant une vision holistique des objectifs et des procédures de l'étude. En complément de cette méthodologie, cette prévision prend en considération les émissions de carbone ainsi que l'analyse des coûts. Dans cette optique, plusieurs étapes seront rigoureusement appliquées dans le but d'obtenir un algorithme optimisé et des prédictions de haute précision.

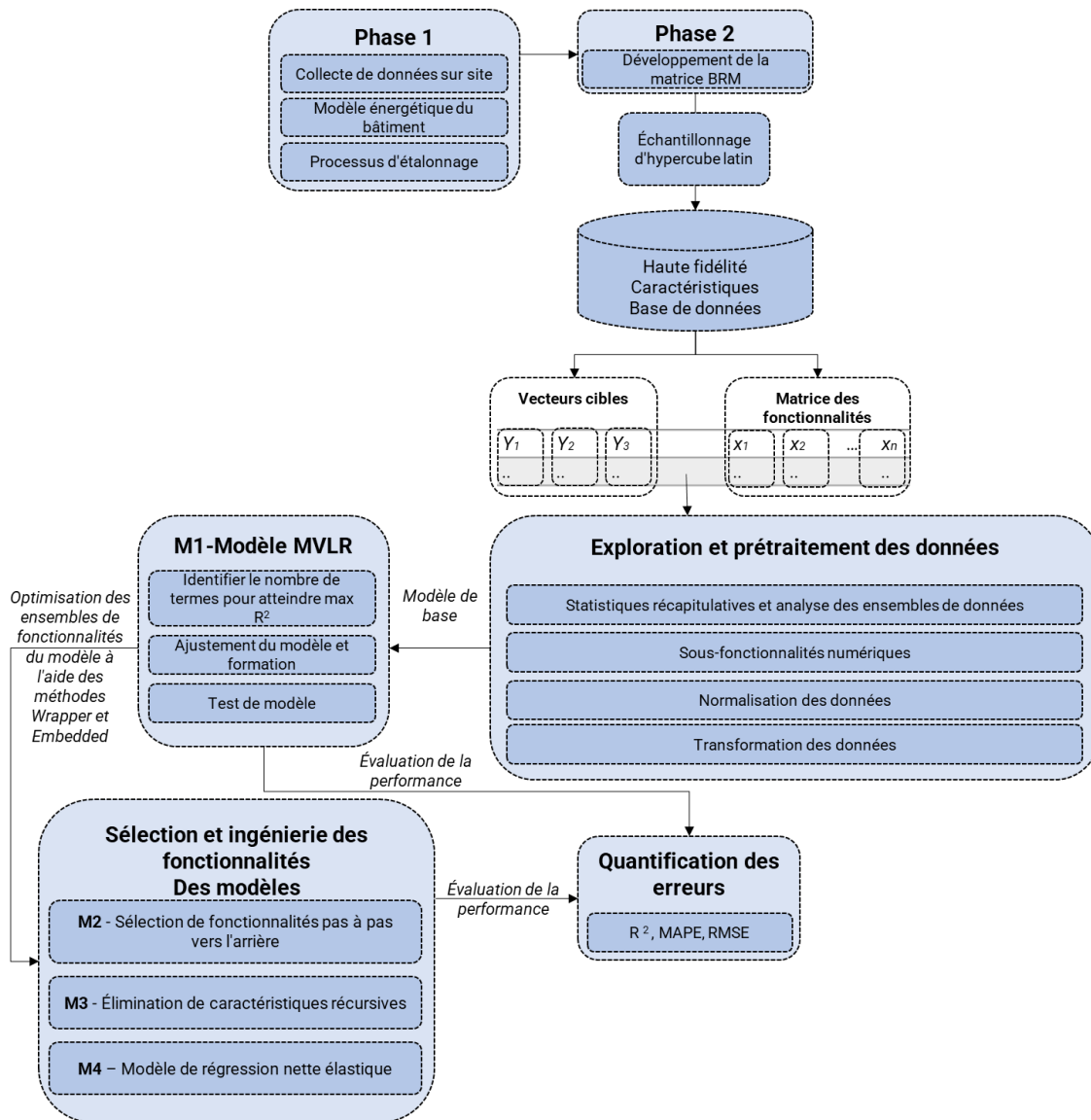


Figure 2 Méthodologie du projet

Étape 1 : Création de l'échantillon de l'ensemble de données à travers l'application de l'échantillonnage hypercube latin résultant de la simulation d'un modèle de bâtiment.

Étape 2 : L'ensemble de données développé fera l'objet d'une harmonisation et d'une normalisation. L'analyse de la colinéarité entre les différentes caractéristiques de cet ensemble de données sera entreprise, dans le but de contrecarrer tout biais potentiel dans les prédictions.

Étape 3 : L'analyse des caractéristiques est la première étape de l'identification d'un potentiel de transformation. À cette fin, les techniques de transformation seront mobilisées pour normaliser les caractéristiques, dans le but de favoriser l'établissement de corrélations statistiques plus précises entre elles.

Étape 4 : Le modèle de régression linéaire multivariée (RLMV) sera examiné dans un premier temps. En utilisant une combinaison de coefficients distincts, attribués à différentes caractéristiques en fonction de leur pertinence au sein de l'ensemble de données et de leurs interrelations, l'objectif de cette étape est d'appréhender la performance d'un modèle présentant des qualités d'interprétation et de transparence.

Étape 5 : Une palette de méthodes de sélection de caractéristiques sera explorée, en ayant recours à des méthodes d'enveloppe et d'intégration pour assurer la sélection des caractéristiques les plus pertinentes en vue d'être incorporées au modèle d'apprentissage automatique affilié.

Étape 6 : Divers modèles basés sur des méthodes de sélection et d'ingénierie des caractéristiques identifiées seront conçus et examinés. Il s'agit notamment de modèles de régression, tels que le Elastic Net, et de modèles de caractéristiques éliminées de manière récursive.

Étape 7 : Une analyse de la précision des modèles conçus sera menée, incluant une évaluation comparative des diverses stratégies mobilisées dans leur élaboration (sélection de caractéristiques, régularisation). L'objectif primordial ici réside dans l'identification de l'algorithme le plus approprié parmi ceux à disposition pour ce type d'étude de cas, tout en produisant une analyse comparative.

Étape 8 : Un examen de la vitesse de calcul résultant des modèles développés sera effectué, ce qui clôturera le processus de développement.

Les résultats de cette étude ouvriront la voie à l'amélioration des prévisions de performance des bâtiments, en utilisant une approche rapide, robuste et précise. Cela permettra aux propriétaires de bâtiments et aux décideurs politiques de prendre des décisions plus éclairées, de manière accessible. Cela contribuera à promouvoir des stratégies de transition adaptées pour la ville de Montréal, conformément à l'objectif de développement durable (ODD) 11.3, qui vise à renforcer l'urbanisation inclusive et durable, ainsi que la capacité de planification participative durable.

3.1 Informations sur l'étude de cas

L'étude s'est concentrée sur un ensemble de bâtiments situés le long du canal de Lachine à Montréal, édifiés au cours du 19e siècle. Cette région recèle une histoire foisonnante, ayant abrité au fil des décennies de multiples complexes industriels. Toutefois, son parc immobilier industriel a subi au fil du temps diverses transformations pour servir des fins variées, englobant une superficie approximative de cinq millions de pieds carrés [21].

Le bâtiment choisi pour cette étude de cas a été construit en 1908 comme usine de production textile, avant d'être converti à un usage commercial, principalement sous la forme de locaux à usage de bureaux. Des données ont été recueillies à partir d'une évaluation structurelle et de visites sur site. Les caractéristiques de l'enveloppe du bâtiment ainsi que des systèmes mécaniques ont également fait partie de ce processus de collecte de données. En outre, l'utilisation de l'espace a été identifiée par des inspections in situ et des informations fournies par l'administration du bâtiment. L'historique de la consommation énergétique du bâtiment a été rassemblé dans le but d'être analysé et de quantifier sa consommation énergétique régulière. La consommation d'énergie sur plusieurs années a été exploitée pour créer un profil de

consommation énergétique. Par la suite, les modèles de consommation énergétique du bâtiment ont été mis à profit pour calibrer le modèle énergétique du bâtiment (MEB).

Un modèle énergétique de bâtiment conventionnel a été développé à partir de données collectées sur le terrain, ainsi qu'un modèle géométrique au niveau de détail 4 (LOD4). En l'absence de données sur le site ou lorsque les capacités d'essai sont insuffisantes, le Code national de l'énergie du Canada (CNEB) [22] a été utilisé en substitution.

3.2 Mesures de rénovation des bâtiments (MRB)

Une analyse plus approfondie de la typologie des bâtiments (représentés par des structures post-industrielles) a révélé plusieurs contraintes spécifiques. Il s'agit notamment de restrictions concernant les éléments de l'enveloppe du bâtiment, les systèmes énergétiques et les systèmes de distribution.

L'évaluation de la typologie du bâtiment mentionnée a révélé plusieurs caractéristiques standard. Tout d'abord, dans la région considérée, les bâtiments industriels prennent généralement la forme de grandes structures non isolées qui ont subi diverses transformations au fil du temps pour une variété d'utilisations. Leur usage initial ainsi que les besoins de production industrielle à haute température ont engendré des systèmes de distribution de chaleur basés sur l'utilisation de la vapeur en tant que fluide caloporteur.

À la lumière des contraintes inhérentes au bâtiment, des mesures de rénovation appropriées ont été identifiées, comme indiqué dans le Tableau 1. L'analyse des mesures fréquemment mises en œuvre a conduit à la sélection d'un ensemble de MRB [23–26] en raison de leur adéquation aux contraintes spécifiques du bâtiment et de leur impact considérable sur l'optimisation de l'efficacité énergétique des immeubles de bureaux dans le contexte climatique étudié. Il

comprend notamment un rajustement des émissions liées aux combustibles pour tenir compte des émissions provenant du réseau local et du gaz naturel, ainsi que des tarifs correspondants. De plus, les émissions résiduelles de carbone agrégées et le rapport coût-efficacité ont été calculés pour chaque scénario, au moyen des méthodes d'Analyse du Cycle de Vie (ACV) et d'Analyse des Coûts du Cycle de Vie (ACCV). Chacune des catégories des mesures de rénovation du bâtiment MRB sera minutieusement exposée dans les sous-sections suivantes, en mettant en exergue la valeur minimale et maximale pour chacune de leurs caractéristiques.

Tableau 1 Catégories d'éléments de construction sélectionnée pour l'élaboration de l'analyse

Catégorie	Options	Description de la Mesure	Paramètre Unité	Min	Max
Murs extérieurs	31	Amélioration des murs pour une meilleure isolation	m ² K/W	0.99	8.00
Toit	22	Amélioration de la toiture par des options plus isolantes	m ² K/W	1.56	6.27
Vitrage	23	Amélioration des fenêtres du bâtiment en optant pour des options à résistance thermique améliorée	W/m ² K	3.09	0.78
Stores pour Fenêtres	5	Utilisation de stores pour fenêtres durant la saison estivale pour moduler l'éclairage diurne.	Largeur des lamelles (m)	0	0.05
Encadrement des Fenêtres	5	Remplacement de l'encadrement des fenêtres par des options alternatives à moindre conductivité	m ² K/W	3.47	5.88
Ratio Fenêtre-Mur	∞	Ajustement du ratio fenêtre-mur pour influencer les besoins de chauffage	%	20	95

Infiltration	∞	Amélioration de l'étanchéité à l'air du bâtiment et réduction du taux de renouvellement d'air entre les environnements intérieur et extérieur	ACH ₅₀	5.8	0.6
Éclairage intérieur	14	Remplacement de l'équipement d'éclairage intérieur pour toutes les zones	NPD (W/m ²)	3.3	7.6
Systèmes CVCA	9	Remplacement du système CVCA existant par un système haut efficacité	Chauffage CoP	0.83	3.3

3.3 Développement du modèle d'émulation

Cette section examine le processus de développement du modèle d'émulation. Ce processus a lieu une fois qu'un modèle de base calibré a été établi pour reproduire le comportement du bâtiment actuel. Westermann et al [19] ont décrit en détail les étapes du développement d'un modèle d'émulation, y compris la clarification de l'intention du modèle, la détermination de l'objectif et l'identification des paramètres de conception appropriés au problème traité. Comme précédemment évoqué, l'objectif central est d'évaluer la viabilité de la création d'un modèle approximatif dirigé par les données, qui se substituerait au modèle énergétique traditionnel du bâtiment pour l'évaluation préliminaire des rénovations. L'objectif consiste à définir trois catégories de modèles reproduisant le calcul de la consommation énergétique de chauffage du bâtiment, les émissions de carbone incorporées et les coûts. Les sous-sections à venir approfondissent chaque étape du processus de développement du modèle d'émulation.

3.3.1 Développement des ensembles de données

Tout d'abord, un ensemble de données comprenant les paramètres des mesures de rénovation du bâtiment (MRB) et les indicateurs de performance correspondants est constitué. Les méthodes

d'échantillonnage jouent un rôle crucial dans l'efficacité et la performance d'un modèle d'émulation lors de la création de l'espace de conception requis. L'efficacité d'une méthode d'échantillonnage peut être définie comme la maximisation des connaissances acquises sur l'ensemble de l'espace et de ses paramètres de conception. Les algorithmes d'échantillonnage couramment utilisés comprennent l'hypercube latin (LHS), Monte-Carlo (MC) et Sobol [19,27]. L'hypercube latin (LHS) est une méthode pseudo-aléatoire qui étend le concept de carré latin et génère un échantillon en créant des hyperplans alignés sur les axes. Dans cette étude, l'échantillonnage de l'hypercube latin (LHS) est une méthode pseudo-aléatoire qui étend le concept du carré latin et génère un échantillon en créant des hyperplans alignés sur l'axe. Dans cette étude, l'échantillonnage de l'hypercube latin a été utilisé en raison de son utilisation fréquente dans la bibliographie et de sa capacité à distribuer des échantillons sur l'ensemble de l'espace de conception tout en nécessitant moins de calculs que l'échantillonnage de Monte-Carlo [28–30]. Le résultat de la méthode d'échantillonnage se compose de vecteurs cibles et d'une matrice de caractéristiques, comme l'illustre l'équation suivante. Les variables indépendantes sont désignées comme caractéristiques, tandis que les variables prédites constituent le vecteur cible.

3.3.2 Ajustement du modèle

Le processus de construction du modèle d'émulation se décompose en plusieurs étapes cruciales : la sélection du type de modèle, l'analyse des caractéristiques, le prétraitement, et les processus de formation et de validation. La sélection du modèle constitue une phase essentielle, ayant des implications déterminantes sur la précision des prédictions et la performance globale du modèle. Dans ce contexte, l'interprétation apparaît comme un critère central, particulièrement

pertinent pour les questions d'ingénierie. Certains modèles, tels que les modèles d'apprentissage profond, peuvent présenter une nature opaque ou « boîte noire », limitant ainsi la capacité à comprendre le comportement intrinsèque du modèle et à guider son évolution future. En revanche, la régression linéaire multivariable (RLMV) se positionne comme un algorithme nettement interprétable. Sa structure offre une flexibilité moyenne, permettant une modélisation précise tout en conservant la transparence et l'intelligibilité nécessaires. Cette étude a mis en œuvre l'algorithme RLMV, en explorant spécifiquement ses performances et son applicabilité dans le domaine d'étude concerné.

3.3.3 Développement du modèle RLMV

Cette sous-section mettra en évidence de plus amples informations sur le développement du modèle RLMV. Le RLMV est essentiellement une équation linéaire telle que $(Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon)$. L'importance de chaque caractéristique est reflétée dans le coefficient (β), appelé coefficient de la caractéristique. La nature de l'algorithme facilite l'identification des principales caractéristiques pour chaque problème. Cependant, les modèles RLMV sont confrontés à plusieurs limitations, telles que la multicollinéarité, la sensibilité aux valeurs aberrantes et la nécessité d'avoir des distributions normales dans les données.

Dans le cadre du développement du modèle, des étapes et des stratégies ont été mises en place pour atténuer et éviter ces faiblesses du modèle. Le processus commence par la normalisation des caractéristiques, afin de les mettre à une échelle de valeurs similaires et d'éviter tout biais du modèle en faveur d'une caractéristique ayant une valeur numérique plus élevée. La normalisation ajuste toutes les caractéristiques pour avoir une moyenne de zéro et un écart type d'un. Les techniques de transformation des caractéristiques sont couramment utilisées dans les problèmes

de régression, en utilisant des formules mathématiques pour améliorer les capacités de prédiction et la précision du modèle. Ces fonctions de transformation permettent au modèle de mieux s'ajuster aux données, ce qui lui permet d'expliquer davantage de variances dans les données. Les formules de transformation incluent les transformations quantile, puissance, logarithme, exponentiel, et racine cubique. Un processus automatisé, incluant une recherche exhaustive de toutes les combinaisons possibles de transformations, a été développé dans le cadre du développement du modèle, atteignant jusqu'à 16 combinaisons par modèle. Ce processus transforme les données et inverse ensuite les transformations dans le processus d'évaluation du modèle.

Une étape cruciale et déterminante dans le développement d'un modèle RLMV robuste est la sélection des caractéristiques qui offrent les meilleures performances, généralisent bien sur de nouvelles données et atténuent la multicollinéarité.

3.3.4 Méthodes de Sélection des Caractéristiques

Une étape importante et déterminante dans le développement d'un modèle RLMV stable consiste à sélectionner les caractéristiques offrant les meilleures performances, une capacité de prédiction généralisée sur des données inconnues, et à traiter la multicollinéarité. Parmi les méthodes, on trouve les méthodes enveloppantes et intégrées, qui proposent deux approches différentes pour pallier ces défauts [31]. La première sélectionne un sous-ensemble plus restreint de la matrice des caractéristiques pouvant être utilisées comme critères indicateurs de performance et élimine les autres caractéristiques. Les méthodes de sélection et de régularisation de caractéristiques choisies sont présentées à la Figure 3 pour le développement des modèles M2-M4.

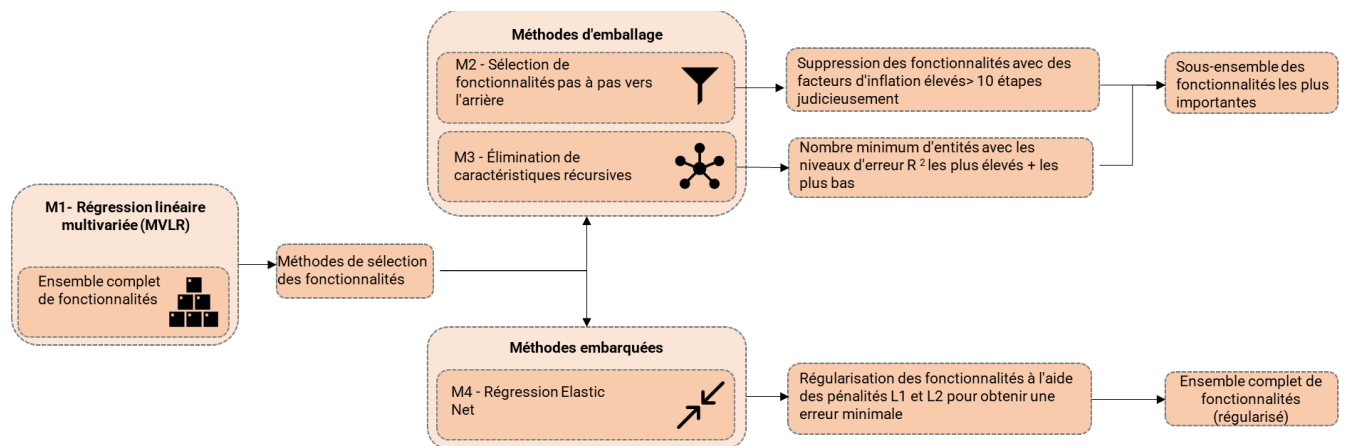


Figure 3 Sélection de caractéristiques et modèles basés sur la régularisation et les objectifs utilisés pour le développement.

3.3.4.1 Sélection de Caractéristiques Rétrogressive Étape par Étape (M2)

L'une de ses méthodes est la sélection rétrogressive pas-à-pas des caractéristiques (BSFS), qui part d'une matrice de caractéristiques complète et élimine la variable la moins significative à chaque étape, améliorant ainsi la précision du modèle. Cette étude a examiné l'efficacité de l'utilisation de BSFS combinée à un facteur d'inflation de la variance (VIF) pour éliminer les caractéristiques hautement colinéaires. Un flux de travail automatisé élimine les caractéristiques par un processus progressif en tenant compte de la multicollinéarité basée sur la valeur de $VIF > 10$. Le processus d'éjection se fait étape par étape, où, lors de chaque itération, le recalcul du VIF a lieu jusqu'à ce qu'aucune caractéristique ne dépasse le seuil maximal. Le sous-ensemble de caractéristiques sélectionné est ensuite utilisé pour ajuster le modèle.

3.3.4.2 Élimination récursive de Caractéristiques (M3)

Une méthode plus exigeante en calcul est l'Élimination récursive de Caractéristiques (RFE), qui utilise une technique d'optimisation combinée à un nombre prédéfini de caractéristiques à éliminer à chaque étape. Le modèle RFE utilise des sous-ensembles de validation croisée pour identifier et éliminer les caractéristiques les moins importantes. Bien que le modèle RFE soit

nettement plus coûteux en calcul, il a été utilisé dans cette étude pour comparaison. Un processus de sélection de caractéristiques RFE innovant a été développé en commençant le développement du modèle avec un nombre optimal de recherches de caractéristiques qui maintenaient le R^2 pour chaque modèle, suivi d'une recherche de la meilleure combinaison de caractéristiques. En général, les méthodes de sélection de caractéristiques assurent la cohérence et l'optimisation du modèle, où moins de caractéristiques peuvent décrire les données plus efficacement.

3.3.4.3 Régression Elastic Net (M4)

Une autre approche consiste en des méthodes intégrées, qui incluent sans effort un processus de régularisation des caractéristiques dans le processus de construction du modèle. La régularisation fait référence à la limitation des coefficients des caractéristiques lors de l'ajustement du modèle pour éviter le surapprentissage et stabiliser sa capacité de prédiction. Elle traite de la multicolinéarité en réduisant les coefficients de caractéristiques corrélées, qui peuvent parfois atteindre zéro, semblable à l'élimination de la caractéristique. Des modèles tels que Lasso, Ridge, ou la forme combinée, Elastic Net, utilisent la régularisation en ajoutant des pénalités à la fonction de perte lors de l'ajustement de l'algorithme. La pénalité aide à une meilleure distribution des valeurs des coefficients, répartissant ainsi l'effet des variables colinéaires sur des coefficients plus petits. Ce processus réduit la variance des estimations, qui est très vulnérable à la multicolinéarité, ce qui permet d'obtenir un modèle plus stable. Cette étude a utilisé Elastic Net pour régulariser le modèle initial et tester sa capacité à fournir un modèle amélioré et plus précis. [32].

3.3.5 Quantification des erreurs

Cette étude a utilisé plusieurs métriques d'évaluation pour évaluer la performance des modèles développés, notamment l'erreur moyenne absolue en pourcentage (MAPE), la racine carrée de l'erreur quadratique moyenne (RMSE), et le coefficient de détermination (R^2). Chacune de ces métriques fournit des informations importantes sur la précision et la capacité explicative des modèles.

Le MAPE est calculé selon l'équation (1). Il évalue l'écart moyen en pourcentage entre les valeurs prédites et les valeurs réelles. Cette mesure relative de l'erreur est avantageuse pour l'analyse des données à différentes échelles, car elle permet d'interpréter et de comparer facilement les résultats entre des modèles ou des ensembles de données hétérogènes.

Le RMSE est calculé selon l'équation (2). Il représente la racine carrée de la moyenne des différences au carré entre les valeurs prédites et les valeurs réelles. La RMSE met surtout l'accent sur les erreurs de grande ampleur, ce qui la rend adaptée aux modèles qui se concentrent sur les erreurs importantes. Par rapport au MAPE, le RMSE est moins influencé par les valeurs aberrantes.

Le R^2 mesure la proportion de la variance de la variable dépendante qui peut être expliquée par les variables indépendantes incluses dans le modèle. Sa valeur varie entre 0 et 1, ou une valeur plus proche de 1 indique une meilleure adéquation du modèle aux données. Le R^2 est calculé selon l'équation (3). Cette métrique évalue à quel point le modèle est capable de capturer la variabilité observée dans les données.

L'utilisation de ces métriques permet d'évaluer différents aspects de la performance des modèles de manière complète. Le MAPE offre une mesure relative de l'erreur moyenne, la RMSE tient

compte des erreurs plus importantes, et le R^2 évalue la capacité du modèle à expliquer la variance des données. L'ensemble de ces métriques offre une vue holistique de la qualité des modèles développés et de leur capacité à prédire avec précision et à expliquer les comportements des mesures de conservation de l'énergie dans les bâtiments.

Formula	Attribute
$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - p_i}{y_i} \right \cdot 100$	Equation (1)
$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}$	Equation (2)
$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$	Equation (3)
\bar{y}	<i>Moyenne des valeurs réelles</i>
y_i	<i>Valeurs réelles</i>
p_i	<i>Valeurs prédites</i>
n	<i>Nombre de points de données</i>

4 Résultats

Les résultats de la méthodologie mise en œuvre sont résumés dans cette partie, sur la base des étapes discutées ci-dessus. Une vue d'ensemble de l'ensemble de données développé est présentée, suivie des différents modèles développés et de leurs performances.

4.1 Analyse des données

Les paramètres décrits dans le Tableau 1 ont été échantillonnés selon la procédure d'échantillonnage LHS, générant un ensemble de données composé de 1000 échantillons. La Figure 4 présente les histogrammes des variables de sortie. La plupart des caractéristiques ont atteint une distribution quasi normale des variables dépendantes, assurant ainsi la représentativité de la taille de l'ensemble de données.

Une distribution asymétrique est observée dans l'analyse de l'histogramme des émissions totales de carbone (kgCO_2e) en raison des différences radicales entre les émissions de combustibles ou le carbone opérationnel. L'alimentation du réseau électrique québécois à partir de sources hydroélectriques propres à 99 % se traduit par un carbone opérationnel pratiquement nul lorsque le système est électrifié. Par conséquent, les émissions de carbone sont nettement inférieures à celles des solutions basées sur les combustibles fossiles.

Comme mentionné précédemment, les modèles émulative de régression linéaire sont vulnérables aux valeurs extrêmes et aux distributions non normales. Par conséquent, le développement d'un modèle capable de prédire le carbone incorporé peut s'avérer plus pragmatique, le carbone opérationnel pouvant être estimé en prédisant la consommation d'énergie de chauffage.

En outre, un modèle de prédiction de la consommation d'énergie a été sélectionné pour être développé comme l'un des modèles émulatifs. L'inclusion des aspects qui affectent la consommation totale d'énergie dans les caractéristiques développées dans toutes ses parties telles que le chauffage, le refroidissement et les charges d'équipement et de prise.

L'analyse du coût du cycle de vie (ACCV) a été utilisée comme variable cible pour le modèle de coût, afin de prédire le coût total associé aux BRM appliqués.

Trois catégories de modèles d’émulation ont été développées : le modèle énergétique, le modèle carbone et le modèle de coût, utilisant respectivement la consommation d’énergie, le carbone incorporé dans les matières et l’ACCV.

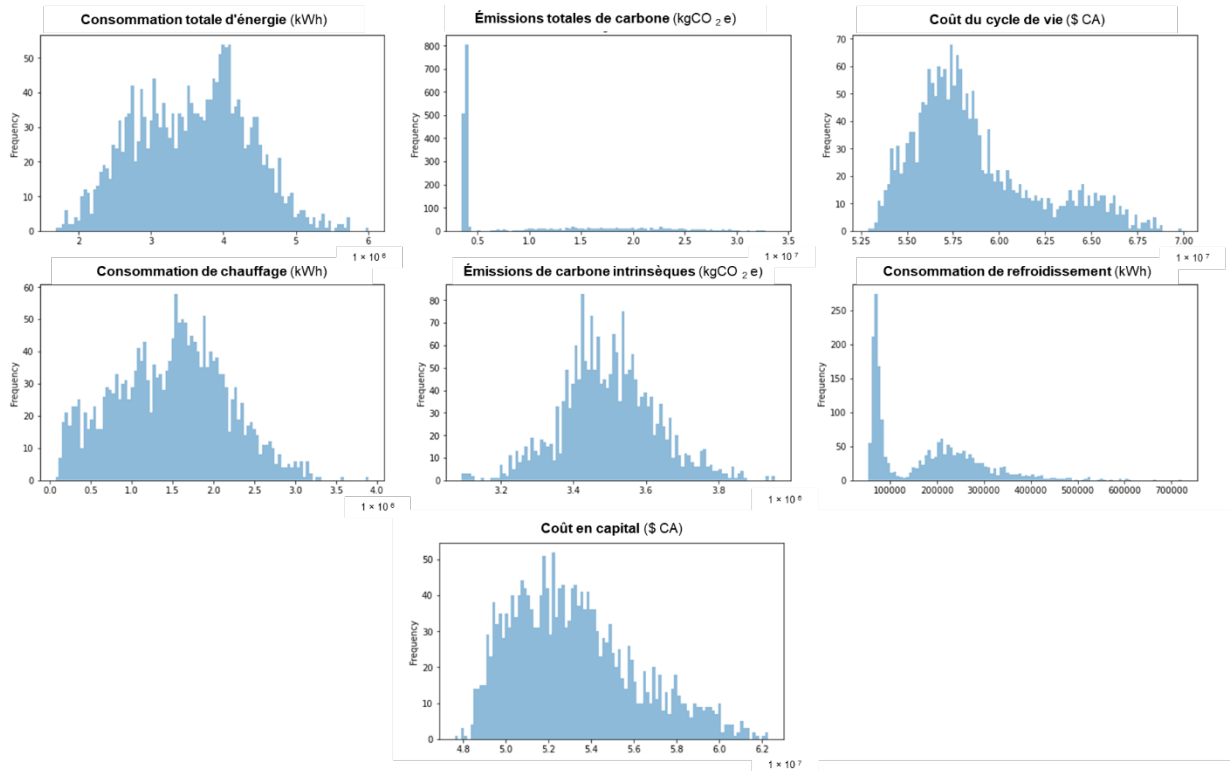


Figure 4 Analyse et distribution des histogrammes des variables de sortie de la simulation

4.1.1 Sous-caractéristiques numériques

Les modèles de régression linéaire à variables multiples (RLMV) font appel à des valeurs numériques pour leur processus d’ajustement. En conséquence, les caractéristiques de nature catégorielle ont été converties en variables numériques, conformément aux données présentées dans le Tableau 2.

La sélection des variables a été guidée par les recommandations établies par les meilleures pratiques [19] documentées dans la littérature ainsi que par une appréhension approfondie du

domaine. Les caractéristiques catégorielles telles que « combustible de chauffage » et « type de contrôle » ont été transformées au moyen d'un encodage d'étiquette et d'un encodage à chaud respectivement.

Tableau 2 sous-attributs choisis pour substituer les caractéristiques catégorielles initiales

	Mur (M)	Toiture (T)	Vitrage (V)	Store (S)	Cadre de Fenêtre (CF)	Éclairage intérieur (EI)	Chauffage, Ventilation et Climatisation (CVC)
I	Caractéristiques Générées						
D							
1	Conductivité Thermique	Conductivité Thermique	Conductivité Thermique	Distance entre le store et le verre	Conductivité Thermique	Densité de Puissance	Combustible de Chauffage
2	Résistance Thermique	Résistance Thermique	transmission de la lumière	Largeur	Résistance Thermique	Fraction Radiante	Énergie Auxiliaire
3	Épaisseur Totale	Épaisseur Totale	transmission solaire	Coût par fenêtre/largeur de lamelle	Chaleur Spécifique	Type de Contrôle (linéaire) — codé	Augmentation de Pression
4	Coût/Surface/Valeur-R	Coût/Surface/Valeur-R	coefficient de gain de chaleur solaire		Épaisseur	Type de Contrôle (arrêt) — codé	Efficacité de Distribution
5	Épaisseur de l'Isolation	Épaisseur de l'Isolation	épaisseur de l'interstice		Coût/Surface/Valeur-U	Type de Contrôle (étagé) — codé	CoP de Chauffage
6	Valeur-U de l'Isolation	Valeur-U de l'Isolation	coût/surface/valeur-U			Coût/Surface/Densité de Puissance	CoP de Refroidissement
7	Chaleur Spécifique de l'Isolation	Chaleur Spécifique de l'Isolation	épaisseur du vitrage				Coût/Surface/CoP de Chauffage
8	Densité de l'Isolation	Densité de l'Isolation	nombre de panneaux				Coût/Charge de Chauffage
9	Coût de l'Isolation/Volume	Coût de l'Isolation/Volume/Valeur-U					
10	Carbone Incorporé de l'Isolation/Valeur-U						

4.2 Résultats du RLMV

La section subséquente dévoile les résultats associés à la régression linéaire multivariée (RLMV), intégrant l'ensemble des modèles élaborés au moyen de diverses méthodes de sélection

et de régularisation des caractéristiques, dans une perspective comparative. De manière générale, pour l'ensemble des modèles RLMV conçus, l'ensemble de données a été scindé en segments dédiés au test et à l'entraînement selon un rapport de 1:4. Plusieurs techniques de validation croisée ont été mises en œuvre au sein des modèles développés, subdivisant ainsi l'ensemble de données d'entraînement en multiples sous-ensembles de validation. Les processus de sélection des caractéristiques mises en avant dans l'étude ont conduit à l'adoption des caractéristiques énumérées dans le Tableau 2.

4.2.1 Performance du modèle de base

Dans une première phase, un modèle de la régression linéaire multivariée (RLMV) préliminaire est ajusté, fournissant une vue d'ensemble des capacités inhérentes au modèle ainsi que des ajustements qui pourraient être nécessaires pour son développement ultérieur. Dans le cadre de ce modèle particulier, aucun ensemble de validation croisée n'a été utilisé. La Figure 5 illustre les performances de trois modèles RLMV de base élaborés pour prédire l'énergie, les émissions de carbone, et les coûts. L'analyse révèle que la structure actuelle des caractéristiques et de l'espace d'échantillonnage a conduit à des performances jugées acceptables pour le modèle des émissions de carbone et le modèle des coûts. Le premier affiche un coefficient de détermination R^2 de 90 % et une erreur moyenne absolue en pourcentage (MAPE) de 0,82 %, tandis que le second présente un R^2 de 96,5 % et une MAPE de 0,74 %.

En revanche, le modèle énergétique a enregistré un R^2 de 78,3 % avec un MAPE de 7,1 %, ce qui reflète un niveau d'erreur considérablement élevé. Toutes les caractéristiques identifiées (51 au total) ont été utilisées dans le développement de ces modèles, une situation qui pourrait potentiellement influencer le surajustement du modèle dans son ensemble. Néanmoins, le

nombre restreint d'échantillons (1 000) rend ce phénomène peu probable. Quant au modèle de coût, il est à noter que les valeurs résiduelles suivent une distribution relativement uniforme au sein de la plage de valeurs considérée.

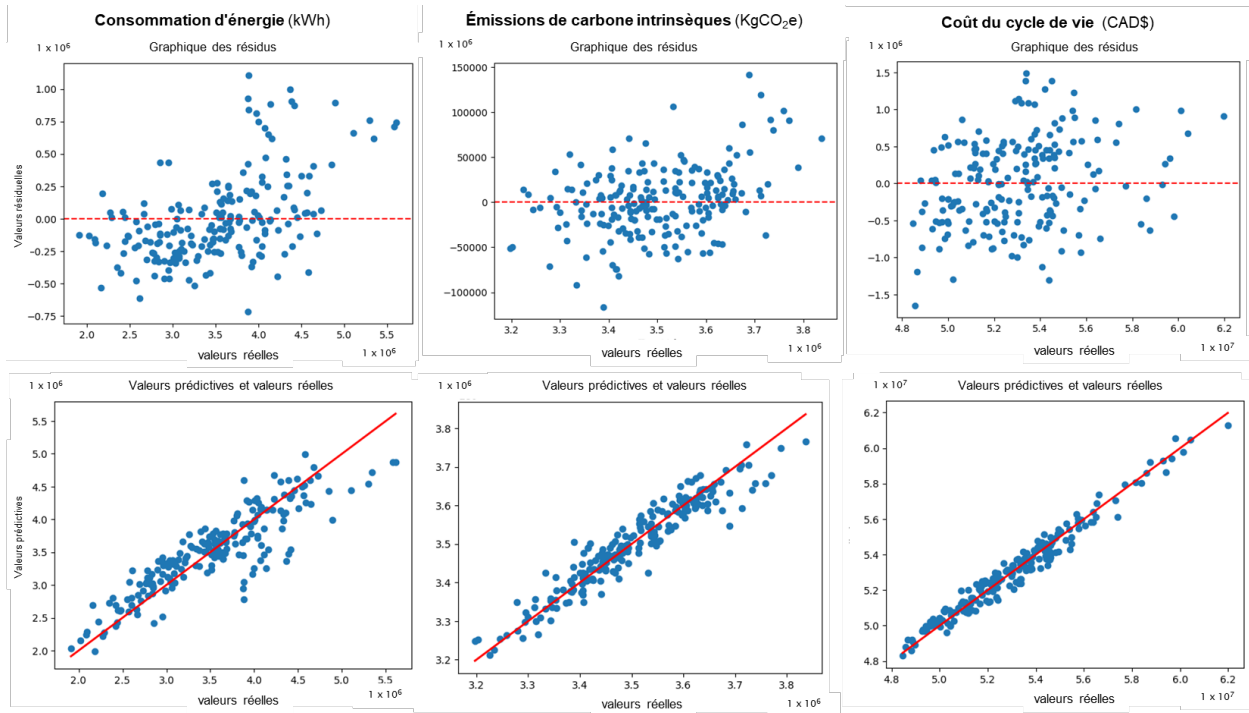


Figure 5 Performance du modèle RLMV dans le cas de base

4.2.2 Sélection de caractéristiques par étapes à rebours

Un processus de sélection des caractéristiques par étapes à rebours a été instauré en tant que première phase dans l'examen des améliorations potentielles des modèles précédemment établis. Un procédé automatisé BSFS (Backward Stepwise Feature Selection) est mis en œuvre, où la sélection des caractéristiques évalue le facteur d'inflation de la variance (VIF) parmi les variables d'entrée, éliminant systématiquement les variables présentant la VIF la plus élevée à chaque étape, jusqu'à l'obtention d'une VIF maximale inférieure ou égale à 10 [71], conformément à d'autres recherches dans ce domaine. La Figure 6 illustre ce processus ainsi que

le nombre final de caractéristiques retenues. À l'issue de ce processus, le nombre de caractéristiques d'entrée s'est réduit à 36. La méthode employée pour le développement du modèle impliquait l'utilisation d'une validation croisée imbriquée à dix niveaux, segmentant l'ensemble de données d'apprentissage en « K » sous-ensembles de taille équivalente. Les ensembles « K-1 » sont combinés pour former l'ensemble d'apprentissages, le sous-ensemble restant étant désigné comme l'ensemble de validation. Cette méthode de validation croisée garantit la généralisation du modèle développé et sa robustesse par rapport à de nouvelles données. Les performances du modèle peuvent être observées dans la Figure 7.

La précision prédictive du modèle s'est améliorée pour tous les modèles. Parallèlement, une amélioration notable a été perçue dans le modèle énergétique, où l'erreur de prédiction a été substantiellement réduite, atteignant une MAPE de 5,8 % et un RMSE de 252155,33 (kWh). Les modèles relatifs aux émissions de carbone et aux coûts ont également vu leur précision de prédiction s'accroître. Le premier modèle atteignant une MAPE de 1,4 % et un RMSE de 62873 (KgCO₂e), et le second une MAPE de 0,73 % et un RMSE de 492000 (\$ CA). Le coefficient de détermination R² de tous les modèles a diminué en raison de la réduction du nombre de caractéristiques. Toutefois, la réduction la plus notable peut être identifiée dans le modèle des émissions de carbone, avec 74,2 %, ce qui suggère soit la nécessité d'un échantillonnage plus substantiel pour développer un modèle plus représentatif, soit l'adoption d'une méthode de sélection des caractéristiques différente pour traiter la multicolinéarité. Pour les modèles de substitution de l'énergie et des coûts, les R² obtenus se chiffrent respectivement à ~88 % et 96,3 %.

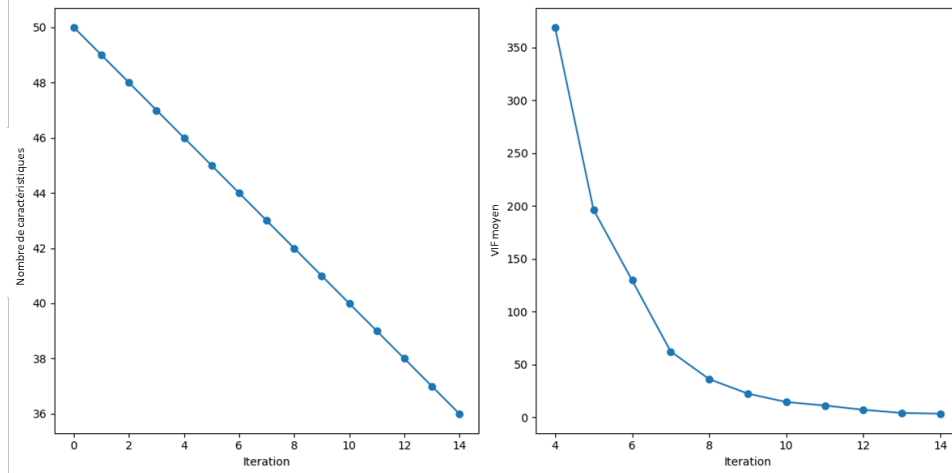


Figure 6 Processus de sélection pas à pas à rebours des caractéristiques

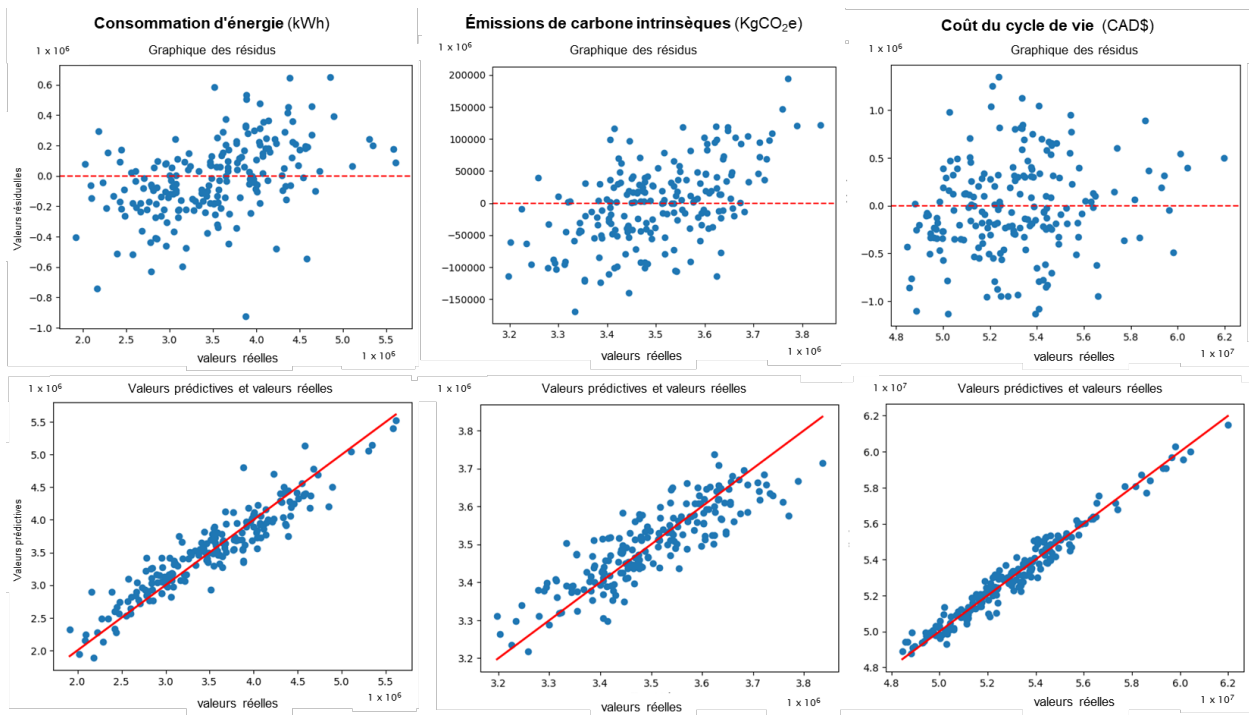


Figure 7 Performance du modèle de sélection pas à pas à rebours

4.2.3 Élimination récursive des caractéristiques

Une méthode alternative de sélection des caractéristiques, l'élimination itérative des caractéristiques, a été mise en œuvre afin de pallier la diminution des performances du coefficient de détermination R^2 . Traditionnellement, la méthode est effectuée suivant un protocole d'élimination visant à obtenir la métrique de performance la plus élevée sans restriction quant au nombre maximal de caractéristiques d'entrée, ce qui peut conduire à l'utilisation de l'ensemble complet de caractéristiques. Dans la présente étude, nous avons employé un critère axé sur l'utilisation du plus petit nombre de caractéristiques permettant de stabiliser la valeur R^2 . Ce processus permet d'harmoniser les performances tout en fournissant une stratégie pour minimiser la multicolinéarité par la réduction du nombre de caractéristiques. Initialement, une démarche de calibrage du modèle et d'identification du nombre optimal de caractéristiques pour stabiliser la valeur R^2 a été entreprise, comme illustré dans la Figure 8. Le nombre optimal de caractéristiques pour les modèles de l'énergie, des émissions de carbone, et des coûts était respectivement de 38, 18, et 14. Il est important de souligner que la technique de validation croisée K-fold imbriqué, similaire à celle précédemment appliquée, a été maintenue dans le modèle afin d'assurer la généralisabilité aux nouvelles données inédites.

Une fois le nombre optimal de caractéristiques déterminé, les modèles ont été ajustés pour une évaluation globale de leurs performances. La Figure 9 détaille les performances réalisées par les différents modèles. L'observation révèle que l'adoption d'un nombre considérablement réduit de caractéristiques d'entrée — spécifiquement, 38 pour l'énergie, 18 pour les émissions de carbone, et 14 pour le coût — n'a pas induit de diminution notable des performances. Plus précisément, par rapport à M2, les modèles de l'énergie et du carbone ont amélioré leurs performances, le

MAPE diminuant respectivement de 0,3 % et de 0,5 %. En outre, une amélioration significative de la valeur R^2 du modèle d'émissions de carbone a été obtenue. Néanmoins, les modèles ont conservé une précision prédictive satisfaisante. Quant au modèle de coût, ses performances ont légèrement diminué par rapport à M2, mais il a conservé une précision significative. De plus, la moyenne de l'erreur absolue en pourcentage (MAPE) du modèle élaboré a enregistré une réduction de 0,2 %, atteignant ainsi une valeur de 0,93 %. Ces résultats témoignent de l'efficacité inhérente à l'exploration de plusieurs techniques de sélection de caractéristiques au sein d'un modèle unique, visant à optimiser les performances. Cette approche systématique et robuste contribue à une meilleure compréhension des relations structurelles dans les données et favorise la création de modèles prédictifs plus précis et informatifs.

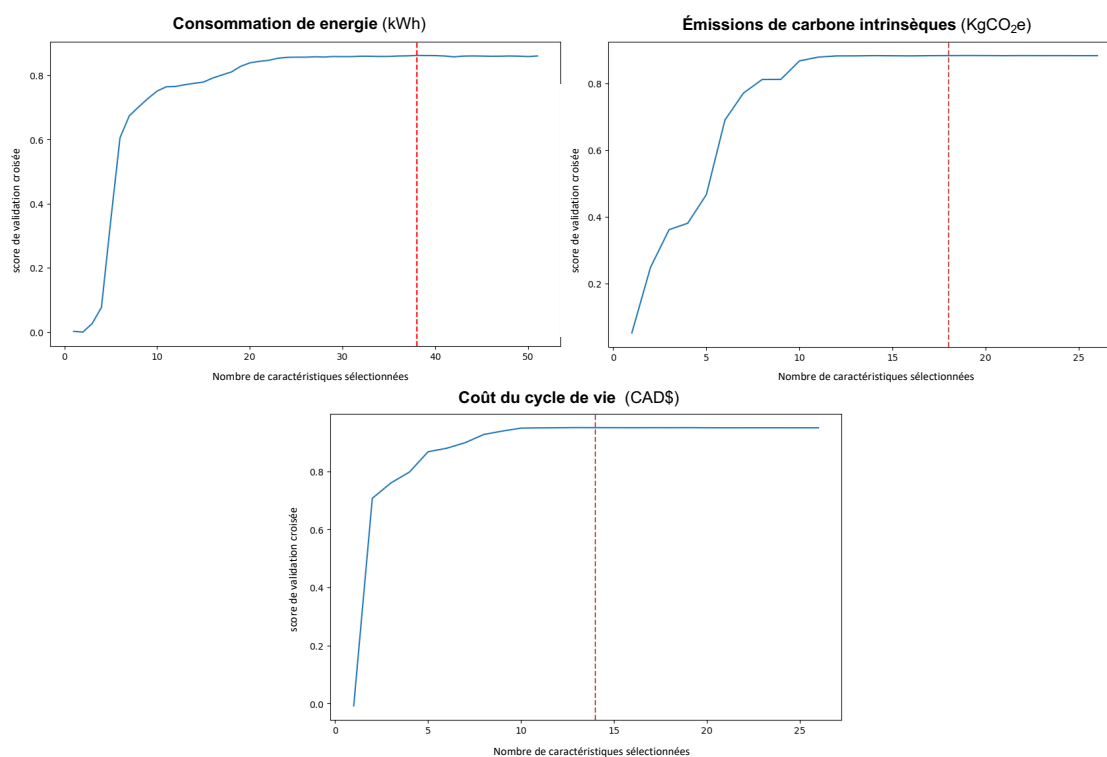


Figure 8 Processus de sélection récurrente des caractéristiques - R^2 en fonction du nombre de caractéristiques

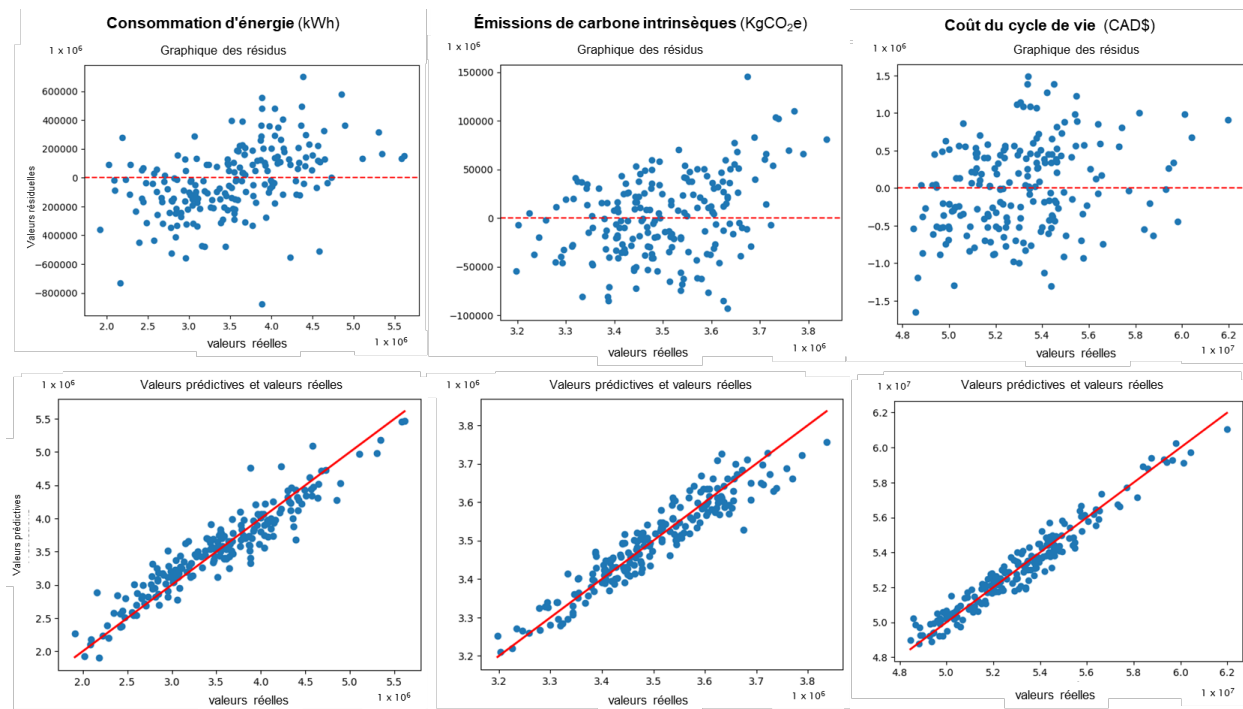


Figure 9 Performance du modèle d'élimination récursive des caractéristiques

4.2.4 Régression Elastic Net

L'application de méthodes intégrées de sélection des caractéristiques est présentée comme une méthode prometteuse pour formuler un modèle généralisable tout en évitant le problème de l'ajustement trop grand. L'une de ces méthodes, Elastic Net, est particulièrement remarquable en ce qu'elle fusionne les techniques de régularisation L1 et L2 pour garantir la robustesse et la stabilité du modèle.

Dans la mise en œuvre de l'Elastic Net, un terme de pénalité (λ) est introduit, contribuant à la somme des coefficients au carré ainsi qu'à la somme des valeurs absolues. Cette contribution réduit les caractéristiques ayant des coefficients significatifs et modifie leurs valeurs respectives.

La Figure 10 offre une visualisation de ce processus de contraction du coefficient, mettant en évidence les différences entre les trois modèles formulés dans le cadre de l'étude.

Ce processus est d'une importance capitale pour la stabilisation des modèles développés, garantissant la préservation de leurs performances lorsqu'ils sont confrontés à des données jamais vues auparavant. Lors de la mise en œuvre de ce modèle, la validation croisée 10 fois a été utilisée pour optimiser la valeur du facteur de pénalisation (λ). Au cours de cette phase de développement, une recherche exhaustive sur la base d'une grille a été effectuée parmi les hyperparamètres — y compris, mais sans s'y limiter, la tolérance, le rapport L1 et la valeur alpha — afin d'identifier la configuration optimale pour le modèle.

Cette approche intégrée incarne une méthodologie rigoureuse et efficace dans la sélection des caractéristiques, offrant un équilibre entre complexité du modèle et capacité prédictive. Le recours à Elastic Net, en particulier, constitue un exemple probant de l'adoption de techniques de régularisation avancées pour accroître la fiabilité et l'applicabilité du modèle dans un contexte d'analyse de données sophistiqué.

La Figure 11 révèle les performances des modèles Elastic Net conçus pour prédire l'énergie, les émissions de carbone et les coûts, avec des améliorations dans toutes les mesures de performance. Les valeurs RMSE pour les trois modèles sont respectivement de $\sim 24,6$ (MWh) pour l'énergie, ~ 41 (TonnesCO₂) pour les émissions de carbone, et 447.6 (Mille \$ CA) pour les coûts, indiquant une haute précision des prédictions. Les valeurs de R² montrent une description efficace de la capacité de variance des modèles, avec 88.5 % pour l'énergie, 89 % pour les émissions de carbone, et $\sim 97\%$ pour les coûts. En outre, les comparaisons d'erreurs entre les ensembles de données de test et de formation révèlent une précision presque similaire pour les modèles d'émission de carbone et de coût, avec des différences de 0,1 % et 0,01 %, tandis que le modèle énergétique présente une légère différence de 0,09 % par rapport aux 0,8 % initiaux.

Malgré l'utilisation de toutes les caractéristiques d'entrée (51 au total), les modèles Elastic Net ont amélioré les performances et démontré une forte généralisation sur les données de formation et de test. Cela souligne l'efficacité de l'approche Elastic Net dans la création de modèles de prédiction robustes et précis qui sont bien adaptés au traitement d'un grand nombre de caractéristiques, confirmant son importance dans le développement de modèles d'apprentissage automatique pour les données complexes et multidimensionnelles.

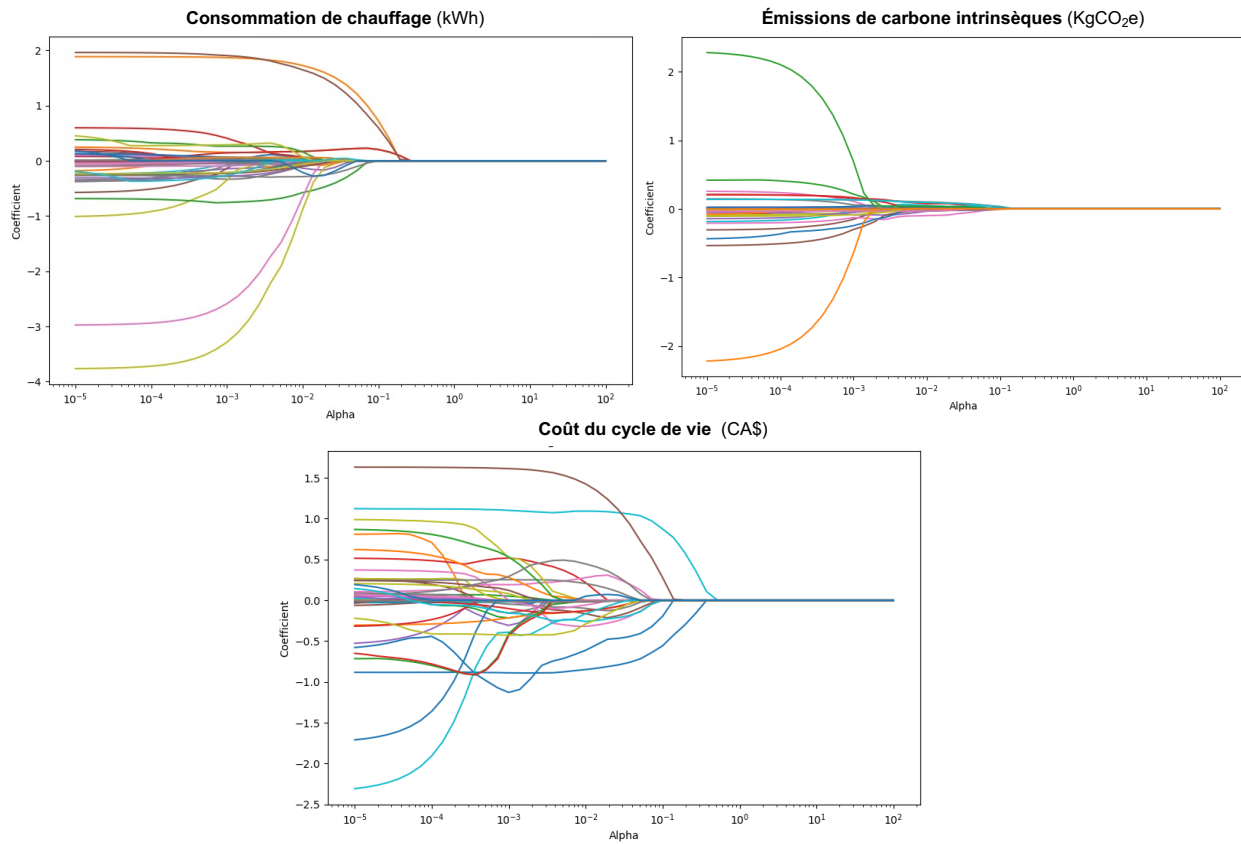


Figure 10 Coefficients de régression Elastic Net processus de rétrécissement

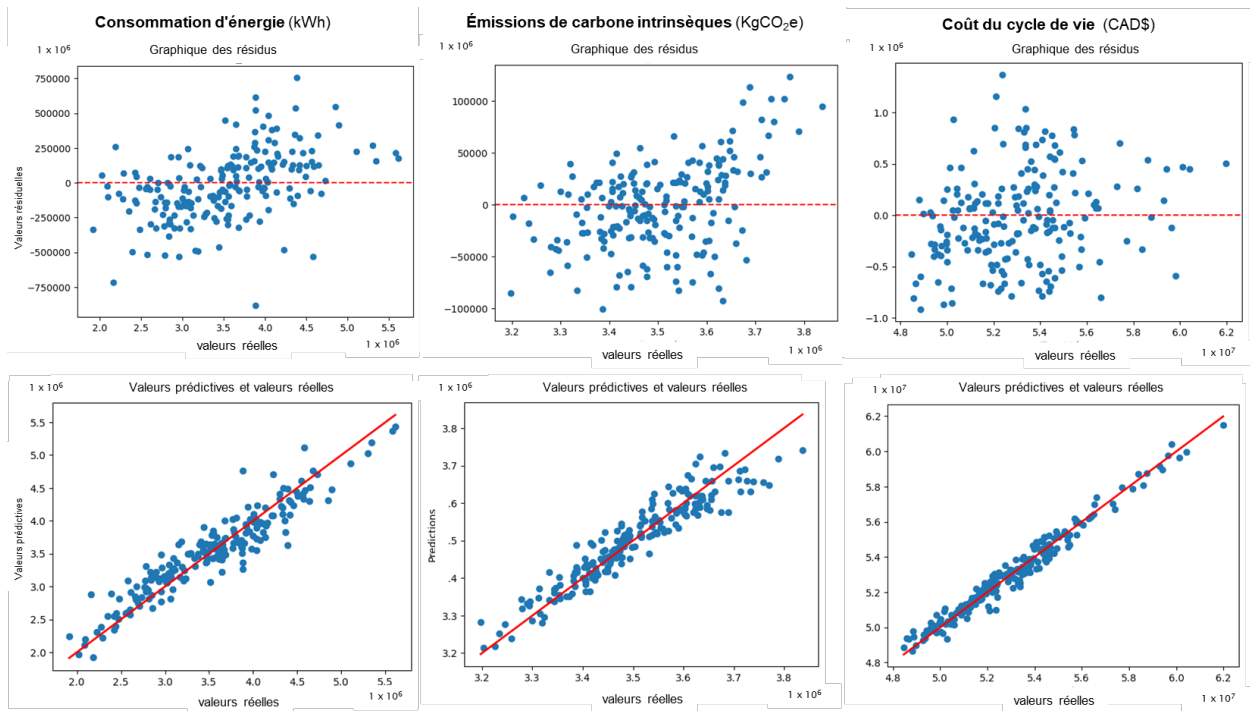


Figure 11 Performance du modèle de régression Elastic Net

4.3 Temps de calcul

Dans cette partie, une évaluation systématique du temps de calcul nécessaire a été réalisée pour les différents modèles développés, en soulignant les avantages et les limites de chaque approche.

Une cible hypothétique d'un million de simulations a été sélectionnée afin de représenter l'échelle complexe des combinaisons possibles des MRB, qui peuvent culminer à près de trois millions de configurations uniques. En utilisant un modèle de simulation conventionnel et un processeur à 12 cœurs, l'exécution de ces simulations nécessiterait approximativement 567 jours.

Le développement de l'ensemble de données utilisé dans cette étude, basée sur 1000 échantillons, a requis 13 heures de temps de calcul comme indiqué dans le Tableau 3.

En ce qui concerne le développement des modèles d'émulation, le modèle Elastic Net RLMV a été ajusté dans un délai de 6,4 secondes, illustrant la faisabilité et l'efficacité de l'approche

d'émulation. Cette efficacité se manifeste clairement dans la prédiction d'un million d'échantillons, où la plupart des modèles ont réussi à fournir des prédictions presque en temps réel.

La Figure 12 illustre cette comparaison, mettant en évidence l'efficacité substantielle qui peut être atteinte en employant une méthodologie de modèle d'émulation. Cette méthodologie permet d'approximer l'impact des MRB sur la performance du bâtiment, en abordant des objectifs multidimensionnels tels que l'énergie, les émissions de carbone, et le coût. L'importance de ces résultats réside dans leur capacité à informer les décideurs sur les méthodes de modélisation optimales, ce qui peut conduire à des économies considérables en matière de temps et de ressources.

Tableau 3 Analyse du temps de calcul (en secondes) entre les modèles de substitution développés

	<i>MEB</i>	<i>M1</i>	<i>M2</i>	<i>M3</i>	<i>M4</i>
<i>Temps de simulation (secondes)</i>	49,000,000	49,000	49,000	49,000	49,000
<i>Ajustement et test du modèle</i>	0	0.3	3.2	0.9	6.4
<i>Prédiction de 1 000 000 d'échantillons</i>	0	0.2	0.3	0.2	0.1

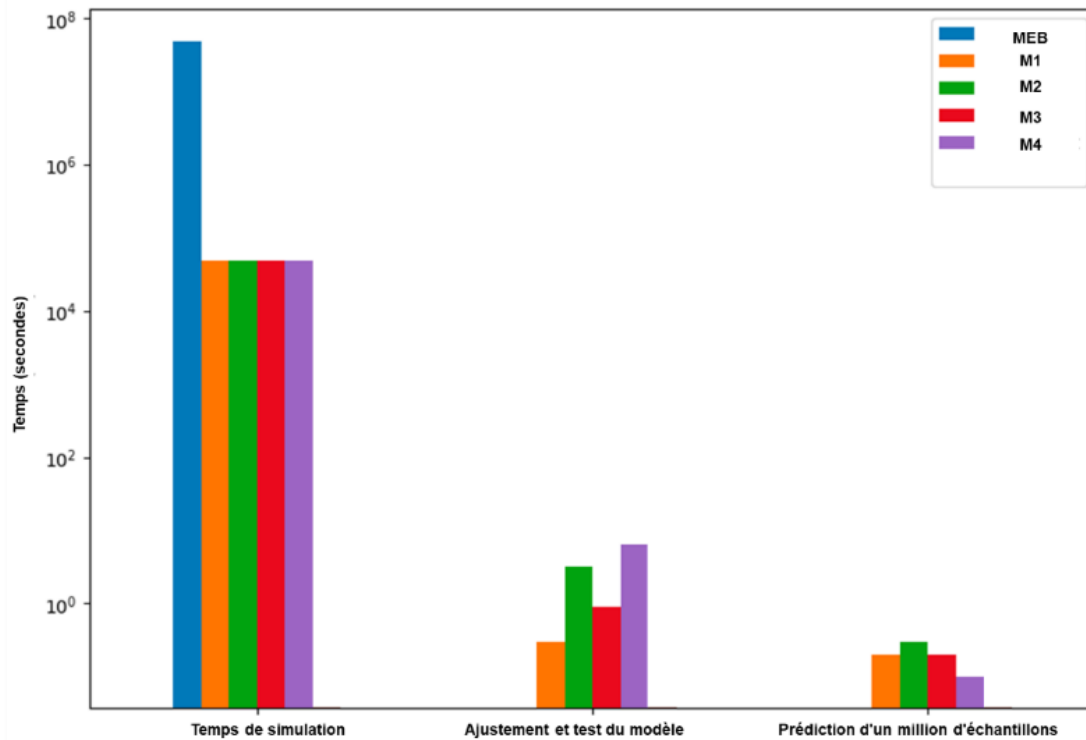


Figure 12 Comparaisons entre les modèles développés et le modèle de simulation conventionnel en termes de temps de calcul

4.4 Comparaison avec d'autres modèles d'apprentissage automatique

Pour démontrer et comprendre la performance des modèles développés. Un modèle parallèle a été développé en utilisant l'algorithme XGBoost, qui est un algorithme plus performant et plus flexible pour s'adapter aux données. L'algorithme XGBoost a permis une amélioration significative de l'adéquation aux données, qui a atteint 91,5 %, 95 % et 95 % pour les modèles d'énergie, de carbone et de coût respectivement. En outre, l'erreur a été réduite à 4,7 %, 0,6 % et 0,8 % respectivement. Les performances du modèle sont présentées à l'adresse Figure 13.

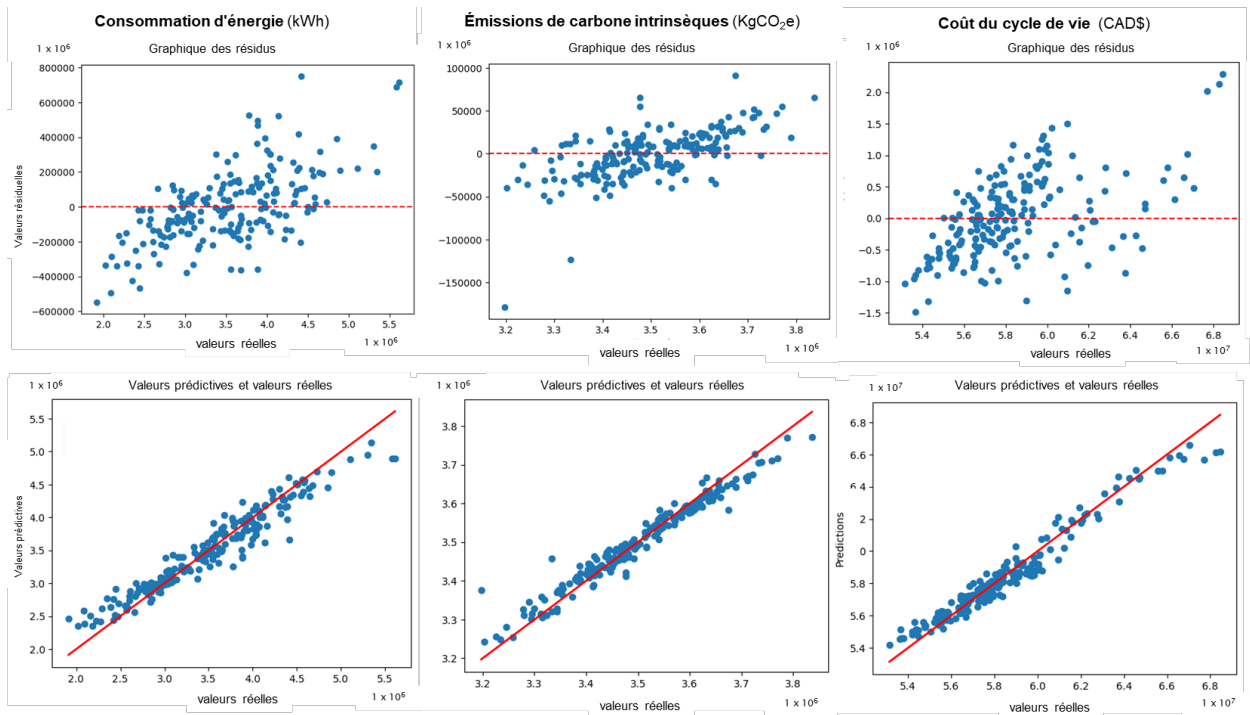


Figure 13 Performance de prédiction des modèles basés sur l'algorithme XGBoost

4.5 Prédiction sur d'autres ensembles de données non vus

Pour démontrer le potentiel d'utilisation des modèles développés sur d'autres cas, un test a été appliqué pour vérifier la capacité à prédire des cas non vus. Par la suite, le modèle a été évalué sur un autre bâtiment présentant des caractéristiques très similaires et situé au même endroit. Le modèle a présenté des performances acceptables avec un pourcentage d'erreur de 12 %.

Cependant, l'adéquation du modèle aux données inédites était proche de R^2 50%. Cela pourrait signifier qu'il faut davantage de données pour entraîner le modèle et qu'il est possible d'utiliser des techniques d'apprentissage par transfert comme indiqué dans Figure 14.

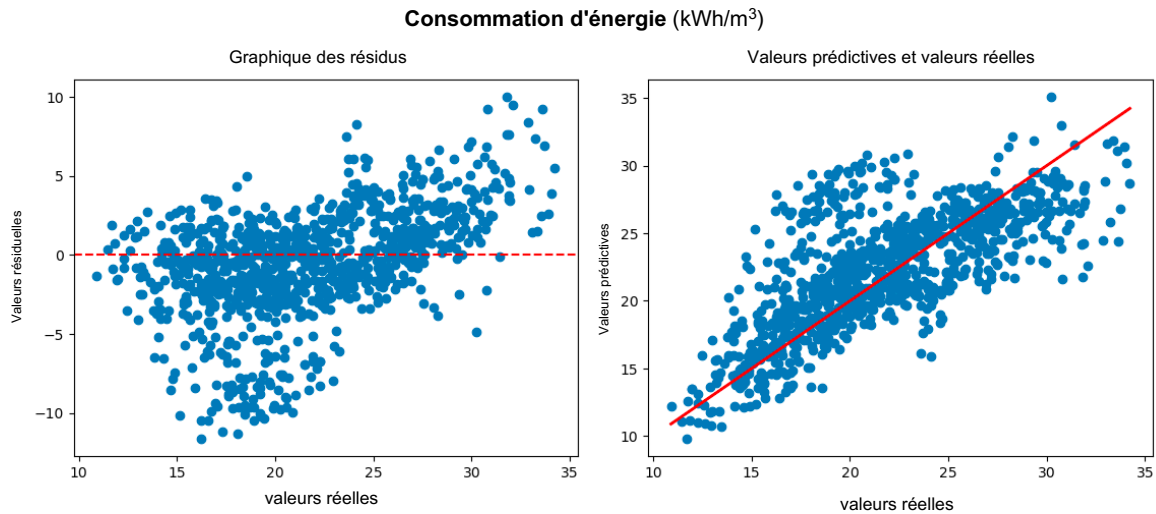


Figure 14 performance du modèle Elastic Net sur des données inédites provenant d'un autre bâtiment

5 Soutien à la prise de décision

L'étude vise à démontrer l'importance et les possibilités de l'aide à la décision dans le processus de rénovation des bâtiments. Par conséquent, les modèles développés ont été utilisés dans un format interactif. Tout d'abord, une plateforme a été développée à l'Université Concordia pour créer un pont entre les différentes parties prenantes, les propriétaires de bâtiments et les décideurs afin d'étudier la ville. Citylayers est une plateforme qui héberge la simulation possible de tous les bâtiments de la ville de Montréal, comme le montre la Figure 15. L'ajout de cette recherche consiste à proposer une couche supplémentaire, où les modèles d'apprentissage automatique sont hébergés pour permettre de tester toutes les combinaisons de mesures de rénovation, comme le montre la Figure 16. L'utilisateur peut prédire l'énergie, les émissions de carbone et le coût du cycle de vie du bâtiment étudié, développer plusieurs scénarios et comparer leurs performances.



Figure 15 capture d'écran de la plateforme développée par l'Université Concordia (Citylayers)

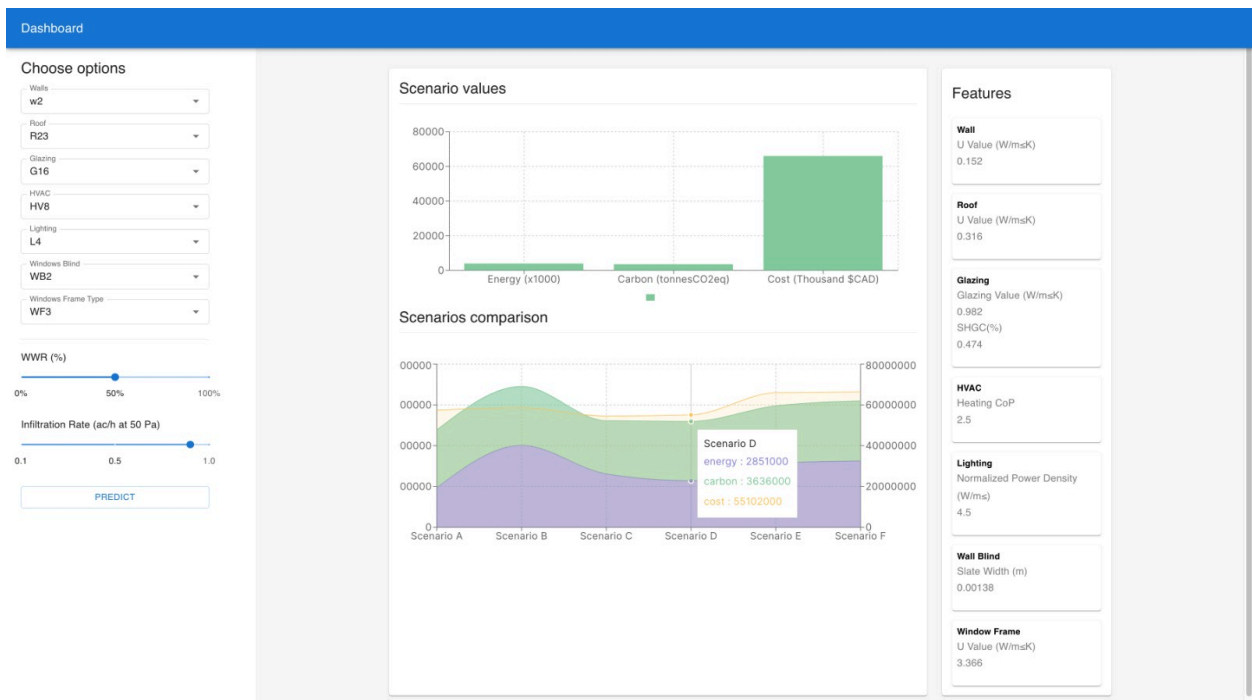


Figure 16 L'interface interactive développée pour les modèles développés

6 Discussion et recommandations

Malgré des résultats apparemment significatifs et positifs issus des rétrofits énergétiques dans le secteur des bâtiments, l'adoption de projets et de pratiques d'efficacité énergétique continue d'être lente. Plusieurs barrières entravent l'adoption de mesures de rénovation énergétique, complexifiant ainsi la mise en œuvre de ces stratégies. La lenteur dans l'adoption de la rénovation énergétique est un phénomène multifactoriel, lié à des obstacles informationnels, économiques et technologiques. La recherche continue et la collaboration entre les différents acteurs du domaine sont nécessaires pour surmonter ces défis et réaliser le plein potentiel des rétrofits énergétiques.

6.1 Recommandations

- 1- Le rôle actif des programmes gouvernementaux pour permettre la création de programmes de mobilisation des connaissances à source ouverte peut représenter une grande opportunité pour atteindre un rythme plus rapide d'adaptation des bâtiments. Les asymétries d'information entre les différentes parties prenantes représentent un obstacle majeur. Les propriétaires, les locataires, les gestionnaires de bâtiments et les prestataires de services énergétiques n'ont pas toujours accès aux mêmes informations. Cela crée une confusion et un manque de coordination qui peut nuire à l'efficacité des projets de rénovation énergétiques constitue une barrière. La complexité et la diversité des solutions disponibles peuvent rendre difficile la sélection de la méthode la plus appropriée pour un bâtiment ou un projet donné. Sans connaissances adéquates, les décisions peuvent être bloquées et entravées. La motivation d'acquérir des gains potentiels sans connaître le

savoir-faire et sans éviter les frais de consultation peut arrêter de nombreux projets de transformation potentiels.

- 2- La construction d'informations dans les ensembles de données publiques est un grand défi. Les décideurs doivent avoir la possibilité de mettre les données sur les bâtiments à la disposition du public et s'y référer. Cela permettra de faciliter les recherches et de mettre au point des méthodes automatisées susceptibles de réduire les coûts et de mobiliser les connaissances à grande échelle. Les stratégies de données ouvertes sont un excellent moyen de permettre la mobilisation des connaissances et, dans le secteur du bâtiment, elles sont extrêmement nécessaires. Le manque de données fiables et complètes peut conduire à des évaluations inexactes des performances et à une adoption plus lente des meilleures pratiques.
- 3- Cette recherche s'attaque à l'accès aux connaissances associées aux obstacles à la rénovation des bâtiments en fournissant une méthode pour obtenir des modèles informatiques abordables qui peuvent ensuite être utilisés pour développer des enquêtes en temps réel. Cette étude s'appuie sur la littérature existante qui étudie l'efficacité du développement de modèles d'émulation, en introduisant une approche méthodologique innovante incluant l'effet de multiples méthodes de sélection des caractéristiques. L'utilisation de modèles d'apprentissage automatique parallèlement aux modèles traditionnels de performance des bâtiments représente une excellente occasion d'accélérer le processus d'examen et de développement de modèles de performance pour les bâtiments existants.

- 4- Le travail présenté s'appuie sur la littérature existante qui étudie l'efficacité du développement de modèles d'émulation, en introduisant une approche méthodologique incluant l'effet de multiples méthodes de sélection des caractéristiques. L'utilisation de modèles d'apprentissage automatique parallèlement aux modèles traditionnels de performance des bâtiments représente une excellente occasion d'accélérer le processus d'examen et de développement de modèles de performance pour les bâtiments existants et les nouvelles constructions. La réalisation de l'analyse de la performance des bâtiments en temps réel permet de déployer davantage de voies de mobilisation des connaissances.
- 5- Les réglementations et les incitations peuvent être utilisées pour encourager l'adoption de la rénovation énergétique et garantir le respect des normes appropriées. Le manque de connaissances et l'absence d'outils accessibles sont des obstacles majeurs qui ralentissent l'adoption de la rénovation énergétique dans le secteur du bâtiment. Les villes et les gouvernements ont un rôle essentiel à jouer pour combler ces lacunes. En investissant dans la formation, la recherche et le développement d'outils, et en promouvant activement la rénovation énergétique par le biais de réglementations et d'incitations, ils peuvent accélérer le processus d'adoption et contribuer à la réalisation des objectifs énergétiques et environnementaux à long terme.

6.2 Limitations de l'études

- 1- L'étude a démontré la capacité des modèles, mais il convient de noter que les modèles sont développés sur la base de données de simulation.

- 2- Les données utilisées pour développer les modèles servent de preuve de concept et démontrent le succès de la méthodologie. Toute recommandation doit être revérifiée en utilisant des données réelles pour calibrer le modèle.
- 3- Le modèle développé est limité au bâtiment étudié et à la typologie de bâtiment faisant l'objet de la recherche, et de nouveaux modèles doivent être développés pour de nouvelles catégories de bâtiments.
- 4- Les simulations ont pris en compte des tarifs constants pour la consommation d'électricité et de gaz naturel et les données météorologiques publiques disponibles. En outre, les valeurs des coûts ont été utilisées à partir de sources disponibles et peuvent ne pas être exactes.
- 5- Les données sur les coûts utilisées pour le développement du modèle de coûts dépendent du logiciel de simulation, qui devrait être remplacé par des données actualisées et localisées pour une application dans le contexte québécois.

7 References

1. Government of Canada Canada's Climate Plan - Canada.Ca Available online: <https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan.html> (accessed on 21 November 2022).
2. Nutkiewicz, A.; Yang, Z.; Jain, R.K. Data-Driven Urban Energy Simulation (DUE-S): A Framework for Integrating Engineering Simulation and Machine Learning Methods in a Multi-Scale Urban Energy Modeling Workflow. *Appl Energy* **2018**, *225*, 1176–1189.

3. Hong, T.; Chen, Y.; Luo, X.; Luo, N.; Lee, S.H. Ten Questions on Urban Building Energy Modeling. *Build Environ* **2020**, *168*, 106,508.
4. Güneralp, B.; Zhou, Y.; Ürge-Vorsatz, D.; Gupta, M.; Yu, S.; Patel, P.L.; Fragkias, M.; Li, X.; Seto, K.C. Global Scenarios of Urban Density and Its Impacts on Building Energy Use through 2050. *Proc Natl Acad Sci U S A* **2017**, *114*, 8945–8950,
doi:10.1073/PNAS.1606035114/SUPPL_FILE/PNAS.1606035114.SD03.XLSX.
5. Creutzig, F.; Baiocchi, G.; Bierkandt, R.; Pichler, P.P.; Seto, K.C. Global Typology of Urban Energy Use and Potentials for an Urbanization Mitigation Wedge. *Proc Natl Acad Sci U S A* **2015**, *112*, 6283–6288,
doi:10.1073/PNAS.1315545112/SUPPL_FILE/PNAS.1315545112.SD01.XLS.
6. Canada Energy Regulator CER – Provincial and Territorial Energy Profiles Available online: <https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/provincial-territorial-energy-profiles/provincial-territorial-energy-profiles-explore.html> (accessed on 21 November 2022).
7. État de l'énergie Au Québec 2023 – Chaire de Gestion Du Secteur de l'énergie Available online : <https://energie.hec.ca/eeq/> (accessed on 18 June 2023).
8. Programme Rénoclimat | Innovation et Transition Énergétiques Available online : <https://transitionenergetique.gouv.qc.ca/residentiel/programmes/renoclimat> (accessed on 18 June 2023).
9. Programme ÉcoPerformance | Innovation et Transition Énergétiques Available online : <https://transitionenergetique.gouv.qc.ca/affaires/programmes/ecoperformance> (accessed on 18 June 2023).

10. Deb, C.; Schlueter, A. Review of Data-Driven Energy Modelling Techniques for Building Retrofit. *Renewable and Sustainable Energy Reviews* **2021**, *144*, 110,990, doi:10.1016/J.RSER.2021.110990.
11. Shen, P.; Braham, W.; Yi, Y.; Eaton, E. Rapid Multi-Objective Optimization with Multi-Year Future Weather Condition and Decision-Making Support for Building Retrofit. *Energy* **2019**, *172*, 892–912, doi:10.1016/j.energy.2019.01.164.
12. Westermann, P.; Welzel, M.; Evins, R. Using a Deep Temporal Convolutional Network as a Building Energy Surrogate Model That Spans Multiple Climate Zones. *Appl Energy* **2020**, *278*, 115,563, doi:10.1016/j.apenergy.2020.115563.
13. Zhang, H. Energy Retrofit Evaluation in Residential Construction : A Life Cycle Thinking Approach. **2021**, doi:10.14288/1.0401894.
14. Feng, H.; Liyanage, D.R.; Karunathilake, H.; Sadiq, R.; Hewage, K. BIM-Based Life Cycle Environmental Performance Assessment of Single-Family Houses: Renovation and Reconstruction Strategies for Aging Building Stock in British Columbia. *J Clean Prod* **2020**, *250*, 119,543, doi:10.1016/J.JCLEPRO.2019.119543.
15. Charles, A.; Maref, W.; Ouellet-Plamondon, C.M. Case Study of the Upgrade of an Existing Office Building for Low Energy Consumption and Low Carbon Emissions. *Article published in Energy and Buildings* **2019**, *183*, 151–160, doi:10.1016/j.enbuild.2018.10.008.
16. Costa-Carrapiço, I.; Raslan, R.; González, J.N. A Systematic Review of Genetic Algorithm-Based Multi-Objective Optimisation for Building Retrofitting Strategies

- towards Energy Efficiency. *Energy Build* **2020**, *210*, 109,690, doi:10.1016/J.ENBUILD.2019.109690.
17. Yigit, S. A Machine-Learning-Based Method for Thermal Design Optimization of Residential Buildings in Highly Urbanized Areas of Turkey. *Journal of Building Engineering* **2021**, *38*, 102,225, doi:10.1016/j.jobbe.2021.102225.
 18. Hey, J.; Siebers, P.; Nathanail, P.; Ozcan, E. Surrogate Optimization of Energy Retrofits in Domestic Building Stocks Using Household Carbon Valuations. **2023**, doi:10.1080/19401493.2022.2106309.
 19. Westermann, P.; Evins, R. Surrogate Modelling for Sustainable Building Design – A Review. *Energy Build* **2019**, *198*, 170–186.
 20. Thrampoulidis, E.; Mavromatidis, G.; Lucchi, A.; Orehounig, K. A Machine Learning-Based Surrogate Model to Approximate Optimal Building Retrofit Solutions. *Appl Energy* **2021**, *281*, 116,024, doi:10.1016/J.APENERGY.2020.116024.
 21. Moser, S.; Fauveaud, G.; Cutts, A. Montréal: Towards a Post-Industrial Reinvention. *Cities* **2019**, *86*, 125–135, doi:10.1016/J.CITIES.2018.09.013.
 22. National Energy Code of Canada for Buildings 2017 - National Research Council Canada Available online: <https://nrc.canada.ca/en/certifications-evaluations-standards/codes-canada/codes-canada-publications/national-energy-code-canada-buildings-2017> (accessed on 13 October 2022).
 23. Rabani, M.; Bayera Madessa, H.; Mohseni, O.; Nord, N. Minimizing Delivered Energy and Life Cycle Cost Using Graphical Script: An Office Building Retrofitting Case. *Appl Energy* **2020**, *268*, 114,929, doi:10.1016/J.APENERGY.2020.114929.

24. Ye, Y.; Lou, Y.; Zuo, W.; Franconi, E.; Wang, G. How Do Electricity Pricing Programs Impact the Selection of Energy Efficiency Measures? – A Case Study with U.S. Medium Office Buildings. *Energy Build* **2020**, *224*, 110,267, doi:10.1016/J.ENBUILD.2020.110267.
25. Lou, Y.; Yang, Y.; Ye, Y.; Zuo, W.; Wang, J. The Effect of Building Retrofit Measures on CO2 Emission Reduction – A Case Study with U.S. Medium Office Buildings. *Energy Build* **2021**, *253*, 111,514, doi:10.1016/j.enbuild.2021.111514.
26. Shen, P.; Braham, W.; Yi, Y. The Feasibility and Importance of Considering Climate Change Impacts in Building Retrofit Analysis. *Appl Energy* **2019**, *233–234*, 254–270, doi:10.1016/j.apenergy.2018.10.041.
27. Sangireddy, S.A.R.; Bhatia, A.; Garg, V. Development of a Surrogate Model by Extracting Top Characteristic Feature Vectors for Building Energy Prediction. *Journal of Building Engineering* **2019**, *23*, 38–52, doi:10.1016/J.JOBE.2018.12.018.
28. Nagpal, S.; Mueller, C.; Aijazi, A.; Reinhart, C.F. A Methodology for Auto-Calibrating Urban Building Energy Models Using Surrogate Modeling Techniques. <https://doi-org.lib-ezproxy.concordia.ca/10.1080/19401493.2018.1457722> **2018**, *12*, 1–16, doi:10.1080/19401493.2018.1457722.
29. Melo, A.P.; Versage, R.S.; Sawaya, G.; Lamberts, R. A Novel Surrogate Model to Support Building Energy Labelling System: A New Approach to Assess Cooling Energy Demand in Commercial Buildings. *Energy Build* **2016**, *131*, 233–247, doi:10.1016/J.ENBUILD.2016.09.033.

30. Dutta, S.; Gandomi, A.H. Design of Experiments for Uncertainty Quantification Based on Polynomial Chaos Expansion Metamodels. *Handbook of Probabilistic Models* **2020**, 369–381, doi:10.1016/B978-0-12-816514-0.00015-1.
31. Zhang, L.; Wen, J.; Li, Y.; Chen, J.; Ye, Y.; Fu, Y.; Livingood, W. A Review of Machine Learning in Building Load Prediction. *Appl Energy* **2021**, 285, 116,452, doi:10.1016/J.APENERGY.2021.116452.
32. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*. **2009**, doi:10.1007/978-0-387-84858-7.
33. Saad, M.M.; Eicker, U. Investigating the Reliability of Building Energy Models: Comparative Analysis of the Impact of Data Pipelines and Model Complexities. *Journal of Building Engineering* **2023**, 71, 106,511, doi:10.1016/J.JOBE.2023.106511.