Biomass Supply Chain Resilience: Integrating Demand and Availability Predictions into Routing Decisions Using Machine Learning

Foad Esmaeili

A Thesis in the Department of
Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of
Master of Applied Science (Building Engineering)

Concordia university
Montreal, Quebec, Canada

August 2022

Concordia University

School of graduate studies

This is to certify that the thesis prepared

By: Foad Esmaeili

Entitled: **Biomass Supply Chain Resilience: Integrating Demand and Availability Predictions into Routing Decisions Using Machine Learning**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Building engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

Dr. Ursula Eicker ....…………………………………………………… Chair

Dr. Ursula Eicker....…………………….………………………………. Examiner

Dr. Yunping Liang …….……………………………………………….. Examiner

Dr. Fuzhan Nasiri……….……………………………………………… Co-Supervisor

Dr. Fereshteh Mafakheri …………………………………………….. Co-Supervisor

Approved by

Dr. M. Nik-Bakht,

Graduate Program Director

Dr. M. Debbabi, Dean

Gina Cody School of Engineering and Computer Science

August 2022

Abstract

Biomass Supply Chain Resilience: Integrating Demand and Availability Predictions into Routing Decisions Using Machine Learning

Foad Esmaeili

Abstract - Renewable energy sources have been pursued as a means of mitigating carbon emission from the energy sector. As biomass resources are a part of natural carbon cycle, they have the potential to mitigate carbon emissions as a renewable source while reducing waste and residues. It shall be noted that biomass has its own challenges as well. Seasonality and disruption risks are some of the disadvantages of biomass resources. Therefore, it is imperative that biomass supply chains be managed such that to withstand disruptions and provide customers with reliable stocks available. In recent years, there has been a growing attention to research on energy supply chain resilience. In case of biomass, most studies have integrated predictions for either supply or demand side of biomass supply chains. This study aims at addressing this gap by formulating biomass supply chain resilience subject to integrating the predictions from both supply and demand dimensions. In doing so, we compare the performance of a host of machine learning techniques combined with routing algorithms. A case study with real (supply and demand) data is considered to present the applicability and usefulness of the proposed methodology accompanied by a results analysis. We then conclude by summarizing the contributions, limitations, and presenting opportunities for future research.

Acknowledgement

I could not have gone through this surprisingly amazing journey without the support of my supervisors, Dr. Fereshteh Mafakheri and Dr. Fuzhan Nasiri, whose help shed light on my path. There are no words to express my gratitude for the opportunity that they provided me with.

I am also grateful for my family’s support and encouragement in every moment of this journey. And the last but not least, I would like to express my most sincere gratitude to my beloved wife, Hadis, for her everlasting love, and support that made my journey as a master’s degree student memorable. We started our master’s degrees at the same time, and I cannot tell how happy I am that we are finishing it alongside each other.

Table of Contents

List of Figures

List of Tables

List of abbreviations

BIMAT	Biomass Inventory Mapping and Analysis Tool
BPS	Broader Public Sector of Ontario
CSCMP	Council of Supply Chain Management Professionals
CVRP	Capacitated Vehicle Routing Problem
DBI	Davies-Bouldin Index
GIS	Geographic Information System
HVAC	Heating, Ventilation, and Air Conditioning
kNN	k-Nearest Neighbours
LiDAR	Light Detection and Ranging
LSTM	Long Short-Term Memory
MPC	Model-Based Predictive Control
PTD	Percent of Total Difference
R^2	Coefficient of Determination / R-Squared
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
VRP	Vehicle Routing Problem
VRPMS	Vehicle Routing Problem with Multiple Synchronisation Constraints
VRPP	Vehicle Routing Problem with Profits
VRPPD	Vehicle Routing Problem with Pick-up and Delivery
VRPTW	Vehicle Routing Problem with Time Windows

Chapter 1: Introduction

1.1 Overview

Fossil fuels still account for most of the end-use energy production in the world. Since these resources are not renewable, decision-makers have tried to encourage end-users to switch to renewable, environmental-friendly resources such as wind, solar, biomass, etc. One of the possible options to replace fossil fuels is biomass. Biomass is the term assigned to any organic matter derived from living organisms, including animals, humans and plants [1].The material can be residues like wood from forests, crops, seaweed, leftovers of agricultural and forestry processes, and organic industrial, human and animal wastes [1].

Biomass could be categorized under six groups; wood and woody biomass, herbaceous and agricultural biomass, aquatic biomass, animal and human biomass wastes, contaminated biomass, and industrial biomass wastes (semi-biomass), and biomass mixtures which is blend of the first five groups. Wood and woody biomass is further sub-categorized as: coniferous or deciduous, and soft or hard. Woody biomass includes stems, branches, foliage, bark, chips, lumps, pellets, briquettes, sawdust, sawmill, etc. Herbaceous and agricultural biomass could be annual or perennial and field-based or processed-based. It includes grasses and flowers, straws, etc. Aquatic biomass is further categorized as marine or freshwater algae, and macroalgae or microalgae. Seaweed, kelp, lake weed, water hyacinth, etc. are considered aquatic biomass. Animal and human biomass waste are Bones, meat-bone meal, chicken litter, various manures, etc. Semi biomass could be municipal solid waste, demolition wood, refuse-derived fuel, sewage sludge, hospital waste, paper-pulp sludge, waste papers, paperboard waste, chipboard, fibreboard, plywood, wood pallets and boxes, railway sleepers, tannery waste, etc.[2].

Use of biomass resources for energy production has several advantages and disadvantages in comparison to other energy sources. Table 1.1 summarizes such advantages and disadvantages as reported in the literature [2][3]. In this sense, it is of particular importance for decision-makers in bioenergy industry to employ best practices in maximizing the amount of energy output (from biomass) delivered to end-users while minimizing environmental footprints. In this sense, implementing an efficient biomass supply chain strategy directing biomass collection from suppliers as well as distribution to end-users is a paramount need.

This study focuses on wood pellets as a source of energy. The interest in wood pellets as a source of bioenergy, either bioelectricity or bioheat, has been globally increasing[4]. It has lower environmental impacts in comparison with fossil fuels[5]. Wood pellets are a product of wood collected from land. Therefore, it is important to analyze the wood pellets supply chain alongside the geography in which the supply chain is operational[6].

Multiple factors affect biomass availability. Defining the factors that have impact on biomass availability is highly dependant on the type of biomass and the environment it is produced in. Zhang, et al, discuss the factors that affect the phytoplankton biomass availability in a large eutrophic lake. They mention Chlorophyll a, water pH, water temperature, water alkalinity, chloride, etc. are the factors the affect the biomass availability[7]. Hoi, et al, discuss that light factors and nitrogen availability alters biomass and C-phycocyanin productivity of Thermosynechococcus[8]. Roll, et al, discuss how water availability controls amount of Melia dubia as biomass in India. They investigate the affects of water availability on biomass increment of Melia dubia and conclude that its growth rate is highly correlated to water availability[9]. While weather fluctuations have effects on the supply levels of biomass, they affect the building demand as well[10]. Tamer, et al, discuss that climate change, including precipitation fluctuations, adaptation and mitigation actions will result in long-term building performance enhancement. The three key-indicators to do so are building energy use, greenhouse gas emissions, and operational cost[10]. In case of precipitation fluctuations, biomass supply chains for buildings will experience changes on both sides of supply and demand. Once supply and demand changes, it is possible that the amounts of biomass delivered to buildings is not sufficient. Therefore, building will either face energy shortage or have to use other sources of energy to meet their demand. If the biomass supply chain is not ready to overcome such circumstances, i.e. disruptions, it is not resilient.

In this study, a resilience assessment model for biomass supply chains is proposed in order to account for supply and demand uncertainties, their possible mismatches, and to establish the impact of such mismatches on continuity of energy production. In doing so, after undergoing a literature review, a biomass supply chain management model using predictive metrics for biomass availability and demand is proposed. These metrics are established and analyzed comparatively under different demand and availability prediction scenarios established using alternative machine learning algorithms. A routing model is then developed using Google Maps API to identify the best distribution routes for delivering biomass from depots to end-users. A resilience index is proposed providing the decision-makers across biomass supply chains with biomass supply, distribution, and demand solutions in coping with climate disruptions affecting the availability of biomass as well as demand for biomass energy.

The rest of this article is organized as follows. First, a literature review on biomass supply chain management is provided with a focus on highlighting the research in resilience and its attributes. Then, the proposed methodology is described in details. A case study is explored to present the applicability of the methodology and to show its practical implications. The results will then be further elaborated and analyzed. The paper concludes by highlighting a summary of the methodology and contributions as well as statements of limitations and avenues for future research.

Table 1. 1: Advantages and Disadvantages of Bioenergy [2][3]

Advantages	Disadvantages
Renewable energy source	Possible soil damage and loss of biodiversity
CO2 neutral conversion	Regional availability
Mitigation of hazardous emissions (CH4,CO2,NOX,SOX,trace elements)	Seasonal availability
Capture of some hazardous components by ash during combustion	Unclear utilization of waste products
Diversification of fuel supply and energy security	Being Perishable
Rural revitalization with creation of new jobs	Possible hazardous emissions during heat treatment
Potential use of oceans and low-quality soils, and restoration of degraded lands Reduction	Potential technological problems during heat treatment

1.2 Problem Statement

Biomass supply chains face various uncertainties in either supply or demand amounts. In addition, weather fluctuations can have a noticeable impact on either side of the supply chain. Lack of real-time or historical biomass availability data is an obstacle that prevents predictive algorithms perform well in terms of predicting the quantities of biomass available on the ground and predicting demand as the each end of biomass supply chains [11]. In addition, a share of operational costs of biomass supply chains originates from the logistics, including warehousing and transportation. Having effective strategies in (1) collecting and processing the sufficient amounts of biomass from the ground within the closest radius of the facilities, in accordance with the end users demand, and (2) finding the best scenarios to distribute biomass from facilities to end users will make biomass supply chains more efficient and reliable [12][13].

Weather fluctuations not only affect the amount of biomass available but also impact energy demand.[14] The mismatch between supply and demand sides reduces resilience of biomass supply chain and causes economic loss or energy insecurity. Having performant predictive models for demand and supply levels enables the decision makers in biomass supply chains to foresee the mismatches and minimize the impact by satisfying the end users' demand through alternative pathways of biomass supply or by other energy sources.

1.3 Research Objectives

In this thesis, the interdependency of biomass supply and demand is investigated, along with the strategies to enhance supply chain logistics identified based on a literature review. Therefore, this study aims at designing a system to consider end users' characteristics and come up with a plan to efficiently collect, process, and deliver biomass to them maximizing biomass supply chain resiliency formulated through minimizing supply and demand mismatches. In this regard, this thesis aims at proposing a methodology consisting of the following steps:

- Predicting buildings' demand (as end users) based on their attributes and levels of biomass availability and supply.
- Clustering buildings and identifying each cluster's centroid as a place for its depot.
- Utilizing routing algorithms to minimize distances traveled by a fleet of vehicles to collect biomass from ground to facilities based on buildings demand and proposing

the best routes to distribute biomass from facilities to end users with different targets.

- Evaluate the amount of biomass delivered to end users under different weather fluctuation scenarios to compare alternative scenarios for improving the resilience of the biomass supply chain.

1.4. Thesis Outline

This thesis is comprised of five chapters. Chapter one, the current chapter is the introduction of the thesis. Chapter two summarizes the state of the literature regarding the supply chain management, its challenges, and the solutions to address them. The literature review contains an overview to prediction algorithms applied to establish predictions of biomass supply and demand levels as well as an overview of the vehicle routing problem, and supply chain resilience concepts. In chapter three, a integrated approach is proposed to coordinate supply and demand level predictions of a biomass supply chain network, simultaneously. Then, collection and distribution algorithms are proposed to minimize targets while meeting constraints. The focus is on a supply chain that provides biomass to end users that are a stock of buildings. In addition, weather fluctuations are formulated to examine the response of the biomass supply chain to these fluctuations and the impact supply chain stock levels and resilience. In chapter four, a case study of Broader Public Sector of Ontario is presented. In chapter five, the proposed method is implemented in the case study and the results are obtained and analyzed. The last chapter presents conclusions with respect to the methodology and the case study as well as future research directions.

Chapter 2: Literature Review

2.1. Biomass Supply Chain

Biomass supply chains are a key component in the management of bioenergy production processes [15]. Council of Supply Chain Management Professionals (CSCMP) defines supply chain management as "the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. Importantly, it also includes coordination and collaboration with channel partners, which can be suppliers, intermediaries, third-party service providers, and customers. In essence, supply chain management integrates supply and demand management within and across companies." [16] Considering this definition, a bio-energy supply chain aims at an integrated management of bioenergy production from harvesting of biomaterials to energy conversion facilities [15] [17]. In this sense, a bioenergy supply chain comprises of five main components, which are harvesting and collection, pre-treatment, storage, transport, and energy conversion [3] [17] as presented in Figure 2.1.

Figure 2. 1: A Typical Bio-energy Supply Chain

In some cases, these components could be merged. For instance, if pre-treatment of raw material is performed at the same place where biomass is collected, and is distributed from this place to end-users, then the process will be further simplified as presented in Figure 2.2.

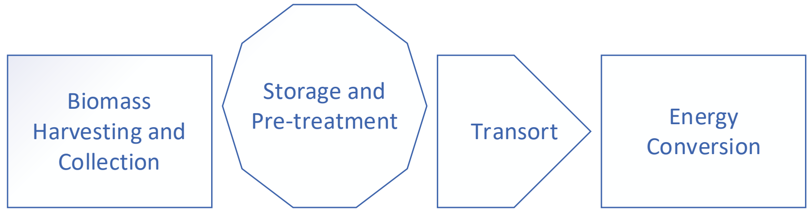


Figure 2. 2: Bio-energy Supply Chain with one Storage and Pre-treatment Unit

Biomass supply chains differ from traditional supply chains in a number of ways. Bioenergy supply chains are designed based on seasonal availability of agricultural biomass, variability of biomass materials, and varied biomass uses, which could require flexibility in choices of biomass transportation and storing strategies [3]. Many researchers have investigated the characteristics of bioenergy supply chains to identify solutions to address the above-mentioned challenges, in reflection of the fact that the interest in using biomass as a source of energy has steadily increased. Mathematical modeling, as a versatile tool, is used extensively to investigate supply chain performance considering environmental, social and economic goals. The mathematical programming tools (for biomass supply chain optimization) such as "Mixed-Integer Linear Programming" and "Mixed-Integer Non-Linear Programming" have been explored [18]. Bioenergy supply chains may undergo several uncertainties. These uncertainties due to variability stemming from biomass production and supply processes, means/routes of transportation, or changes in demand for biomass. Various approaches have been explored to model uncertainties including stochastic programming, robust optimization, and fuzzy mathematical programming [19].

In addition, optimally locating the facilities is of great importance. The main reason is that the trucks are driving around to collect and distribute biomass while emitting greenhouse gasses.

If locating facilities follows optimum criteria such as minimizing emissions, then a more environmental-friendly supply chain is established. In addition, cost targets or objectives could be considered as well to ensure economic feasibility of biomass supply chains. A possible solution to achieve these targets is to use geographic information system (GIS) in concert with decision support systems when maximizing the productivity of biomass plant, minimizing environmental impacts and minimizing costs [20]. Some studies have particularly investigated the amount of greenhouse gasses emitted into environment across supply chains [21]. Studies on green supply chain design, aiming at low-waste production, environmentally friendly action, and social awareness, have also shown that decentralized supply chain networks are more promising in terms of reducing environmental impacts [22][12].

2.2. Supply Chain Disruptions and Resilience

Supply chains are vulnerable to disruptions due to natural disasters, transportation problems and delays, and many other operational issues [23] [24] [25]. In many previous studies, it was assumed that biomass energy facilities, including pre-treatment, collection, and storage units, would operate constantly. However, these facilities are subject to disruptions as well. The incidents that can affect these facilities are water scarcity, flooding, routine maintenance, or adverse weather condition [24]. In addition, there are disruptions to supply chains that are classified as demand, supply, process, control, and environmental disruptions. The demand risk is the difference between actual and predicted demand. Supply risk is defined as potential fluctuations in availability of supply stocks. Process risk is considered as any kind of deviation from quality and quantity at which a process must be fulfilled at any point of time. This risk encompasses disruptions in internally owned assets or reliability of supporting communication and transmission systems. Control risks include assumptions, rules, systems, and procedures that administer an organization's processes. Environmental risks are the ones external to the firm. In this study, the risk is fluctuations in the supply levels of biomass due to the fluctuations in precipitation and is measured by the amount of fluctuations of supply levels. If the levels of biomass available decrease, the buildings, i.e., end users will receive less biomass; as a result of this, they might undergo periods when their share of energy required is not met. To address the problem when buildings do not receive the amount of biomass they need, this study proposes a method to predict levels of biomass availability in case of a weather fluctuation and building demand simultaneously. With having these amounts in hand, any mismatch between supply and demand sides could be analyzed, so the buildings will have plans to substitute the share of biomass with other sources of energy.

The literature points to four supply chain resilience enhancers on mitigating the disruptions that are flexibility, agility, collaboration and redundancy [26]. They could affect any of the supply chain components [26]. 'Flexibility refers to the ability of a firm to respond to long-term or fundamental changes in the supply chain and market environment by adjusting the configuration of the supply chain' [27]. Agility is defined as the capability of changing operational conditions as a response to environmental or market uncertainty [27]. Collaboration encompassed the ability to work efficiently with other components of a supply chain for mutual benefit in terms of forecasting, postponement, and risk sharing[26]. 'Redundancy involves the strategic and selective use of spare capacity and inventory that can be invoked to cope with a crisis, such as demand surges or supply shortages' [26].

After a disruption takes place, additional operational costs will arise because the suppliers need to use alternative pathways or facilities to constantly satisfy end-users demand [28]. Therefore, it is critical that these facilities and their supporting supply chains remain functional. This maximum functionality is referred to as resilience. Supply chain resilience is defined as the capacity of a supply chain to surmount stress, disruption or system failure and mitigate the impact of disruptions as much as possible [29] [30] [31].Nevertheless, information flow between supply chain components is crucial to continuously seek most feasible action plans, as one user's supply resilience could result in another user's loss and disruption.

Establishing risk management strategies can increase the supply chain resilience. The main challenge to supply chain risk management is that it needs to encompass to all components of a supply chain. Thus, a risk management action plan should include not only the separate characteristics of supply chain components, but also their interdependency [32].

2.3. Supply and Demand Prediction in Supply Chains

Among the methods employed for supply chain management, predictive models using machine learning techniques have gained more interest during the past decade. This is due to the facts that there have been advances in computer hardware and computational capability of algorithms, and increase in the amount of data collected recently. Many researches have investigated the use of machine learning predictions in supply chain. Such algorithms could be applied to predict either supply or demand.

2.3.1. Demand Prediction

Karimi, et al, introduce a cost-function- based prediction markets as means of sharing demand forecasts. In their study, a decentralized two-stage supply chain, including one supplier, benefits from demand predictions. They also discuss that prediction market has been studied less in supply chain management in comparison with that of project management [33]. Raiyani, et al, apply time-series forecasting models on sales data to detect abnormality in the buying pattern of customers. They utilize five years of historical data of 100 items in 10 stores. They firstly train the models on the first four years of data; secondly, they validate the model with the fifth year of time-series data[34]. Zougagh, et al, defines the goal of supply chain demand prediction as a means of inventory cost decrease and service level increase. They train prediction models to forecast market demand[35]. Dorostian, et al, firstly identify the weaknesses of applying model-based predictive control (MPC) to supply chain management. These weaknesses contribute to uncertainties in the models and tuning of MPC. Their goal is to providing end-users with robust supply chains when facing high fluctuations in demand. They define the goal of their study as "meet customer demand in the shortest time, with minimum cost and best quality in the presence of all targets, uncertainties, and delays"[36]. Dai, et al, investigate a grey prediction model to optimize stock levels. Their reason for choosing such a model is lack of data. Since they did not have data good enough to predict demands in a garment supply chain, they used Monte Carlo simulation to simulate demand [37]. Ibrahim, et al, set the scope of their work as data analysis in the reservation process on ordering the drug stocks. They have surveyed 300 people to collect data. They try to define patterns for a drugstore such that it can place an order to its supplier to restock quickly[38]. Yang, et al, provide a predictive control strategy for inventory management considering uncertainties in demand and time delays[39].

Biomass supply chains provide stocks to factories, power plants, buildings, etc. In this study, the end users are buildings. Therefore, the investigators of the current study went through the literature to review the practices for building demand prediction. The literature review shows that not all regression algorithms are suitable for building energy consumption prediction, e.g. linear regression finds a straight line fitted to the training data. On the other hand, building energy consumption is non-linear and relies on many parameters. Meng, et al, proposes a methodology to use neural networks for load forecasting to address the imbalance between supply and demand in smart grids[40]. They also factor energy storage and release time prediction. They compare performance metrics of back propagation neural network, radial basis function neural network, general regression neural network, Elman neural network, and support vector regression. They investigation reveals that Elman neural networks yields the best performance metrics for next hour predicition. Chen, et al, investigate a hybrid support vector regression algorithm to forecast short-term electric demand. They improve the performance of their model by applying multi-resolution wavelet decomposition as a pre-processing for signal analysis[41]. Bassi, et al, commence they study by discussing the importance of building energy consumption in global warming. They discuss further how machine learning algorithm and building demand prediction can mitigate the impacts of energy consumption on global warming. They propose a comprehensive comparison between Catboost regression, Light gradient boosting regression, and eXtreme Gradient Boosting. They finally conclude that eXtreme Gradient Boosting had the best performance of building demand prediction[42]. In a similar study, Haque, et al, compare the results of support vector regression, random forest regression, and kNN regression for building demand prediction. They conclude that multi-variant nature among the independent variables corresponding to the dependent one decreases the performance of the algorithms[43]. These studies have mostly showed how well the predictive algorithms could forecast the amounts of building demands in accordance with their energy system and building configurations, i.e., storage, etc., and what challenges and limitations each algorithm have had. However, the relation between the predictive algorithms and the way energy supply chain reacts to building demands predictions needs more investigation.

2.3.2. Supply Prediction

Various material could be categorized as biomass stock. Predicting each kind of biomass has its own complexities. Pan, et al, propose an image processing approach using Light Detection and Ranging (LiDAR) to predict the levels of above-ground biomass. Their motivation to do so is that measuring process of the biomass levels of a cereal plot needs cutting, drying and weighing. This process is costly and laborious. They propose a biomass prediction model that takes into account plant structure, whereas previous models would only consider density and height of the plants. Therefore, their state-of-the-art model not only is less costly, but also it produces better prediction performance over the previously introduced models[44]. Huy, et al, propose a deep learning algorithm to predict the levels of above-ground tree biomass and compare its result with regression algorithms. They conduct destructive sampling on 968 individual trees distributed across five ecoregions of Vietnam. They collect a dataset of tree predictors of diameter at breast height, tree height, wood density and the response variable of above ground biomass along with forest stand factors of basal area and density; ecological and environmental variables such as ecoregion, slope, altitude, soil type, averaged annual temperature , average annual rainfall and average dry season length. They train sixteen deep learning models, each of which is fed with one to nine predictors and conclude that the model with the nine inputs outperforms the others[45]. Another application of machine learning, including regression and deep learning, in biomass is predicting the amount of biomass yield from waste treatment. Hu, et al, propose a model which is fed with the carbon, hydrogen, nitrogen, and oxygen contents of the waste on a dry-ash-free basis, and the proximate analysis including volatile matter, fixed carbon and ash contents on a dry basis, as well as operational temperature and time. They finally conclude that their model is capable of predicting the amounts of biomass yield with 97% of coefficient of determination[46]. Katongtung, et al, propose a machine learning algorithm to predict the amount of the biocrude yields and higher heating values from hydrothermal liquefication of wet biomass and waste. They feed the algorithm with 17 inputs; including feedstock characteristics (biological and elemental properties) and operating conditions. They apply four algorithms to their dataset: xtreme gradient boosting, kernel ridge regression, random forest regressor, and support vector regression. The results of their investigation reveals that although xtreme gradient boosting is more performant in terms of prediction accuracy result, its execution time is more than kernel ridge regression and support vector regression[47]. Masjedi, et al, propose a model integrating time-series data and recurrent neural networks to predict the levels of sorghum biomass. In this study, they first utilize unsupervised feature learning through a fully connected auto-encoder system. The inputs of feature learning process are hyperspectral and LiDAR remotely sensed data. The input features if the encoder are turned into learned features through the encoder network; then the learned features are passed to the decoder to reconstruct the output features. In other words, the outputs of the encoder, namely learned features, are the inputs of the decoder. Then, they pass the newly constructed features to two predictive algorithms: support vector regression and recurrent neural network. Their compare the results of these algorithms for each sneario of the feature selection algorithm. They conclude that there are challenges for each scenario regarding small sample sizes, including weather and sensitivity to the associated ground reference information[48]. Zhang, et al, propose a hybrid parallel neural network, integrating a feed forward neural network and a recurrent neural network, to predict the levels of biomass concentration in fermentation. They synthesize the data with a fed batch model of a streptomyces actuosus fermentation. The first four hundred samples are used to train the hybrid network and the send one hundred are used to test the previously trained network. They conclude that the hybrid model yields a smaller amount of mean square deviation in comparison with radial basis function neural networks and Elman neural networks; they imply that their hybrid model outperforms the other two models.

The literature review reveals that the ways the predictions on levels of biomass availability affect the consumers is studied less often. In addition, neural networks, especially recurrent neural networks, are a decent nominee in terms of predictive modeling of biomass stock availability. Recurrent neural networks can capture the underlying patterns of sequential data and predict their behaviour over the next sequences.

2.4. Vehicle Routing Problem

Vehicle Routing Problem (VRP) algorithms are combinatorial optimization algorithms where a set of vehicles have to start from a depot and traverse between given destinations. The aim of these algorithms is to minimize travelled distances. The first variant of these problems was introduced in 1959 by Dantzig and Ramser. [49] They investigated the gasoline delivery trucks' route optimization between terminals and service stations. Other variants with different objectives were then studied. The variants include, but are not limited to, Vehicle Routing Problem with Profits (VRPP), Vehicle Routing Problem with Pickup and Delivery (VRPPD), Vehicle Routing Problem with Time Windows (VRPTW), Capacitated Vehicle Routing Problem (CVRP), Vehicle Routing Problem with Multiple Synchronisation Constraints (VRPMS), etc.

Vehicle routing problems are computationally expensive. Therefore, various exact and heuristic algorithms have been proposed to solve the VRP. Once the number of nodes, either delivery points or depots, increase, the level hardness to solve the problem drastically increases[50]. VRP is applied to various industries; one of which is supply chain management. A part of the supply chain complexity arises from managing the logistics costs of supply chain fleet. The costs will decrease if optimal routes are taken by the fleet drivers. In addition, other parameters, such as freshness of perishable goods upon delivery, is of great importance. Finding the optimal routes to deliver perishables results in higher quality goods at the point of delivery[51]. Utama, et al, have reviewed the literature of VRP for perishable goods. They categorize the literature as (1) single objective problems and (2) hybridizations and simulation; each of which solved through heuristics, metaheuristics, exact methods, or hybridizations and simulation. They conclude that genetic algorithm is widely used to solve both single and multi objective problems. In multi objective problems, minimizing costs while maximizing goods freshness was investigated[51]. Another area of investigation for VRP in supply chain management is bioenergy. A portion of the biomass cost, as a product which is delivered to the end user, arises from its transportation costs. Efficient supply chains should have strategies to minimize transportation costs. Therefore, studying the strategies leading to lower transportation costs, alongside minimizing distances travelled by the fleet, emissions, etc., is of great importance. Generally, a fleet of vehicles are dispatched from a storage or a facility where biomass is stored, to deliver biomass to end users. Soares, et al, investigate a problem in which a set of interconnected trucks for delivery and pick up need to be synchronised to minimize travelled distances and non-productive times [52]. The VRPMS is defined tries to find the routes with minimum costs for a fleet of vehicles, which have to synchronise in some nodes to accomplish common tasks [53]. The principal parameter distinguishing this variant from conventional VRPs is the dependency of one vehicle's route on another one's. They apply a metaheuristic approach based on the fix-and-optimise principles methodology to a wood chips supply chain in Southern Finland[52]. Their methodology has shown lower logistics costs for deliveries and pick-ups. Cao, et al, propose a methodology to optimally locate facilities and route vehicles. They discuss that the two objectives have been studied separately, yet they try to integrate the two objectives. They name their proposed methodology the location- routing problem for biomass supply chains (LRP-BSCs) and solve the problem integrating a mixed-integer programming model with a hierarchical heuristic algorithm based on Tabu Search[54]. The problem that they face is that as the number of nodes in their model increases, finding an exact optimal solution within an acceptable time frame gets harder. Our literature review shows that the number of studies considering biomass supply chain for implementing VRP algorithms is scarce.

Giallanza, et al, study a three-echelon regional agri-food supply chain. Their vehicle fleet and distribution centers have defined capacities. They propose a fuzzy time-dependant algorithm to generate the customers' demands; they discuss that the fuzzy nature of customers' demands had not been investigated before this study. The objective of their work is to minimize total costs and emissions by utilizing a non-dominated sorting genetic algorithm alongside multiple-criteria decision-making ELECTRE III method to find the best solutions[11]. Al Theeb, et al, propose a method to minimize total cost using mixed integer programming combining inventory allocation and finding best routes for the fleet. Their case study is a cold supply chain Jordan responsible for transporting chicken, meat, vegetables, and fruits. The products are mostly locally produced. They claim that no clear cold supply chain study had been done in Jordan to estimate the costs of such supply chains although energy prices are relatively high in Jordan. One important aspect of their work is that they could come up with a solution which saves 9.25% of total distribution cost compared with the cost paid by the organization of the case study[13]. Ransikarbum, et al, investigate safety of food supply chain network using making decisions on routing of the distribution fleet. They aim at investigating food-safety system under the green supply chain's scope. They firstly consider travelling distances, vehicles capacity, supply amount, and locations of stakeholders. Then, they propose an optimization formulation to come up with the sequence of deliveries for the vehicles[12].

Reviewing the literature in 2.1, 2.2 and 2.3, given the examples discussed above, reveals that many researches have investigated supply chains predictive control based on either demand or supply amount predictions; therefore, the number of articles considering both supply and demand predictions is scarce. Having reviewed the literature in 2.4, the authors conclude that the number of studies considering transportation scenarios for both supply side and demand side, e.g. supply collection into facilities and distribution to end users, is scarce[12][11][13]. Having a comprehensive model capable of predicting supply and demand levels, and biomass collection and distribution enables decision makers to monitor a biomass supply chain during its life cycle and come up with strategies to manage supply chains efficiently.

In recognition of such a gap in the literature, this study first proposes a three-step model which intelligently learns to (1) predict energy demand, (2) predict supply stock levels, and (3) formulate the biomass collection and distribution provided the above predictions. In addition, a feed-back loop between the supply and demand is established to ensure that the level of biomass availability and building demands prediction are synchronised. This feedback loop ensures that the amount of biomass collected from ground does not surpass the buildings demands, and at the same time, it ensures the buildings will receive the biomass they need if their demands change over time. Then, resilience metrics are proposed to establish supply chain performance under different circumstances in demand, supply, and distribution components. Using these metrics, decision-makers could have a better predictive understanding of biomass supply chains in order to adapt purchasing and distribution decisions accordingly.

Chapter 3: Methodology

It is vital that an efficient supply chain be able to have strategies to meet the users' demand by having the necessary levels of supply. It also noteworthy that logistics costs play a crucial role in operational costs of the supply chain. Therefore, the aim of the methodology is to find the optimal routes for (1) biomass collection from ground to facilities/depots, and (2) biomass distribution from depots to end-users. The location of each depot is found through clustering; each depot provides service to the buildings in its cluster. Each depot has biomass collected from ground by trucks. The trucks will not collect raw material any more than the buildings in their clusters need, because raw material conversion and storage is costly and the processed material will lose their energy intensity as time goes by. The trucks follow a collection strategy to minimize the distances they are traversing. Then, biomass is distributed to end users under three different scenarios, distance target, duration target, and capacity target. The following schematic figure represents the supply chain network.

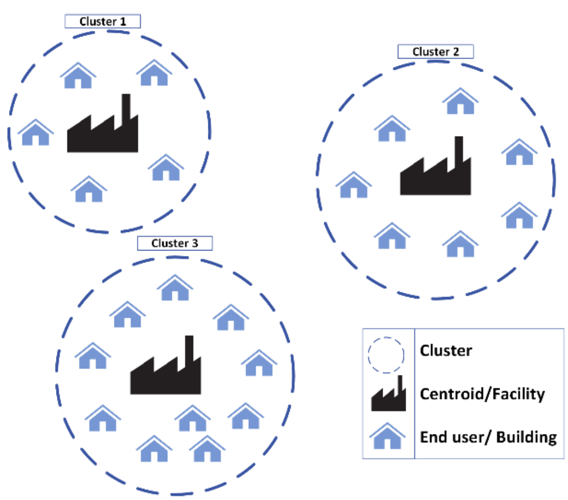


Figure 3. 1: Schematic Representation of the Supply Chain Network

As mentioned in the above, methodology developed in this study consists of three steps. Since the bioenergy supply chain in this study is supposed to be managed by a model capable of forecasting supply and demand levels, steps one and two could benefit from machine-leaning prediction algorithms. Step three benefits from a VRP formulation and solution approach for collection of biomass from land, transport it to depots, and distribute it from depots to end-users. Figure 3.2 shows a general overview of the prediction methodology.

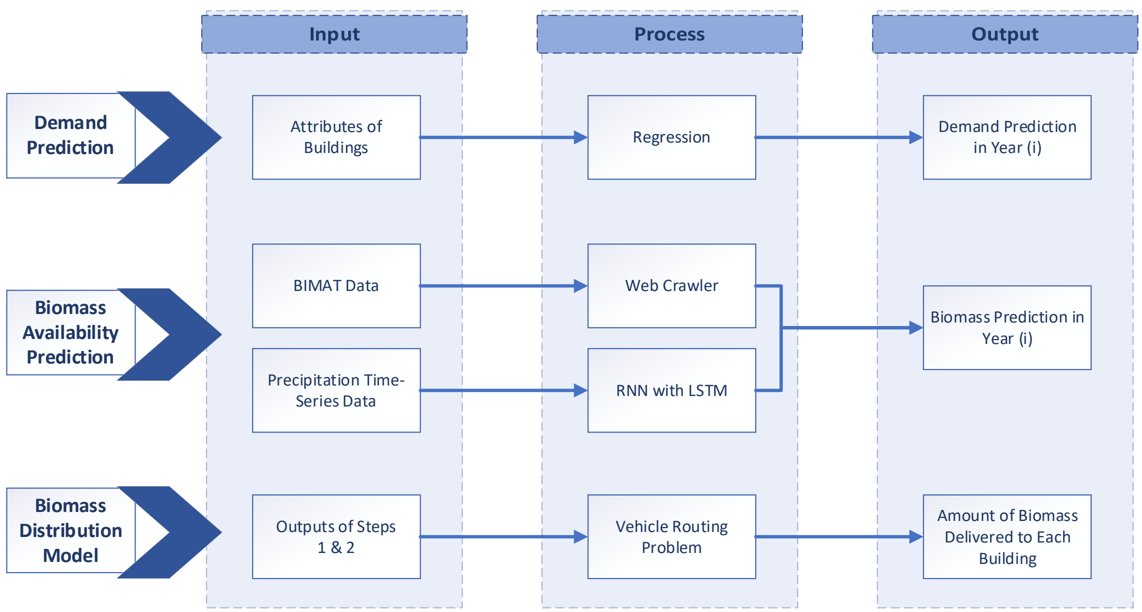


Figure 3. 2: General Overview of the Methodology

3.1. Demand Prediction

Building energy consumption prediction is significant in terms of building energy management[55]. The factors that influence the prediction goodness are ambient weather conditions, building structure and characteristics, the operation of sub-level components like lighting and HVAC systems, occupancy and their behavior[56]; each of the parameters mentioned could have momentous effects on the accuracy of the predictive model. In cases, where multiple parameters start to fluctuate simultaneously, the results of the prediction algorithms could be highly altered. Therefore, it is important to train robust algorithms to minimize the impact of the parameters fluctuating. Buildings, which use biomass as a source of energy, need to benefit from the cutting-edge energy consumption prediction algorithms to consider the fact that biomass availability on its own might undergo fluctuations; therefore, it is harder to come up with a predictive management model which balances out all the fluctuations in both building consumption and biomass demand. Step one aims at predicting the demand at the scale of buildings. Since historical data is a requisite for prediction algorithms, the models are trained on a set of buildings. This allows the algorithms to be fed with higher amounts of data and yield more accurate predictions. Once the prediction results for a single building is needed, the attributes of the given building are fed into the trained algorithm to predict its energy consumption. When the dataset is fed into the prediction algorithm, the results might be satisfactory. If they are not, different strategies and algorithms are used to improve the results. Firstly, the dataset of buildings consumption is cleansed. Cleansing is referred to detection and correction of corrupt or inaccurate observations. These observations could be deleted, modified, or removed. Secondly, demand prediction algorithms are tuned and trained to yield the best results possible for each algorithm. Simultaneously, feature selection algorithms are used to decrease the feature space size as much as possible; models with fewer input features are more agile in terms yielding results from time and memory perspective. The process of feature space reduction might either improve or impair the prediction performance. Therefore, the goal is to find the set of input features which yield the highest prediction accuracy. In case of building energy consumption prediction, the features could be building accuracy, occupancy and behaviour pattern, weather data, etc. based on the case study.

The predictive algorithms are applied on either the whole dataset or homogenous subsets of the dataset. The aim of the training algorithms on the homogenous subsets is that if better fits are yielded from these subsets, the algorithms will be faster as these subsets have fewer data points. Lastly, the algorithm with the highest and most reliable building consumption prediction performance is chosen as the main predictor.

In order to measure the goodness of fit for the predictive algorithms, two metrics are used. The first one is r-squared, or the coefficient of determination, or R^2. It is defined as the following formula [57]:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{1}$$

Where:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

The second metric is root mean square error (RMSE). It is the root value of equation 2 [57]:

$$\text{MSE}\,(\,y, \hat{y}) = \frac{1}{n_{\text{samples}}}\sum_{i=0}^{n_{\text{samples}}-1}(y_i - \hat{y}_i)^2 \tag{2}$$

Please note that higher values of R^2 and lower values of RMSE correspond to better fits for predictions.

The investigators started the demand prediction by examining some algorithms and keeping track of their results. If they could not reach satisfactory levels of prediction results, they would explore more algorithms. In order to improve the results of each predictive algorithm, the following three steps are taken[58]:

1. Hyper Parameter Optimization: Algorithms are integrated to find the best sets of hyperparameters to yield the most accurate results. The algorithms are: Random Search, Bayesian Optimization (Gaussian Processes, and Tree-structured Parzen Estimators), Multi-fidelity optimization, and Genetic Algorithms.
2. Feature Engineering: Algorithms are integrated to impute the missing data, categorical variables encoding, variable transformation, discretization, outlier removal, and feature scaling.
3. Feature Scaling: Algorithms are integrated to reduce the feature space and improve prediction performance, as well as training time reduction. The categories of algorithms are: filter, wrapper, embedded and hybrid.

These algorithms are integrated to prediction algorithms in a "try and error" manner; meaning that each combinatorial set of predictive algorithms, hyper parameter optimization, feature engineering, and feature selection is trained and if any given algorithms results in a better performance greater than a threshold, that set is moved to the next level of combination.

The authors start the training process with the algorithms frequently used for building demand prediction and combine them with the algorithms from [58] to yield the most performant results. If the algorithms did not yield satisfactory results, other algorithms would be studied. Based on the literature review in 2.3.1, the regression algorithms to start examination in this study are support vector regression, kNN regression, decision tree regression, random forest regression, Gradient Boosting regression and its variants [59], Catboost regression[60], and neural network regression[61]. It will be discussed in results that CatBoost algorithm reaches satisfactory levels of results in terms demand prediction; therefore, no further investigation is needed.

3.2. Biomass Availability Prediction

In this section, a model is developed to predict the amount of available biomass in a given region. It uses the "Biomass Inventory Mapping and Analysis Tool" (BIMAT) [62] to extract the values of available biomass. BIMAT is developed and maintained by "Agriculture and Agri-Food Canada", offered by government of Canada. This web-based Geographic Information System (GIS) map is capable of reporting availability of biomass at any geographical coordinate within a radial distance. It stores the data in a zonal statistical system. Zonal statistical functions utilized in a GIS are employed to summarize the biomass availability into BIMAT's 10 km x 10 km reporting framework [63].

Since demand prediction (3.1) identifies how much biomass is needed to satisfy buildings consumption, the system needs to predict how much biomass is available at each point to meet that demand. Having the availability of biomass enables decision-makers to come up with biomass collection strategies based on the buildings consumption. Having predictions of building demand and supply levels enables to system to dynamically match the balance between supply and demand sides. This dynamic interdependency is controlled via two feedback loops between supply and demand. The reason behind this feedback loop is that if a mismatch between supply and demand occure, two scenarios are foreseen:

1. Supply level is lower than buildings demand. In this case, the supply chain safety is put at risk and buildings will need to replace the share of biomass with (an)other source(s) of energy.

2.Buildings demand is lower than supply level. In this case, the supply chain will not be at its most efficient state since a part of the stock are left in the facilities/depot.

Therefore, these feedback loops control the interdependency between supply and demand levels, and ensure that the facilities have the adequate amounts of supply stored to meet buildings demand.

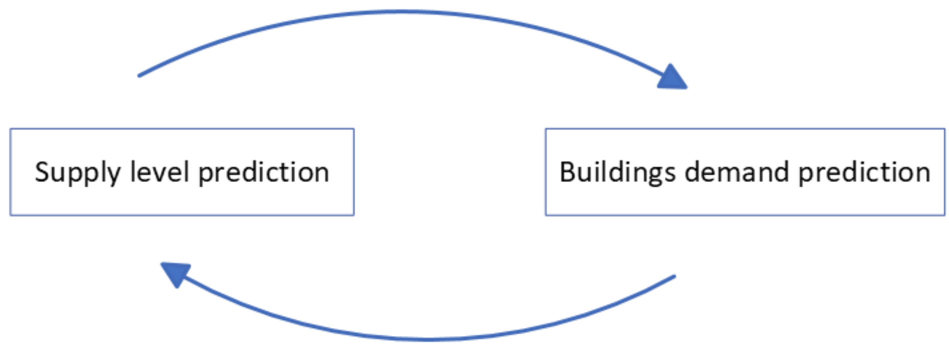


Figure 3. 3: Demand and Supply Feedback Loop

As manually collecting data of supply levels from BIMAT, based on the buildings demand, is tedious and fallible, a web crawler (i.e., scraper) is developed to collect data from BIMAT. This crawler collects the amounts of biomass, based on pre-set parameters, i.e., defined by thecase study, and stores them for further processing. The arguments to be passed in the crawler are coordinates of each depot and the amount of biomass to be collected based on the buildings demand prediction. Since the radial resolution of the BIMAT is 10 km, the crawler uses this number as its maximum resolution. Please note that every 10 km of distance, either horizontal or vertical distance, is roughly equal to 0.09009 degrees on the geographical coordinate system based on the coordinates of the location, as the earth is not a perfect sphere.

Biomass is transported to the pre-processing units, where it can be processed and stored. The location of the pre-processing units, which can be called “depots”, is correlated with the location of the end-users. The reason behind this assumption is that the main goal of biomass collection and distribution is to minimize distribution costs. Therefore, it is assumed that depots are located at the same location as the centroids of the building clusters. Building clustering is performed based on k-means Neighbour algorithms. [64] The k-means clustering algorithms are trained based on different numbers of k. Then, the Davies-Bouldin index is used to measure the performance of each k and find the optimal number of k. [65]

As presented in Figure 3.4, the crawler scans the eight adjacent squares to find the cell with the highest amount of biomass available and starts collection of biomass from that cell. The truck is sent to that cell from the depot to collect biomass. If the capacity of the truck is full, it will return to the depot to offload the biomass. If the capacity is not full, it will choose the one of the two adjacent cells, which has more biomass availability, to fill the rest of the capacity. As presented in Figure 3.4, if the truck starts from a white cell, then the two adjacent cells are black. Once the first eight cells are collected, the scraper will add the next sixteen adjacent cells in the next layer to be scraped (Figure 3.5). This process stops when the required amount of biomass is collected at depot.

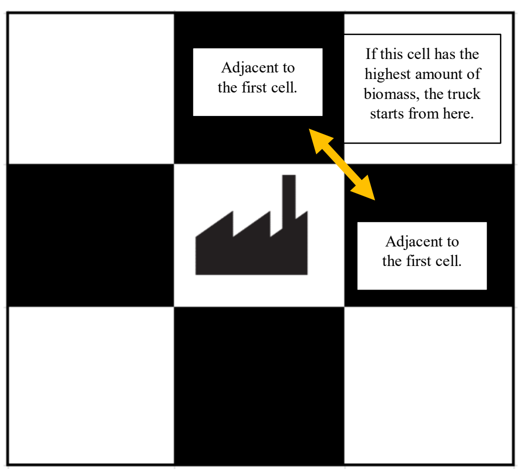


Figure 3. 4: Depot in the center of the 10 km * 10 km mesh

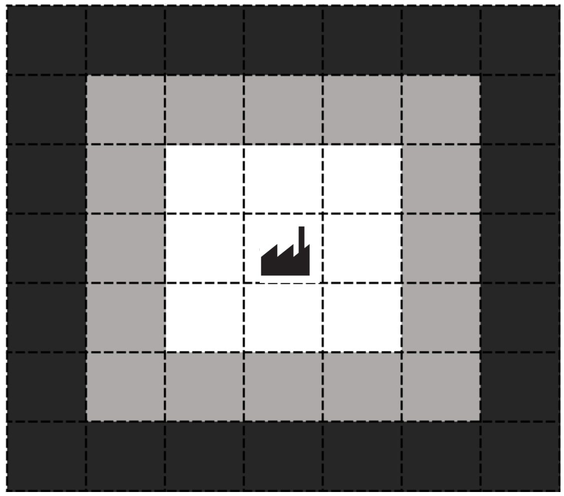


Figure 3. 5: Sequence of biomass collection (1st: White , 2nd: Grey and 3rd: Black)

It shall be noted that the amount of biomass collected in each depot is equal to the amount of biomass needed by buildings in the same cluster. The advantage of such a system is that based on the building demand prediction (step one), depots decide about how much biomass they need to deliver in the upcoming year; such that they can collect and process biomass in advance to avoid disruptions. The biomass collection process can be summarized as the flowchart presented in Figure 3.6.

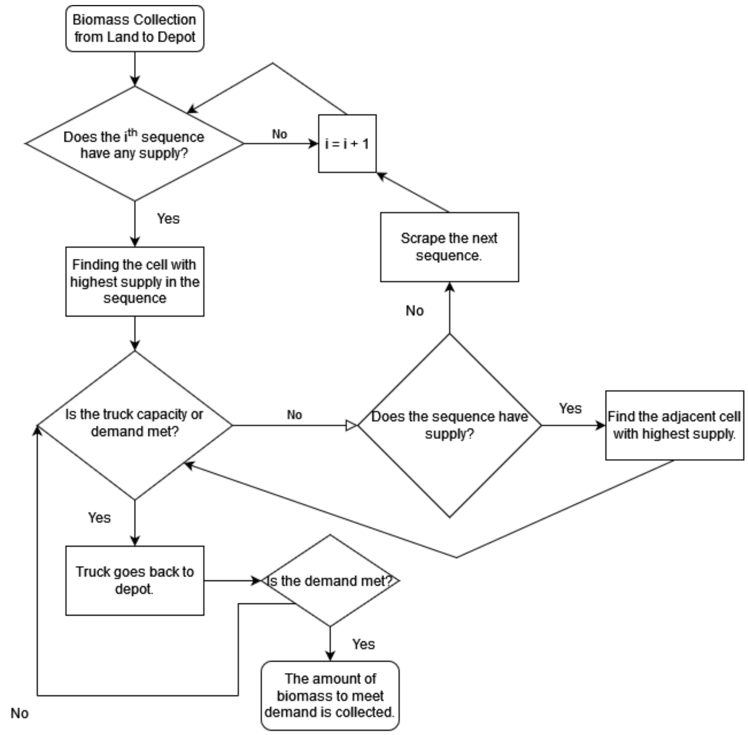


Figure 3. 6: Biomass collection process

In this study, wood pellets are considered as biomass. Therefore, data is extracted from BIMAT for hardwood and softwood of Roadside Harvest Residue, Mill Residue (Wood), Mill Residue (Bark) and urban wood waste (both residential and non-residential). These residues are turned into pellets in the processing units. Pellets are a preferred type of biomass due to having higher levels of standardization and energy density. In addition, transportation and storage of pellets are easier [66]. On the other hand, wood is a residue of forestry and urban processes. It is a part of the carbon cycle and its emissions are absorbed back into the cycle. In BIMAT, the average annual production of woody biomass is calculated based on forestry activities for years 2013-2014. [62] In this study, it is estimated that the amount of biomass is directly correlated to levels of precipitation. The reason behind this assumption is that step two of the methodology tries to predict supply stock levels. The algorithm in this step is “Recurrent Neural Network” (RNN), which utilizes “Long Short-Term Memory”(LSTM) [61] to predict time-series data. The reason behind using LSTM is that predictions of RNNs using LSTM has a memory and considers behaviour of previous data points in establishing future trends (predictions). Therefore, RNNs need to be fed with adequate amounts of data, i.e., number of observations, to be able to perform the prediction well. Otherwise, the LSTM will not capture the underlying trend of the data and will result in faulty predictions. After searching for a dataset that can provide adequate number of data points, the precipitation dataset was chosen. The reason is that this dataset is available from [67] and contains monthly values from 1939 to 2020. This dataset is the most consistent open-access dataset to the best knowledge of the authors. As BIMAT calculator works based on yearly data, the outputs of the RNN model, e.g., 12 values representing each month of the year, are summed to calculate the yearly precipitation. A performance metric to represent the goodness of fit is calculated as follows for a given year (i):

$$\text{Percent of Total Difference(i)} = \frac{\text{Total Precipitation(i)} - \text{Total Predicted Precipitation(i)}}{\text{Total Precipitation(i)}} * 100\%$$

This indicator reflects the fact that the main point of interest in of the above prediction is that we minimize the differences between actual and predicted values in a yearly scale. The “Percent of Total Difference” (PTD) metric lets the positive and negative residual values cancel out each other. If a fluctuation happens in year (i), its immediate impact will be seen in the next year. This is called “primary effect”. In other words, when fluctuation happens, the values of precipitation in year (i) change and the RNN weights get updated according to new values. Therefore, the prediction of year (i+1) is based on the new RNN weights. Moreover, it is possible to use the predicted values of year (i+1) to predict the values of year (i+2). This is called “secondary effect”, where fluctuation in year (i), affects year (i+1) and (i+2) accordingly.

All in all, precipitation levels are chosen as indicators of biomass availability. Now, the aim of predictive models in step two is to help suppliers understand how much biomass is available for sale in a given year.

3.3. Biomass Distribution Model

In step three, a model for distribution of biomass from depots to end-users is developed. Firstly, the buildings are clustered. Then centroid of each building is assigned as the depot. Each depot is responsible for satisfying the demands of buildings in its cluster. The algorithm used in this step is VRP. As a gap of the literature review, in this study, our focus is on vehicle routing problems with either distance or duration targets, and capacitated vehicle routing problems, where trucks have limited capacities. It is important to note that VRP algorithms work based on the distance matrix. A distance matrix shows distances between depots and destinations. To form the distance matrix, this study uses Google Distance Matrix API. [68] A limitation of this matrix is that it accepts up to 100 queries at any time (as a 10*10 matrix). Thus, in each iteration, provided that one depot is always assigned, the maximum number of buildings will be limited to 9. Therefore, to adapt to this limitation, this study chooses a sample of 9 buildings in a cluster, and formulated the distribution problem for a network comprised of one depot and 9 buildings. The algorithm can be applied to a bigger network provided that the computation problem is solved.

VRP algorithms try to minimize a target, i.e., distance, time, etc., given a set of locations to traverse. In this study, the targets are distance and time. Minimizing distances using VRP algorithms ensures that the trucks travel the minimum distances. The minimum distances does not ensure that the amount of time trucks spend on the road is minimized. As an example, driving on a street with a length of 100 kms and speed limit of 50 km/h takes two hours; while the same distance on a highway with a 100km/h speed limit takes one hour. On the other hand, some cases might require the drivers to drop off their deliveries as soon as possible. An example of such a case is hospitals in winters. Hospitals are among the most critical infrastructure assets and should be operational all the time. Therefore, it is vital the trucks arrive at the hospitals as soon as possible, even thought they travel longer distances. Therefore, three VRP models are studied in this article.

In the first two models, which are not a variant of capacitated VRP, all the trucks have the same capacity. Therefore, we assume that the suppliers deliver equal amounts of biomass to end-users. This equal amount of biomass is proportional to the maximum allowable payloads of the trucks and the number of buildings they will deliver to on their routes. The advantage of duration target model is in a case when buildings lose their connections to the grid or natural gas pipelines. In this case, a duration constrained model will help suppliers to reach out to buildings in as minimum time as possible using least number of trucks. Keep in mind that duration targets may or may not correspond to shortest routes possible (when Google API tries to minimize time). As an example, google might navigate the truck through highways where speed limit is higher than streets. However, these highways may be longer than other routes with lower speed limits.

In the third model, amount of biomass delivered to each end-user is proportional to its share of energy demand in the total energy demand (of the whole cluster of buildings). In this this sense, the amount of deliverables should not surpass the capacity of the trucks. After each of three models yields results, the optimal routes are determined. It shall be noted that empty trucks shall go back to depots.

Based on the sequence of the depot and buildings on each route and payloads of the trucks, emissions and costs can be calculated. In order to calculate emissions based on distance and payloads, "Guidelines for Measuring and Managing CO2 Emission from Freight Transport Operations" is used. [69]. Table 3.1 shows Carbon emission factors (gCO2/tonne-km) subject to varying payloads and levels of empty running trucks [69].

We assume an empty thirty-tonne truck emits 600 gCO2 per km from the study done by Seo, et al[70]. It is estimated that natural gas and wood pellets emit 223 gCO2e/kWh and 54 gCO2e/kWh respectively. Wood pellets on average produce 4,900 kWh of energy per tonne[71][72][73]. These numbers, alongside transportation emissions, could help us understand how much CO2 can be saved by burning pellets instead of natural gas. The price of natural gas for non-residential buildings is $0.028/kWh in Canada. This price for pellets is $0.067/kWh[74][75]. By having these numbers, the cost difference between burning pellets and natural gas can be calculated.

Table 3. 1: Amounts of Emissions for Truck Payloads per km

Payload Tonnes	% of trucks run empty										
	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
10	81.0	84.7	88.8	93.4	98.5	104.4	111.1	118.8	127.8	138.4	151.1
11	74.8	78.2	81.9	86.1	90.8	96.1	102.1	109.1	117.3	127.0	138.6
12	69.7	72.8	76.2	80.0	84.3	89.2	94.7	101.1	108.6	117.5	128.1
13	65.4	68.2	71.4	74.9	78.9	83.4	88.5	94.4	101.3	109.5	119.3
14	61.7	64.4	67.3	70.6	74.2	78.4	83.2	88.7	95.1	102.7	111.8
15	58.6	61.0	63.8	66.8	70.3	74.2	78.6	83.7	89.7	96.8	105.3
16	55.9	58.2	60.7	63.6	66.8	70.5	74.6	79.5	85.1	91.7	99.7

17	53.5	55.7	58.1	60.8	63.8	67.2	71.2	75.7	81.0	87.2	94.7
18	51.4	53.5	55.8	58.3	61.2	64.4	68.1	72.4	77.4	83.3	90.4
19	49.6	51.5	53.7	56.1	58.8	61.9	65.4	69.5	74.2	79.8	86.5
20	48.0	49.8	51.9	54.2	56.8	59.7	63.0	66.9	71.4	76.7	83.0
21	46.6	48.3	50.3	52.5	54.9	57.7	60.9	64.5	68.8	73.9	80.0
22	45.3	47.0	48.8	50.9	53.3	55.9	59.0	62.5	66.5	71.4	77.2
23	44.2	45.8	47.6	49.6	51.8	54.3	57.2	60.6	64.5	69.1	74.7
24	43.2	44.7	46.4	48.3	50.5	52.9	55.7	58.9	62.7	67.1	72.4
25	42.3	43.8	45.4	47.3	49.3	51.7	54.3	57.4	61.0	65.2	70.3
26	41.5	42.9	44.5	46.3	48.3	50.5	53.1	56.0	59.5	63.6	68.5
27	40.8	42.2	43.7	45.4	47.3	49.5	52.0	54.8	58.1	62.1	66.8
28	40.2	41.5	43.0	44.6	46.5	48.6	51.0	53.7	56.9	60.7	65.3
29	39.7	41.0	42.4	44.0	45.7	47.8	50.1	52.7	55.8	59.5	63.9

The outcome of these steps provides the blue sky or baseline scenario, where no disruption has affected the supply chain. Then, black sky scenarios, where disruptions take place, are simulated. The supply chain disruptions in this study are limited to be precipitation fluctuations (step two). When the disruption takes place, the PTD of the RNN model in each year changes; consequently, the levels of biomass stocks change as a consequence of precipitation changes. The resilience index will reflect the reductions in the amount of biomass delivered to nodes in comparison with the blue-sky scenario. Decision-makers and suppliers can take into account the results of these scenarios as benchmarks to cope with climate disruptions and provide their end-users with adaptive capacities.

Chapter 4: Case Study

For the case study the "Broader Public Sector of Ontario" (BPS) buildings are selected. BPS buildings are those buildings entitled to receive public funding from the government of Ontario, but they are not serving the government of Ontario itself [76]. BPS buildings include four main sectors as presented in Table 4.1 where the number of existing data points related to each sector is listed [34]:

Table 4. 1: Sectors of BPS and their data points

Sector Name	**Number of Data points**
Public Hospital	341
Post-Secondary Educational Institution	733
Municipal	10,052
School Board	4,925

These buildings host eight subsectors, 548 organizations, and 34 operation types. All buildings are scattered over 1,248 cities [34].

BPS buildings reported their energy usage, emissions and types of used energy resources in the open data catalogue of Ontario for years 2011-2018 [77]. However, some inconsistencies exist in the way data were reported. To tackle these inconsistencies and to produce regression models with maximum accuracy, first, the units are converted to a uniform format, and then, missing values and outliers are identified and cleansed.

The cleansed dataset includes 9,485 unique postal codes. The postal codes of buildings were reported using a six-character format of strings. This format is not suitable for clustering as it is not possible to calculate distances between locations based on these postal codes. Therefore, these strings shall be converted to their equivalent coordinates. To convert postal codes to coordinates, the “Geolocation Service” available on “Natural Resources Canada” website is used [78]. The resulting coordinates are shown in Figure 4.1.

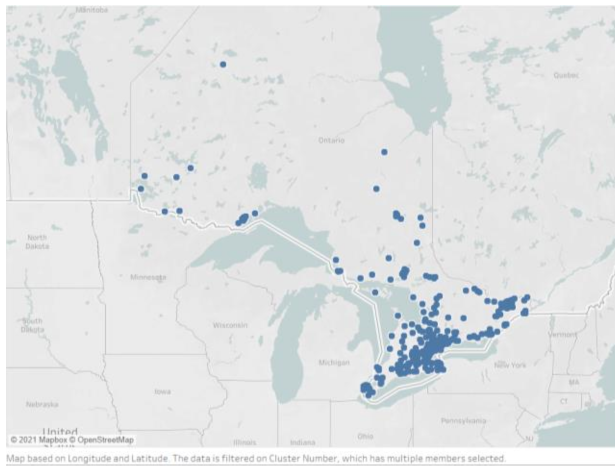


Figure 4. 1: BPS Map of Buildings

These buildings are then geographically clustered. The algorithm used in this step is k-means clustering. This algorithm tries to minimize inertia or within-cluster sum-of-squares[79]:

$$\sum_{i=0}^{n} \min_{\mu_j \in C}(|| x_i - \mu_j ||^2)$$

Where a set of (n) samples of (x) are divided into (k) clusters, each of which is described by a mean μj corresponding to the samples in the cluster [64].

Then, the Davies-Bouldin index is used to measure the performance of each *k* in the clustering algorithm [80]. “The index is defined as the average similarity between each cluster Ci for i=1,...,k and its most similar one Cj.

A simple choice to construct Rij so that it is nonnegative and symmetric is:

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

Where

si, the average distance between each point of cluster i and the centroid of that cluster – also know as cluster diameter.dij, the distance between cluster centroids i and j.

Then, the Davies-Bouldin index is defined as:

$$DB = \frac{1}{k}\sum_{i=1}^{k} \max_{i \neq j}^{R_{ij}}$$

Figure 4.2 shows the values of Davis-Bouldin index for k-means clustering with 1<k<16:

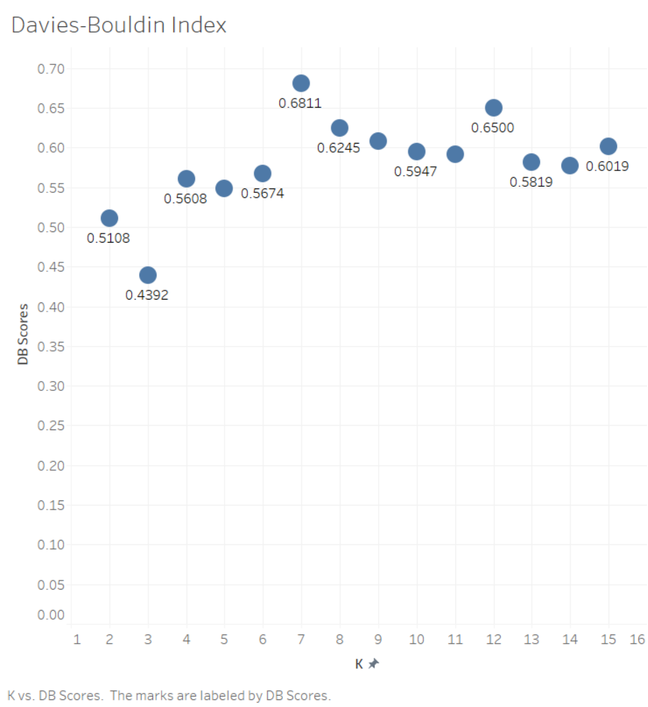


Figure 4. 2: DBI Results

Since the minimum Davies-Bouldin Index score yielded is 0.4392 with k=3, the clustering is based on k=3. Figure 4.3 shows the clusters with their centroids. The coordinates of the centroids are provided in Table 4.2.

Table 4. 2: Cluster Centroids

Cluster Number	Latitude	Longitude
0	43.87886	-80.081
1	48.9969	-90.5125
2	45.00884	-76.1556

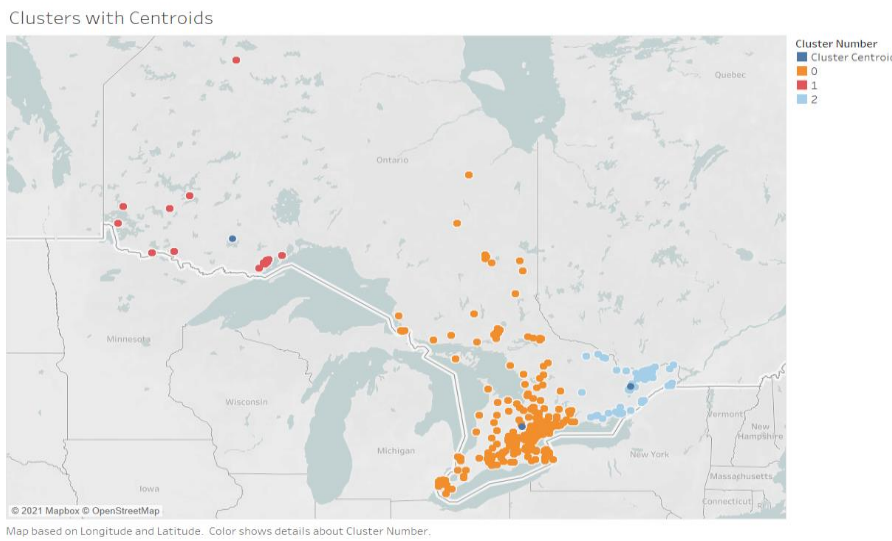


Figure 4. 3: Clusters of Buildings

Chapter 5: Results Analysis

5.1. Demand Prediction

In this step, the goal is to train the regression algorithms to predict the values of energy intensity per square foot of the case study buildings. Categorical features, e.g., sectors, subsectors, operation types, cities, etc., are encoded. In the process of cleansing the dataset, missing values and outliers are taken care of. Then, multiple regression algorithms and feature selection algorithms are applied to the dataset to find the best regressor with the minimum number of selected features. The aim of this process is to train high-accuracy models as fast as possible.

Alongside these regression algorithms, filter, wrapper, embedded, and hybrid feature selection algorithms are used to reduce the feature space size. The advantage of feature space size is that algorithms will get faster and less over/under-fitted. The performance of each prediction algorithm is evaluated after each feature reduction method is applied to the model. If the space with a lower number of features results in a model with higher performance metrics, the subset of features is selected as the new feature space. This process is repeated until no performance improvement is observed. Generally, the process is unsupervised and human interaction has no control over the performance metrics. Based on feature selection algorithms, the following features are selected to be the input of the regression algorithms:

- Total Indoor Space (m^2)
- Weekly Average Operational Hours (Hours)
- Electricity Quantity (kWh)
- Natural Gas Quantity (m^3)
- GHG Emissions (Kg)

The predicted value is "Energy Intensity Per Square Foot". To make sure the model is neither over-fitted nor under-fitted, the dataset is divided into training and testing set. As a rule of thumb, the proportion of training to testing set is 4 to 1. Firstly, five-fold cross validation is performed on the training set. Then, the trained algorithm resulted from the training set is applied to the testing set, as new, never-seen-before data points. This process is summarized in Figure 5.1 [81].

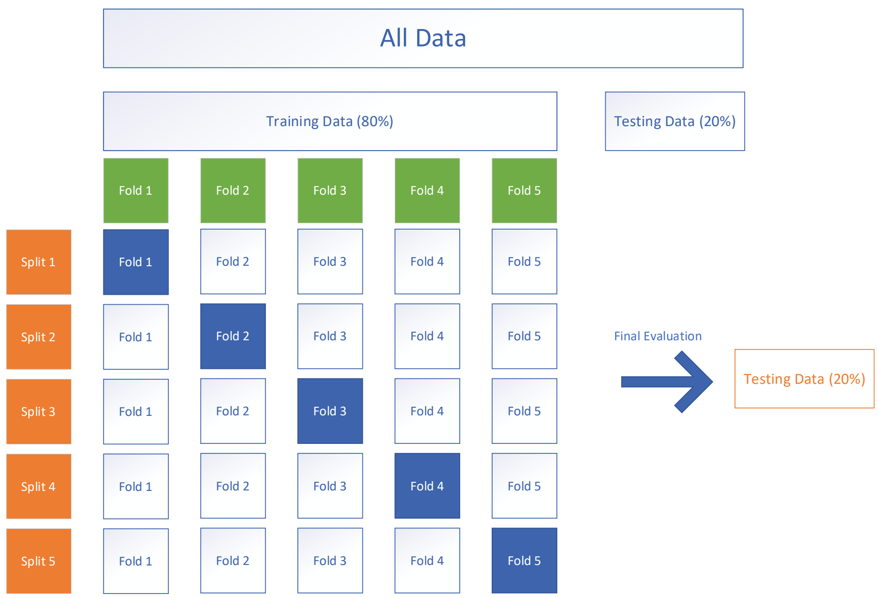


Figure 5. 1: Five-Fold Cross-Validation

Different homogenous subsets, for each of the clusters or sectors, of the datasets are trained. The highest predictive performance is yielded by CatBoost regressor on the whole dataset. CatBoost benefits from gradient boosting on decision trees[82]. Table 5.1 shows the CatBoost regressor results on five-fold cross-validation, training set, and testing set. It can be seen that the difference between the training set R^2 score and testing set R^2 score is less than 0.25%. This difference between the Five-fold cross-validation and testing set is 0.45 %. Conclusively, the model is neither over-fitted nor under-fitted. Figure 5.2 shows the histogram of the residuals of the Catboost regressor. Figure 5.3 shows the residuals versus predicted values of the training and testing sets.

Table 5. 1: CatBoost Regressor Results

Subset	**R^2 Score (%)**	**RMSE (kwh/sqft)**
Five-fold cross validation on Training Set	98.43 (Standard Deviation: 0.26)	-
Training Set	99.12	1.45
Testing Set	98.88	1.64
Difference between CV and Testing Set	0.45	-
Difference between Training and Testing Set	0.24	0.19

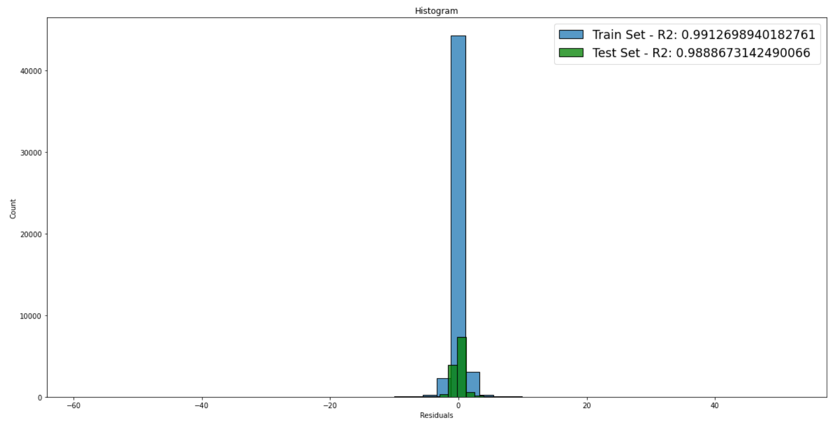


Figure 5. 2: Catboost Regressor Residuals Histogram

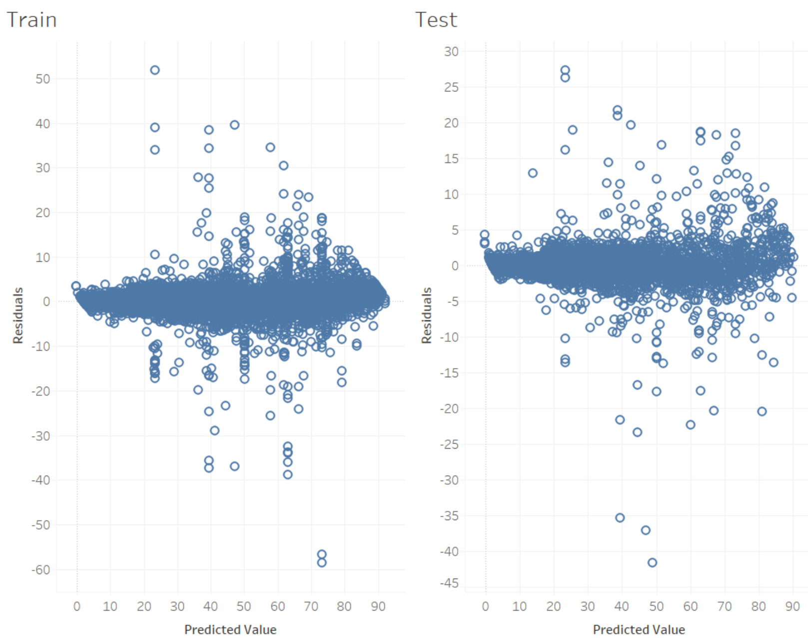


Figure 5. 3: CatBoost Regressor Residuals Scatter Plot

To demonstrate how Catboost regressor outperforms other algorithms, its results are compared to a number of alternative algorithms as follows.

The results of a gradient boosting regressor is presented in Table 5.2. The residuals histogram and plot of the gradient boosting regressor is also presented in Figure 5.4.

Table 5. 2: Gradient Boosting Regressor Results

Subset	**R^2 Score (%)**	**RMSE (kwh/sqft)**
Five-fold cross validation (Training Set)	88.20 (Standard Deviation: 0.3)	-
Training Set	89.26	5.10
Testing Set	88.67	5.25
Difference between CV and Testing Set	0.47	-
Difference between Training and Testing Set	0.59	0.15

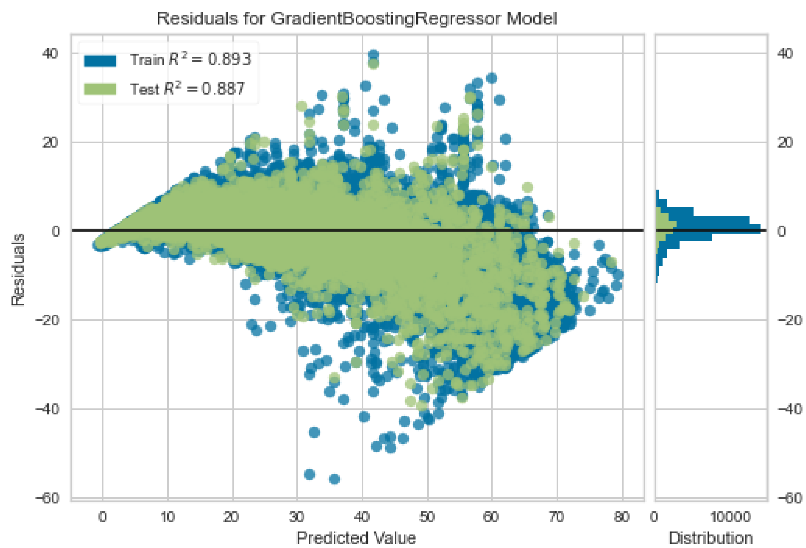


Figure 5. 4: Gradient Boosting Regressor Residuals

In case of this algorithm, the difference between R^2 in different subsets is not large; however, the accuracy is not as high as that of CatBoost regressor as R^2 score and RMSE are both lower.

The results obtained from the Random Forest regressor are shown in Table 5.3. Even though the R^2 score of the Random Forest regressor for training set is higher than that of CatBoost regressor, the difference between R^2 and RMSE scores of training and testing set for Random

Forest is more than those of CatBoost regressor. Therefore, it can be concluded that CatBoost regressor is subject to less over-fitting. The residuals plot and histogram of Random Forest regressor is presented in Figure 5.5.

Table 5. 3: Random Forest Regressor Results

Subset	**R^2 Score (%)**	**RMSE (kwh/sqft)**
Five-fold cross validation (Training Set)	98.13 (Standard Deviation: 0.14)	-
Training Set	99.65	0.92
Testing Set	98.53	1.89
Difference between CV and Testing Set	0.4	
Difference between Training and Testing Set	1.12	0.97

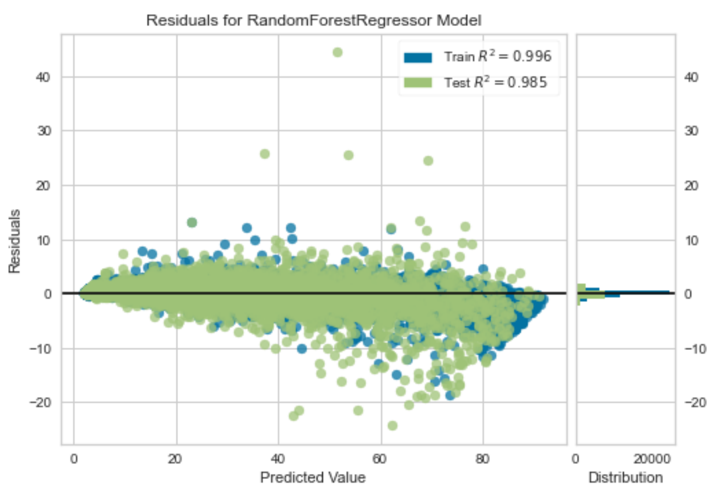


Figure 5. 5: Random Forest Regressor Residuals

The results of kNN regressor are reported for k = 2 and k = 8. The reasons are that k =2 yields the highest accuracy and k = 8 yields the lowest difference between training and testing set accuracy. Table 5.4 shows these results.

Table 5. 4: kNN Regressor Results

K	Subset	R^2 Score (%)	RMSE (kwh/sqft)
2	Training Set	96.32	2.99
	Testing Set	87.95	5.42
	Difference between Training and Testing Set	8.37	2.43
8	Training Set	88.80	5.21
	Testing Set	85.21	6.00
	Difference between Training and Testing Set	3.59	0.79

The residuals plots for k=2 and for k=8 are shown in Figures 5.6 and 5.7, respectively. It can be concluded that k=2 is over-fitted in comparison with k=8. In addition, k = 8 is not capable of outperforming the Catboost regressor.

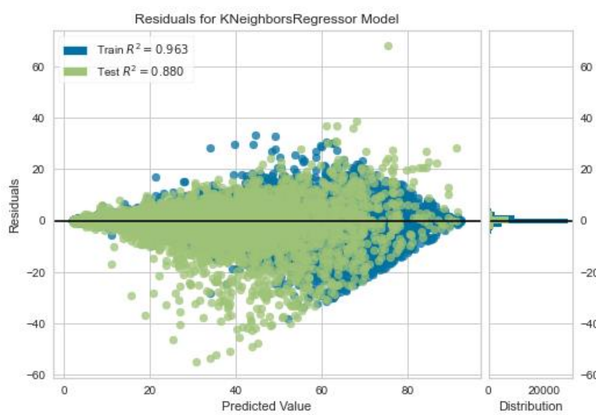


Figure 5. 6: kNN Regressor Residuals (k=2)

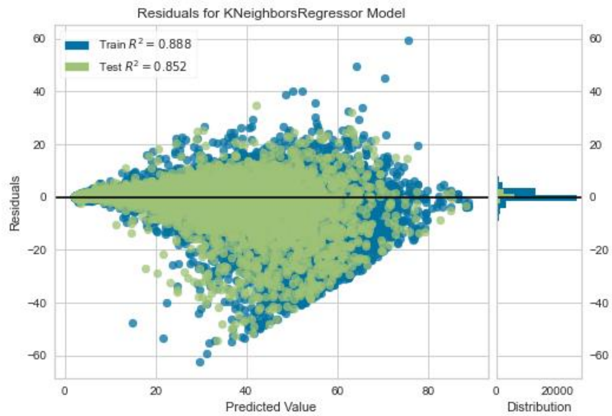


Figure 5. 7: kNN Results (k=8)

Overall, comparing the above predictive models, the CatBoost regressor is selected based on the following reasons:

1. R^2 and RMSE scores of this algorithm for five-fold cross validation, training set and testing set show a higher and more stable accuracy.

2. It is not computationally costly in comparison with the other algorithms.

3. Training on the whole dataset, rather than homogenous subsets, is faster.

5.2. Biomass Availability Prediction

The biomass yield dataset adopted from [67] contains monthly data from 1939 to 2020. Since the final reporting year in BPS dataset is 2018, the foundation of precipitation amounts and biomass supply stocks are also based on the year 2018. The following graph shows the time-series data of the precipitation dataset from 1939 to 2017[67]. These data (Figure 5.8) will serve as the inputs for the prediction model.

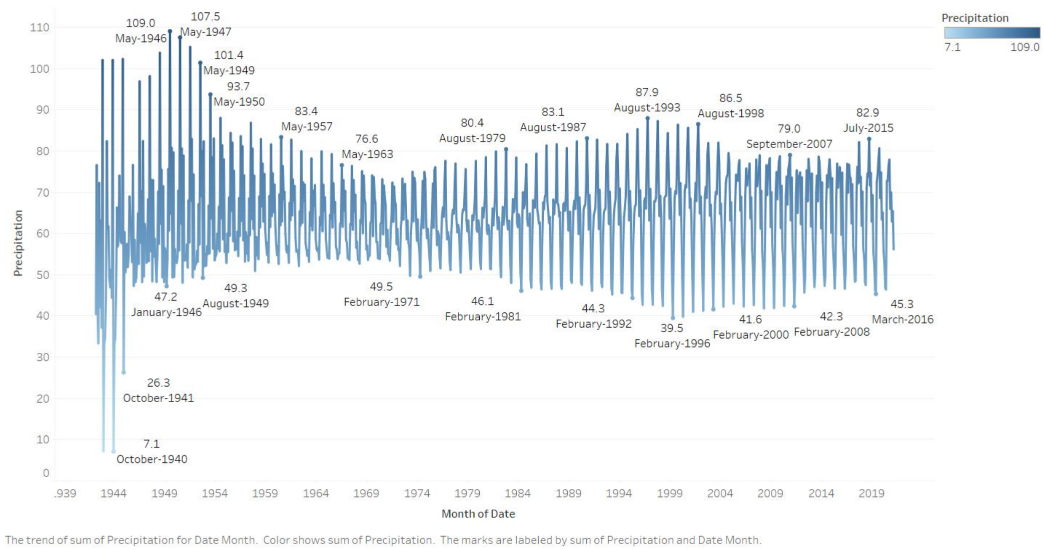


Figure 5. 8: Time-series Historical Data of Precipitation

The RNN which utilizes LSTM could predict the amount of monthly precipitation for the year 2018. Figure 5.9 shows the predicted and real values of precipitation in the year 2018. The real and predicted yearly precipitations are 771.03 and 763.67, respectively. The percent of the total difference for year 2018 is -0.9%. The average precipitation level for the years 2013 and 2014 is 782.87. Therefore, the coefficient to convert the level of biomass from averages of years 2013 and 2014 to 2018 is (763.67 / 782.87 = 0.97).

As an example to show how the availability estimations are carried out, consider centroid 2 that is located at (Latitude: 45.00884, Longitude: -76.1556). The amount of biomass scraped off BIMAT shows 482 tonnes of biomass was available within 10km radial range of this point on average between the years 2013 and 2014. Therefore, the amount of biomass available in 2018 is 482 * 0. 97 = 467.54. After this initial availability estimation, the impacts of weather fluctuation on the levels of precipitation (and thus biomass availability) are then assessed. Table 5.9 shows the amount of precipitation based on different levels of weather fluctuation.

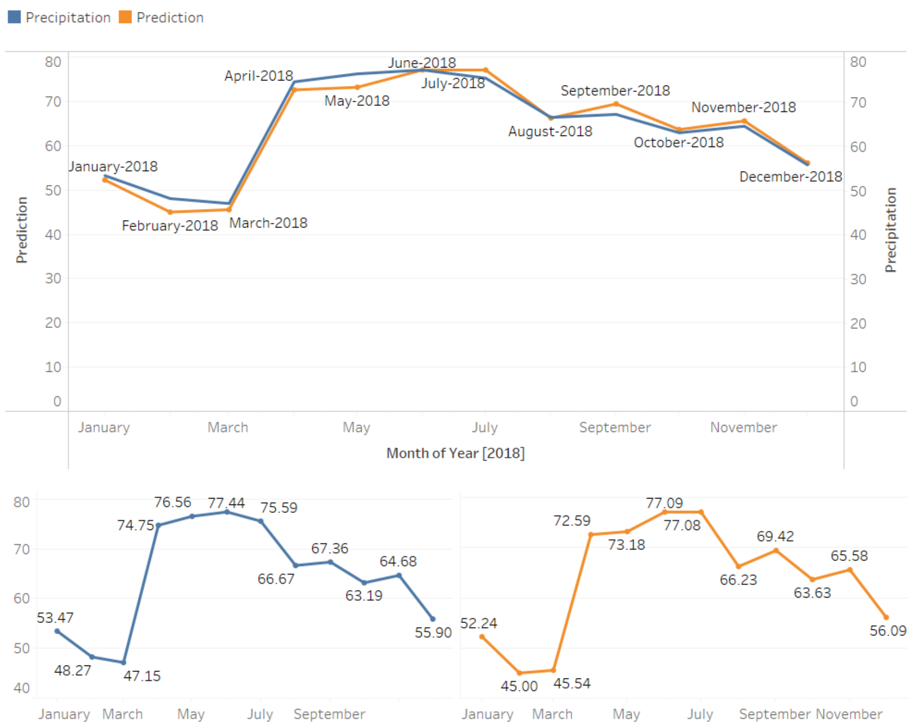


Figure 5. 9: Year 2018 Precipitation Real and Predicted Values

Table 5. 5: PTD for Different Fluctuations

Precipitation Fluctuation	Real Precipitation	Predicted Precipitation	PTD
-0.1	780.41	745.61	-4.45
-0.2	780.41	754.52	-3.31
-0.3	780.41	763.80	-2.12
+0.1	780.41	776.72	-4.72
+0.2	780.41	798.549	2.32
+0.3	780.41	818.01	4.81

In addition, the secondary effects of fluctuations can be assessed during the second next year. In this situation, precipitation fluctuation in year 2018 (as the primary effect) affects the estimation of values of year 2019. Similarly, the predicted values of 2019 serve as the input of the RNN model (as a secondary effect from 2018 values) to generate the estimations for year 2020. Table 5.6 presents an example of establishing primary and secondary effects (numbers in headers show the extent of fluctuation).

Table 5. 6: Primary and Secondary Effects

Time	**0%**	**-0.1%**	**Time**	**Primary Effect of 2018**	**Time**	**Real Values of 2020**	**Secondary Effect of 2018**
2018-12-01	55.9	55.341	2019-12-01	58.527	2020-12-01	58.16	57.59
2018-11-01	64.68	64.0332	2019-11-01	66.081	2020-11-01	64.99	64.93
2018-10-01	63.19	62.5581	2019-10-01	65.161	2020-10-01	65.27	64.52
2018-09-01	67.36	66.6864	2019-09-01	68.065	2020-09-01	66.75	66.57
2018-08-01	66.67	66.0033	2019-08-01	67.710	2020-08-01	69.53	66.41
2018-07-01	75.59	74.8341	2019-07-01	74.768	2020-07-01	75.19	72.95
2018-06-01	77.44	76.6656	2019-06-01	76.594	2020-06-01	78.74	75.50
2018-05-01	76.56	75.7944	2019-05-01	75.953	2020-05-01	78.03	75.00
2018-04-01	74.75	74.0025	2019-04-01	74.635	2020-04-01	79.67	73.55
2018-03-01	47.15	46.6785	2019-03-01	52.331	2020-03-01	48.34	52.53
2018-02-01	48.27	47.7873	2019-02-01	50.425	2020-02-01	50.2	49.12
2018-01-01	53.47	52.9353	2019-01-01	54.844	2020-01-01	56.21	52.83

In this case, PTD of the year 2020 as the secondary effect of the fluctuation in year 2018 is -2.47%, while this number (as the primary effect) for year 2019 is -4.45%.

5.3. Biomass Distribution

In this step, VRP is used to model the optimal distribution of biomass. Three variants of VRP are considered as VRP with distance targets, VRP with duration targets, and Capacitated VRP. To select the buildings that are included as nodes in VRP (i.e. a maximum of 9 buildings as per limitations of the chosen VRP algorithm), criticality level of buildings is taken into account. Four sectors available in the buildings dataset are hospitals, Post-Secondary Educational Institutions, Municipal buildings, and school boards. It is assumed that hospitals have the highest criticality level. Nine hospital buildings are selected as the demand nodes. The geographical locations of the buildings are shown in Figure 5.10. The red dot represents the depot, and the black dots represent hospitals. The distance matrix for "distance target" is formed as presented in Table 5.7 (and in form of a heat map as presented in Figure 5.11).

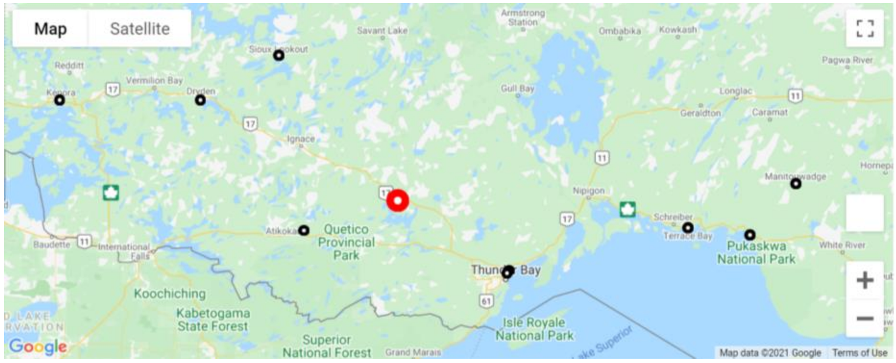


Figure 5. 10: Geographical Locations of Hospitals

Table 5. 7: Distance Matrix for Distance Target

	0	1	2	3	4	5	6	7	8	9
0	0	419859	132192	510238	134376	340793	215780	222132	260890	358620
1	419714	0	292276	95814	289605	79601	484805	631524	670282	768012
2	132606	292351	0	382731	2838	213285	197697	344416	383175	480904
3	510152	95917	382714	0	380043	170039	575243	721962	760721	858450
4	134497	289360	2823	379740	0	210294	199588	346307	385066	482795
5	340648	79601	213210	169980	210539	0	405739	552458	591216	688946
6	217967	484933	197265	575312	199450	405867	0	206801	245559	343289
7	222106	631643	343975	722022	346160	552577	206775	0	99811	140532
8	260862	670399	382731	760778	384916	591333	245531	99808	0	236296
9	358525	768062	480394	858441	482579	688996	343194	140426	236230	0

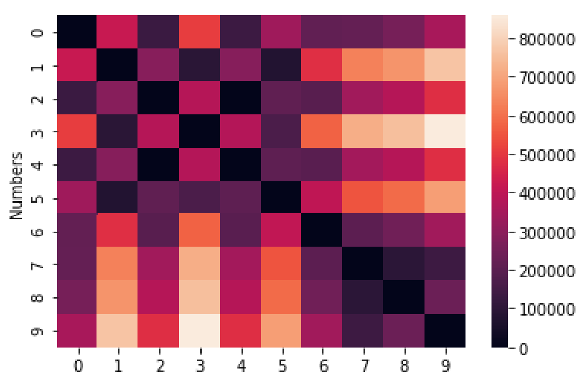
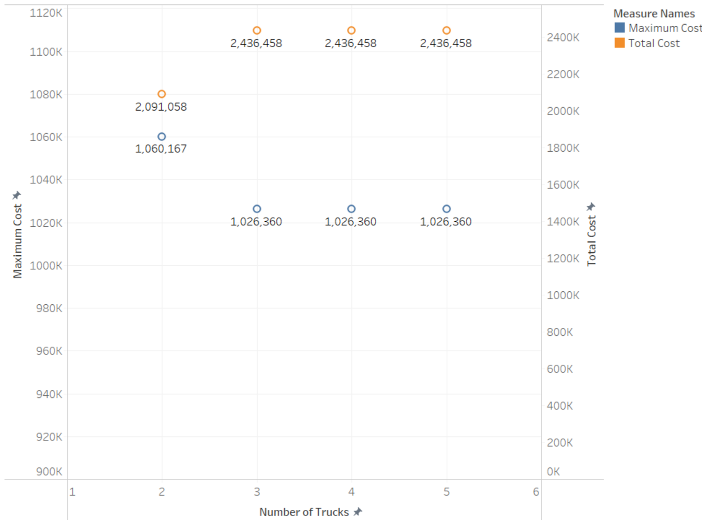


Figure 5. 11: Distance Matrix Heat Map for Distance Target

With distance target, the VRP is performed for a set of two to five trucks. The results are presented in Table 5.8. It can be concluded that if the number of trucks exceeds three, the additional trucks will not be dispatched in the "distance target" scenario. The "maximum route cost" in table 5.8 represents the maximum distance one of the trucks need to traverse in each dispatchment. Figure 5.12 shows the total costs obtained for each scenario.

Table 5. 8: VRP Results with Distance Target

Number of Trucks	Vehicle #	Sequence of the Nodes	Route Cost
2	0	0 -> 2 -> 4 -> 5 -> 1 -> 3 -> 0	1,030,891
	1	0 -> 6 -> 8 -> 9 -> 7 -> 0	1,060,167
	Maximum Route Cost		1,060,167
3	0	0 -> 8 -> 9 -> 7 -> 0	859,718
	1	0 -> 5 -> 1 -> 3 -> 0	1,026,360
	2	0 -> 6 -> 2 -> 4 -> 0	550,380
	Maximum Route Cost		1,026,360
4	0	0 -> 8 -> 9 -> 7 -> 0	859,718
	1	0 -> 0	0
	2	0 -> 5 -> 1 -> 3 -> 0	1,026,360
	3	0 -> 6 -> 2 -> 4 -> 0	550,380
	Maximum Route Cost		1,026,360
5	0	0 -> 8 -> 9 -> 7 -> 0	859,718
	1	0 -> 0	0
	2	0 -> 0	0
	3	0 -> 5 -> 1 -> 3 -> 0	1,026,360
	4	0 -> 6 -> 2 -> 4 -> 0	550,380
	Maximum Route Cost		1,026,360



Number of Trucks vs. Maximum Cost and Total Cost. Color shows details about Maximum Cost and Total Cost. For pane Maximum Cost: The marks are labeled by Maximum Cost. For pane Total Cost: The marks are labeled by Total Cost.

Figure 5. 12 : Maximum and Total Costs for Distance Targets

In the case where three trucks are dispatched, the first route starts from the depot and delivers biomass to buildings number eight, nine and seven, and finally returns to the depot. The cost of this route is 859,718 based on the table 5.8. On the other hand, if we use three trucks in the same route, meaning that each truck loads the biomass for one distinct building, delivers it to the building, and returns to the depot; the total cost will be 1,683,135.

The distance matrix for “duration target” is formed as presented in Table 5.9 and Figure 5.13. With duration target, the VRP is similarly performed for a set of two to five trucks. The results are presented in Table 5.10. The “maximum route cost” in table 5.10 represents the maximum time one of the trucks need to spend in each dispatchment. The Figure 23 presents the total costs obtained for each scenario.

Table 5. 9: Distance Matrix for Duration Target

	0	1	2	3	4	5	6	7	8	9
0	0	16702	5660	20179	5703	13531	8501	8895	10369	14459
1	16693	0	11723	3884	11600	3315	19169	24647	26121	30211
2	5706	11695	0	15172	310	8524	8182	13660	15134	19224
3	20192	3916	15222	0	15099	6814	22668	28146	29620	33710
4	5719	11583	336	15059	0	8411	8195	13673	15147	19237
5	13469	3266	8499	6743	8376	0	15945	21423	22897	26987
6	8806	19152	8109	22629	8153	15981	0	8992	10466	14556
7	8901	24751	13708	28228	13752	21580	8999	0	4415	6033
8	10408	26258	15215	29734	15259	23086	10506	4446	0	10010
9	14503	30354	19311	33830	19355	27182	14601	6021	10018	0

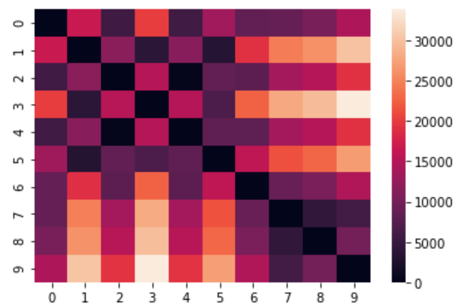


Figure 5. 13: Distance Matrix Heat Map for Duration Target

Table 5. 10: VRP Results with Duration Target

Number of Trucks	Vehicle #	Sequence of the Nodes	Route Cost
2	0	0 -> 2 -> 4 -> 1 -> 3 -> 5 -> 0	41,720
	1	0 -> 6 -> 8 -> 9 -> 7 -> 0	43,899
		Maximum Route Cost	43,899
3	0	0 -> 6 -> 2 -> 4 -> 0	22,639
	1	0 -> 1 -> 3 -> 5 -> 0	40,869
	2	0 -> 8 -> 9 -> 7 -> 0	35,301
		Maximum Route Cost	40,869
4	0	0 -> 2 -> 4 -> 1 -> 5 -> 0	34,337
	1	0 -> 6 -> 0	17,307
	2	0 -> 3 -> 0	40,371
	3	0 -> 8 -> 9 -> 7 -> 0	35,301
		Maximum Route Cost	40,371
5	0	0 -> 2 -> 4 -> 1 -> 5 -> 0	34,337
	1	0 -> 6 -> 0	17,307
	2	0 -> 0	0
	3	0 -> 3 -> 0	40,371
	4	0 -> 8 -> 9 -> 7 -> 0	35,301

	Maximum Route Cost	40,371

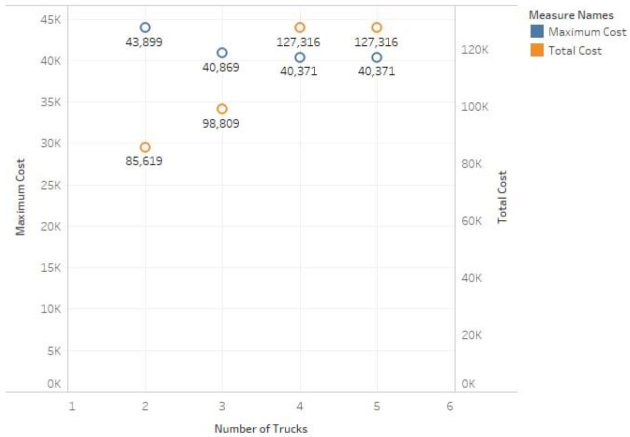


Figure 5. 14: Maximum and Total Costs for Duration Targets

In the scenario with “distance targets”, the optimal number of trucks is three. Because based on the routes report, if we assume that end-users are going to receive a fixed, equal amount of biomass, we can merge this criterion with maximum capacity of trucks (Table 3.1) and set the capacity of trucks as thirty. In this case each building will receive ten tonnes of biomass. In this sense, the following map (Figure 5.15) shows the optimal routes.

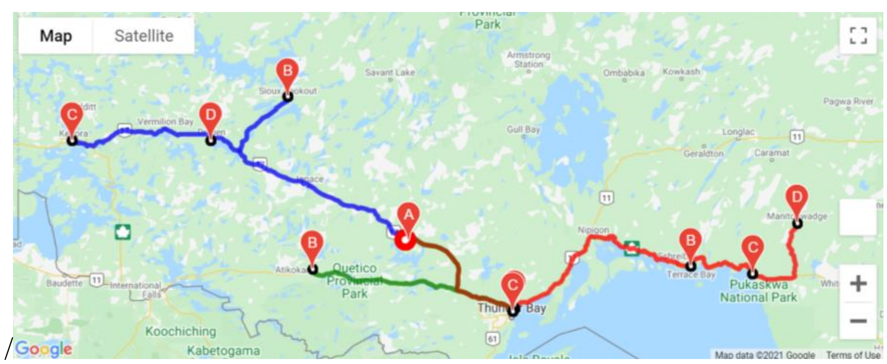


Figure 5. 15: Routes for VRP with Distance Targets

In “Distance Target” scenario, each truck starts the route with full payload of forty-five tonnes, delivers 15 tonnes to each building and goes back to the depot empty. As a sample, the total distribution and building emissions can be calculated, using payload emission factors (Table 3.1), and distances between each two destinations (using the distance matrix). An example of calculations based on the route for vehicle 0 (0 -> 8 -> 9 -> 7 -> 0) presented in Table 5.11. On that basis, the total emission for this route is estimated as 41.628 kg of eCO2.

Table 5. 11: Route Emissions Calculations

Building Number	0	8	9	7	0
Payload at the Destination	-	45	30	15	0
Distance from the Previous Point	-	261	236	140	222
Emissions	-	10440	12508	9800	8880

If we assume that these buildings burn biomass that they receive to replace natural gas, then the total amount of CO2 saved by these buildings is calculated as follows:

Total Emissions Saved = 45 tonnes of biomass * 4,900 kWh/tonne * (Emissions of Biomass/tonne – Emissions of Natural Gas / Tonne)

Then buildings can save 24,843 kg of eCO2 emissions per year. Considering 41.628 kg of emissions for distributions, a net 24,801 kg of emissions is saved. On the other hand, the buildings need to pay about five thousand seven hundred (5,700) dollars more for switching to burning pellets; assuming that they already have stoves. If the government provides the buildings with incentives equal to this amount, as a reward for saving emissions, this scenario will be financially feasible for these buildings. In this case, the total biomass available on the ground for this cluster is 1,150,000 tonnes in 2018. If each of 285 buildings of the cluster 1 receives 10 tonnes of biomass, the total biomass amount distributed among them would be 2,850 tonnes. Then, 1,147,150 tonnes of biomass will be available to be either sold to other clusters or exported.

On the other hand, if three trucks are responsible for delivering biomass to the three buildings, each truck to one building, the payload of each truck will be fifteen tonnes. After each truck delivers fifteen tonnes of biomass to each building, it will turn around and return to the depot. In this case, the total amount of 79.342 kg of eCO2 will be emitted into the environment, whereas the VRP algorithm distributes the same amount of biomass with 41.628 km of eCO2.

In the scenario with "duration targets", the optimal number of trucks is chosen as four because the results will not improve with five trucks. The map in Figure 5.16 shows the optimal routs.

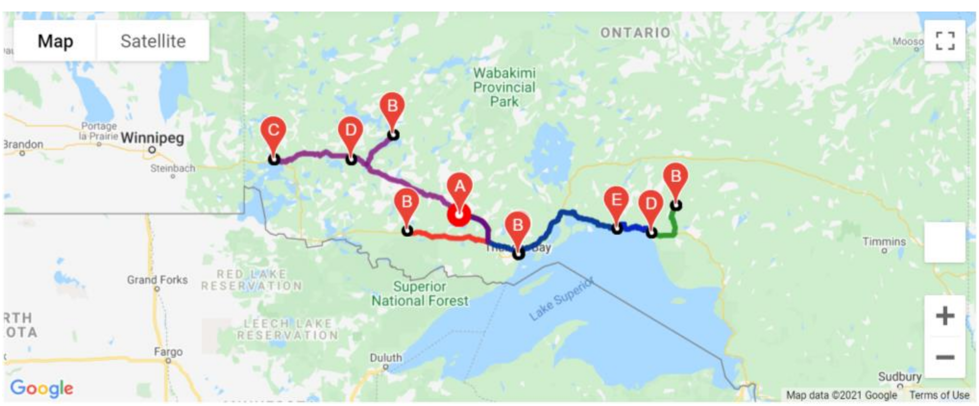


Figure 5. 16: Routes for VRP with Duration Targets

We now turn to CVRP model. Nine hospitals receive the biomass amount proportionate to their total energy demand. Table 5.12 shows the characteristics of these hospitals (the first row corresponds to the depot). Because the minimum truck capacity is 40 tonnes in table 3.1, CVRP uses 40-tonne trucks. After forming the distance matrix, the results of the CVRP show that three 40-tonne trucks can deliver the biomass to the end-users. The total distance traversed in this scenario is 2531285. Figure 5.17 shows the map of these routes.

Table 5. 12: Depot and Nine Hospitals Attributes

	latitude	longitude	demand
0	48.996900	-90.512500	0
1	48.729196	-86.359268	5
2	48.432008	-89.229466	5
3	49.125499	-85.827681	5
4	48.449019	-89.205107	5
5	48.786665	-87.098590	10
6	48.761339	-91.620874	10
7	49.770029	-92.841063	15
8	50.104298	-91.924269	25
9	49.767572	-94.499809	30

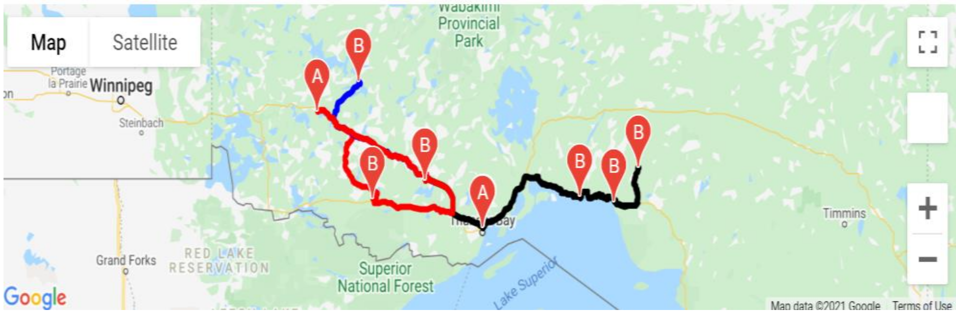


Figure 5. 17: Routes for Distance Targets in CVRP

If the level of precipitation (demand for biomass) in the year 2017 drops by 10% (Table 5.13), the RNN model predicts that the total amount of precipitation in the year 2018 would be 692.41. This means that the amount of biomass for this year is multiplied by a coefficient of (692.41 / 782.87= 0.884). Therefore, 11.6 percent of the biomass is lost due to weather conditions. Consequently, this coefficient could be applied to establish the demands in CVRP, by assuming that biomass delivery to the buildings will drop by 11.6 percent. As a result, the capacity of trucks can be decreased to 35. In this case, three trucks with the moving distance of 3,158,823 can supply the buildings with biomass. Figure 5.18 shows the map of these routes.

Table 5. 13: Depot and Nine Hospitals Attributes

	latitude	longitude	demand
0	48.996900	-90.512500	0.0
1	48.729196	-86.359268	4.5
2	48.432008	-89.229466	4.5
3	49.125499	-85.827681	4.5
4	48.449019	-89.205107	4.5
5	48.786665	-87.098590	9.0
6	48.761339	-91.620874	9.0
7	49.770029	-92.841063	13.5
8	50.104298	-91.924269	22.0
9	49.767572	-94.499809	26.0

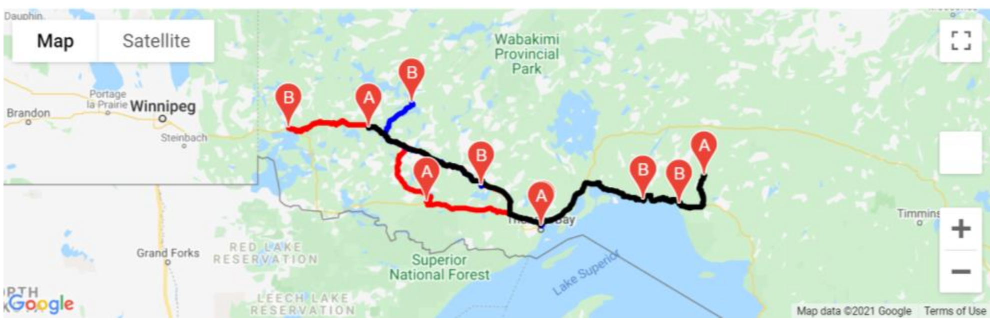


Figure 5. 18: Routes for Distance Targets in CVRP

In summary, in the blue-sky scenario, 110 tonnes of biomass is delivered to the buildings. In the black-sky scenario, 97.5 tonnes of biomass is delivered. Therefore, the percentage of decrease in biomass delivery is calculated as:

((97.5 – 110) / 110) * 100 = - 11.3%

A smaller decline in biomass delivery shows a supply chain that is more resilient to climate conditions. It shall be noted that the percentage of decrease in biomass delivery can be calculated for secondary effects as well; because once a weather fluctuation occurs in an area, its effects last for years. By calculating the amount of biomass delivered to end users, because of secondary effects, we can estimate how much indirect impact the weather fluctuation has.

In this research, predictive algorithms, including CatBooost, Random Forest, Neural Networks, etc., are fed with buildings attributes, such as , to predict buildings demand. Secondly, a recurrent neural network is trained to approximate the availability of biomass. Then, trucks are dispatched from facilities to collect biomass from land based on buildings demand of each cluster, while the trucks traverse the minimum distances. Lastly, a fleet of trucks are dispatched to deliver biomass to buildings. The results of the proposed model reveals that carbon footprint of the buildings will be mitigated by replacing fossil fuels with biomass, whereas the buildings need financial assistance to afford the biomass prices. Lastly, weather fluctuations affect the supply levels and show how biomass levels, and consequently building demands, will be altered. The results of the proposed model could be used by decision makers to come up with plans of replacing fossil fuels with biomass.

Chapter 6: Conclusions

This study aims at addressing two objectives: (1) optimally routing biomass collection from ground into depot and distribution from depot to end users, and (2) the affects that weather fluctuation have on biomass availability and its supply chain resilience. A group of buildings are taken as a case study. By applying clustering algorithms to the geographical coordinates of the buildings, centroid of each cluster is considered as the biomass facility/depot. A number of predictive models were developed and tuned to forecast the values of building energy demand and stock availability of biomass. Then, an allocation algorithm was formulated to direct the optimal collection of biomass from land into depots while trucks' traveling the minimum distances. Lastly, Google Maps API was employed to find the best distribution routes for delivering biomass from depots to end-users. Different targets were considered in modeling of biomass distribution from depots to end-users. Lastly, it was investigated how weather fluctuation affects the biomass availability and its supply chain resilience.

To formulate these models, a number of challenges needed to be addressed. The predictive model of energy demand required many building data points with accurate information. In addition, Google API had its limitation of accepting ten nodes on the map. If an API and map with larger query sizes were available, the number of buildings in step three could have been increased. In this sense, if steps two and three could be iterated for a larger set of parameters, e.g., varied values of precipitation fluctuations, more generic prediction benchmarks could be developed for future reference.

In this study, the number of observations in BPS dataset was not large enough for neural networks to capture the underlying patterns of the data set. In addition, neural networks are computationally more expensive than the competitive regressors trained in this study. Therefore, if a data set with more observation points becomes available, use of neural networks could be investigated.

For future research, the deliveries could be handled by multiple carriers. Therefore, their characteristics and pricing policies could be considered in the model. If the building demand data were reported with more levels of details, e.g., hourly, daily, etc., biomass delivery schedules could be established in weekly or monthly basis. Moreover, this study made a number of generic assumptions in regard to conversion technologies located at end-users' locations. In case, the buildings could disclose their conversion technologies, and their corresponding coefficients of performance, this information could be factored in the prioritizations of buildings with higher energy intensity returns on each unit of biomass delivered to them. In this study, depots are located first, then then routing problem is modeled. A dependent, multi-objective optimization algorithm to locate depots and find optimal routes, simultaneously, could be investigated. The current model is not optimized to work with country-wide case studies. The challenges could be driving over days and nights, supplier selection, hub and spoke location, etc. The current model is not optimized to work with country-wide case studies. The challenges could be driving over days and nights, supplier selection, hub and spoke location, etc. The challenges and advantages of integrating "Graph Neural Networks" could be investigated. In this study, price is a linear function of distance from the one depot. With dynamic pricing of multiple suppliers, supplier selection and competitiveness could be investigated using "Game Theory". Lastly, other variants of VRP, e.g., VRP with time windows for supplying biomass to remote communities which have limited access to roads, VRP with pickups and deliveries for cases where supply chains have hubs, etc. could be investigated.

This study did not investigate the impact of biomass on social criteria such as job creation and its incentives. By considering these factors in communities which use decentralized energy sources, the feasibility of using biomass as an alternative source of energy could be further justified.

References

[1] R. Saidur, E. A. Abdelaziz, A. Demirbas, M. S. Hossain, and S. Mekhilef, “A review on biomass as a fuel for boilers,” *Renewable and Sustainable Energy Reviews*, vol. 15, no. 5. Pergamon, pp. 2262–2289, 01-Jun-2011.

[2] S. V. Vassilev, D. Baxter, L. K. Andersen, and C. G. Vassileva, “An overview of the chemical composition of biomass,” *Fuel*, vol. 89, no. 5. pp. 913–933, May-2010.

[3] E. Iakovou, A. Karagiannidis, D. Vlachos, A. Toka, and A. Malamakis, “Waste biomass-to-energy supply chain management: A critical synthesis,” *Waste Manag.*, vol. 30, no. 10, pp. 1860–1870, Oct. 2010.

[4] F. Schipfer, L. Kranzl, O. Olsson, and P. Lamers, “The European wood pellets for heating market - Price developments, trade and market efficiency,” *Energy*, vol. 212, Dec. 2020.

[5] Y. P. Lin, W. H. Wang, S. Y. Pan, C. C. Ho, C. J. Hou, and P. C. Chiang, “Environmental impacts and benefits of organic Rankine cycle power generation technology and wood pellet fuel exemplified by electric arc furnace steel industry,” *Appl. Energy*, vol. 183, pp. 369–379, Dec. 2016.

[6] L. Visser, R. Hoefnagels, and M. Junginger, “Wood pellet supply chain costs – A review and cost optimization analysis,” *Renew. Sustain. Energy Rev.*, vol. 118, Feb. 2020.

[7] J. Zhang, M. Zhi, and Y. Zhang, “Combined Generalized Additive model and Random Forest to evaluate the influence of environmental factors on phytoplankton biomass in a large eutrophic lake,” *Ecol. Indic.*, vol. 130, Nov. 2021.

[8] S. K. Hoi, B. N. R. Winayu, H. T. Hsueh, and H. Chu, “Light factors and nitrogen availability to enhance biomass and C-phycocyanin productivity of Thermosynechococcus sp. CL-1,” *Biochem. Eng. J.*, vol. 167, Mar. 2021.

[9] A. Röll *et al.*, “Water availability controls the biomass increment of Melia Dubia in south India,” *Forests*, vol. 12, no. 12, Dec. 2021.

[10] T. Tamer, I. Gürsel Dino, and C. Meral Akgül, “Data-driven, long-term prediction of building performance under climate change: Building energy demand and BIPV energy generation analysis across Turkey,” *Renew. Sustain. Energy Rev.*, vol. 162, p. 112396, Jul. 2022.

[11] A. Giallanza and G. L. Puma, “Fuzzy green vehicle routing problem for designing a three echelons supply chain,” *J. Clean. Prod.*, vol. v 259, Jun. 2020.

[12] K. Ransikarbum, C. Chaiyaphan, and R. Pataratanased, “Analysis of Logistical Aspect of Food-Safety System in the Green Supply Chain using Vehicle Routing Problem Model,” *Proc. - 2021 Res. Invent. Innov. Congr. Innov. Electr. Electron. RI2C 2021*, pp. 48–53, Sep. 2021.

[13] N. Al Theeb, H. J. Smadi, T. H. Al-Hawari, and M. H. Aljarrah, “Optimization of vehicle routing with inventory allocation problems in Cold Supply Chain Logistics,” *Comput.* , vol. v 142, Apr. 2020.

[14] L. Van Thillo, S. Verbeke, and A. Audenaert, “The potential of building automation and control systems to lower the energy demand in residential buildings: A review of their performance and influencing parameters,” *Renew. Sustain. Energy Rev.*, vol. 158, Apr. 2022.

[15] S. Gold and S. Seuring, "Supply chain and logistics issues of bio-energy production," *J. Clean. Prod.*, vol. 19, no. 1, pp. 32–42, Jan. 2011.

[16] "SCM Definitions and Glossary of Terms." [Online]. Available: https://cscmp.org/CSCMP/Academia/SCM_Definitions_and_Glossary_of_Terms/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms.aspx?hkey=60879588-f65f-4ab5-8c4b-6878815ef921. [Accessed: 06-May-2021].

[17] F. Mafakheri and F. Nasiri, "Modeling of biomass-to-energy supply chain operations: Applications, challenges and research directions," *Energy Policy*, vol. 67, pp. 116–126, 2014.

[18] M. Fattahi, K. Govindan, and M. Farhadkhani, "Sustainable supply chain planning for biomass-based power generation with environmental risk and supply uncertainty considerations: a real-life case study," *Int. J. Prod. Res.*, vol. 59, no. 10, pp. 3084–3108, 2021.

[19] K. Govindan and M. Fattahi, "Investigating risk and robustness measures for supply chain network design under demand uncertainty: a case study of glass supply chain," *Int. J. Prod. Econ.*, vol. v 183, pp. 680–699, Jan. 2017.

[20] F. Frombo, R. Minciardi, M. Robba, F. Rosso, and R. Sacile, "Planning woody biomass logistics for energy production: A strategic decision model," *Biomass and Bioenergy*, vol. v 33, n 3, no. 3, p. p 372-383, Mar. 2009.

[21] M. Brandenburg, "Low carbon supply chain configuration for a new product-a goal programming approach," *Int. J. Prod. Res.*, vol. v 53, n 21, no. 21, p. p 6588-6610, Nov. 2015.

[22] S. Elhedhli and R. Merrick, "Green supply chain network design to reduce carbon emissions," *Transp. Res. Part D (Transport Environ.*, vol. v 17, n 5, no. 5, pp. 370–379, 2012.

[23] M. Peng, Y. Peng, and H. Chen, "Post-seismic supply chain risk management: A system dynamics disruption analysis approach for inventory and logistics planning," *Comput. Oper. Res.*, vol. 42, pp. 14–24, Feb. 2014.

[24] M. Marufuzzaman, S. D. Eksioglu, X. Li, and J. Wang, "Analyzing the impact of intermodal-related risk to the design and management of biofuel supply chain," *Transp. Res. Part E Logist. Transp. Rev.*, vol. 69, pp. 122–145, Sep. 2014.

[25] R. Gedik, H. Medal, C. Rainwater, E. A. Pohl, and S. J. Mason, "Vulnerability assessment and re-routing of freight trains under disruptions: A coal supply chain network application," *Transp. Res. Part E Logist. Transp. Rev.*, vol. 71, pp. 45–57, Nov. 2014.

[26] M. Shekarian and M. Mellat Parast, "An Integrative approach to supply chain disruption risk and resilience management: a literature review," *Int. J. Logist. Res. Appl.*, vol. v 24, n 5, no. 5, pp. 427–455, 2021.

[27] M. Shekarian, S. V. Reza Nooraie, and M. M. Parast, "An examination of the impact of flexibility and agility on mitigating supply chain disruptions," *Int. J. Prod. Econ.*, vol. v 220, Feb. 2020.

[28] Y. Bai, X. Li, F. Peng, X. Wang, and Y. Ouyang, "Effects of disruption risks on biorefinery location design," *Energies*, vol. 8, no. 2, pp. 1468–1486, Feb. 2015.

[29] S. N. Emenike and G. Falcone, "A review on energy supply chain resilience through optimization," *Renewable and Sustainable Energy Reviews*, vol. 134. Elsevier Ltd, 01-Dec-2020.

[30] R. Carvalho, L. Buzna, F. Bono, M. Masera, D. K. Arrowsmith, and D. Helbing, "Resilience of natural gas networks during conflicts, crises and disruptions," *PLoS One*, vol. 9, no. 3, p. 90265, Mar. 2014.

[31] R. Hoggett, "Technology scale and supply chains in a secure, affordable and low carbon energy transition," *Appl. Energy*, vol. 123, pp. 296–306, Jun. 2014.

[32] D. H. Syahchari, D. Sudrajat, L. Lasmy, M. G. Herlina, F. Estefania, and E. Van Zanten, "Achieving Supply Chain Resilience through Supply Chain Risk Management and Supply Chain Partnership," *ACM Int. Conf. Proceeding Ser.*, p. p 209-212, Jan. 2022.

[33] M. Karimi and N. Zaerpour, "Put your money where your forecast is: Supply chain collaborative forecasting with cost-function-based prediction markets," *Eur. J. Oper. Res.*, vol. 300, no. 3, pp. 1035–1049, Aug. 2022.

[34] A. Raiyani, A. Lathigara, and H. Mehta, "Usage of time series forecasting model in Supply chain sales prediction," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1042, no. 1, p. 012022, Jan. 2021.

[35] N. Zougagh, A. Charkaoui, and A. Echchatbi, "Prediction models of demand in supply chain," *Procedia Comput. Sci.*, vol. 177, pp. 462–467, Jan. 2020.

[36] M. Dorostian and A. Moradmand, "Hierarchical Robust Model-based Predictive Control in Supply Chain Management under Demand Uncertainty and Time-delay," *2021 7th Int. Conf. Control. Instrum. Autom. ICCIA 2021*, Feb. 2021.

[37] J. Dai, Y. Zhang, M. Zhang, and Q. Zhang, "Optimization of stock basing on improved grey prediction model: a case study on garment supply chain," *Adv. Comput. Sci. Eng. Educ. III. Adv. Intell. Syst. Comput. (AISC 1247)*, vol. 1247 AISC, pp. 465–474, 2021.

[38] A. Ibrahim, E. Irawan, N. Kartika Dewi, Nurani, R. Filaresy, and Yusmaniarti, "The Implementation of Supply Chain Management and Big Data to Accelerate Stock Order in Mega Drug Store," *J. Phys. Conf. Ser.*, vol. v 1196, no. 1, Apr. 2019.

[39] Y. Yang, C. Peng, and Q. Li, "Predictive Control of Inventory Management in Supply Chain Systems with Uncertain Demands and Time Delays," *Chinese Control Conf. CCC*, vol. 2021-July, pp. 450–455, Jul. 2021.

[40] Q. Meng, Y. Xi, X. Ren, H. Li, L. Jiang, and L. Yang, "Thermal Energy Storage Air-conditioning Demand Response Control Using Elman Neural Network Prediction Model," *Sustain. Cities Soc.*, vol. v 76, Jan. 2022.

[41] Y. Chen and H. Tan, "Short-term prediction of electric demand in building sector via hybrid support vector regression," *Appl. Energy*, vol. v 204, pp. 1363–1374, 2017.

[42] A. Bassi, A. Shenoy, A. Sharma, H. Sigurdson, C. Glossop, and J. H. Chan, "Building Energy Consumption Forecasting: A Comparison of Gradient Boosting Models," *ACM Int. Conf. Proceeding Ser.*, vol. 9, Jun. 2021.

[43] H. Haque, A. K. Chowdhury, M. N. R. Khan, and M. A. Razzak, "Demand analysis of energy consumption in a residential apartment using machine learning," *2021 IEEE Int. IOT, Electron. Mechatronics Conf. IEMTRONICS 2021 - Proc.*, Apr. 2021.

[44] L. Pan *et al.*, "Biomass Prediction with 3D Point Clouds from LiDAR," *Proc. - 2022 IEEE/CVF*

Winter Conf. Appl. Comput. Vision, WACV 2022, pp. 1716–1726, 2022.

[45] B. Huy, N. Q. Truong, N. Q. Khiem, K. P. Poudel, and H. Temesgen, “Deep learning models for improved reliability of tree aboveground biomass prediction in the tropical evergreen broadleaf forests,” *For. Ecol. Manage.*, vol. 508, p. 120031, Mar. 2022.

[46] Z. Hu *et al.*, “Yield prediction of ‘Thermal-dissolution based carbon enrichment’ treatment on biomass wastes through coupled model of artificial neural network and AdaBoost,” *Bioresour. Technol.*, vol. v 343, Jan. 2022.

[47] T. Katongtung, T. Onsree, and N. Tippayawong, “Machine learning prediction of biocrude yields and higher heating values from hydrothermal liquefaction of wet biomass and wastes,” *Bioresour. Technol.*, vol. v 344, Jan. 2022.

[48] A. Masjedi, N. R. Carpenter, M. M. Crawford, and M. R. Tuinstra, “Prediction of sorghum biomass using uav time series data and recurrent neural networks,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 2695–2702, Jun. 2019.

[49] G. B. Dantzig and J. H. Ramser, “The Truck Dispatching Problem,” *Manage. Sci.*, vol. 6, no. 1, pp. 80–91, Oct. 1959.

[50] M. Nazari, A. Oroojlooy, M. Takáč, and L. V Snyder, “Reinforcement Learning for Solving the Vehicle Routing Problem.”

[51] D. M. Utama, S. K. Dewi, A. Wahid, and I. Santoso, “The vehicle routing problem for perishable goods: A systematic review,” *Cogent Eng.*, vol. v 7, n 1, no. 1, Jan. 2020.

[52] R. Soares, A. Marques, P. Amorim, and J. Rasinmäki, “Multiple vehicle synchronisation in a full truck-load pickup and delivery problem: a case-study in the biomass supply chain,” *Eur. J. Oper. Res.*, vol. v 277, n 1, no. 1, pp. 174–194, Aug. 2019.

[53] M. Drexl, “Synchronization in Vehicle Routing—A Survey of VRPs with Multiple Synchronization Constraints,” *https://doi.org/10.1287/trsc.1110.0400*, vol. 46, no. 3, pp. 297–316, Mar. 2012.

[54] J. X. Cao, Z. Zhang, and Y. Zhou, “A location-routing problem for biomass supply chains,” *Comput. Ind. Eng.*, vol. v 152, Feb. 2021.

[55] K. Li, W. Xue, G. Tan, and A. S. Denzer, “A state of the art review on the prediction of building energy consumption using data-driven technique and evolutionary algorithms,” *Build. Serv. Eng. Res. Technol.*, vol. v 41, n 1, no. 1, p. p 108-127, Jan. 2020.

[56] H. X. Zhao and F. Magoulès, “A review on the prediction of building energy consumption,” *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, Aug. 2012.

[57] “3.3. Metrics and scoring: quantifying the quality of predictions — scikit-learn 0.24.2 documentation.” [Online]. Available: https://scikit-learn.org/stable/modules/model_evaluation.html#r2-score. [Accessed: 07-May-2021].

[58] “solegalli (Soledad Galli) · GitHub.” [Online]. Available: https://github.com/solegalli. [Accessed: 09-Aug-2022].

[59] “1. Supervised learning — scikit-learn 0.24.2 documentation.” [Online]. Available: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning. [Accessed: 07-May-2021].

[60] "CatBoostRegressor - CatBoost. Documentation." [Online]. Available: https://catboost.ai/docs/concepts/python-reference_catboostregressor.html. [Accessed: 07-May-2021].

[61] "The Sequential model | TensorFlow Core." [Online]. Available: https://www.tensorflow.org/guide/keras/sequential_model. [Accessed: 07-May-2021].

[62] "Biomass Inventory Mapping and Analysis Tool." [Online]. Available: https://www.agr.gc.ca/atlas/bimat. [Accessed: 07-May-2021].

[63] "Biomass Inventory Mapping and Analysis Tool - Agriculture and Agri-Food Canada." [Online]. Available: https://www.agr.gc.ca/atlas/apps/aef/main/index_en.html?emafhelp=bimat_ocib. [Accessed: 02-Aug-2022].

[64] "2.3. Clustering — scikit-learn 0.24.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/clustering.html#k-means. [Accessed: 06-May-2021].

[65] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979.

[66] Z. Vazifeh, F. Mafakheri, and C. An, "Biomass supply chain coordination for remote communities: A game-theoretic modeling and analysis approach," *Sustain. Cities Soc.*, vol. 69, Jun. 2021.

[67] "Weather Dashboard for Toronto." [Online]. Available: https://toronto.weatherstats.ca/. [Accessed: 08-May-2021].

[68] "Overview | Distance Matrix API | Google Developers." [Online]. Available: https://developers.google.com/maps/documentation/distance-matrix/overview. [Accessed: 09-May-2021].

[69] CEFIC and ECTA, "Guidelines for Measuring and Managing CO2 Emission from Freight Transport Operations," *Ecta Rc*, vol. march, no. 1, p. 19, 2011.

[70] J. Seo, J. Park, Y. Oh, and S. Park, "Estimation of total transport CO2 emissions generated by medium- and heavy-duty vehicles (MHDVs) in a sector of Korea," *Energies*, vol. 9, no. 8, 2016.

[71] "Energy density - Energy Education." [Online]. Available: https://energyeducation.ca/encyclopedia/Energy_density. [Accessed: 03-Aug-2022].

[72] N. Resources Canada, "Natural Resources Canada solid biofuels bulletin series #2 – Primer for solid biofuels."

[73] "Carbon factor for wood fuels for the Supplier Obligation Final report."

[74] "Historical natural gas rates | Ontario Energy Board." [Online]. Available: https://www.oeb.ca/consumer-information-and-protection/natural-gas-rates/historical-natural-gas-rates. [Accessed: 03-Aug-2022].

[75] B. Gagnon, H. Macdonald, E. Hope, M. J. Blair, and D. W. McKenney, "Impact of the COVID-19 Pandemic on Biomass Supply Chains: The Case of the Canadian Wood Pellet Industry," *Energies*, vol. 15, no. 9, p. 3179, May 2022.

[76] "Broader public sector accountability | Ontario.ca." [Online]. Available: https://www.ontario.ca/page/broader-public-sector-accountability. [Accessed: 06-May-2021].

[77] "Energy use and greenhouse gas emissions for the Broader Public Sector - Datasets - Ontario Data Catalogue." [Online]. Available: https://data.ontario.ca/dataset/energy-use-and-greenhouse-gas-emissions-for-the-broader-public-sector. [Accessed: 06-May-2021].

[78] "Topographic Information." [Online]. Available: https://www.nrcan.gc.ca/topographic-information/10785. [Accessed: 06-May-2021].

[79] A. Strehl and J. Ghosh, "Cluster ensembles - A knowledge reuse framework for combining multiple partitions," in *Journal of Machine Learning Research*, 2003, vol. 3, no. 3, pp. 583–617.

[80] "sklearn.metrics.davies_bouldin_score — scikit-learn 0.24.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html. [Accessed: 06-May-2021].

[81] "3.1. Cross-validation: evaluating estimator performance — scikit-learn 0.24.2 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html. [Accessed: 07-May-2021].

[82] "CatBoost - open-source gradient boosting library." [Online]. Available: https://catboost.ai/. [Accessed: 02-Aug-2022].