PREDICTIVE DEMAND RESPONSE MODELING FOR LOGISTIC SYSTEMS INNOVATION AND OPTIMIZATION

A Dissertation
Presented to
The Academic Faculty

by

Zeynab Bahrami Bidoni

In Partial Fulfillment
of the Requirements for the Degree
Machine Learning in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2022

PREDICTIVE DEMAND RESPONSE MODELING FOR LOGISTIC SYSTEMS INNOVATION AND OPTIMIZATION

Approved by:

Dr. Benoit Montreuil, Advisor
School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Kamran Paynabar
School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Yao Xie
School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Frederick Benaben
School of Industrial & Systems Engineering
Georgia Institute of Technology

Dr. Nico Schmid
Department of Operations Management
IÉSEG - School of Management

Date Approved: December 5, 2022

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Benoit Montreuil, for guiding and supporting me over the years. He has set an example of excellence as a researcher, mentor, instructor, and project manager. I also want to thank my committee members, Dr. Kamran Paynabar, Dr. Yao Xie, Dr. Frederick Benaben, and Dr. Nico Schmid for all their collaboration and guidance throughout this process; their discussion, ideas, and feedback have been invaluable.

I especially thank my amazing family and colleagues for the love, support, and constant encouragement I have gotten over the years. In particular, I thank my parents. I undoubtedly could not have done this without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

BM	Bass Model
CAGR	Compound Annual Growth Rate
DMFs	Decision Making Factors
ETS	Exponential Triple Smoothing
GDP	Gross Domestic Product
GoF	Goodness-of-Fit
GWMCA	Geographically Weighted Modular Construction Attractiveness
IDB	International Data Base
KPIs	Key Performance Indicators
LSP	Logistic Service Provider
MC	Modular Construction
MiC	Modular integrated Construction
MSA	Metropolitan Statistical Area
OSC	Off-Site Construction
POD	Physics of Decision
US	United States
WoM	Word of Mouth
WoS	Web of Science
YOY	Year on Year (growth rate)

SUMMARY

In the ever-increasing dynamics of global business markets, logistic systems must optimize the usage of all possible sources to continually innovate. Scenario-based demand prediction plays an important role in the effective economic operations and planning of logistics. However, many uncertainties and demand variability, which are associated with innovative changes, complicate demand forecasting and expose system operators to the risk of failing to meet demand. This dissertation presents new approaches to predictively explore how customer preferences will change and consequently demand would respond to the new setup of services caused by an innovative transformation of the logistic layout. The critical challenge is that the responses from customers in particular and demand in general to the innovative changes and corresponding adjustments are uncertain and unknown in practice, and there is no historical data to learn from and directly support the predictive model.

In this dissertation, we are dealing with three different predictive demand response modeling approaches, jointly shaping a new methodological pathway. Chapter 1 provides a novel approach for predictive modeling probabilistic customer behavior over new service offers which are much faster than ever done before, based on the case of a large Chinese parcel-delivery service provider. Chapter 2 introduces an approach for predicting scenario-based erection-site demand schedules under uncertainty of disruptive events in construction projects whose logistics transformed from traditional to modular style, based on the case of a USA-based innovative leader in modular building production. For such a leader to advance in its logistics design innovations and associated capacity adjustments, and also to enhance its capability for taking more market share, it is crucial to estimate potential future demand for modular construction and corresponding probable projects in terms of their potential location, size, and characteristics. For this purpose, Chapter 3 introduces a methodological approach for estimating scenario-based future demand for modular construction projects to be implemented over the US metropolitan statistical areas. The correlation among these three chapters' characteristics has been shown in Table 1.

Table 1. Brief introduction on correlation among chapters characteristics

Topic/Question	Chapter 1	Chapter 2	Chapter 3
What type of Innovation or optimization on logistic industry happened and its impact on demand has been studied and modeled?	A transformation in parcel-delivery logistic for offering new faster delivery service levels that never have done before.	A transformation in the construction logistic industry from traditional to modular building construction.	A transformation in the construction logistic industry from traditional to modular building construction.
What data has been used and what are the inputs to the model/ application?	**Dataset:** One-year real waybill and barcode scanning dataset (June 2016 to May 2017) for Shenzhen (a Chinese City) **Application Inputs:** Extracted seasonality on time factors, customer behavior patterns and origin/destination types/rates from Historical data.	**Application Inputs:** 1- Historical weather information for each project's city location (Avg rain/wind, precipitations). 2- Sequence layout of module types (in order) for each building/floor. 3- Experience-based knowledge from a group of experts which were modeled mathematically and embedded in the application.	**Input datasets from the us census:** 1-MSA Population rates (2010-2021) 2- MSA GDP per capita (2010-2021) 3- MSA construction labor costs 4- MSA yearly rates of construction demand with characteristics of area (sqf), value, number of floors/units, number of projects (2017-2021) **Expert estimates for MSA on:** 1- Incentive/Restrictive regulations 2- Access to the materials
What assumption have been made to shape different scenarios of predicted demand?	1-Year on year growth rates for potential orders in intracity/intercity 2- Market share percentage among competitors 3- Overall yearly percentage of intracity customer who are seeking new faster services 4- change or discount on price offering for some service types	1- Number of available cranes in each building/project. 2- Number of working days per week 3- maximum capacity of daily erection rates. 4- Number of working shifts, and working hours per day 5- Planned start/finish dates for each building/project.	1- Predicted values for targeted future time horizon for: • MSA Population rates • MSA GDP per capita • MSA regulations 2- Yearly percentage of the US potential demand which is modular 3- Set of module types and probability dist. for usage of each one on different type of buildings.
What are the outputs of the model/application?	**Scenario-based App (in MATLAB) with user interface, able to make wide range of scenarios, its outputs are:** 1- Forecasted demand logs in csv format for any arbitrary time horizon. 2- Probabilistic patterns (csv format) which are feedable to the logistic simulation for testing performance and sales attributes. 3- Dashboard of results visualization with managerial insights.	**Scenario-based App (in PYTHON-TKinter package) with user interface, able to make wide range of scenarios, its outputs are:** 1- Scenario-based modular demand schedule logs (on multi-erection sites) in csv format. 2- Dashboard of visual resulted graphs on completion time windows, daily probabilistic patterns for required quantity of different module types, and cost/revenue/profit analysis.	**Scenario-based App (in PYTHON-TKinter package) with user interface, able to make scenarios, its outputs are:** 1- Scenario-based Potential/modular demand logs for yearly/monthly aggregated level and daily level (in csv format). 2-Tableau Dashboard of visual results on map and graphs.
What is the contribution of application results?	1- Comparison multi-scenarios demand logs, providing deep managerial insights on demand shape and its geographical distribution in terms of volume and service types which is helpful to be prepared for future risks/challenges with suitable policies. 2- Comparison multi-scenarios with simulated sales after feeding probabilistic patterns leads to: • Understanding of Sale shape and its geographical distribution in terms of volume and service types for each scenario. • Required capacities or lack of resources in different hubs-locations or different hours of day which result to lost sales. • Comparing performance of alternative logistic layouts by monitoring their simulated sales with the level of lost sale and on-time delivery. 3- Testing third party LSP bidders' capability under divers' demand scenarios.	1-Using scenario-based demand schedule logs for integrated assessment of the disruptions impact on consumption rates of different module types at the erection sites, and estimating projected changes on required production rates, storage capacity, and production start/finish dates. 2- Visual graphs on dashboard guide managers and executives through risks caused by disruptions with providing managerial insights to better understand the different decision variables' impact on project/building completion time windows, and probabilistic time windows which different type of modules will be needed in multi-projects. In overall, guiding managers toward appropriate contracting and actionable recourse options and policies.	1- Estimation of potential demand and its characteristics on MSA helps to find the best location for setting the production hubs and storages, in the way to catch more market share with a lower cost of transportation and supply erection sites. 2- Predicted future number of potential projects and their size/value helps for long term resource and capacity assignments planning. 3- Better understanding of where geographically and when we will have demand and we can expect potential projects may pop up, it helps for optimal investment and advertisement/incentive planning and contracting policies.

CHAPTER 1. PREDICTIVE SCENARIO-BASED DEMAND AND CUSTOMER BEHAVIOR MODELING FOR NEW SERVICES IN HYPERCONNECTED URBAN PARCEL LOGISTICS.

Abstract: Rapid demand growth and fierce competition are encouraging logistics service providers towards expanding their competency and capability in terms of offering novel and faster services and reinventing their logistics system so as to profitably and sustainably gain market shares. However, analyzing customer behavior and the underlying causes of demand variability for new services are complex tasks. This chapter is dealing with customer behavior modeling for a service provider who wants to extend its offering system to much faster delivery service than ever done before. To adjust its logistic capacities with future demand, it needs to estimate the volume and geographical distribution of demand for newly offered services. By capturing customers' sensitivities to the delivery-time observed in historical sales data and geo-categorization of orders in different time factors, a scenario-based demand generation methodology and tool is introduced for generating a wide range of demand scenarios with probabilistic patterns for customer behavior over all service offers with dynamic pricing, aimed to feed a simulator which models large-scale urban logistics networks service and offerings. The chapter describes the application of the methodology for the case of a major China-based logistic service provider, to enable testing service capability improvements achievable by leveraging Physical Internet aligned transformation toward hyperconnected urban logistics in a Chinese megacity.

1.1 Introduction

In fast-developing industries with increasing global competition in the introduction of new and innovative services, proper assignment of marketing and operational resources, and best use of the available data to explain the demand and sales dynamics of a new service are crucial in the success and profitability of new offerings. The research leading to this chapter stems from a large-scale collaborative industry-university project (see Campos et al., 2021), notably its work package dealing with demand prediction and customer behavior modeling. The partnering company is a leading Chinese parcel logistic service provider (LSP) that is facing rapid market growth and endeavors to improve efficiency and extend service offers to the less than few hours intra-city delivery services. Such fast delivery is way beyond their current market offers. Therefore, for future demand prediction, it is required to model potential/current customer behaviors and preferences to estimate what portion of current customers may shift to order newly offered faster services, how many potential new customers are to consider this company because of its new offers, and how customers are to react if their promised offers are not satisfactorily fulfilled. In this chapter, we propose approaches to model customer behavior for new urban parcel logistic services. A key target application of these approaches relates to the comprehensive scenario-based probabilistic modeling of demand and customer behavior for parcel logistics, as introduced by Bahrami-Bidoni & Montreuil (2021). Such modeling notably allows to generate demand to feed the urban parcel logistics simulator developed through the project and to test the simulation performance of alternative logistic systems and offerings, and to monitor the level of customer satisfaction for multiple scenarios (Kaboudvand et al., 2021).

The rest of the chapter is structured as follows. Section 2 presents bibliometric analysis results over publications on the area of this study, and also provides a brief literature review. Section 3 focuses on historical demand and behavior data analytics. Section 4 tackles demand and customer behavior modeling for new services. Section 5 introduces a creative methodology inspired by the generalized Bass model to analytics of price-based demand response. Section 6 presents a proposed scenario-based user-interactive application and its uses. Finally, section 7 concludes the paper through a synthesis of the contribution, limitations, and avenues for further research.

1.2 Bibliometric Analysis and Literature Review

To provide an overview of the existing research in customer behavior and demand data analytics, a bibliometric analysis was conducted on 11/20/2022 using the well-established and acknowledged, Web of Science (WoS) database. The used query for WoS is as follows: TS= ((“sales” OR “orders” OR “consumption” OR “demand”) AND “data” AND (“household” OR “industrial” OR “individual” OR “customer” OR “consumer”)

AND (”Analytics” OR “modeling” OR “predicting” OR “demand response” OR “clustering” OR “forecasting” OR “profiling” OR “classification” OR “abnormal” OR “anomaly”) AND (“logistic system” OR “market” OR ”competitor” OR ”company” OR ”parcel delivery”)). Figure 1.1 shows the number of publications and total citations indexed by WoS 1990 to 2022. In total, 2919 publications were found in WoS. Before 2013, the number of publications was at a relatively low level, then it increased rapidly to reach 360 in the year 2021 on WoS.

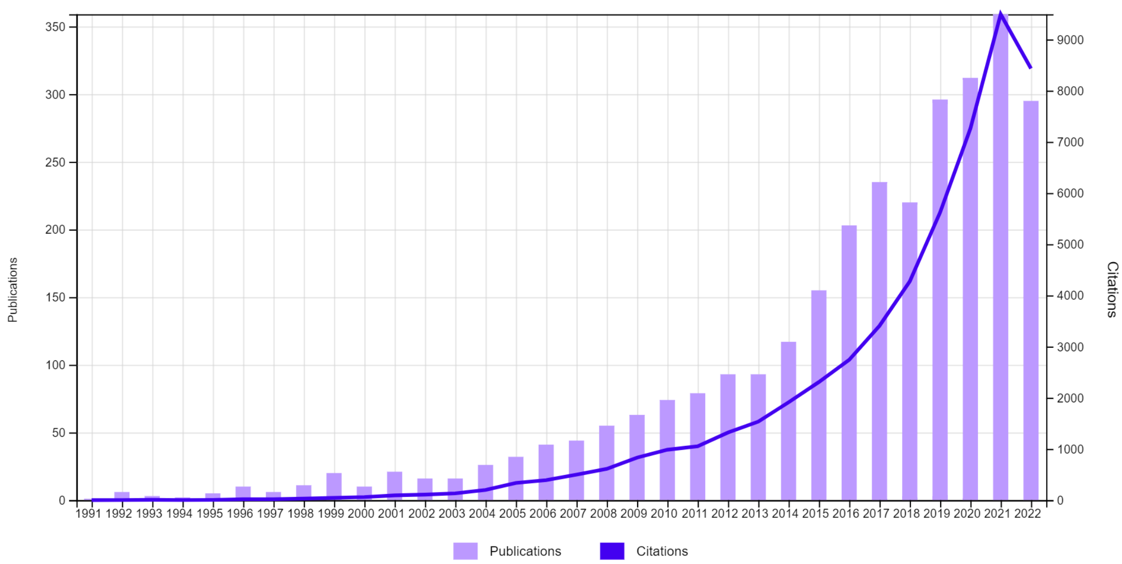


Figure 1.1 Annual scientific publications and total citations on customer behavior and demand data analytics indexed by WOS

Figure 1.2 shows the most productive countries and collaborations in these publications. Figure 1.3 draws the Publications trends of top authors over the time; the size of circles represents number of publications, and the color darkness defines the total earned citations at that related year. In addition, the rest of bibliometric results have been presented in Appendix A. For instance, Figure A.1 shows average article citations per year, which is keep increasing until now, and Figure A.2 is mapping the publications area distributions, showing most to be categorized in the business area. Figure A.3 provides keyword co-occurrences network in which the size of nodes represents frequency of word’s appearance and edges display co-occurrences.

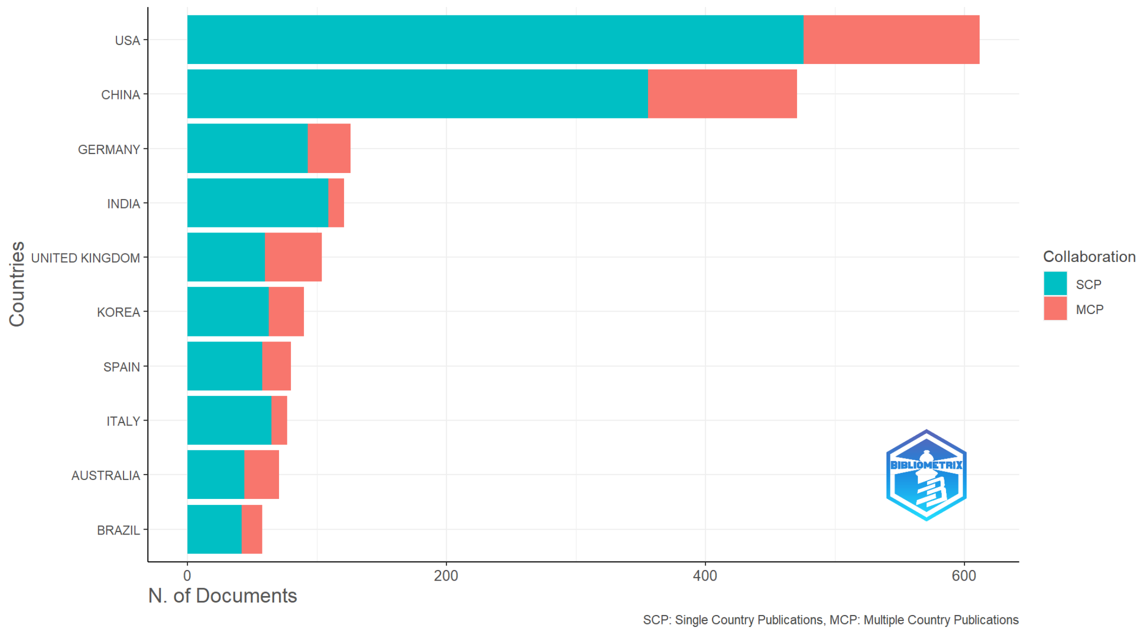


Figure 1.2 Most productive countries and collaborations on publications about customer behavior and demand data analytics

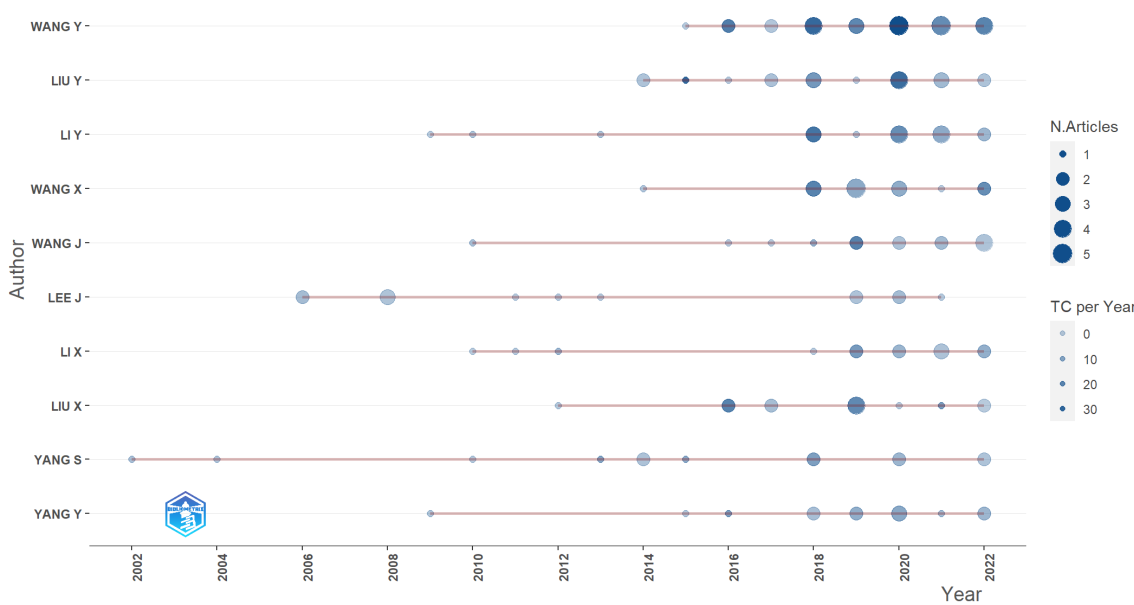


Figure 1.3 Top author's production trends on publications about customer behavior and demand data analytics

Also, Figure A.4 provides the topic dendrogram from extracted literature in WoS. Hierarchical clustering produces a dendrogram, a structure that is a nested sequence of clusters which look like a tree as depicted in Figure A.4. The y-coordinate of the horizontal line is the similarity of the two clusters that were merged, where the objects being clustered are viewed as singleton clusters. Hierarchical clustering uses either a divisive (top-down) or an agglomerative (bottom-up) algorithm which has implications on the way in which the data is separated by the algorithm.

The detailed taxonomy illustrated in Figure 1.4 as a result of this study provides a comprehensive application-oriented overview of the existing literature on demand and customer data analytics. Analytics is known as the scientific process of transforming data into insights for making better decisions, and at a conceptual perspective level, it is commonly divided into four stages (Wang et al., 2018): (1) descriptive data summarization and visualization for exploratory purposes, (2) explanatory diagnostic models that estimate relationships between variables and allow for hypothesis testing, (3) predictive models that enable forecasts of variables of interest and simulation of the effect of marketing control settings, and (4) prescriptive optimization models that are used to determine optimal levels of control variables. Some of the techniques and methodologies adopted or developed to address each application have been summarized in Figure 1.4.

Moreover, Figure 1.4 shows that the size and degree of structure in data increases from right to left, and the feasibility of the higher level of analysis decreases as a function of big data dimensions. It illustrates that the information value of the data grows as its volume, variety, and velocity increase, but that the decision value derived from analytical methods increases at the expense of increased model complexity and computational cost. Wang et al. (2018) and Chicco G. (2016) in their review papers provided a critical examination of analytics methods, with a marketing focus, by tracing their historical development, examining their applications to structured and unstructured data generated within or external to a firm, and reviewing their potential to support marketing decisions.

The application of smart meter data analytics on customer behavior (Wedel & Kannan (2016)) is particularly revealing in regard to our paper due to its emphasis on customer behavior. As summarized in Hong & Fan (2016), there are three ways to modify the workflow to generate probabilistic forecasts: 1) generating multiple input scenarios to feed to a point forecasting model; 2) applying probabilistic forecasting models, and 3) augmenting point outputs to probabilistic outputs by imposing simulated or modeled residuals or making ensembles of point forecasts. The scenario generation method was also used to develop a probabilistic view of power distribution system reliability indices (Black et al. (2018)).

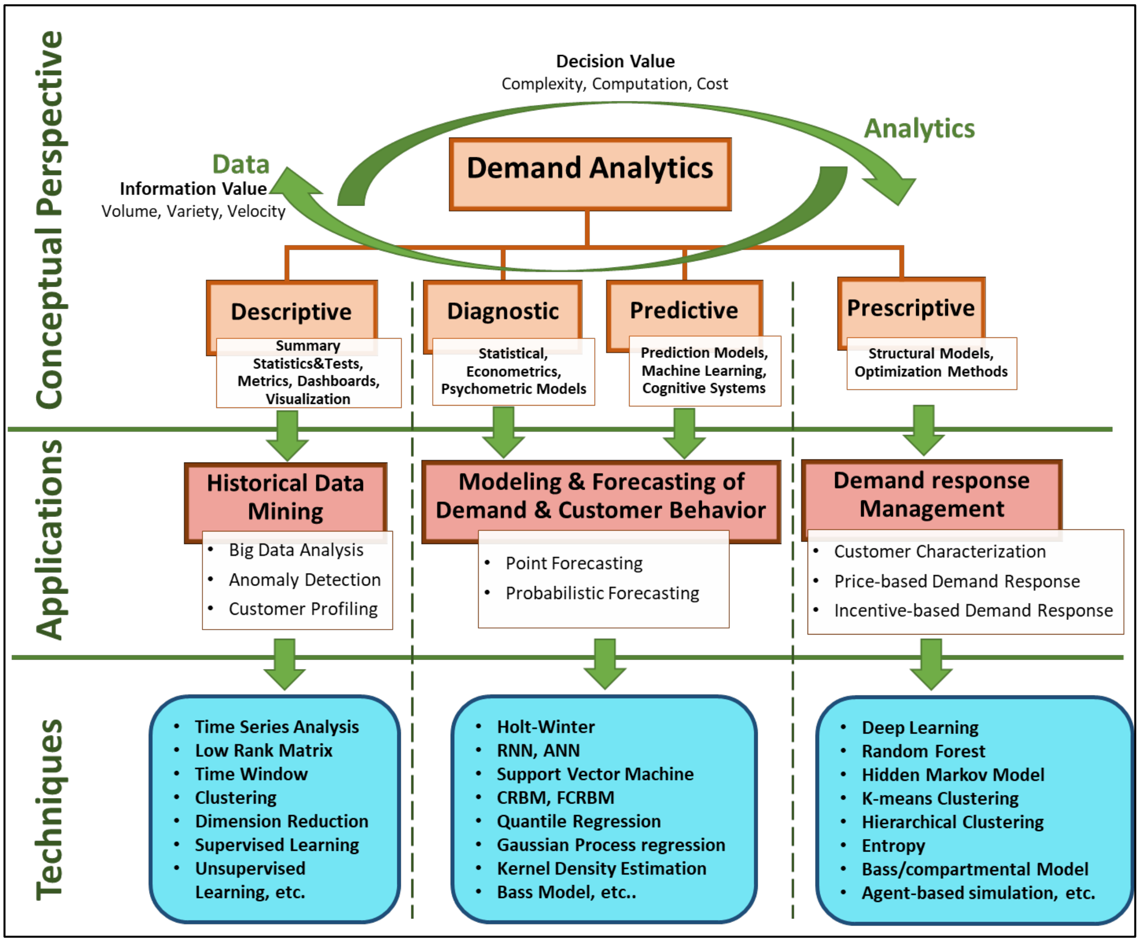


Figure 1.4 Literature review taxonomy for predictive demand and customer data analytics

The Bass model was first to forecast demand for new products which have been widely used in new product forecasting mainly because of its concise conceptual framework, parameter setting and good explanatory ability (Bass, 1969). It was defined using a differential equation in which purchases are initiated by mass communication and further driven by word of mouth (WoM) from past purchases. The Bass model and its various extensions (e.g., Ramírez-Hassan & Montoya-Blandón, 2020) typically explain the customer purchase process at a highly aggregate level of a market which does not allow for capturing customer heterogeneity. Moreover, they assume that all customers can communicate with all the others (perfect mixing), an assumption often violated in observed network of customers. Consequently, their estimates are highly biased and hard to relate to customer behavior (Abedi, 2019). The Bass model and its extensions require accuracy and completeness of sales data to ensure prediction accuracy, and these types of models are also difficult to adapt to the complex and changeable market environment.

Following the Bass model and its variations to model adoption of a new product, a variety of methods have been proposed: agent-based simulation models (e.g., Rand & Rust, 2011; Kiesling et al., 2012); equation-based models, particularly, compartmental modeling (e.g., Rahmandad & Sterman, 2008; Hariharan et al., 2015). Nonlinear differential equation models can easily encompass a wide range of feedback effects, but typically aggregate agents into a relatively small number of states (compartments). For instance, innovation diffusion models may aggregate the population into categories including unaware, aware, in the market, adopters, and so on (Urban et al. 1990, Mahajan et al. 2000). These have in common modeling the adoption pattern at the level of each compartment. Compartmental models add spatial and psychographic segments to Bass-type models, allowing modelers to approximate the heterogeneity, clustering, and communication network of customers, while retaining an analytical structure. These models have been mainly explored in the context of forecasting demand and describing customer behavior, but their suitability as a scalable modeling tool to support large scale marketing and/or operational decision making is not explored.

Abedi (2019) proposed a flexible compartmental model and assessed its suitability in terms of use of data, adherence to micro-level customer behavior, and use in large scale decision making. The paper reveals that the introduced compartmental model results in estimates that are less biased and can predict the shape of the adoption curve significantly better than the Bass model. A hybrid new product demand forecasting model proposed by Yin et al. (2020) combines clustering analysis, using a fuzzy clustering-rough set method, and deep learning, using a neural network model. First, they make a primary prediction based on the classical Bass model, where fuzzy clustering and rough set methods are used to obtain the attributes of the new product through historical sales data analysis of related products. Then, they correct the prediction error by a neural network model to form robust predictive results. Temporal sales patterns are often dynamically influenced by word of mouth, previous experience, and loyalty as well as by marketing/advertising and distribution support (Abedi, 2019). Demand estimates for new products in the existing methodologies mostly rely on adoption patterns through customers' network connections and products' similarities. Until now, existing prediction models for new services have rarely been fully validated in terms of quality and consistency. Yet the need for such models is significant in parcel logistic systems, as making such predictions is part of the critical tasks faced by LSP managers on a regular basis.

To address this problem, we construct three connected models hereafter presented. First is a model that predicts the total number of customers who consider the LSP to order (customers who come to the LSP's business centers or website- and receive offers). This is a scenario-based mathematical model proposed for hourly long-term demand, which consists of a nested combination of three subsections using 2-D representations for modeling specific events and the hourly variation within a week. Second is a model of customer behavior and sensitivities to the delivery time. This model, for any scenario assumption, estimates the probability of selecting a specific offer among all offers, for each type of customers. Third is a model providing probabilistic parameter values of the above two models in different scenarios.

In our reported research experimentation, the first two models are implemented inside the parcel routing simulator. For the third model, we provide an AI-based App with a user interface to compute the value for input parameters of the above two models based on the specific assumed scenario. Then we feed values of these parameters to the comprehensive parcel routing simulation model (in CSV files) and investigated the impact of customer preferences on demand. During simulation runs, this enables us to see how the source and destination of orders is generated in the area and how customers are to select among offers. We also test the performance of a simulated logistic system for that specific scenario. The rest of the chapter presents four stages of analytics in this study.

1.3 Historical data analytics

Predicting demand faces many information challenges. For instance, in the logistics context, estimating potential intracity demand for parcel logistics in an overall megacity market depends on a lot of factors such as population and microeconomic growth factors to name a few. Most challenging is the lack of knowledge about potential customers who approach diverse competitors and those who reject all offers. In most cases, the only direct information one can get comes from the historical sales data of one or a few LSPs and is about those customers who approached and ordered from a given LSP. Rarely are logged rejected offers at sale time, nor the fact that the customer considered several LSPs.

Another challenge is that LSP defines some cut off times after which a service may not be available anymore. As an example, a cut off time for the next-morning delivery is at noon, and therefore after noon no parcel is accepted for the next-morning delivery service. Thus, the historical sales pattern is impacted by cut-offs, as shown in Figure 1.5, where a high volume of sales is depicted prior to the two cut-off times. So, we need to consider this fact and weigh the rest of day hours to balance the demand volume patterns by pushing proportional of that toward hours after cut-offs.

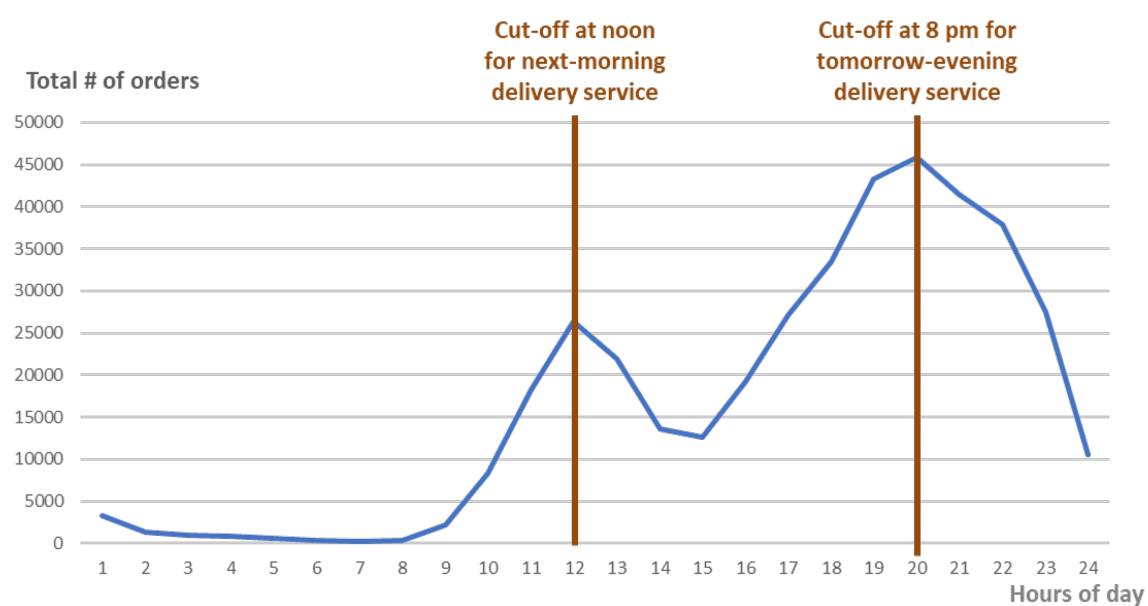


Figure 1.5 An example of cut-off time impacts on sales volume pattern

1.3.1 Big data analysis and anomaly detection

The growth of customers in LSP's market has resulted in the use of big data analytics to understand customers' behavior in predicting the demand for items. It uses a complex process of examining large amounts of data to uncover hidden patterns in the information. It is established on the basis of finding correlation between various parameters that are recorded, understanding purchase patterns and applying statistical measures on collected data. For this purpose, we have leveraged as a bench test the large-scale real-world Chinese megacity database of a LSP collaborating to our research. The database notably includes one-year waybills and Barcode Scanning streaming (in terabytes volume) as well as some multivariate data fusion such as latitude/longitude geographical information about hubs and customers, and socio-demographic information. This has been used for profiling different types of customers and clustering them in terms of their preferences over services and their sensitivities relative to offered delivery time. Cleaning this big data, detecting anomalies/outliers, and estimating null/missed data was a challenging part of this project in the early steps.

1.3.2 Customer profiling

Customers' diversity can be modeled through geo-categorization. For instance, consider a city as a grid of small units such as in Figure 1.6. These units can for example represent zip codes. Based on the type of buildings, facilities or departments, it is possible to categorize customers who are living in each unit. Grid units can be categorized as entirely or partially being business, industrial, residential, etc. Thus, to model customer behavior we categorize potential and loyal customer's preferences based on their order types and geographical- sociodemographic characteristics and probabilistically modeling their offer selection or rejection behaviors as presented in section 1.4.3.

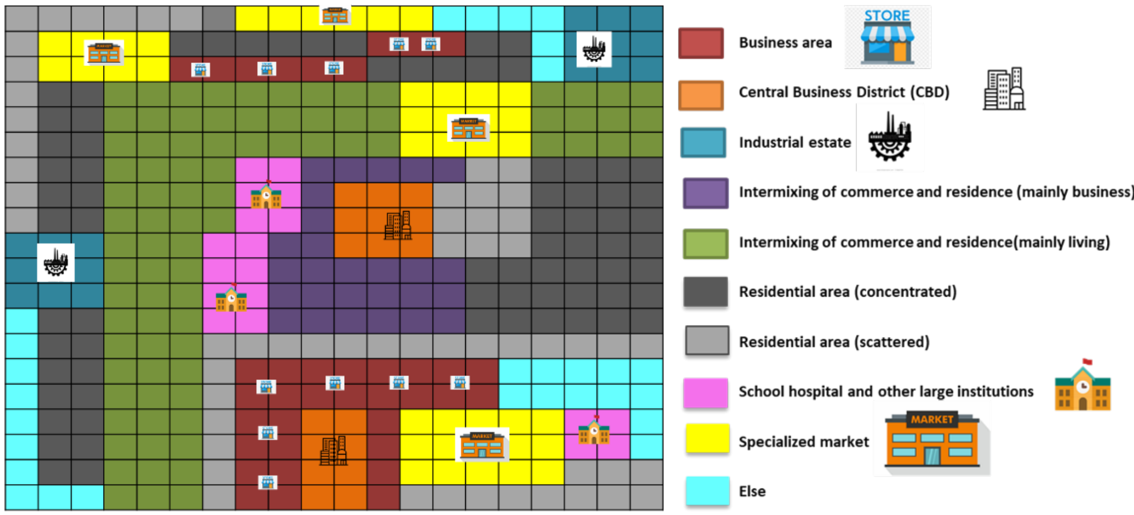


Figure 1.6 City as a grid of units with the corresponding customer categories based on building types

1.4 Modeling demand and customer behavior for new services

1.4.1 General perspective of proposed model

The proposed model for implementing the demand and customer behavior model as well as feeding its outputs to the given simulated logistic system have been summarized in Figure 1.7. In the collaborative research project having introduced in this chapter's contribution, we have used an urban logistics simulator developed using the Anylogic tool to simulate the parcel routing logistics with offering service levels (Kaboudvand et al., 2021). Probabilistic modeling and analytics, based on multivariate data, customer profiles, and historical sales trend over past time intervals, enables generating probabilistic demand patterns, customer preferences, and price-based demand response patterns into the simulated virtual world which are explained in sections 1.4.2, 1.4.3, and 1.5. This sets the stage for the agents operating the logistic system in the simulator.

Demand generation of potential orders requires knowing the feasible service types and also estimating their costs to find the feasible domain area for the offers. As an LSP may have some predefined routing cycles in the logistic simulation and estimates of how long each trip between different logistics hubs during different times of day will take, it becomes possible to estimate for each order feasible piecewise trajectory trips, and the duration of parcel journeys from their source and their destination. For doing this, the offering manager collaborates with the logistic model to get a real-time robust estimate on the delivery time of the parcel from pickup point to the delivery point through the system. Based on this information, the LSP filters offers by removing those that are infeasible with respect to the fastest possible delivery time. After LSPs provide the set of feasible service offers to the customer, he/she/it rejects them all or chooses an offer from one of them, in a way that can be estimated through the probabilistic model of customer behavior computed by models in forthcoming sections 1.4.3 and 1.5. If the customer's preference is not in the list of LSP offers, the customer will choose or not to substitute for the next best fitting alternative feasible offer, which can be modeled based on the cumulative probability function of the corresponding order category.

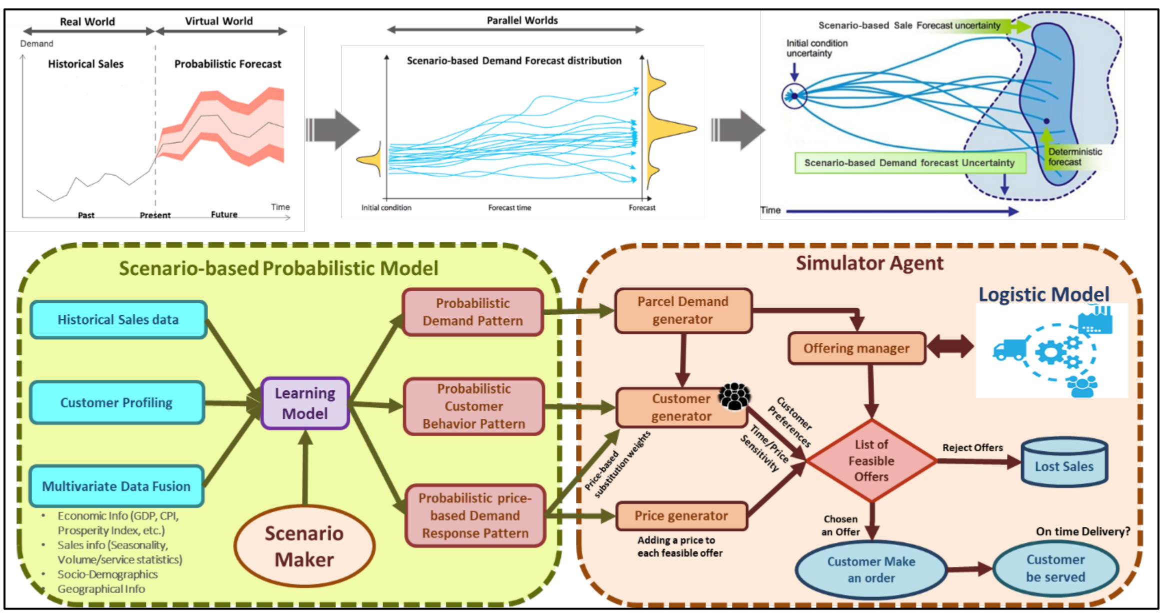


Figure 1.7 Scenario-based predictive modeling of demand and customer behavior

After enough iterations of feeding many demand scenarios and corresponding probabilistic customer behavior, this leads to adequately estimating the distribution of scenario-based forecasts for potential demand, sales, and lost sales. The reference Kaboudvand et al. (2021) provides more details regarding how the simulator parameters are obtained from probabilistic customer behavior models as inputs, how they interact with each other, and how verification and validation of the simulator are conducted.

1.4.2 Modeling probabilistic pattern for scenario-based demand volume and geographical source/destination distribution

To generate the patterns for demand volume, we use a mathematical model inspired by the methodology proposed in Filik et al. (2011) which contains a nested combination of three subsections using weekly residual load variations and two-dimensional representations for modeling specific events and the hourly variations within a week. Data-driven learning and fitting surface functions to this template structure allow getting the base pattern of demand volume with lower/upper bounds on the probabilistic range at the hourly granularity over a given forecast period. We have introduced and explained this proposed approach in more detail in Bahrami-Bidoni & Montreuil (2021).

1.4.3 Modeling probabilistic customer behavior pattern for demanding services

First, a category based on four attributes (*h, s, r, d*) has been considered for each potential order: *h* is the time window of placing order (which hour of a day), *s* and *r* are category types of shipper and receiver of corresponding parcel (type referring to geo-categorization in section 1.3.2), and d is the categorical distance variable between order's source and destination (e.g. $d1 < 10$ km, 10 km $\leq d2 < 40$ km, 40 km $\leq d3$). For instance, a potential order category could be a parcel from a business to a residential area with *d2* distance at 2 PM. Moreover, the proportional demand (pd) weight for faster services (w_{pd}^{c}

) is computed for order category c by $w_{pd}^{c} = w_{TS}^{c} * w_{fd,t_1^{old}}^{c}$, where $w_{fd,t_1^{old}}^{c}$ is the fractional historical demand volume weight for the t_1^{old} in category c among all other categories (Table 1.2), and the delivery-time sensitivity weights (w_{TS}^{c}) obtained by equation (1), (Table 1.1 shows heatmaps of delivery-time sensitivity weights over pairs of source/destination types).

$$w_{TS}^{c} = R^{c}(t_1^{old}) \big/ [R^{c}(t_1^{old}) + R^{c}(t_2^{old})] \tag{1}$$

where $R^{c}(t_1^{old})$ and $R^{c}(t_2^{old})$ are the average demand rates in category c for the first and second fastest delivery-time offered on historical sales data. This network relies on a multi-tier meshing starting at the lowest tier with (1) unit zones, (2) local cells as clusters of unit zones, (3) urban areas as clusters of local cells.

The cumulative probability function over the previously available offering set of delivery times is extracted from historical sales data as a base (e.g., left diagram of Figure 1.8). Then, based on assumptions in given scenario and using the algorithm hereafter described, a new piecewise cumulative probability function will be simulated over all the new sets of delivery time services and finally lead to completing the probabilistic customer behavior model over a new set of service levels for a given scenario (e.g., right diagram of Figure 1.8).

Table 1.1 Heatmaps of delivery-time sensitivity weights

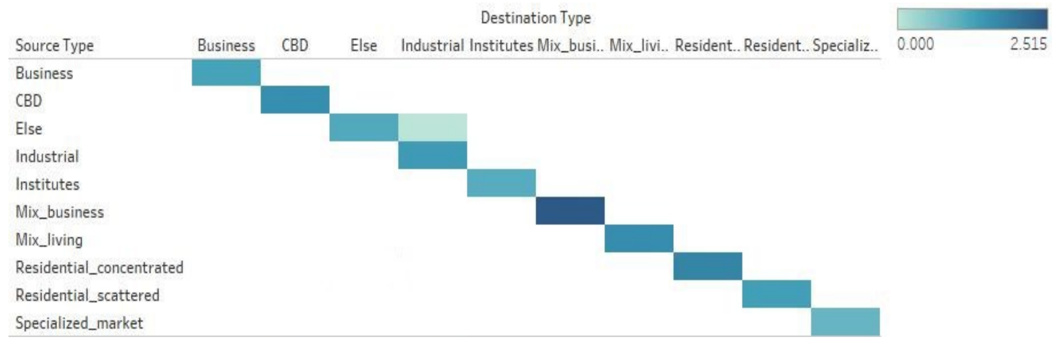
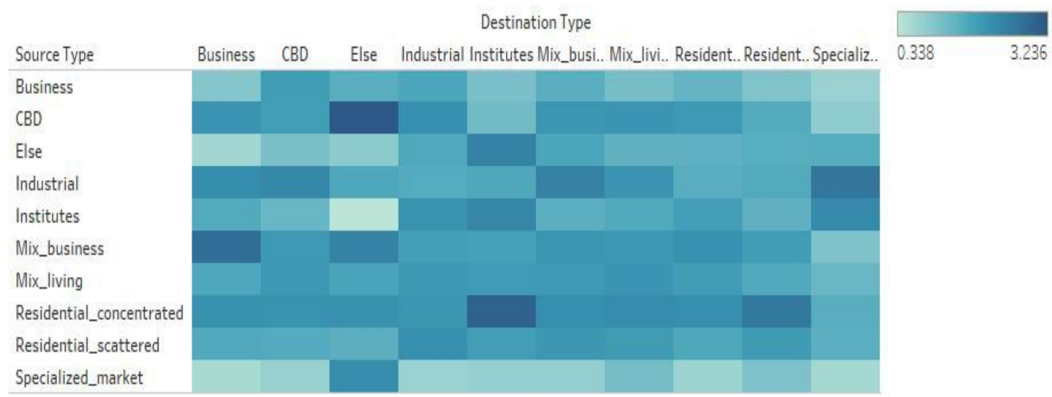
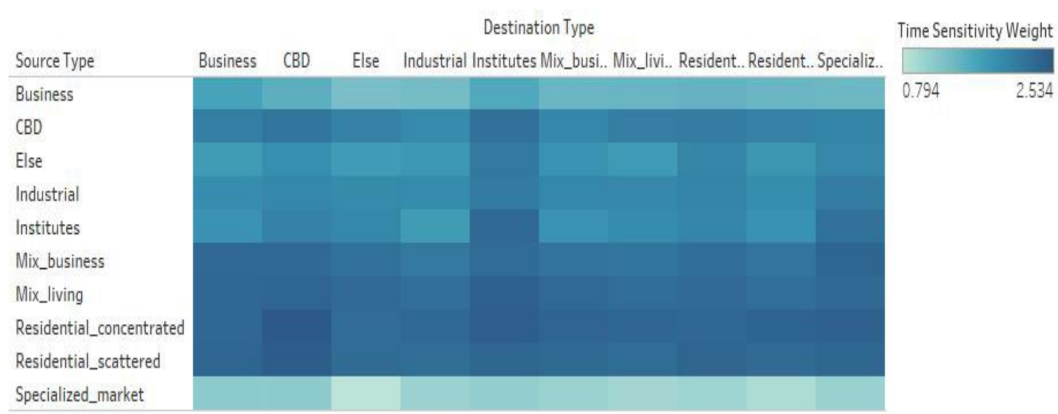
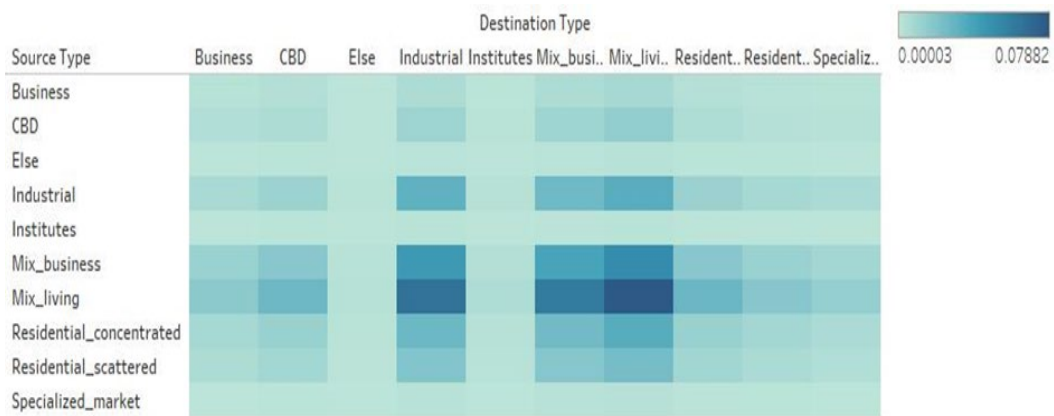
Within Zones	
Within Cell	
Within Area	

Table 1.2 Heatmaps of fractional historical demand volume weight for next morning

Within Zones	
Within Cell	
Within Area	

Algorithm for simulation of customer preference over all potential continuous delivery-time:

Assumptions:

D: Average daily demand forecasted by the first model presented in section 1.5.2.

P: Total percentage of potential customers seeking new faster services (one of scenario assumption).

$O^{Old} = \left\{ t_1^{Old}, t_2^{Old}, ..., t_s^{Old} \right\}$: The set of all promised delivery-times in the old service offering system.

$O^{New} = \left\{ t_1^{New}, t_2^{New}, ..., t_r^{New} \right\}$: The set of all faster delivery-times in the new service offering system.

$V^c = \left(v^c_{t_1^{Old}}, v^c_{t_2^{Old}}, ..., v^c_{t_s^{Old}} \right)$, $\sum_{i=1}^{s} v^c_{t_i^{Old}} = 1$: Probability vector of demand over the set of old offering services.

Goal: Computing continues cumulative probability function of category c for selecting over all delivery-time services.

Step 1: Compute the vectors $K^c = \left(K_0^c, K_1^c, K_2^c, ..., K_s^c \right)$, where $K_0^c = P * D * w_{pd}^c$ and $K_i = \left(1 - P\right) * D * w_{pd}^c * v^c_{t_i^{Old}}$, $i = 1, ..., s$.

Step 2: Normalize K^c to get the vector $K^{c\prime} = \dfrac{1}{\sum_{i=0}^{s} K_i^c} K^c$ where its first component $K_1^{c\prime}$ is the cumulative probability of demand over all new offering services.

Step 3: $L = K_1^{c\prime}$

$$\begin{aligned} & Do \;\; for \;\; i = 1:s \\ & L = L + K_{i+r}^{c'} \\ & \hat{F}(t_i^{Old}) = L \\ & End \end{aligned} \qquad (2)$$

Step 4: Fitting function $F_{p,q}^c(t) = \left(1 - \exp(-(p+q)t)\right) \big/ \left(1 + (q/p)\exp(-(p+q)t)\right)$ inspired by Bass diffusion Model and compute the optimal p^*, q^* coefficients with lower function error to the data sets $\left\{ (t, \hat{F}(t)) \;\; | \, t \in O^{Old} \right\}$.

Return $F_{p^*,q^*}(t)$ As the cumulative probability function of category c for selecting over all delivery-time services. $(t \in T = O^{New} \cup O^{Old})$.

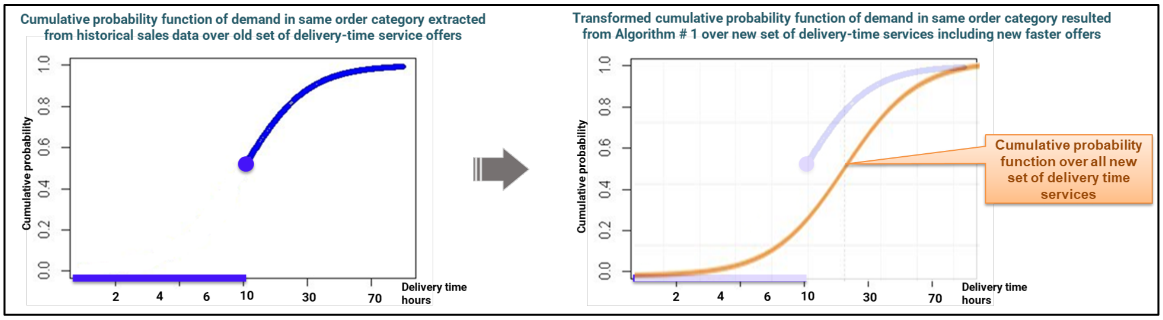


Figure 1.8 The cumulative probability demand estimation for new faster services

In the first two steps of the algorithm, the cumulative probability demand over all new offers is obtained and demand probability over old offer set calibrated based on value of P in scenario assumption and d from section 1.4.2). In stage 3, a rough estimate for F(t) over new delivery-time services is computed and then the optimal value of p and q parameters will be obtained such that the continue Bass model function is fitted to the estimated discrete cumulative probability in corresponding order category.

1.5 Price/Incentive-based probabilistic demand response analytics

Research shows that customer satisfaction, advertisements, word of mouth (WoM), and incentive offers/plans have a positive effect on increasing the potential customers (Chen, 2017). Thus, modeling the customer sensitivities to the price changes helps to construct a dynamic pricing system that optimizes profits while increasing demand market share and keeping high the level of satisfaction. Here, the functional framework of a generalized Bass Model, presented by Ramírez-Hassan & Montoya-Blandón (2020) incorporating market effort into the Bass Model, has been used for modeling demand adoption diffusion rate due to internal/external influences over the delivery-time axis. F(t) and f(t) are the cumulative and non-cumulative proportions of demand at offered delivery-time t, and Y(t) is the total number of potential customers demanding faster services up to but not including offers with delivery time t. The coefficient (p) is the rate of spontaneous demand adoption, and the coefficient (q) is the rate of imitation of demand adoption that the optimal p and q parameters for any order category computed by the proposed algorithm (in section 1.4.3). T is the set of all available offered delivery-time services.

$$\frac{f(t_i)}{1-F(t_i)} = (p+qY(t_i))x(t_i), \quad x(t_i) = 1+\alpha_1 \frac{P(t_i)-P(t_{i-1})}{P(t_{i-1})} + \alpha_2 \frac{\max\{0, A(t_i)-A(t_{i-1})\}}{A(t_{i-1})}, \ t_i \in T \tag{3}$$

Where $x(t_i)$, $P(t_i)$, and $A(t_i)$ are respectively the market effort, the price and the advertising for the i^{th} service offer with t_i promised delivery-time. These variables enter the market effort equation as percentage increases. The sale on offered service with delivery-time t_i will be computed by

$$S(t_i) = F(t_i) - F(t_{i-1}) + e \tag{4}$$

Where *e* is an additive normally distributed error term with variance σ^2 and *F(t)* is given by equation (3).

$$F(t_i) = \frac{1-\exp\{-\bar{X}(t_i)(p+q)\}}{1+(q/p)\exp\{-\bar{X}(t_i)(p+q)\}}, \quad \bar{X}(t_i) = t + \alpha_1 Ln\left(\frac{P(t_i)}{P^{\min}}\right) + \alpha_2 Ln\left(\frac{\tilde{A}(t_i)}{A(0)}\right), \quad t_i \in T \tag{5}$$

Where $\bar{X}(t_i)$ is the cumulative market effort, found by transforming equation (3) into continuous delivery-time and integrating from 0 to t_i, and $\tilde{A}(t_i)$ is the last value for which there was a positive change in advertising ($P^{\min}$ is the minimum price that was offered among available services). Moreover, the time of peak sales defined as the time of the highest diffusion rate *S(t)* can be calculated by using the equation

$$t^* = (\ln q - \ln p)/(p+q) \tag{6}$$

Which means that decreasing/increasing in price, advertisement, and any other market effort on offering a service with promised delivery-time by t^* would have highest impact to increase demand on the corresponding order category.

1.6 Proposed scenario-based user-interactive application

We have conceived and developed a scenario maker application in MATLAB allowing us to set scenario assumptions through an interactive user interface, and to run the model under these assumptions to generate scenario-based probabilistic distributions of customer behavior. All mathematical models explained in sections 1.4 and 1.5 have been embedded in this user-interactive application. This includes the probability distribution of customer preferences on different available offers for all categories and substitution probability for each offer when it is not feasible. Aggregated, this results in scenario-based probabilistic patterns for hourly total demand volume for intracity, inbound, and outbound flows, to be transposed into orders in a simulated logistics system.

Using the scenario-based probabilistic models presented in sections 1.4.2, 1.4.3, and 1.5, and the probability distributions for source/destination of parcels on different time factors, enables to generate future scenario-based forecasted demand logs for the given time horizon which includes every single parcel request, as part of a demand scenario to be simulated. As depicted in Figure 1.7, alternative future logs based on alternative scenarios can be simulated in parallel worlds, and the results from these simulations combined to reveal and analyze logistics outcome and performance distributions. Using the demand and customer behavior models as drivers for the logistics simulator allows performing simulations jointly enabling to compare, beyond logistics costs and environmental impacts, the sales and customer satisfaction outcomes from different scenarios.

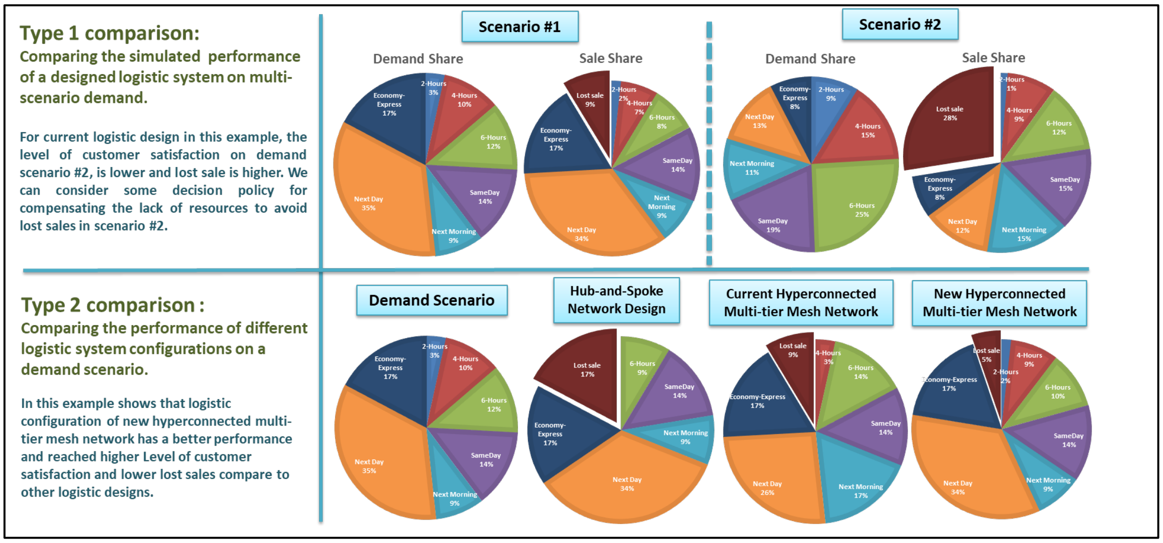


Figure 1.9 Beneficial outcomes of analysis over simulator results by feeding scenario-based probabilistic patterns for demand and customer behavior.

Figure 1.9 provides examples for two types of how feeding probabilistic patterns for demand and customer behavior helps to evaluate logistics system performance in a holistic manner. In type 1 comparison, we compare the simulated performance of a designed logistic system in two different demand scenarios. For each scenario, the left and right pie charts contrast the modeled demand share for each offer and the resulting simulated sales share for each of these offers. Here, the results show that the current logistic model would be less efficient in face of scenario #2 demand pattern and would lead to a lack of resources to answer the demand, inducing lost sales. In type 2 comparison, we assume a demand scenario, and feeding its probabilistic pattern to three designed simulated logistics models enables us to examine their performance in terms of both revenues and costs, and to demonstrate that design #3 is better poised to meet demand.

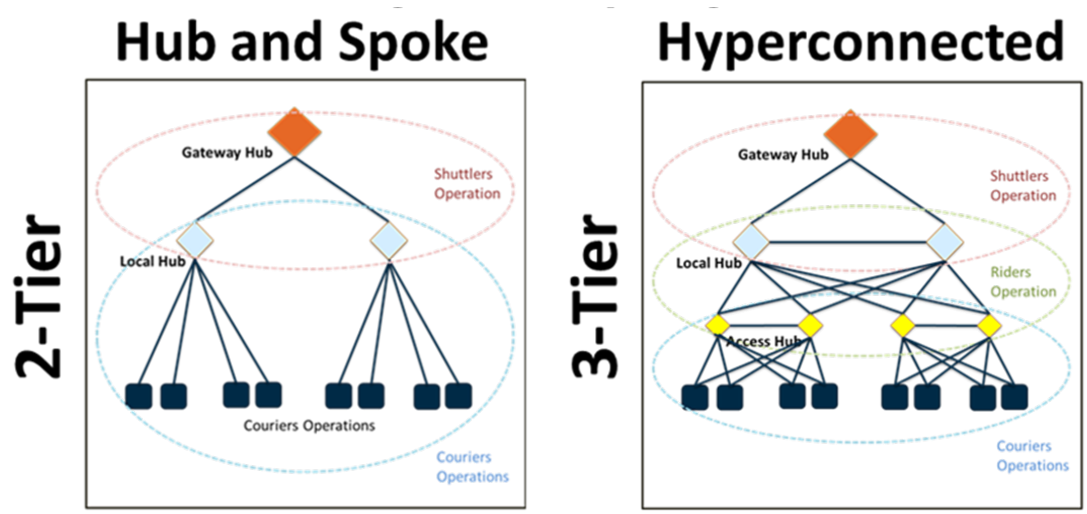


Figure 1.10 Comparison of hyperconnected and hub and spoke logistics topologies (Campos and Montreuil et al., 2021)

Similarly, Figure 1.11 provides examples of how feeding probabilistic patterns for demand and customer behavior to the simulation helps holistically evaluate logistics system performance for two different configuration layouts of logistics called hub and spoke and hyperconnected (Figure 1.10). As a brief introduction about these two layouts, in a traditional hub-and-spoke network each unit zone is served by a single access hub, each local cell is served by a single local hub, and each urban area is connected to a single gateway hub. Moreover, no direct shipment between neighboring hubs at the same tier is allowed and each access hub in 1-tier is connected to single local hub in 2-tier, itself connected to a single gateway hub in 3-tier. Alternatively, in a hyperconnected logistic web, each unit zone can be potentially served by more than one access hub (access hubs are commonly located close to the intersection of unit zones).

Likewise, each local cell can be served by more than one local hub and each urban area can be assigned to more than one gateway hub. Also, lateral flow between hubs at the same tier is allowed in a mesh network way. Parcel shipment is allowed between nearby access hubs in the same local cell and nearby local hubs in the same area, which helps avoid unnecessary travel to higher tiers for shipping between nearby hubs. Furthermore, each hub at a lower tier may be connected to a few nearby hubs at the higher tier as is more convenient given the direction of parcel destinations.

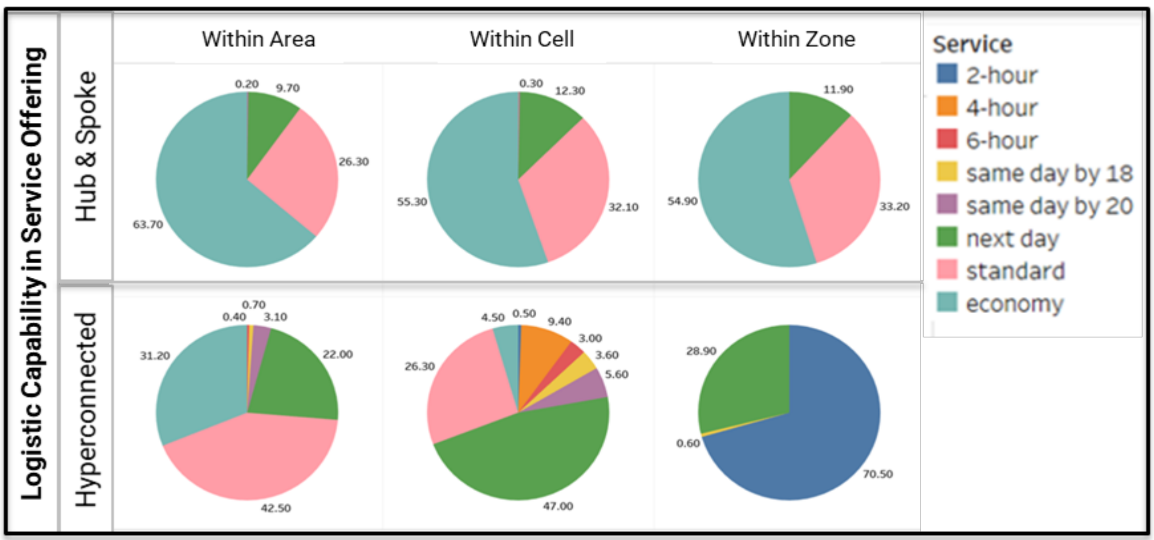


Figure 1.11 Compare simulated sale rates of two logistic layouts using probabilistic patterns for demand and customer behavior

In this experiment, we have a scenario assumption that a total of 15% of customers may seek new same-day faster service offers. You can see in the sales rate that hyperconnected layout was able to perform better and can takes some portion of demand for same-day faster intra-cell and intra-area services. In the simulation, customers choose among available offers which are identified as feasible given the state of the logistic system at the customer demand time and its projected state between pickup and delivery times. Our modeling feeds the simulation with probabilistic customer behavior patterns from which are dynamically computed the probability of choosing among feasible offers for each customer demand. Scenario-based demand logs and the probabilistic patterns which are generated by the proposed methodology and our prototype application are useful to test performance of different logistic configurations.

1.7 Conclusion

The research reported in this chapter is a part of an industry-university project for a large urban parcel logistic system, focused on modeling and forecasting the intracity demand and customer behavior to extend service offers for faster delivery. We have proposed a scenario-based probabilistic modeling approach for demand generation to feed the parcel routing simulation model used for testing the performance of current/alternative logistic systems and monitoring the level of customer satisfaction. Moreover, a scenario-based application with an interactive user-interface has been described to make a wide range of various demand scenarios and to simulate customer preference on new services. This tool is resulting in scenario-based probabilistic patterns for demand and customer behavior which are feedable to the logistic simulator for testing performance and sales attributes as well as providing scenario-based forecasted demand logs for any arbitrary duration. Comparing multiple scenarios by analysis on demand logs is providing deep managerial insights on demand shape and its geographical distribution in terms of volume and service types, which is helpful to be prepared for future risks/challenges with suitable policies. Also, multi-scenario comparison by analysis of simulated sales after feeding probabilistic patterns leads to a better understanding of sales shape and its geographical distribution in terms of volume and service types and a better understanding of required capacities or lack of resources in different locations or different hours of the day which result in lost sales. Moreover, it helps compare logistic performance for different hub/routing designs and the level of on-time delivery and customer satisfaction. Beyond the need for further empirical investigation of the paper's contribution, a key avenue for further research with a high potential for model improvement is to explicitly account for customer satisfaction and its positive/negative impacts on future demand. Customer satisfaction measurement and impact is twofold, first relative to the offers made by the LSP in terms of time and price, and second relative to the actual delivery performance relative to the promised delivery time and price.

1.8 References

Abedi, V. S. (2019): Compartmental diffusion modeling: Describing customer heterogeneity & communication network to support decisions for new product introductions. Physica A: Statistical Mechanics and its Applications, 536, 120964.

Bahrami-Bidoni Z., B. Montreuil (2021): Enabling Scientific Assessment of Large Scale Hyperconnected Urban Parcel Logistics: Scenario-based Demand and Customer Behavior Modeling. Proceedings of IISE 2021 Conference.

Bahrami-Bidoni Z., B. Montreuil, "Predictive Demand Modeling for New Services in Hyperconnected Urban Parcel Logistics", Proceeding of IPIC 2021, 8th International Physical Internet Conference, June 2021.

Bass, F. M. (1969): A new product growth for model consumer durables. Management science, 15(5), 215-227.

Campos M., B. Montreuil, L. McGinnis, S. Kaboudvand, S. Kwon, Z. Bahrami-Bidoni, L. Faugere, S. Buckley (2021): Enabling Scientific Assessment of Large Scale Hyperconnected Urban Parcel Logistics: System Configuration and Assessment. Proceedings of IISE 2021.

Chen, Zoey (2017), "Social Acceptance and Word of Mouth: How the Motive to Belong Leads to Divergent Wom with Strangers and Friends," Journal of Consumer Research, 44 (3), 613-32.

Chicco G. (2016): Customer behavior and data analytics. in 2016 International Conference and Exposition on Electrical and Power Engineering (EPE). IEEE, 2016, pp. 771–779.

Filik, Ü. B., Gerek, Ö. N., Kurban, M. (2011): A novel modeling approach for hourly forecasting of long-term electric energy demand. Energy Conversion and Management, 52(1), 199-211.

Hariharan, V. G., Talukdar, D., Kwon, C. (2015): Optimal targeting of advertisement for new products with multiple consumer segments. Internat. J. of research in marketing, 32(3), 263-271.

Hong T., S. Fan (2016): Probabilistic electric load forecasting: A tutorial review. International Journal of Forecasting, vol. 32, no. 3, pp. 914–938, 2016.

Kaboudvand S., M. Campos, B. Montreuil (2021): Enabling Scientific Assessment of Large Scale Hyperconnected Urban Parcel Logistics: Agent-Based Simulator Design", IISE 2021 Conference.

Kiesling, E., Günther, M., Stummer, C., Wakolbinger, L. M. (2012): Agent-based simulation of innovation diffusion: a review. Central European Journal of Operations Research, 20(2),183-230.

Mahajan, V., E. Muller, Y. Wind. (2000). New-Product Diffusion Models. Kluwer Academic Publishers, Boston

Urban, G., J. Hauser, J. Roberts. (1990). Prelaunch forecasting of new automobiles. Management Sci. 36(4) 401–421.

Rahmandad H., Sterman J. (2008): Heterogeneity & network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. Management Science, 54(5), 998-1014.

Rand W., R. T. Rust (2011): Agent-based modeling in marketing: Guidelines for rigor. International Journal of research in Marketing, 28(3), 181-193.

Ramírez-Hassan, A., Montoya-Blandón, S. (2020). Forecasting from others' experience: Bayesian estimation of the generalized Bass model. International Journal of Forecasting, 36(2), 442-465.

Wang Y., Q. Chen, T. Hong, C. Kang (2018): Review of smart meter data analytics: Applications, methodologies, and challenges," IEEE Transactions on Smart Grid, v10, no3, 3125–3148.

Wedel M., P. Kannan (2016): Marketing analytics for data-rich environments," Journal of Marketing, vol. 80, no. 6, pp. 97–121, 2016.

Yin, P., Dou, G., Lin, X., & Liu, L. (2020): A hybrid method for forecasting new product sales based on fuzzy clustering and deep learning. Kybernetes.

CHAPTER 2. RISK-SENSITIVE SCENARIO-BASED PREDICTION OF ON-SITE DEMAND AND COMPLETION TIMES IN DISTRIBUTED MODULAR CONSTRUCTION PROJECTS

Abstract: Implementation of large-scale construction projects is risky as many disruptive events may impact planning schedules. A systematic approach for assessing the associated risks of such projects is a critical success factor. In collaboration with an industrial partner, we have focused on modular construction projects with cuboid modules built in the factory and then erected on site. We have developed a predictive scenario-based on-site demand model for modules to be erected and for project completion times. We use scenario-based estimation for projected demand volume of each type of modules and the probabilistic time windows during which they are needed for buildings on active sites in a multi-project context. We consider disruptive events impacting module demand and erection completion times, such as rain, wind, crane malfunctions, module shortage, reduced labor/equipment availability, delays with foundations or with getting legal permits, non-respect of contractual agreements, and discrepancies between plant and actual production.

These risks and disruptive events are modeled mathematically using statistical modeling, then we generate a wide range of most probable scenarios from daily probabilistic distributions of risks, and finally, by mining over these scenarios we come up with analytical results guiding managerial and executive decisions. As tested in several real large-scale modular construction projects, the proposed approach and the developed modeling and technology supporting it have proven capable of enabling managers and executives to better understand the different variables' impact on project

and building completion time windows, and module demand patterns, so as to guide them through risks and disruptions toward appropriate contracting and actionable recourse options and policies.

2.1 Introduction

The construction sector is currently undergoing a shift from onsite stick-built construction to distributed modular construction that takes advantage of modern prefabrication techniques. As a result of this transformation, current construction supply chains, which have focused on the delivery of raw materials to sites, are no longer apt and need to make way for new, strengthened, and time-critical logistics systems (Hsu et al., 2018). The benefits of modular construction come from manufacturing building components in a factory environment, where higher efficiencies and quality can be achieved, the need for space to store materials or environment on erection sites is reduced, and the assembly process is significantly shortened (Rogers & Bottaci (1997), Lawson et al. (2014)).

Previous studies have indicated that erection site delays constitute the largest cause of demand schedule deviations. Gunduz et al. (2013) listed 83 distinct factors causing delays in building projects with over 90% traceable to activities within construction sites. When delays occur, the actual progress of the project lags behind the original schedules. Material demand will decrease, and project duration will have to be extended, incurring additional costs (Sweis et al., 2008). Because delays in construction schedules are almost inevitable (Sambasivan & Soon, 2007), and changes in the demand often have a severe impact on upstream logistics, their effect must be carefully taken into consideration

(Assaf & Al-Hejji (2006), Hsu et al. (2017)). Zou et al. (2007) proposed a technique for predicting demand uncertainties on the construction site by using a probabilistic analysis on historical data.

The chapter introduces a mathematical model for scenario-based estimation of projected demand volume of each type of modules and the probabilistic time windows which they are needed on different erection sites and multi-projects considering disruptive events. The research underpinning this chapter's contribution has been realized as part of a distributed modular-construction research project in collaboration with an industry partner.

The chapter is structured as follows. Next, section 2.2 presents a bibliometric analysis on the topic of this chapter, then proceeds with a focused review of the literature. Section 2.3 introduces the methodology including assumptions and model formulation. Section 2.4 covers model implementation by developing a user-interactive application. Finally, Section 2.5 provides conclusive remarks and avenues for further research.

2.2 Bibliometric Analysis and Literature Review

To provide an overview of the existing research in risk analytics in modular construction, a bibliometric analysis was conducted on 11/21/2022 using the well-established and acknowledged databases, Web of Science (WoS). The query for WoS is as follows: TS= ((risk OR hazard OR barrier OR uncertainty OR uncertainties OR delay) AND ('disruption' OR 'Disruptive' OR 'Delay') AND ('offsite construction' OR 'off-site construction' OR 'offsite production' OR 'off-site production' OR 'offsite manufacturing' OR 'off-site manufacturing' OR prefabrication OR prefabricated OR prefab OR pre-fab OR pre-fabricated OR 'industrialized building system' OR 'modular construction' OR modular OR 'off-site fabrication' OR modularization OR 'prefabricated refinished volumetric construction' OR 'modern method off construction' OR 'industrialized construction')). Figure 2.1 shows the number of publications and total citations indexed by WoS 1977 to 2022. In total, 2,576 publications were found in WoS. Before 2012, the number of publications was at a relatively low level, while it increased rapidly beginning in 2013 and reached 200 in the year 2020 on WoS.

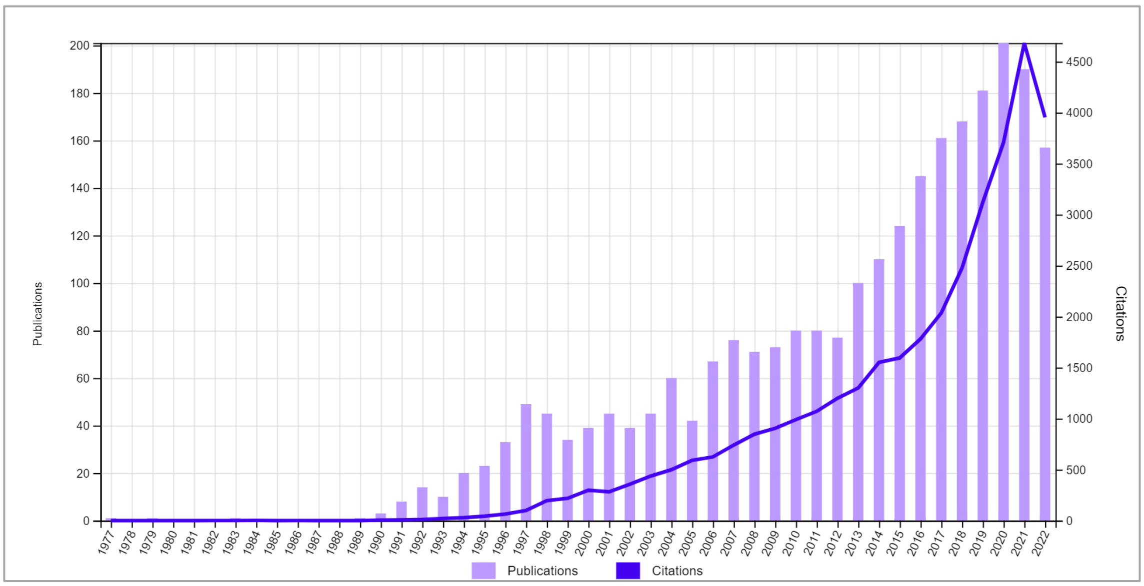


Figure 2.1 Annual scientific publication and total citation on risk analytics in modular construction indexed by WOS

Figure 2.2 shows the most productive countries and collaborations in these publications. Figure 2.3 draws the publications trends of top authors over the time; the size of circles represents the number of publications, and the color darkness defines the total earned citations in a specified year.

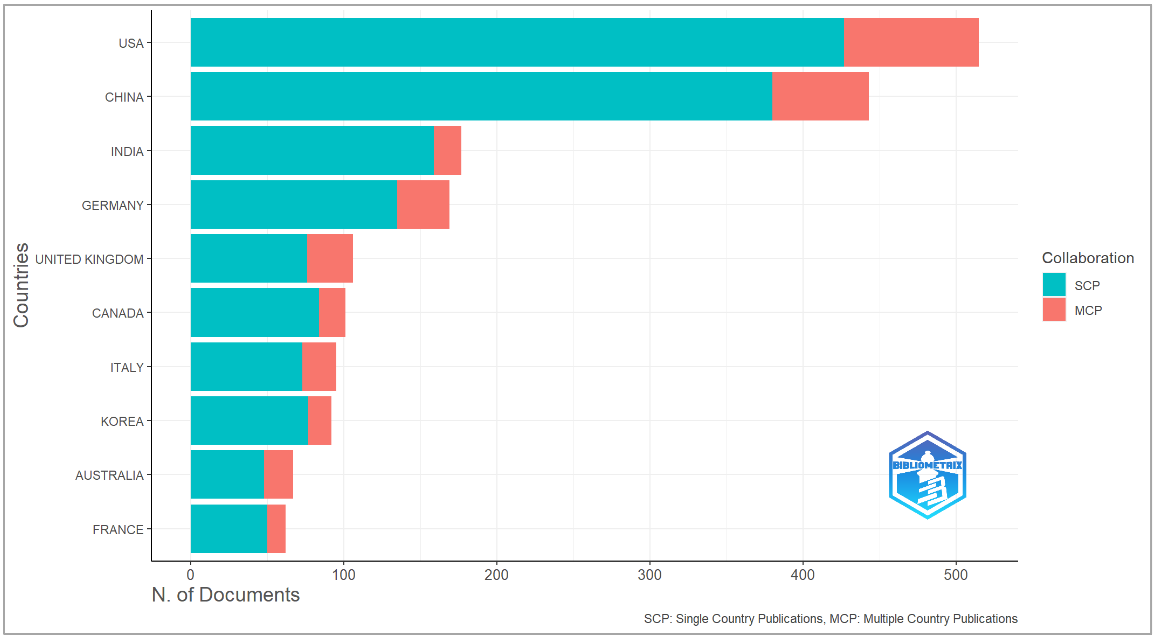


Figure 2.2 Most productive countries and collaborations on publications about risk analytics in modular construction

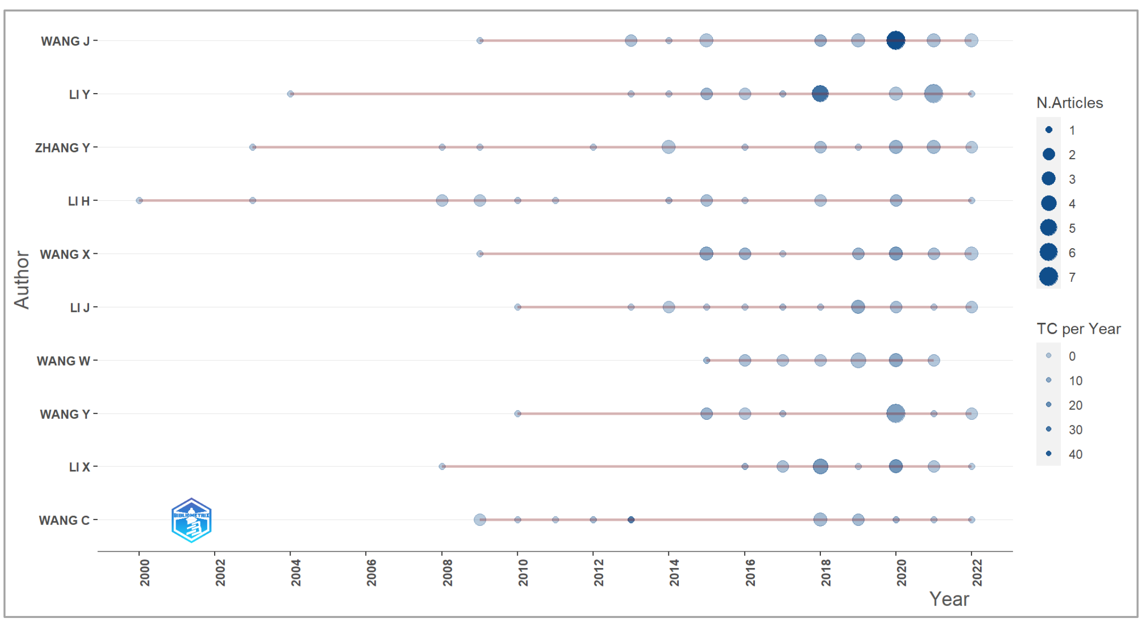


Figure 2.3 Top author's production trends over the time on publications about risk analytics in modular construction

In addition, the rest of bibliometric results have been presented in Appendix B. For instance, Figure B.1 depicts the yearly-increasing average number of article citations. Figure B.2 maps the distribution of publications areas, revealing that most of the publications are categorized as being in the engineering and computer science areas. Figure B.3 provides keyword co-occurrences network in which the size of nodes represents frequency of word's appearance and edges display co-occurrences.

Existing literature provides useful evidence on the presence of risk factors in various construction projects. Though some risk factors are mutual in several construction projects, other risk factors are different for every construction project due to different requirements such as resources, technologies and management skills (Luo et al. 2019). Addressing the risk in the large-scale projects requires systematic modelling tools and the visualization could help decision-makers to focus on the highlighted risk factors in a project portfolio. Kimiagari and Keivanpour (2019) provided an interactive risk visualization tool for large scale and complex engineering and construction projects under uncertainty and interdependence. They have used joint application of fuzzy group decision making, analytic network process and mapping the resulting network of dependencies together with proximity information, graph theory, and mutual information theory.

Projects involving a modular construction method have specific risks. Modular construction requires a significant amount of initial investment which has been highlighted in several studies (Kamali & Hewage 2016; Blismas & Wakefield 2009; Pan et al. 2007; Sun et al. 2020). Other important risk factors affecting the implementation of modular construction are transportation constraints (Wuni & Shen 2020; Jaillon & Poon

2009; Pan et al. 2007; Sun et al. 2020), change order due to defective design (Li et al. 2016 ; Hassim et al. 2009; Li et al. 2013), complexity of modular building design (Hassim et al. 2009; Wu et al. 2019), unskilled labor (Rahman 2014; Zhang et al. 2014) and technology incompetence (Jaillon et al. 2008; Jaillon & Poon 2009; Luo et al. 2015). Pervez et al. 2022 had done a comprehensive literature review on critical risk factors in the implementation of modular construction and proposed a risk assessment framework for identification, evaluation and prioritization of critical risk factors affecting the implementation of modular construction in Pakistan.

Intelligent systems with the ability to detect, predict, and make decisions are very useful in the field of management science (Macdonald et al. 2010). In the risk management area, due to dealing with unforeseen events, the necessity of such systems is undeniable. Moradkhani et al. (2021) and Feldkamp et al. (2021) leverage the Physics-of-Decision (POD) intelligent risk management framework introduced in (Benaben et al. 2021). This original framework considers that risks can be seen as physical forces applied to the system which may push or pull it in its performance space by varying the system's KPIs (Key Performance Indicators) (Bénaben et al. 2020).

Recently, Hsu et al. (2018) have developed a two-stage stochastic programming model to capture possible demands on the erection site. They set of possible disruptive events and for each several levels of severity such as normal, low, medium, and high, then assigned a discounted daily rate of module assembly to any single of these individual risk levels, and then through a few steps, their model considered minimum discounted performance rate for combinational permutations of risk statuses and aggregated the probability of the same performance rates to come up with a probability distribution of possible erection rates.

The contribution of Hsu et al. (2018) is pertinent and significant yet also makes clear facets prone to be improved. First, they assign a discounted daily erection rate to any single risk level which does not account for the fact that performance for erection rate is the result of all sets of risk states. Second, they aggregate performances by considering the minimum performance rate among all risk statuses, ignoring dominance relationships between risks and between risk levels. Addressing the observed strengths and limitations of previous models, the model proposed in this chapter aggregates severity levels of n-dimensional risks considering the weight of dominancy between risks, and then computes a discount factor for performance by a defined function based on the uncertainty of risk statuses and combinational severity level.

2.3 Methodology

The model we introduce is simultaneously considering demand variations on erection sites incurred by delay factors such as weather, crane malfunctions, module shortage, reduced resources (crew/labor/equipment), delays with foundations or with getting legal permits, non-respect of contractual agreements, and plan vs, production discrepancies.

Since weather forecasts most commonly only provide the probability of precipitation (PoP) on a certain day, the probabilities for the amount of rainfall to exceed the cancellation point need to be further calculated. Wilks (1995) & Applequist et al. (2002) have established that the special gamma distribution, known as the exponential distribution, provides a reasonable approximation for the frequency distribution of rainfall amounts. An estimator is defined using the exponential distribution probability density function to approximate the unconditional probability of exceeding (uPoE) a selected rainfall amount (x) on a certain day (Hsu et al., 2018):

$$uPoE(x) = PoP \times e^{(-x/\mu)} \tag{7}$$

where μ is the predicted average rainfall amount on that day. Weather forecasts usually provide average wind speed estimates. Consulting these predictions, we can calculate the probability for wind speeds to exceed the turning and cancellation points by employing the Weibull distribution (Morgan et al., 2011; Conradsen et al., 1984; Azad et al., 2014). The latter can be used to determine how often winds of different speeds will be encountered at a location, with a probability density function defined as follows (Hsu et al., 2018):

$$p(v) = \left(\frac{k}{c}\right)\left(\frac{v}{c}\right)^{k-1} \exp\left[-\left(\frac{v}{c}\right)^{k}\right] \tag{8}$$

where *p(v)* refers to the probability for the selected wind speed *v* to happen on a certain day, *c* is the Weibull scale parameter, with a unit equaling that of the wind speed (km/h), and *k* is the unit-less Weibull shape parameter. The cumulative distribution function is written as:

$$P_c(v) = 1 - \exp\left[-\left(\frac{v}{c}\right)^{k}\right] \tag{9}$$

The shape factor k and scale factor c can be calculated by using the moment method, a common technique widely used in the field of parameter estimation (Kumar & Gaddada, 2015). The two parameters can be evaluated by the following equations:

$$k = \left(\frac{0.9874}{\sigma / U} \right)^{1.098} \tag{10}$$

$$c = \frac{U}{\Gamma(1 + \frac{1}{k})} \tag{11}$$

where σ is the standard deviation of the wind speed on a given day in the area of interest, U is the average wind speed on that day given by the weather forecast, and Γ represents the Gamma function.

2.3.1 Performance discount factor due to k-dimensional risk analysis

For model formulation of k-dimensional risk analysis in erection site, we assume $R = \{r_1, ..., r_k\}$ is the set of potential risks corresponding *k* type of disruption events and $L_i = \{l_{i1}, ..., l_{im_i}\}$, $i = 1, ..., k$ is the set of defined potential m_i levels of severity for *i'* th risk and $n = \sum_{i=1}^{k} m_i$ is the total number of possible severity levels on all k risks.

As shown in Table 2.1, these severity levels can be such as normal, low, Medium, etc. Moreover, $P_i = \{P_{l_{i1}}, ..., P_{l_{im_i}}\}$, $\sum_{j=1}^{m_i} P_{l_{ij}} = 1$, $i = 1, ..., k$, is the probability distribution of happening of *i*'th risk on its severity levels, where $P_{l_{ij}}$ is the probability of happening *i*'th risk on its *j*'th severity level.

Table 2.1 Severity levels and distribution probabilities for each type of risk

Risk	Severity level	Severity score	Probability
r_1	Normal	0	0.98
	low	0.3	0.01
	medium	0.7	0.005
	high	1	0.005
…	….	….	….
r_k	Normal, low	0	0.95
	medium	0.5	0.01
	high	0.8	0.02
	Very high	1	0.02

To consider dominance relationship among risks on their different severity levels, we rely on a dominance weight matrix (A), leveraging the knowledge of experts to estimate a_{ij} , the dominancy score of risk level i relative to risk level j in range of 1 to 5. It is 1 when they both have the same impact on performance and 5 when i has highest dominance over j in terms of impact on overall performance. This matrix is symmetric with $\forall_{i,j} \quad a_{ij} = \dfrac{1}{a_{ji}}$. Matrix A is formulated as below:

$$A = \left[a_{ij}\right]_{n\times n} = \begin{array}{c} \\ l_{11} \\ \vdots \\ l_{1m_1} \\ \vdots \\ l_{k1} \\ \vdots \\ l_{km_k} \end{array} \begin{array}{c} \quad l_{11} \;\cdots\; l_{1m_1} \;\cdots\; l_{k1} \;\cdots\; l_{km_k} \\ \begin{bmatrix} a_{11} & \cdots & \cdots & & \cdots & \cdots & a_{1n} \\ \vdots & \ddots & & \cdots & & & \vdots \\ \vdots & & \ddots & & & & \vdots \\ & \vdots & & \ddots & & \vdots & \\ \vdots & & & & \ddots & & \vdots \\ \vdots & & & \cdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & & \cdots & \cdots & a_{nn} \end{bmatrix} \end{array} \tag{12}$$

For computing the importance weight of risk dimensions when we have an occurrence of severity levels for all k risks $\rho = (l^{r_1}, l^{r_2}, \dots, l^{r_k})$, we assume that w_i^ρ is the importance of risk i in occurrence of ρ, and $W^\rho = (w_1^\rho, w_2^\rho, \dots, w_k^\rho)$, $\sum_{i=1}^{k} w_i^\rho = 1$. Also $A_{k\times k}^\rho \subset A_{n\times n}$ is a submatrix of A corresponding to a specific levels of risk severity which occur on k risks' dimensions. Any component of this new submatrix of $A_{k\times k}^\rho = \left[a_{ij}^\rho\right]_{k\times k}$ representing the expert beliefs for dominance comparison between occurred severity levels of risk i and risk j, so this value can give approximate importance weight of risk i over risk j, as $a_{ij}^\rho \sim \frac{w_i^\rho}{w_j^\rho}$. Since A^ρ is a symmetric matrix, we should have a real eigenvector for that which can be an estimation for our importance weight vector of risk severity levels. Below equation proves that if $a_{ij}^\rho \sim \frac{w_i^\rho}{w_j^\rho}$, then W^ρ would be an eigenvector of A^ρ.

$$A^\rho \times W^\rho \approx \begin{array}{c} r_1 \\ \vdots \\ r_k \end{array} \begin{bmatrix} \frac{w_1^\rho}{w_1^\rho} & \cdots & \frac{w_1^\rho}{w_k^\rho} \\ \vdots & \ddots & \vdots \\ \frac{w_k^\rho}{w_1^\rho} & \cdots & \frac{w_k^\rho}{w_k^\rho} \end{bmatrix} \times \begin{bmatrix} w_1^\rho \\ \vdots \\ w_k^\rho \end{bmatrix} = kW^\rho \tag{13}$$

To analyze the impact of all k types of risks on performance, we proposed a novel method to compute a combinational severity level which we call Multi-Dimensional Risk Convergence score (MDRC). In this method we assume a k-dimensional space of risk severity levels, and two reference points which are the least desired point and the most desired point (as shown in Figure 2.4).

Every occurrence $\rho = (l^{r_1}, l^{r_2}, ..., l^{r_k})$, $\forall i \quad 0 \le l^{r_i} \le 1$ is a k-dimensional point in this space. The most desired point is a point with value zero on all risk severity, and the least desired point has value one on all dimensions. For a given occurrence, S^- is defined as Weighted Euclidean distance to the most desired point and S^+ is Weighted Euclidean distance to the least desired point. If at least one of the severity levels is 1, the multi-dimensional risk Convergence score S_ρ^{MDRC} is also equal to 1, but if all of them are less than 1, we can use the following equations (14 and 15) to compute the convergence score of multi-dimensional risk severity.

$$S_\rho^{MDRC} = \begin{cases} 1 & if \ \exists_{i \in \{1,...,k\}} \ l^{r_i} = 1 \\ \dfrac{S^-}{S^- + S^+} & if \ \forall_{i \in \{1,...,k\}} \ l^{r_i} < 1 \end{cases} \tag{14}$$

$$S^- = \sqrt{\sum_{i=1}^{k} w_i^\rho (l^{r_i})^2} \quad , \quad S^+ = \sqrt{\sum_{i=1}^{k} w_i^\rho (1 - l^{r_i})^2} \quad , \sum_{i=1}^{k} w_i^\rho = 1 \tag{15}$$

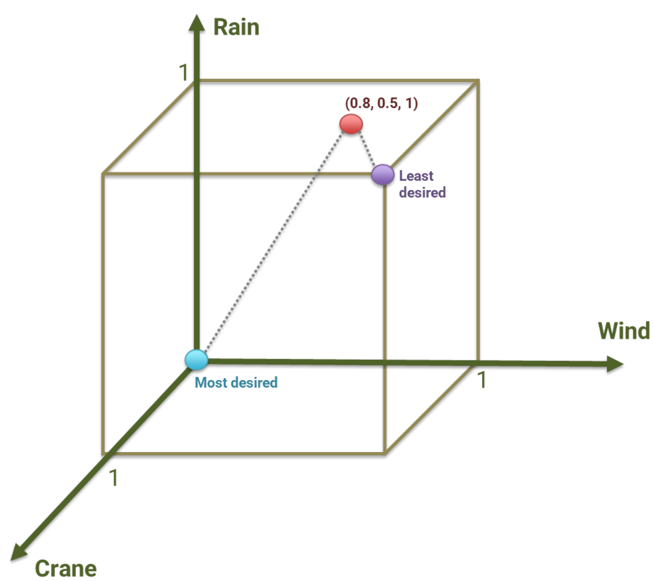


Figure 2.4 Three-dimensional risk space and distance to most/least desired points

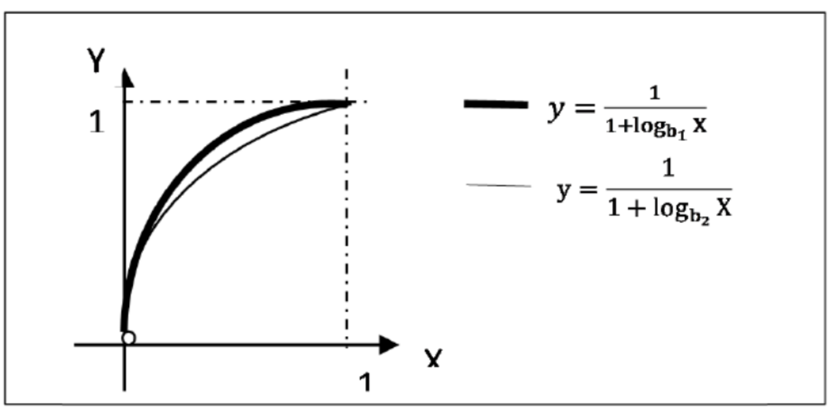
In order to estimate erection rate for specific occurrence of $\rho = (l^{r_1}, l^{r_2}, \dots, l^{r_k})$, the estimated discount factor should be multiplied by the max capacity of daily erection. The following logarithmic function provides a logical estimation of the discount factor on performance.

Estimated rate of erection = (Discount Factor) * (Max capacity of daily erection)

$$DF = \frac{1}{1 + \log_{(1-C_\rho)}^{(1-S_\rho^{MDRC})}} \tag{16}$$

Where C_ρ is uncertainty level of occurrence ρ that is equal to entropy of probabilities on the occurrence risk levels and can be computed as below:

$$C_\rho = \frac{1}{\ln(k)} \sum_{i=1}^{k} P_{l^{r_i}} * \log(P_{l^{r_i}}) \tag{17}$$



The graph of $Y=\frac{1}{1+\log_b X}$, *when* $0 < b_1 < b_2 < 1$

Figure 2.5 Logarithmic function used for discount factor, illustrated with two base values

2.3.2 *Generation of daily scenarios of erection rate and completion time windows*

Using the methodology presented in section 2.3.1, we compute the daily erection rate for any possible combinational permutation of statuses on k risks $\rho = (l^{r_1}, l^{r_2}, \ldots, l^{r_k})$. The distribution probability of occurrence $\rho \sim f(\rho) = \prod_{i=1}^{k} P_{il^{r_i}}$ on a day, and this distribution is variate on different dates because for instance the probabilistic pattern of weather status such as rain, snow, storm would vary over different dates. Thus, for any specific date we have a distribution probability and estimated erection rate over all possible permutations of k-dimensional risk status. As shown in Table 2.2, for each date, we come up with a different distribution probability of occurrence over any combinatorial status of multi-dimensional risks.

Table 2.2 A sample for computed daily distribution probability of occurrence for any combinational status of multi-dimensional risks

	Rain	Wind	Crane	Labor	ModAvailibility	SiteNotReady	Discrepancy	ModErectionRate	8/1/22	8/2/22	8/3/22	8/4/22	8/5/22	8/6/22	8/7/22	8/8/22	….	1/31/23	2/1/23	2/2/23	2/3/23	2/4/23	2/5/23	2/6/23	2/7/23
0	Normal	Normal or low	Normal	Normal	Normal	Normal	Normal	8	0.29467114	0.442007	0.498135	0.406927	0.357815	0.245559	0.301687	0.245559	...	0.491119	0.371847	0.259591	0.378863	0.364831	0.470071	0.589342	0.512167
1	Normal	Normal or low	Normal	Normal	Normal	Normal	Low	7	0.009208473	0.013813	0.015567	0.012716	0.011182	0.007674	0.009428	0.007674	...	0.015347	0.01162	0.008112	0.011839	0.011401	0.01469	0.018417	0.016005
2	Normal	Normal or low	Normal	Normal	Normal	Normal	high	0	0.003069491	0.004604	0.005189	0.004239	0.003727	0.002558	0.003143	0.002558	...	0.005116	0.003873	0.002704	0.003946	0.0038	0.004897	0.006139	0.005335
3	Normal	Normal or low	Normal	Normal	Normal	Low	Normal	0	0.015673997	0.023511	0.026497	0.021645	0.019033	0.013062	0.016047	0.013062	...	0.026123	0.019779	0.013808	0.020152	0.019406	0.025004	0.031348	0.027243
4	Normal	Normal or low	Normal	Normal	Normal	Low	Low	0	0.000489812	0.000735	0.000828	0.000676	0.000595	0.000408	0.000501	0.000408	...	0.000816	0.000618	0.000432	0.00063	0.000606	0.000781	0.00098	0.000851
5	Normal	Normal or low	Normal	Normal	Normal	Low	high	0	0.000163271	0.000245	0.000276	0.000225	0.000198	0.000136	0.000167	0.000136	...	0.000272	0.000206	0.000144	0.00021	0.000202	0.00026	0.000327	0.000284
6	Normal	Normal or low	Normal	Normal	Normal	high	Normal	0	0.003134799	0.004702	0.005299	0.004329	0.003807	0.002612	0.003209	0.002612	...	0.005225	0.003956	0.002762	0.00403	0.003881	0.005001	0.00627	0.005449
7	Normal	Normal or low	Normal	Normal	Normal	high	Low	0	9.79625E-05	0.000147	0.000166	0.000135	0.000119	8.16E-05	0.0001	8.16E-05	...	0.000163	0.000124	8.63E-05	0.000126	0.000121	0.000156	0.000196	0.00017
8	Normal	Normal or low	Normal	Normal	Normal	high	high	0	3.26542E-05	4.9E-05	5.52E-05	4.51E-05	3.97E-05	2.72E-05	3.34E-05	2.72E-05	...	5.44E-05	4.12E-05	2.88E-05	4.2E-05	4.04E-05	5.21E-05	6.53E-05	5.68E-05
9	Normal	Normal or low	Normal	Normal	Late	Normal	Normal	7	0.015842534	0.023764	0.026781	0.021878	0.019237	0.013202	0.01622	0.013202	...	0.026404	0.019992	0.013957	0.020369	0.019615	0.025273	0.031685	0.027536
10	Normal	Normal or low	Normal	Normal	Late	Normal	Low	5	0.000495079	0.000743	0.000837	0.000684	0.000601	0.000413	0.000507	0.000413	...	0.000825	0.000625	0.000436	0.000637	0.000613	0.00079	0.00099	0.00086
11	Normal	Normal or low	Normal	Normal	Late	Normal	high	0	0.000165026	0.000248	0.000279	0.000228	0.0002	0.000138	0.000169	0.000138	...	0.000275	0.000208	0.000145	0.000212	0.000204	0.000263	0.00033	0.000287
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2910	high	high	Crash	Low	Stop	Low	Normal	0	1.56322E-10	2.42E-08	1.4E-07	6.91E-08	8.22E-10	7.98E-08	6.75E-09	4.55E-08	...	1.95E-12	3.03E-10	1.76E-09	8.63E-10	1.03E-11	9.97E-10	8.44E-11	5.69E-10
2911	high	high	Crash	Low	Stop	Low	Low	0	4.88507E-12	7.57E-10	4.39E-09	2.16E-09	2.57E-11	2.49E-09	2.11E-10	1.42E-09	...	6.51E-13	1.01E-10	5.85E-10	2.88E-10	3.42E-12	3.32E-10	2.81E-11	1.9E-10
2912	high	high	Crash	Low	Stop	Low	high	0	1.62836E-12	2.52E-10	1.46E-09	7.19E-10	8.56E-12	8.31E-10	7.03E-11	4.74E-10	...	1.25E-11	1.94E-09	1.12E-08	5.53E-09	6.58E-11	6.38E-09	5.4E-10	3.64E-09
2913	high	high	Crash	Low	Stop	high	Normal	0	3.12644E-11	4.85E-09	2.81E-08	1.38E-08	1.64E-10	1.6E-08	1.35E-09	9.11E-09	...	3.91E-13	6.06E-11	3.51E-10	1.73E-10	2.05E-12	1.99E-10	1.69E-11	1.14E-10
2914	high	high	Crash	Low	Stop	high	Low	0	9.77013E-13	1.51E-10	8.78E-10	4.32E-10	5.14E-12	4.98E-10	4.22E-11	2.85E-10	...	1.3E-13	2.02E-11	1.17E-10	5.76E-11	6.85E-13	6.65E-11	5.62E-12	3.79E-11
2915	high	high	Crash	Low	Stop	high	high	0	3.25671E-13	5.05E-11	2.93E-10	1.44E-10	1.71E-12	1.66E-10	1.41E-11	9.49E-11	...	6.25E-11	9.7E-09	5.62E-08	2.76E-08	3.29E-10	3.19E-08	2.7E-09	1.82E-08

Table 2.3 A sample of daily erection rates computed for running scenarios

Scenario	8/1/22	8/2/22	8/3/22	8/4/22	8/5/22	8/6/22	8/7/22	8/8/22	….	1/31/23	2/1/23	2/2/23	2/3/23	2/4/23	2/5/23	2/6/23	2/7/23	Completion date
1	8	7	8	7	7	7	8	7	….	8	7	0	0	6				2/4/2023
2	5	0	0	0	8	7	7	7	….	8	7	5						2/2/2023
3	7	8	8	8	7	8	8	0	….	0	0	0	7	7	8	0	7	2/7/2023
……	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	….	……
999	7	8	8	0	0	0	0	4	….	6	7	7	8	7	8	8		2/6/2023
1000	6	6	7	8	8	5	0	0	….	8	8	0	0	5	8			2/5/2023

One of the most used ways to generate a random variable with specific discrete probabilistic distribution builds on the fact that when variable x has probability distribution f, $x \sim f(x)$, the cumulative probability function of x, F(x) has a uniform distribution, $F(x) \sim U(0,1)$. Thus, for running the scenarios, a random value is generated between zero and one and then finding the position of that value in the vector of cumulative probability distributions on that specific day, and then the first combinatorial k-dimensional risk state that has higher cumulative probability value, becomes the selected scenario of k risk statuses on that day. When we run this process from the start date of the project and reduce estimated erection rates for each day until reaching to the date that all modules are erected, we come up with a scenario of demand schedule and its completion date (as shown in matrix form in Table 2.3).

Figure 2.6 shows an example of two parallel projects' modular demand schedules that have different characteristics in terms of location, size, module types, and start time. Based on the projects' current progress and the delay caused by disruptive events in the past demand logs, we can have a new projected demand logs (modular onsite demand schedules) for the future assuming a scenario of risk statuses on future days.

A visual example shown in Figure 2.7 presents a sample set of probable scenarios for future projected modular onsite demand schedules to the end of a specific project. These scenarios are made by assuming probabilistic pattern of disruptive events on the given future dates. Each of these scenarios happens with a probability and ending to a different completion date for a project.

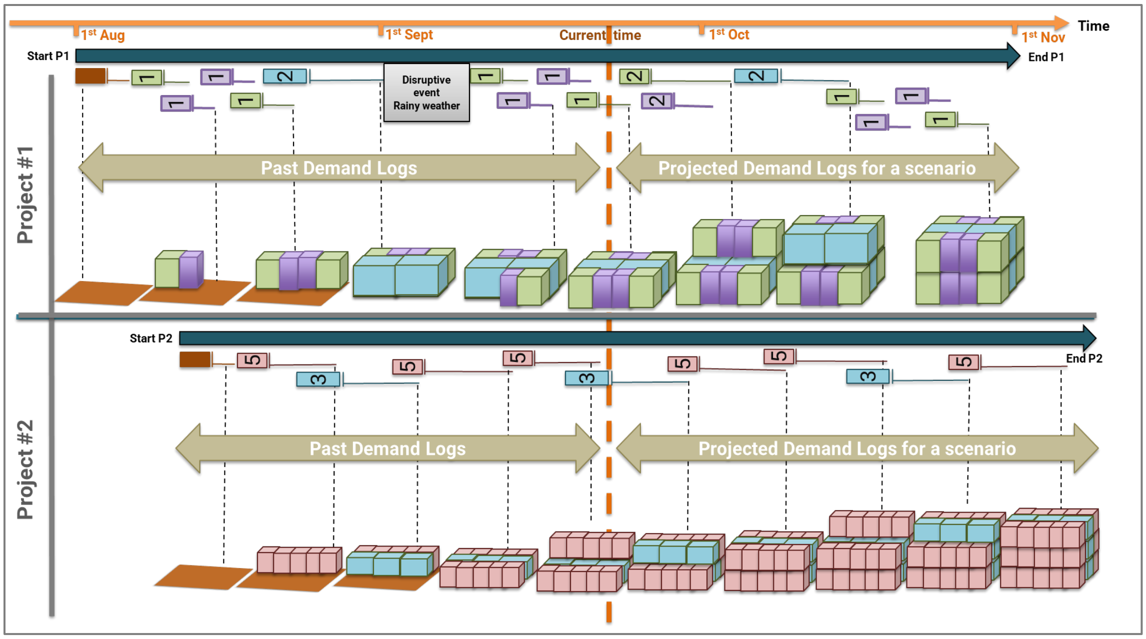


Figure 2.6 Aiming to update projected demand logs for possible scenarios while running projects

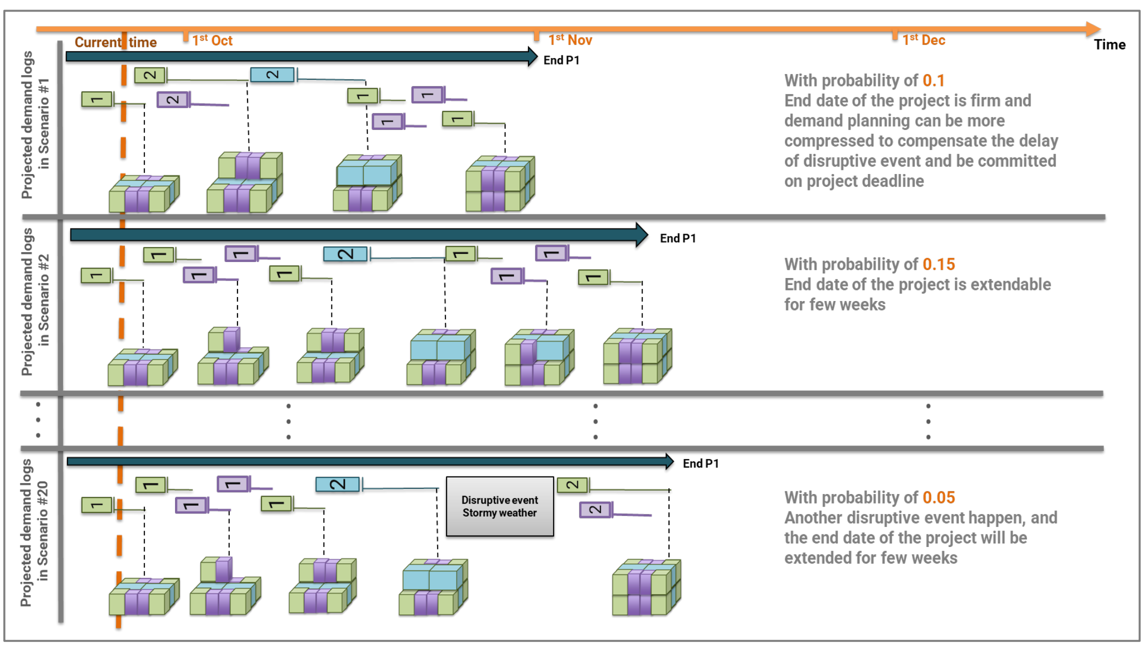


Figure 2.7 Generating a wide set of probable scenarios for projected future demand logs

2.4 User-Interactive Application for risk-sensitive demand scenario generation

For implementation of the proposed model formulation as presented in the previous sections, a user interactive application prototype has been developed for risk-sensitive analysis and scenario-based demand generation. This application prototype developed in Python language with user interface (see Figure 2.9) and dashboard for visualization of results. In terms of computation speed, for instance its run time for generating and analyzing 100 and 1000 scenarios for a case provided by our industrial collaborator are respectively 68 and 525 seconds. Its analytical results have already been used in several real large-scale modular construction projects of our industry collaborator.

Figure 2.8 displays inputs and outputs of the prototype. There are three main inputs, historical weather information of erection site location over past decades; sequence and layout of module types for each building and floors; and expert-based knowledge on type of disruptive events, their level of severity and their impact on performance (see Table 2.4). We mathematically model these risks and disruptive events using the proposed model, then we generate a wide range of most probable scenarios from daily probabilistic distributions of risks, and finally, by mining over these scenarios we come up with analytical results guiding managerial and executive decisions. The application has two type of outputs (see Figure 2.10), first scenario-based demand schedules for different module types in detail of dates and hours which can be used as synthetic input data for other optimization parts of project, and second some visual reports which enables managers to better understand the different variables' impact on project completion time windows, and module demand patterns guiding them toward appropriate contracting and actionable recourse options and policies.

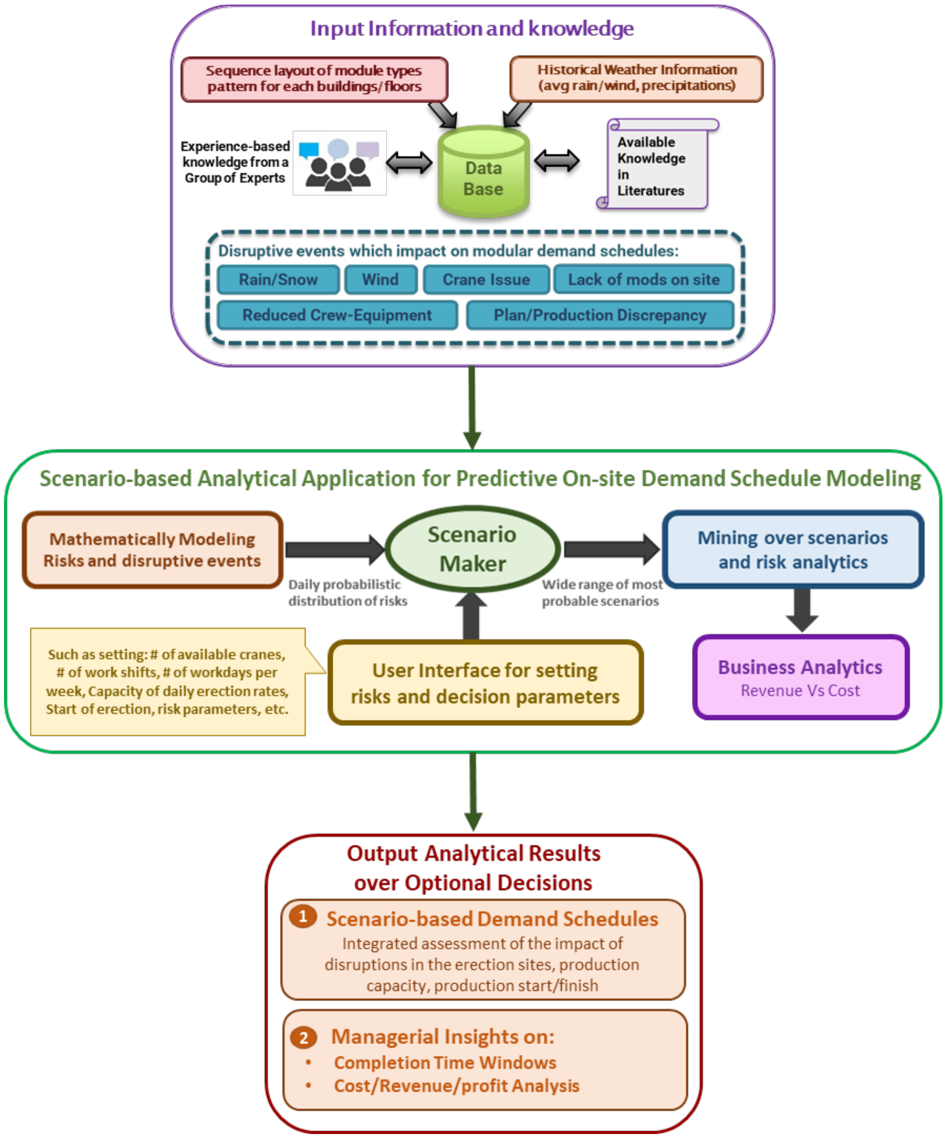


Figure 2.8 Representation diagram for input/output of the developed user-Interactive App

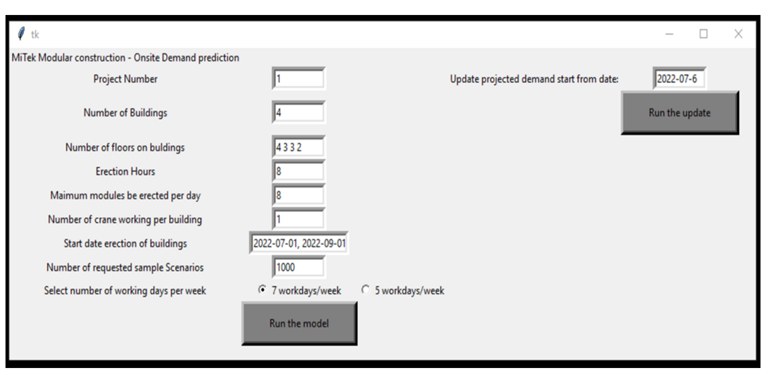


Figure 2.9 A sample of App's user interface page

Table 2.4 List of disruptive events and their characteristics based on expert believes

Disruption risk	Impact Level	Severity level	Probability of happening	Delays	Notes
Rain	Normal	0	Daily perception	None	Clear
	low	0.5	Daily perception	2 Hour Delay	Scattered Light Rain
	medium	0.8	Daily perception	4 Hour Delay	Overcast - Scattered Showers
	high	1	Daily perception	No Work This Day	Rain
Wind	low	0	Daily perception	None	0-10 mph
	medium	0.5	Daily perception	4 Hour Delay	10-20 mph
	high	1	Daily perception	No Work This Day	over 20 mph
Crane Issue	Normal	0	0.98	None	No Issues
	Down	0.5	0.015	4 Hour Delay	Mechanical Breakdown/No-show - Quickly fixed/replaced
	Crash	1	0.005	No Work This Day	Crane must be replaced
Labor/ Equipment Efficiency	Normal	0	0.98	None	No Issues
	Medium Low	0.8	0.015	2 Hour Delay	Reduced Crew - Equipment
	Low	1	0.005	No Work This Day	Crew Not Available
Mods Available	Normal	0	0.99	None	No Issues
	Low	0.5	0.009	4 Hour Delay	Shipments delayed - Modify Plan
	High	1	0.001	No Work This Day	No mods on Site
Site Not Ready	Normal	0	0.98	None	No Issues
	Low	1	0.015	No Work This Day	Waiting on Supporting Structure (Foundation - Steel Erection)
	High	1	0.005	2-3 days	Major delays in Supporting Structure
Plan/Production Discrepancy	Normal	0	0.98	None	No Issues
	Low	0.5	0.015	4 hours	Unique issue involving few modules
	High	1	0.005	2-3 days	Significant issue requiring rework

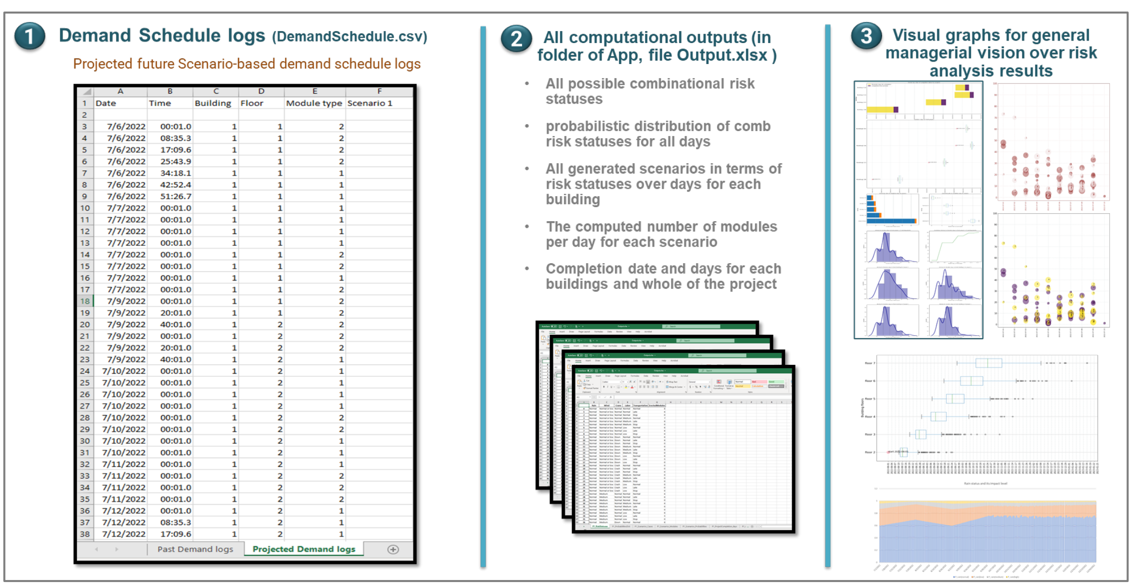


Figure 2.10 A sample of application outputs

Some examples of application visual results in its final dashboard as shown in Figures 2.11 to Figure 2.15. For instance:

1. Overall box plot of completion time windows for building floors that each box plot shows the quantiles of completion dates among all scenarios (Figure 2.11)

2. Histogram of project/building construction durations (in days) among all scenarios (Figure 2.12)

3. Cumulative Density for duration days to finish building which it is the Cumulative curve of previous diagram (Figure 2.13)

4. Normalized aggregated scenarios' probabilities on their completion dates that in this diagram we sum probabilities of scenarios on each completion date (Figure 2.14)

5. Probabilistic pattern of demand for different module types on specific erection site and specific date (see Figure 2.15). By aggregating probabilities on all scenario-based modular schedules, we come up with a probabilistic pattern of required quantity for each type of module on each specific date. Tag number on circles show the quantity and y axis is showing probability of consumption. For instance, in Figure 2.15, given specific date and erection site, with 0.7 probability we will need two modules of type X-B1, with 0.05 probability we need one of that and with 0.25 probability we do not need any quantity of X-B1 type. When we compute expected quantity on each erection site and aggregate these on multi projects, we can have expected number of required quantities for each type of modules on each specific date which is very helpful for production and storage planning.

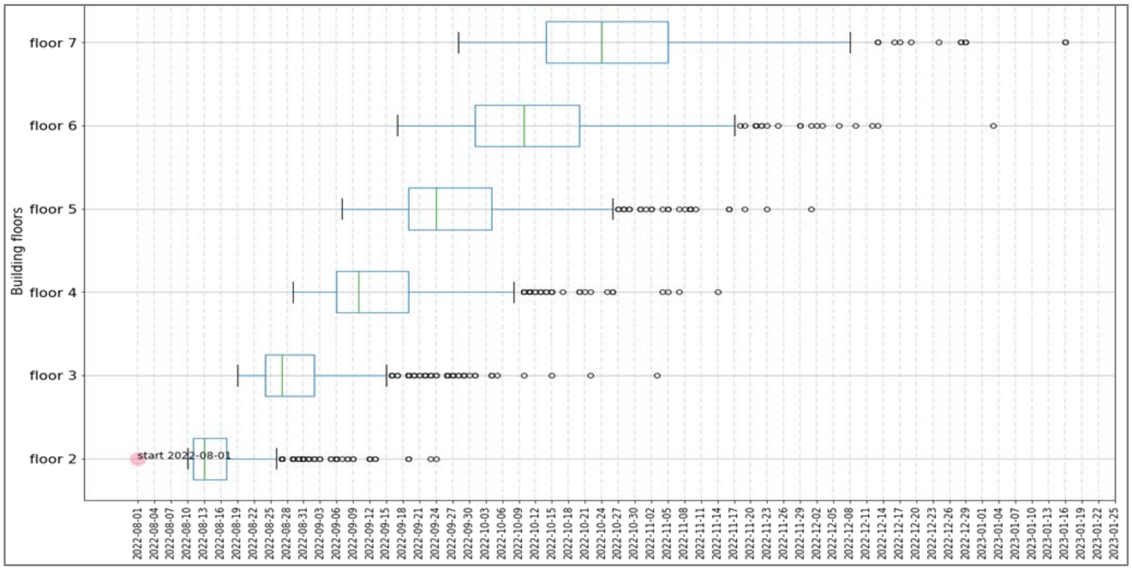


Figure 2.11 Overall box plot of completion time windows for building floors

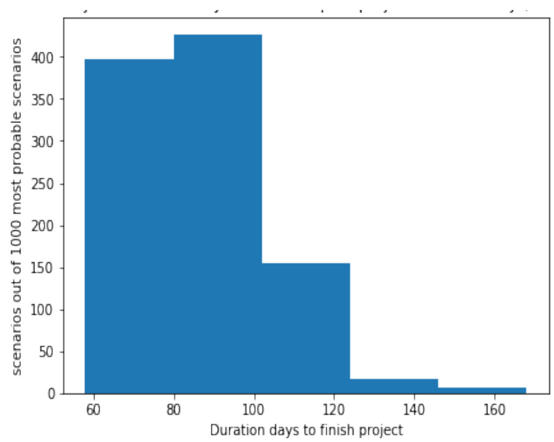


Figure 2.12 Density for duration days to finish building

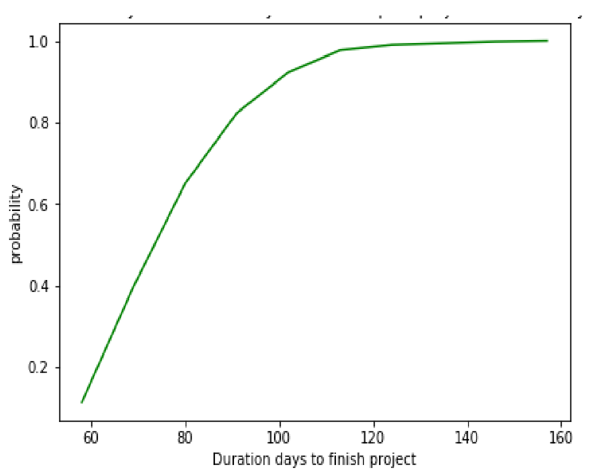


Figure 2.13 Cumulative density for duration days to finish building

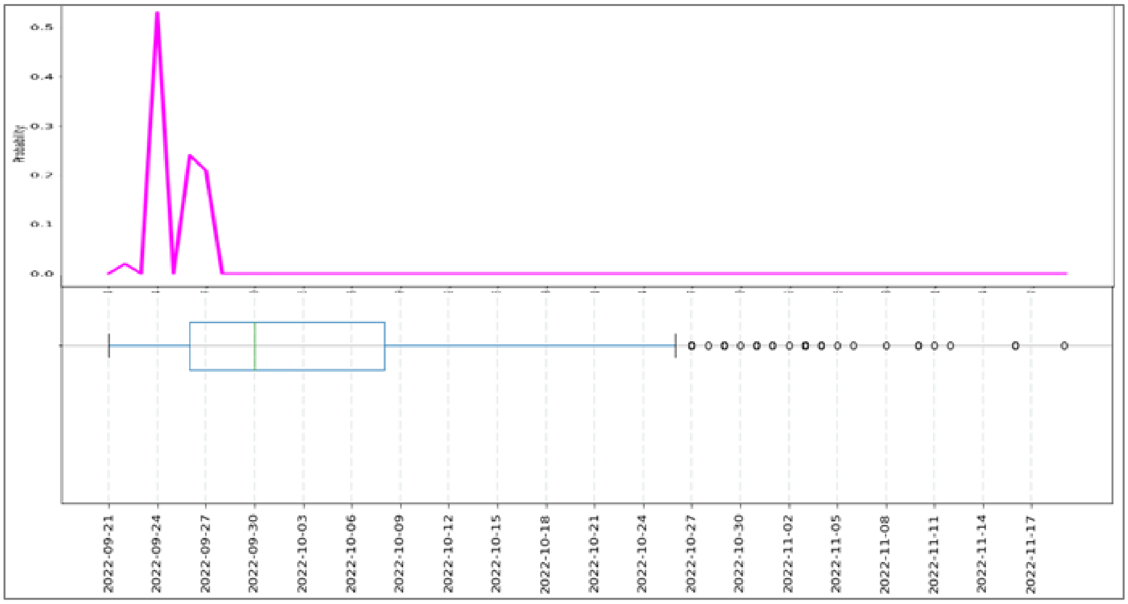


Figure 2.14 Normalized aggregated scenarios' probabilities on their completion dates

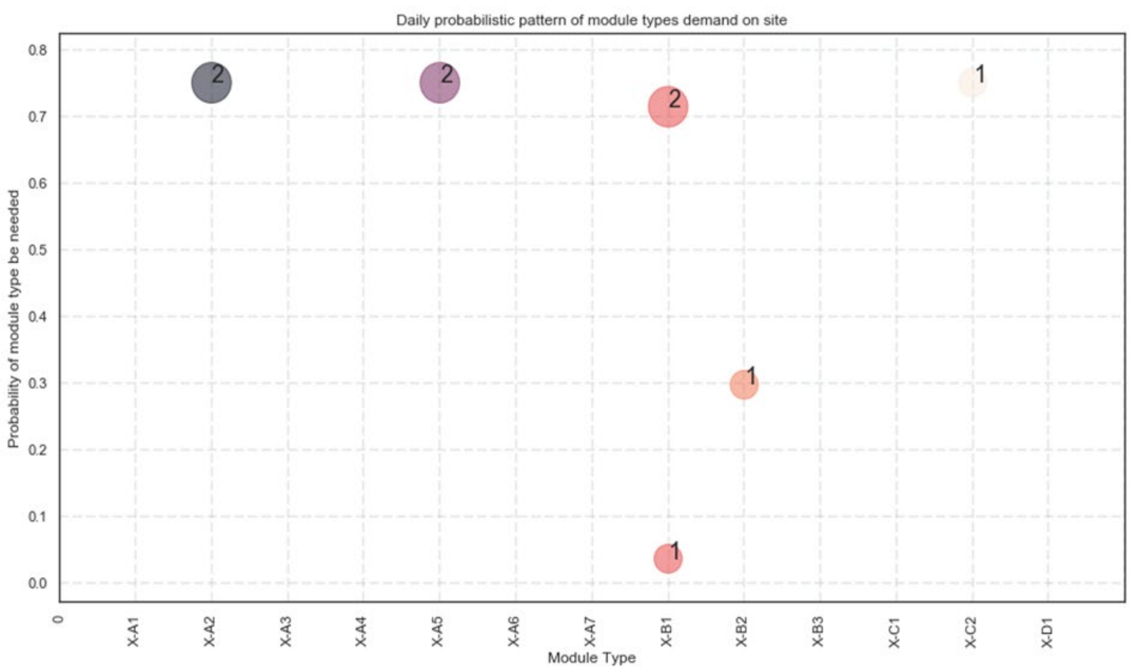


Figure 2.15 An example of probabilistic pattern of module types' demand on specific erection site and specific date

2.5 Conclusion

The adoption of distributed modular construction is increasing in many countries due to its benefits; however, for developing countries, this method is novel, and the current rate of adoption is low due to various risk events and uncertainties associated with it. Since risks are inevitable, it is necessary for practitioners to identify, evaluate, prioritize, control and monitor these risks. The systematic approaches and empirical studies related to the visualization and communicating multi-dimensional risks of these megaprojects with large-scale engineering remain missing.

In this study, we introduced a risk sensitive scenario-based prediction tool with user interface for assessing the risk of erection sites' disruptive events in the complex projects considering uncertainty and dominant impact between the risk elements. This tool enables making a wide range of scenarios using probabilistic patterns of daily risk distribution over decision variables such as number of cranes available, number of working days or shifts, start date of erections, and etc. Then by mining over the results of simulated scenarios, providing managerial insights for better understanding of decision variables' impact on buildings/ projects completion time windows and probabilistic modular demand patterns on a given time horizon, and also guiding managers toward appropriate contracting and economic resource usages. The other output of this application tool is generating the synthetic demand logs of scenario-based modular demand schedules for multi-erection sites and multi-projects which is useful for other optimization purposes such as estimate projected demand to production sites, required storage capacity, etc.

The results of this study can be used by potential modular manufacturers while implementing an offsite construction project to look for most prominent risks and devise corresponding risk mitigation strategies. To extend the scope of this research, we recommend further research in the following streams. This work can be extended to consider the risk sharing mechanism between the parties. Additionally, other major categories of risks such as macroeconomics and finance risks, contract risk, regulations and disputes resolution risk, design risks, and production site' disruptions risk which all impact on planning and the performance of modular construction projects, can be incorporated.

2.6 References

Applequist S., Gahrs G.E., Pfeffer R.L., Niu X.F., (2002), Comparison of methodologies for probabilistic quantitative precipitation forecasting, Weather Forecast. 17 (4), 783–799

Assaf S.A., Al-Hejji S., (2006) "Causes of delay in large construction projects", Int. J. Proj. Manag., 24(4), 349.

Azad A.K., M.G. Rasul, M.M. Alam, S.M. Ameer Uddin, Sukanta Kumar Mondal, "Analysis of wind energy conversion system using Weibull distribution", Proc. Eng. 90 (2014) 725–732.

Bénaben, F. & Montreuil, B. & Faugere, L. & Lauras, M. & Gou, J. & Mu, W.. (2020). A Physics-Based Theory to Navigate Across Risks and Opportunities tn the Performance Space: Application to Crisis Management. 10.24251/HICSS.2020.267.

Benaben, F., Faugere, L., Montreuil, B., Lauras, M., Moradkhani, N., Cerabona, T., ... & Mu, W. (2021). Instability is the norm! A physics-based theory to navigate among risks and opportunities. Enterprise Information Systems, 1-28.

Blismas N, Wakefield R. (2009), "Drivers, constraints and the future of offsite manufacture in Australia". Arif M, editor. Constr Innov. 2009 Jan 1; 9(1):72–83.

Conradsen K., L.B. Nielsen, L.P. Prahm, (1984) "Review of Weibull statistics for estimation of wind distribution", J. Clim. Appl. Meteorol. 23, 1173–1183.

Feldkamp, N., Benaben, F., Montreuil, B., Barenji, A. V., & Nazzal, D. (2021, December). Physics of decision: application to polling place risk management. In 2021 Winter Simulation Conference (WSC) (pp. 1-12). IEEE.

Gunduz M., Nielsen Y., Ozdemir M., “Quantification of delay factors using the relative importance index method for construction projects in Turkey”, J. Manag. Eng., 2013, 29 (2), 133-139.

Hassim S, Jaafar MS, Sazalli SA. (2009), The contractor perception towers industrialised building system risk in construction projects in Malaysia. Am J Appl Sci.; 6(5):937.

Hsu P.Y., Aurisicchio M., Angeloudis P.,(2017) ”Investigating schedule deviation in construction projects through root cause analysis,” Proc. Comput. Sci., 121, 732-739.

Hsu, P. Y., Angeloudis, P., & Aurisicchio, M. “Optimal logistics planning for modular construction using two-stage stochastic programming”. Automation in Construction, 2018, 94, 47-61.

Jaillon L, Poon CS. (2008), Sustainable construction aspects of using prefabrication in dense urban environment: a Hong Kong case study. Constr Manag Econ. 2008 Sep; 26(9):953–66.

Jaillon L, Poon CS. (2009), The evolution of prefabricated residential building systems in Hong Kong: A review of the public and the private sector. Autom Constr. 2009 May; 18(3):239–48.

Kamali M, Hewage K. (2016), Life cycle performance of modular buildings: A critical review. Renew Sustain Energy Rev. 2016 Sep; 62:1171–83.

Kimiagari S. & Keivanpour S. (2019) An interactive risk visualization tool for large-scale and complex engineering and construction projects under uncertainty and interdependence, International Journal of Production Research, 57:21, 6827-6855.

Kumar K.S.P., S. Gaddada, (2015) “Statistical scrutiny of Weibull parameters for wind energy potential appraisal in the area of northern Ethiopia”, Wind Water Solar,2 (14).

Lawson M., Ogden R., Goodier C., (2014) “Introduction to planning of modular buildings”, design in Modular Construction, CRC Press, Boca Raton.

Li H.X., Al-Hussein M., Lei Z., Ajweh Z. (2013), Risk identification and assessment of modular construction utilizing fuzzy analytic hierarchy process (AHP) and simulation. Can J Civ Eng. 2013 Dec; 40(12):1184–95.

Li CZ, Hong J, Xue F, Shen GQ, Xu X, Luo L. (2016), SWOT analysis and Internet of Things-enabled platform for prefabrication housing production in Hong Kong. Habitat Int. 2016 Oct; 57:74–87.

Luo LZ, Mao C, Shen LY, Li ZD. (2015), Risk factors affecting practitioners’ attitudes toward the implementation of an industrialized building system: A case study from China. Eng Constr Archit Manag. 2015 Nov 16; 22(6):622–43.

Luo L., Qiping Shen G., Xu G, Liu Y., Wang Y. (2019), Stakeholder-Associated Supply Chain Risks and Their Interactions in a Prefabricated Building Project in Hong Kong. J Manag Eng. 2019 Mar; 35 (2):05018015.

Macdonald G., Macdonald K., (2010) Safeguarding: A Case for Intelligent Risk Management, The British Journal of Social Work, Volume 40, Issue 4, June 2010, Pages 1174–1191, https://doi.org/10.1093/bjsw/bcq041.

Moradkhani, N., Benaben, F., Montreuil, B., Barenji, A. V., & Nazzal, D. (2021, June). Physics of decision for polling place management: a case study from the 2020 USA presidential election. In ICMS 2021-15th International Conference on Modeling and Simulation (Vol. 15, No. 9, pp. 440-448).

Morgan E.C., M. Lackner, R.M. Vogel, L.G. Baise, "Probability distributions for offshore wind speeds", Energy Convers. Manag. 52 (2011) 15–26.

Pan W, Gibb AGF, (2007), Dainty ARJ. Perspectives of UK housebuilders on the use of offsite modern methods of construction. Constr Manag Econ. 2007 Feb; 25(2):183–94.

Pervez, H., Ali, Y., Pamucar, D., Garai-Fodor, M., & Csiszárik-Kocsir, Á. (2022). Evaluation of critical risk factors in the implementation of modular construction. Plos one, 17(8), e0272448.

Rahman M.M. (2014), Barriers of Implementing Modern Methods of Construction. J Manag Eng. 2014 Jan; 30 (1):69–77.

Rogers G.G., Bottaci L., "Modular production systems: A new manufacturing paradigm", J. Intell. Manuf., 1997, 8 (2), 147-156.

Sambasivan M., Soon Y.M., "Causes and effects of delays in Malaysian construction industry", Int. J. Proj. Manag., 2007, 25 (5), 517-526.

Sun Y., Wang J., Wu J., Shi W., Ji D., Wang X, et al. (2020), Constraints Hindering the Development of High-Rise Modular Buildings. Appl Sci. 2020 Oct 14; 10(20):7159.

Sweis G., Sweis R., Hammad A. A., Shboul A., "Delays in construction projects: the case of jordan", Int. J. Proj. Manag., 2008, 26 (6), 665-674.

Wilks D.S.,(1995) "Parametric probability distributions", in: J. Hele (Ed.), Statistical Methods in Atmospheric Sciences, Academic Press, San Diego, 9780080456225.

Wuni I.Y., Shen G.Q. (2020), Barriers to the adoption of modular integrated construction: Systematic review and meta-analysis, integrated conceptual framework, and strategies. J Clean Prod. 2020 Mar; 249:119347.

Wu P., Xu Y., Jin R., Lu Q., Madgwick D., Hancock C.M. (2019), Perceptions towards risks involved in off-site construction in the integrated design & construction project delivery. J Clean Prod. 2019 Mar; 213:899–914.

Zhang X., Skitmore M., Peng Y. (2014), Exploring the challenges to industrialized residential building in China. Habitat Int. 2014 Jan; 41:176–84.

Zou P.X., Zhang G., Wang J., "understanding the key risks in construction projects in China", Int. J. Proj. Manag., 2007, 25(6), 601-614.

CHAPTER 3. PREDICTIVE POTENTIAL DEMAND MODELING FOR MODULAR CONSTRUCTION OVER THE US METROPOLITAN STATISTICAL AREAS

Abstract: A growing interest in applying modular construction has been seen as a game-changing method to transform traditional site-based construction in recent decades. However, the adoption of modular construction is still slow within high-density city contexts. A significant reason for that is the real or perceived low capability of the supply chain of delivering modular high-rise buildings in built urban areas. New disruptive innovators are at work improving drastically such capability and in enabling distributed modular building construction at large scale over vast territories. In order to guide such innovators, this chapter proposes an approach enabling predictive modeling of potential demand for modular construction.

The approach leverages making scenarios on overall aggregated level of assumptions values and accurately predicting the patterns of geographical and volume distribution of demand over geo-markets. The chapter uses as testbed modular construction demand for multi-family housings, hotels/motels, dormitories, clinics, and convalescent centers over the set of Metropolitan Statistical Areas (MSA) in the United States. The chapter demonstrates how the results of the proposed approach can notably guide where and how much capital investment in building module production capacity should be made to be best poised for persistently meeting growing demand with both customer and shareholder satisfaction.

3.1 Introduction

Off-site construction (OSC) is a construction technique involving the planning, design, production, and assembly of building components at a location other than their final installed location to support the rapid and efficient construction of structures (Goodier & Gibb 2007). An innovative OSC method is modular integrated construction (MiC) where building components are built in an assembly line in a factory, transported to an erection site in modules (sections), set in place with crane (s), and then joined together to form a complete building (Wuni et al., 2019).

Previous studies have suggested various advantages of completing modules in factory conditions including (1) a more stable rate of production output, (2) a higher level of quality control, (3) greater consistency and accuracy in production, (4) enhanced safety performance, (5) more opportunities for leveraging automation and information technology, (6) reducing onsite labor requirements, (7) reducing construction waste, (8) reducing carbon emissions, (9) shortening construction time, and (10) decreasing theft (Gibb, A.G. 1999; McGraw Hill Construction 2013; McKinsey Global Institute 2017; Arcadis 2018; Construction Industry Council 2018; Pan & Hon 2018). MiC may also be a superior choice for projects in remote locations with harsh weather conditions and climates where construction labor with the requisite knowledge and skills is not readily available or just too expensive to warrant traditional construction (Rentschler et al. 2016). If building construction companies are strategic about where and when they establish their manufacturing plants, and innovative in how performing they engineer these plants, they can capitalize on the skyrocketing trends induced by enabling the above benefits.

Accurately predicting where new construction demand will occur before it occurs is not easy but doing so can be a distinct competitive advantage (Barker et al., 2021).

By leveraging scenario-based demand predictions of MSAs within the United States primed for growth, a construction company could lower costs by establishing operations closer to the most probable project locations and building activities improving delivery times and customer service. In this study, the proposed predictive model estimates potential construction demand and the scenario-based probable modular portion of that and generates synthetic demand logs over MSAs, by leveraging historical data such as Gross Domestic Product (GDP), population growth, historical pattern of building permit requests, and labor/expertise costs data at the MSA level. Based on these scenario-based results, managers will have a better understanding of investments in infrastructure and capabilities that should be considered in MSAs primarily due to the projected growth and demand for modular construction.

This chapter is structured as follows. Section 3.2 provides a Bibliometric analysis and the related literature review. Section 3.3 introduces the proposed methodology, including assumptions and model formulation. Section 3.4 covers model implementation and results of the proposed application. Finally, section 3.6 concludes the chapter through a synthesis of the contributions, limitations, and avenues for further research.

3.2 Bibliometric Analysis and Literature Review

To provide an overview of the existing research in potential demand prediction for modular construction, a bibliometric analysis was conducted on 11/22/2022 using the well-established and acknowledged, Web of Science (WoS) databases. The query for

WoS is as follows: TS= (('predict' OR 'forecast' OR 'estimate' OR 'predictive' OR 'modeling' OR 'model') AND 'demand' AND (‘offsite construction’ OR 'off-site construction’ OR ‘offsite production’ OR ‘off-site production’ OR ‘offsite manufacturing’ OR ‘off-site manufacturing’ OR prefabrication OR prefabricated OR prefab OR pre-fab OR pre-fabricated OR ‘industrialized building system’ OR ‘modular construction’ OR modular OR ‘off-site fabrication’ OR modularization OR ‘industrialized construction’)). Figure 3.1 shows the number of publications and total citations indexed by WoS 1989 to 2022. In total, 2,237 publications were found in WoS. Before 2013, the number of publications was at a relatively low level, while it increased rapidly beginning in 2014 and reached 240 in the year 2021 on WoS.

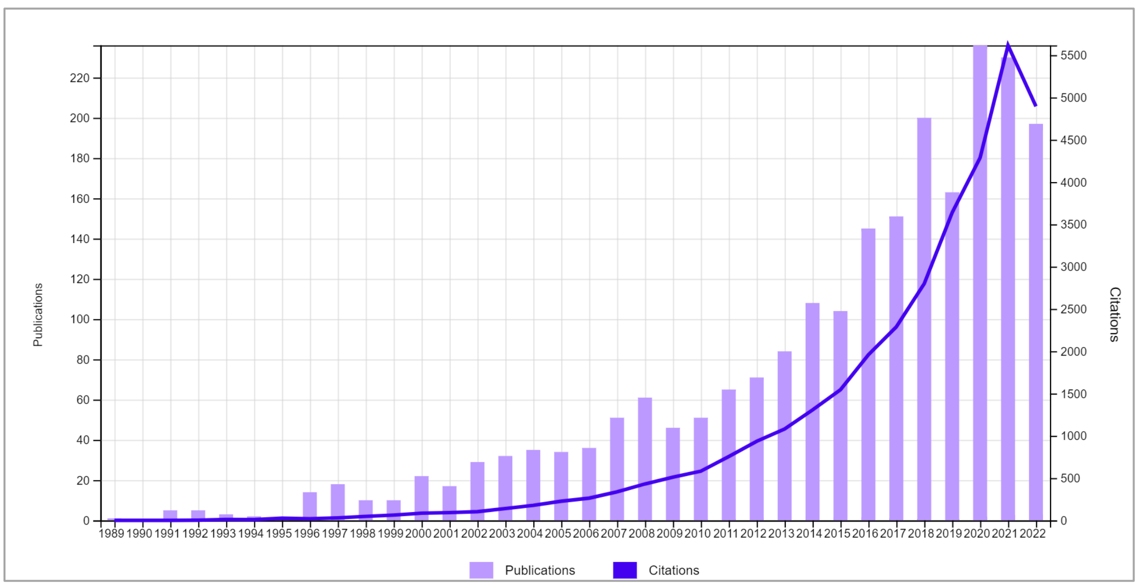


Figure 3.1 Annual scientific publication and total citations on potential demand prediction for modular construction as indexed by WOS

Figure 3.2 shows the most productive countries and collaborations in these publications. Figure 3.3 draws the publications trends of top authors over the time; the size of circles represents number of publications, and the color darkness defines the total earned citations at that related year.

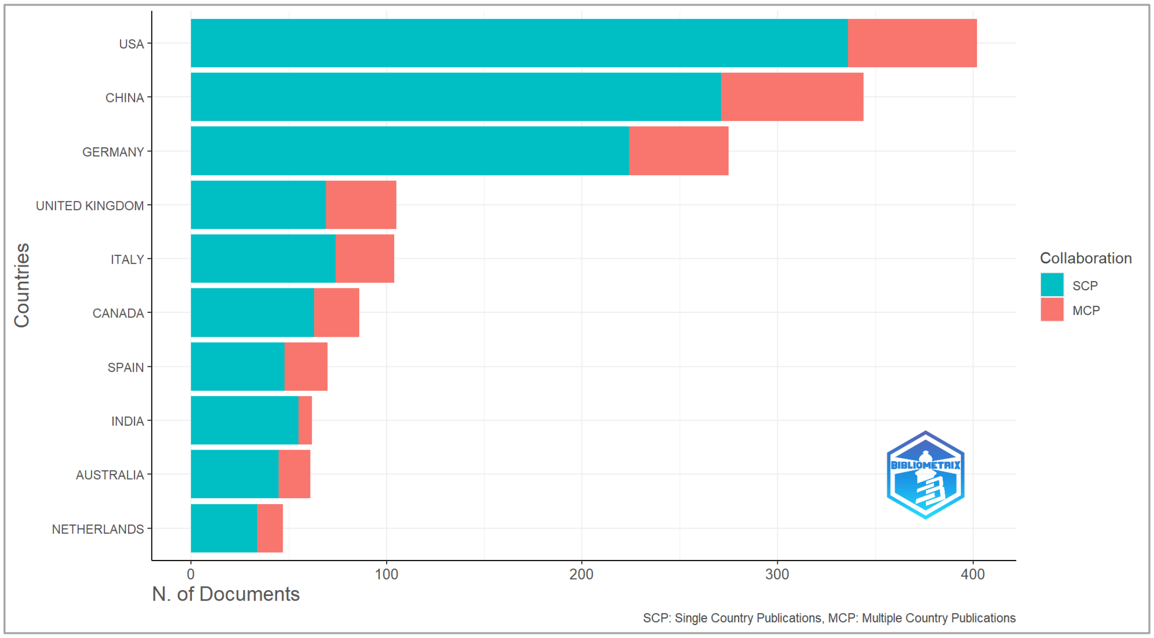


Figure 3.2 Most productive countries and collaborations on publications about potential demand prediction for modular construction

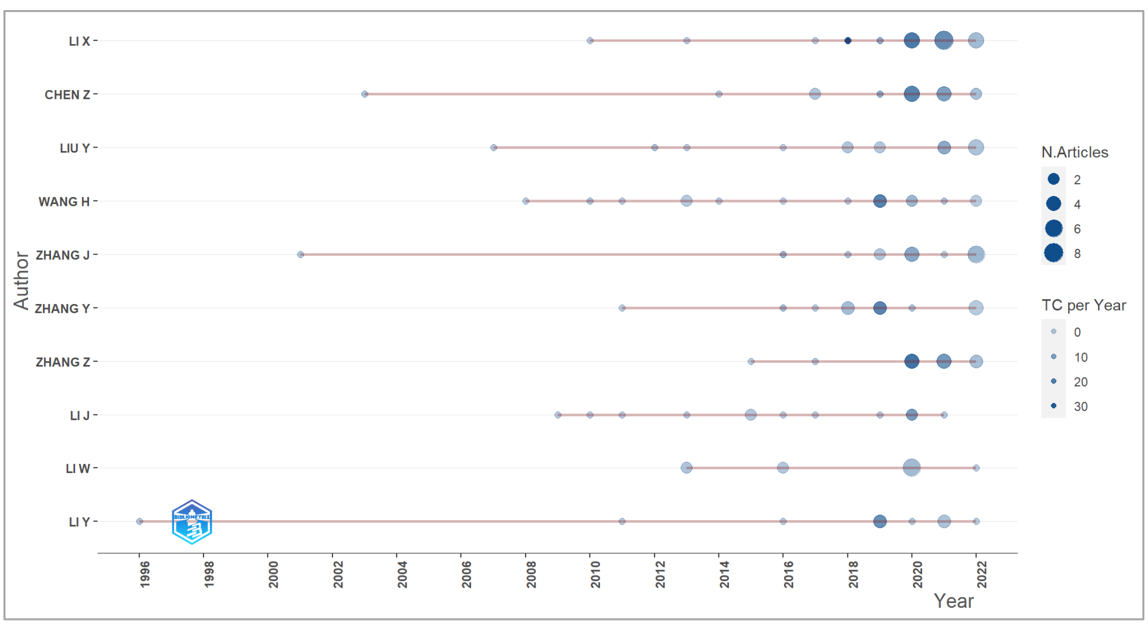


Figure 3.3 Top author's production trends over the time on publications about potential demand prediction for modular construction

In addition, complementary bibliometric results have been presented in Appendix C. For instance, Figure C.1 shows the increasing average number of article citations per year and Figure C.2 displays a mapping of the distribution of publications areas that reveals that most of these publications are categorized in the engineering and computer science areas. Figure C.3 provides a keyword co-occurrences network in which the size of nodes represents frequency of word's appearance and edges display co-occurrences.

Review of the literature reveals on one side that many approaches have been proposed to predict construction growth for a variety of reasons, and on the other side that there has been a limited set of studies specifically targeted at informing decision-making for MiC companies on where their next capital investment should be to take advantage of potential demand targeting specific geographic regions. Wuni & Shen (2022) have recently published a review article towards an understanding of the primary identified decision-making factors (DMFs) for implementing modular integrated construction (MiC). They proposed a conceptual framework for the identified DMFs consisting of labor considerations, project characteristics, location and site attributes, and organizational factors. In another review paper, Yang et al. (2017) provides a systematic review of existing academic perspectives and suggest future research directions to improve building module manufacturing systems. Their review explores critical research issues from five aspects: process and activities, organization and people, factory configuration, technology, and information and control systems.

Pan et al. (2018) reviewed the 10-year journey (2007-2017) of high-rise modular building and interviewed with seven representative cases for verification and with industry stakeholders for consultation. The cases were selected worldwide from the U.K., U.S., Singapore, Australia, and China. The case studies jointly enabled a longitudinal examination of the adoption of high-rise modular buildings, providing answers to the following three questions: (1) what were the benefits of the adoption of a modular approach for high-rise buildings? (2) What were the challenges to the adoption of a modular approach for high-rise buildings? (3) How were the challenges addressed?

In another paper, Yang et al. (2019) examined the supply chains adopted for modular high-rise buildings in high-density city contexts and explored the scenarios of developing MiC supply chains for high-rises in Hong Kong. Their analytical framework has been developed based on supply chain management theory and systems thinking in the contexts of the modular building supply chains. This framework addresses five interconnected aspects, namely, supplier capability, logistics feasibility, regulatory compliance, organizational integration competency, and process risk manageability.

Fullerton et al., (2000) used univariate ARIMA and random-walk models to evaluate the accuracy of residential construction forecasts. The data used in their study were derived from quarterly forecasts between the first quarter of 1985 and the second quarter of 1996. The forecasted single-family start rates were compared with univariate time series and random-walk alternatives. Their results showed that the accuracy of the historical estimates for regional single-family construction does not compare well to forecast accuracy from univariate ARIMA equations or random-walk predictions. Krylovas et al. (2011) and Linné & Cirincionne (2010) have constructed a model to identify features that are most critical for predicting when and which MSAs will experience growth in construction demand. In another study, Gupta et al. (2012)

predicted the housing demand in eight Southern California MSAs by examination of the time-series relationship between house prices and demand.

Mao et al. (2014) proposed a forecasting model based on a neural network of genetic algorithm optimization, then used for predicting housing demand in China by using the macro data on the housing market in Hangzhou during 1999-2012. Their results got an acceptable level of accuracy, however, there were fluctuation years in the prediction that were caused by housing control policies in China (some regulations issued by the State Council in 2011 to curb rising housing prices which caused consumers to assume a wait-and-see posture to bring about low turnover). The wait-and-see approach of Chinese consumers can be similar to United States consumers, an attitude that can influence the housing market.

Another influencing factor that impacts consumers when searching for a new home is property tax rates. In the study of McGibany (1991), county data from Wisconsin over twelve years has been studied, leading to discover that property tax rate differentials harm the construction of single-family houses. They recognize that the property tax rates for MSAs are a key factor to consider in a model when accurately predicting new housing builds. Applying machine learning techniques is another way that has been used to predict housing price movements across the United States. Gupta et al. (2021) used random forests consisting of 50, 75, or 100 random trees to accommodate multiple predictors and nonlinearities. They suggested based on their experience that future research should go beyond the state-level analysis and explore the specific MSA levels to obtain the national and regional housing factors for the predictive analysis.

Recently, Barker et al. (2021) have presented a predictive modeling approach that focuses on single-family housing in the United States. Their approach aims to guide where capital investments such as truss plants and lumber yards should be constructed before housing demand in a particular 5 top Metropolitan Statistical Area (MSA) with higher rates of demand.

In McKinsey Global Infrastructure Institute (2019), seven factors were identified to determine whether modular construction is likely to penetrate a given market (Figure 3.4). Labor dynamics and demand are at the top of the list of factors driving the adoption of modular construction; Labor shortages and an inability to keep up with demand stand out as the most decisive. A limited supply of skilled labor, which in turn drives up wages and costs, often sets the stage for modular construction solutions. Shifting to offsite manufacturing work may be cheaper and it may even attract new people into the workforce who do not wish to move from one construction site to another following projects.

Several additional drivers can have an important impact on the attractiveness of modular construction, including supply chain and logistics. Transport regulations constrain the size of modules that can be moved by road in some markets (including some US states), and access may also be limited in some dense urban locations. The second and related point involves other types of local constraints. For example, in Scandinavia, limited daylight in winter makes it particularly attractive to reduce onsite construction. In other cases, compact sites may make it desirable to deliver and rapidly install modules without requiring significant storage of materials. Third, geographies with ample access to low-cost materials (such as timber) are markets with less attractiveness to modular construction (MGI, 2019).

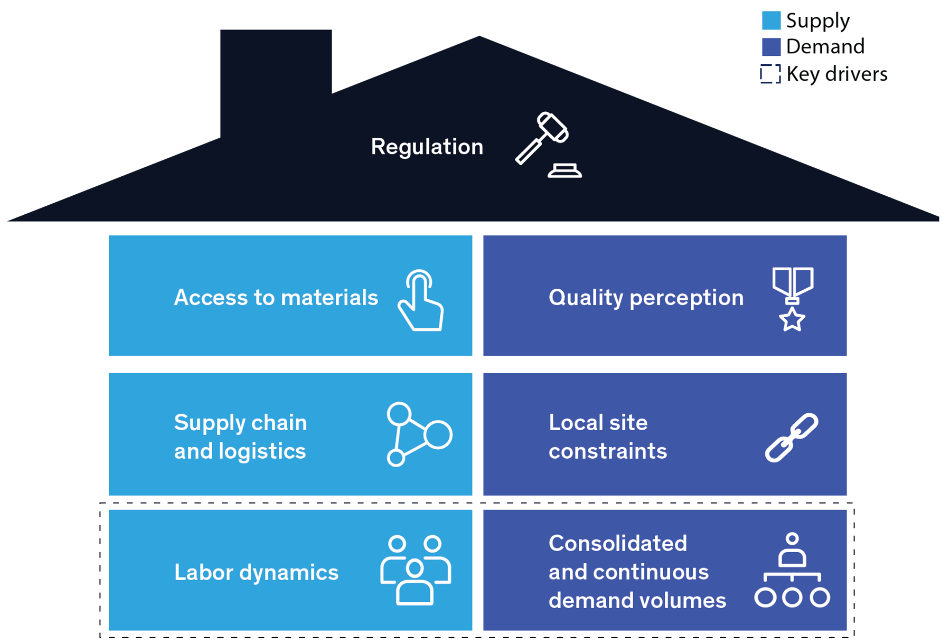


Figure 3.4 Factors determine the attractiveness of a market for modular construction McKinsey Global Infrastructure Institute (2019)

One major factor is quality perception (MGI, 2019). In some markets, the industry will need to overcome lingering perceptions from the post-war era that prefab housing is only a poor-quality solution for the masses. One route is to emphasize sustainability and future savings on energy and repair bills. Another route would be to focus on the appeal of modular construction in parts of the housing market where consumers already expect standardized offerings at scale, such as hostels, public-housing projects, retirement communities, and hotels.

The final determinant is regulation (MGI, 2019). Quality certification standards and warranties are big drivers that can inform customers and give them confidence. These certifications and warranties also facilitate the provision of financing as development financiers and mortgage providers need them to agree on loans. Governments can additionally help to drive adoption by including offsite manufacturing targets in public projects. For example, in Singapore, all government housing projects must use prefinished volumetric modules. Sustainability requirements and incentives will also help to drive the industry toward the most carbon-neutral products and practices. Another option is to support mortgages for the purpose of offsite manufactured homes. Similarly, building standards will have an important role in driving modular construction. The more that governments can move towards harmonization across different geographies and sectors, the more that suppliers will be able to drive scale into their pipeline.

One of the key drivers of cost savings comes from economies of scale (MGI, 2019). This requires large enough factories as well as sufficient output to ensure repeatability, learning, and volume savings on procurement.

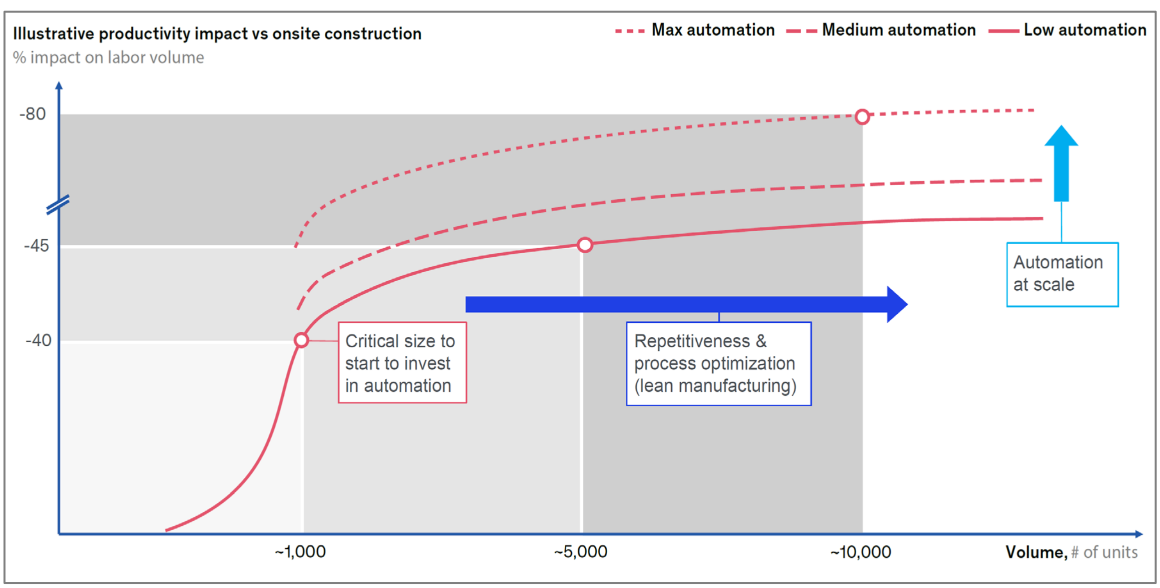


Figure 3.5 Achieving economies of scale in modular construction (MGI, 2019)

The (MGI, 2019) interviews indicate that companies achieve a rapid and substantial step-up in productivity when they begin turning out approximately 1,000 units per year. Another step-up, typically associated with another 5 percent boost in productivity, seems to be reached at about 5,000 units per annum (Figure 3.5). The fundamental dilemma facing many modular suppliers at this stage of their evolution is whether they can tap into a reliable pipeline of work within geographic reach to justify these larger-scale and more productive plants.

As shown in Figure 3.6, the modular construction market was valued at more than USD 85 billion in 2021, and the market is projected to register a CAGR of over 7.2% during the forecast period (2022-2028). Compound annual growth rate (CAGR) is the mean annual growth rate of an investment over a specified period of time longer than one year. It represents one of the most accurate ways to calculate and determine returns for individual assets, investment portfolios, and anything that can rise or fall in value over time.

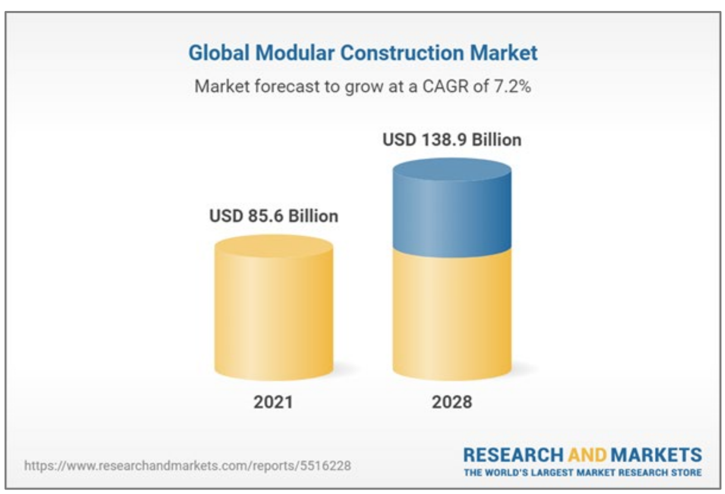


Figure 3.6 Global perspective for investment on modular construction market 2022-2028

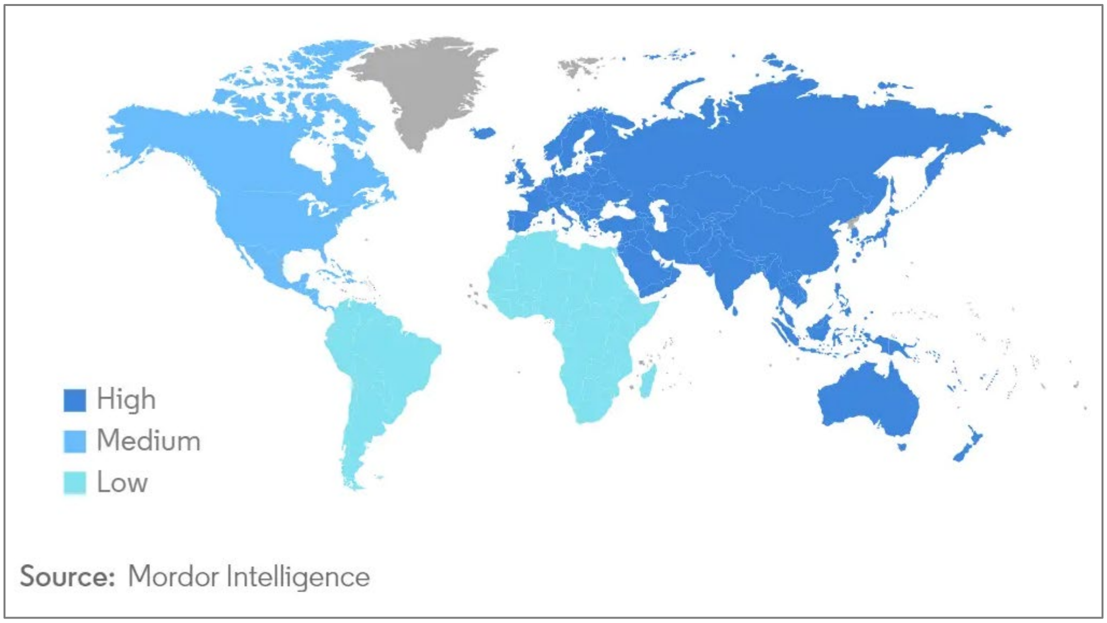


Figure 3.7 Modular construction market-expected growth rate by region, 2022-2028

Figure 3.7 compares the expected growth rate of modular construction in different regions of the world in the period of 2022 to 2028, highlighting clearly that modular construction has an international market potential.

3.3 Methodology

Our proposed approach for modeling scenario-based potential and modular part of future construction demand is summarized in Figure 3.8, emphasizing the modeling and computational levels. A wide range of historical databases are used as inputs including historical MSAs' population and GDP per capita rates, MSAs' construction labor/experts' costs, MSAs' yearly demand rates for different building types of construction with characteristics of area (ft^2), the building value ($$), the number of story classes and units, the number of projects, and MSAs' historical monthly demand seasonality on different type of buildings.

When making scenarios, there are two categories of variables and assumptions as mutable inputs which always can be set with new values or estimations and running the model for a new scenario. The first category is including estimates values on below variables by a group of experienced expertise in the field of modular construction:

1- MSAs' Incentive/Restrictive regulation level for MC (e.g., low, moderate, high)
2- MSAs' level of Access to the materials (e.g., low, moderate, high)
3- Yearly percentage of future demand which is likely to be modular (Figure 3.9)
4- Module types and the probability distribution for their usage on each type of building (see example Table 3.1)

The second category of scenario maker input assumptions consists of the predicted values for future MSAs' population and GDP per capita in the targeted future time horizon. Section 3.3.1 and section 3.3.2 introduce some existing approaches for estimation of these input assumptions.

Table 3.1 An example for scenario assumptions on module type usage distribution for different type of buildings

Type of Building	Module 1 (10*25 sqf)	Module 2 (15*30 sqf)	Module 3 (12*30 sqf)
Multifamily	0.3	0.3	0.4
Hotels and motels	0.4	0.2	0.4
Clinics/Nursing	0.2	0.3	0.5
Dormitories	0.6	0.25	0.15

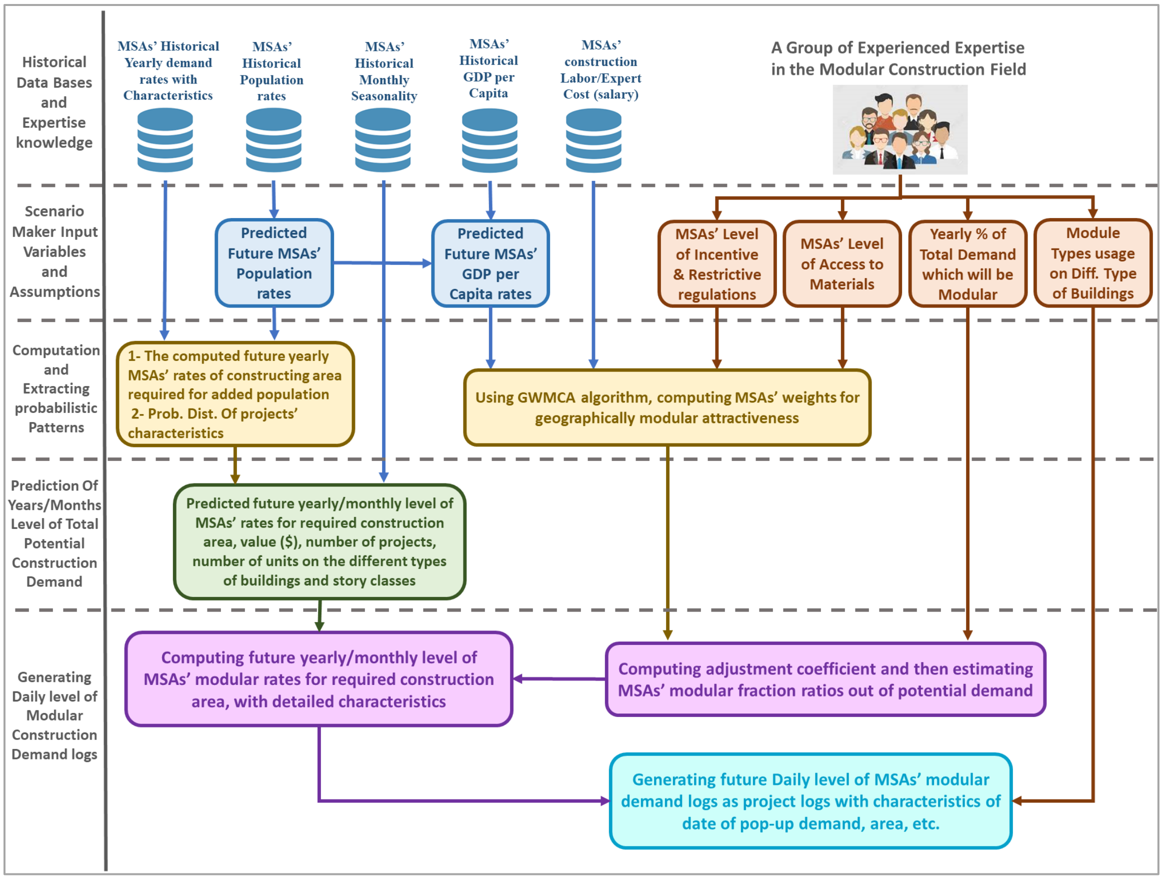


Figure 3.8 The proposed approach for generating scenario-based modular demand logs

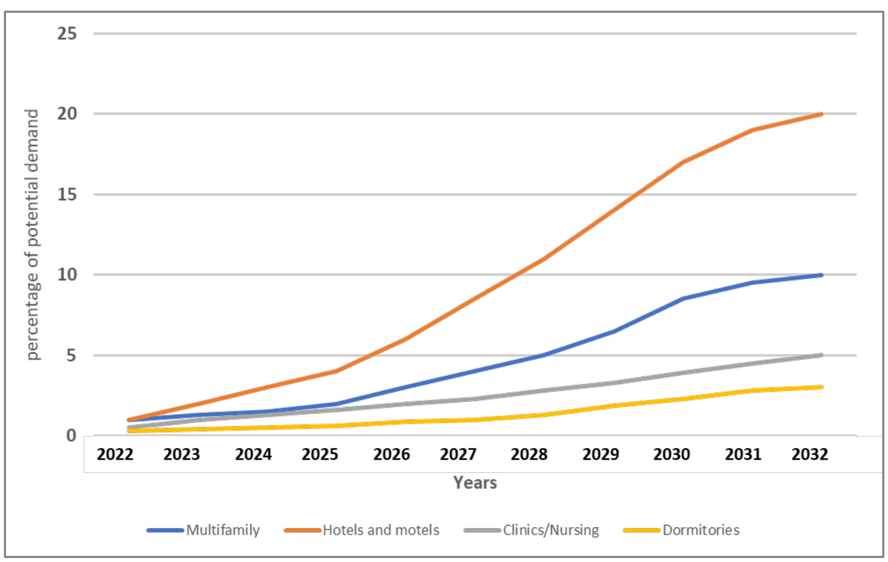


Figure 3.9 An example for scenario assumptions on yearly trends of total construction demand fractional percentage which will be modular

In the third level of the framework, MSAs' weights for geographically modular attractiveness can be computed using GWMCA algorithm (see section 3.3.3). In addition, the MSAs' ratio of construction area (ft^2) required for a population increase of 1,000 inhabitants is extracted from historical permit spatial data and historical population data. Then, using linear regression or the Exponential Triple Smoothing (ETS) algorithm on these extracted ratios enables us to estimate the future yearly MSAs' rates of construction area (ft^2) required for predicted adding population. Moreover, probabilistic distribution of projects' characteristic variables such as area per project, area per unit, value per square feet must be computed in this level.

By using the above results and monthly seasonality patterns extracted from historical data base, we can generate demand logs for MSAs in the years and months level with characteristics of required construction area (ft^2), value ($), number of projects, number of units on the different types of buildings and story classes. Finally, generating the daily level of potential construction demand logs as project logs with characteristics of date of pop-up demand, area (ft^2), type of building, story class, number of units, the number of different module types needed, expected start erection, and expected duration days of erection. Moreover, by computing adjustment of the coefficient and then estimating MSAs' modular fraction ratios out of potential demand, we can construct demand logs for the modular fraction of demand.

3.3.1 MSAs' population rates prediction

Governments use population forecasts at all levels (national, regional, city, international) for planning purposes, broadly defined. Traditional methods are deterministic using scenarios, but probabilistic forecasts are desired to get an idea of accuracy, assess changes, and make decisions involving risks. In a significant breakthrough, since 2015, the United Nations haves issued probabilistic population forecasts for all countries using a Bayesian methodology that has been reviewed by Raftery & Ševčíková (2021). In another study, Talkhabi et al. (2022) have developed some investigations over spatial and temporal population change in the Metropolitan Regions and its consequences on urban decline and sprawl. They investigated the process of urban sprawl as a spatial and visual manifestation of Tehran's periphery expansion between 1976 and 2016. Their findings revealed a link between the Tehran Metropolitan Region's pattern and manner of sprawl development and urban deterioration.

Since the focus of our experimental testbed is on the US MSAs, for estimation of future population as the model's input, we have leveraged directly results of Vespa et.al (U.S. Census Bureau, 2018) on US population estimations and projections from 2020 to 2060. Moreover, the U.S. Census Bureau released new population estimates and projections and other demographic data up to the year 2100 for 30 countries and areas in the International Database (IDB). The IDB consists of estimates and projections of demographic indicators, including population size and growth (by sex and single year of age up to 100-plus) and components of change (mortality, fertility and net migration) for more than 200 countries and areas. Figure 3.10 and Figure 3.11 have shown the aggregated whole US population and its growth rates up to the year 2060 based on report of the U.S. Census Bureau' IDB[1] in September 2020.

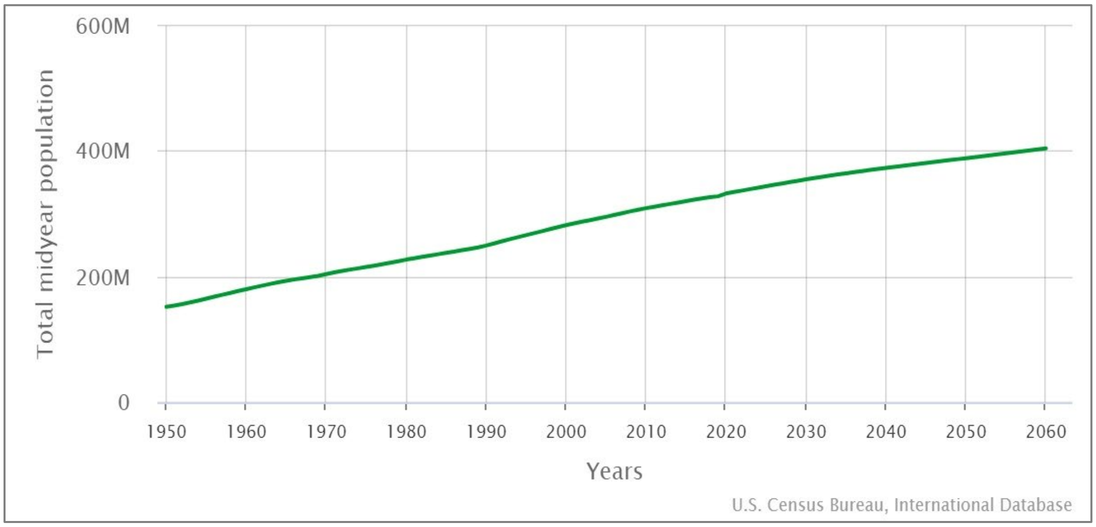


Figure 3.10 Total forecasted midyear population rates for the United States (the US Census Bureau' report in Sept 2020)

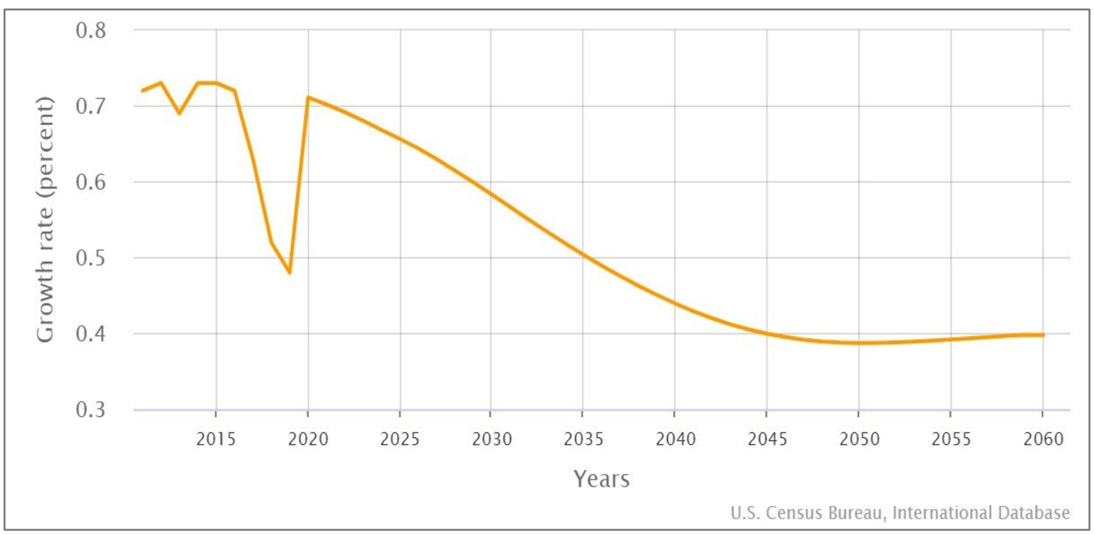


Figure 3.11 Forecasted population growth rates for the United States (the US Census Bureau' report in Sept 2020)

Another simple way to predict future MSAs' population is to use the methodology of Exponential smoothing forecast on historical population rates, for example the AAA

[1]https://www.census.gov/data-tools/demo/idb/#/country?COUNTRY_YR_ANIM=2060&COUNTRY_YEAR=2022&FIPS_SINGLE=US

version (additive error, additive trend, and additive seasonality) of the Exponential Triple Smoothing (ETS) algorithm, with smoothing out minor deviations in past data trends by detecting seasonality patterns and confidence intervals. The ETS algorithm has three parameters: (1) Alpha specifies the coefficient for the level; a higher value gives more weight to recent data points. (2) Beta specifies the coefficient for the trend smoothing; a higher value gives more weight to the recent trend. (3) Gamma specifies the coefficient for the seasonal smoothing; a higher value gives more weight to the recent seasonal period. There is also a parameter for the type of seasonality called Additive seasonality, where each season changes by a constant number. This forecasting method is best suited for non-linear data models with seasonal or other recurring patterns. Figure 3.12 shows aggregated prediction results of the implementation of ETS algorithms for the US MSAs using historical rates from 2010 to 2022.

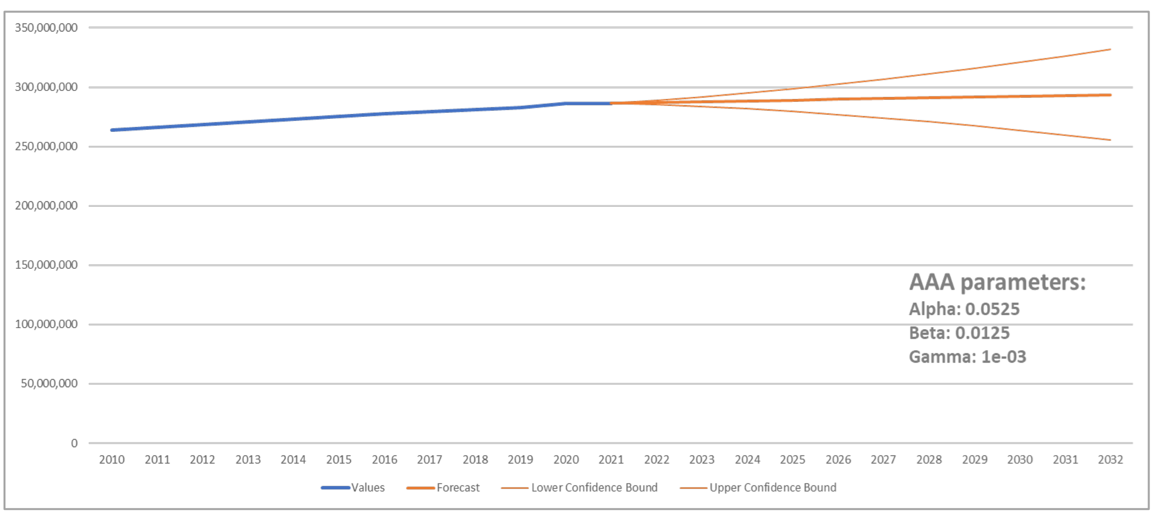


Figure 3.12 Total forecasted population rates aggregated over the US MSAs using ETS

Another approach to deal with population prediction is fitting functions to historical yearly added population rates and choosing the function with higher R square score as the Goodness-of-Fit (GoF) Measure (Little, 2004). For instance, in Figure 3.13 fitting different functions tested over total added population for 2010 to 2021 and linear function got higher score in R square. If we remove the outlier point corresponding to the year 2020, we can have higher rates of R square score and better fitting as shown in Figure 3.14. When this fitting function over added population has been done for each of MSAs separately, Figure 3.15 shows the aggregated rate of predicted adding population. As it is clear in Figure 3.15, for more than %95 of MSAs fitting a linear function got a better R square score and has been chosen for estimate future added population. Any of the mentioned ways can be used for setting values of population rates as input scenario assumptions to the model.

Figure 3.13 Fitting different functions to added population rates for forecast the trend

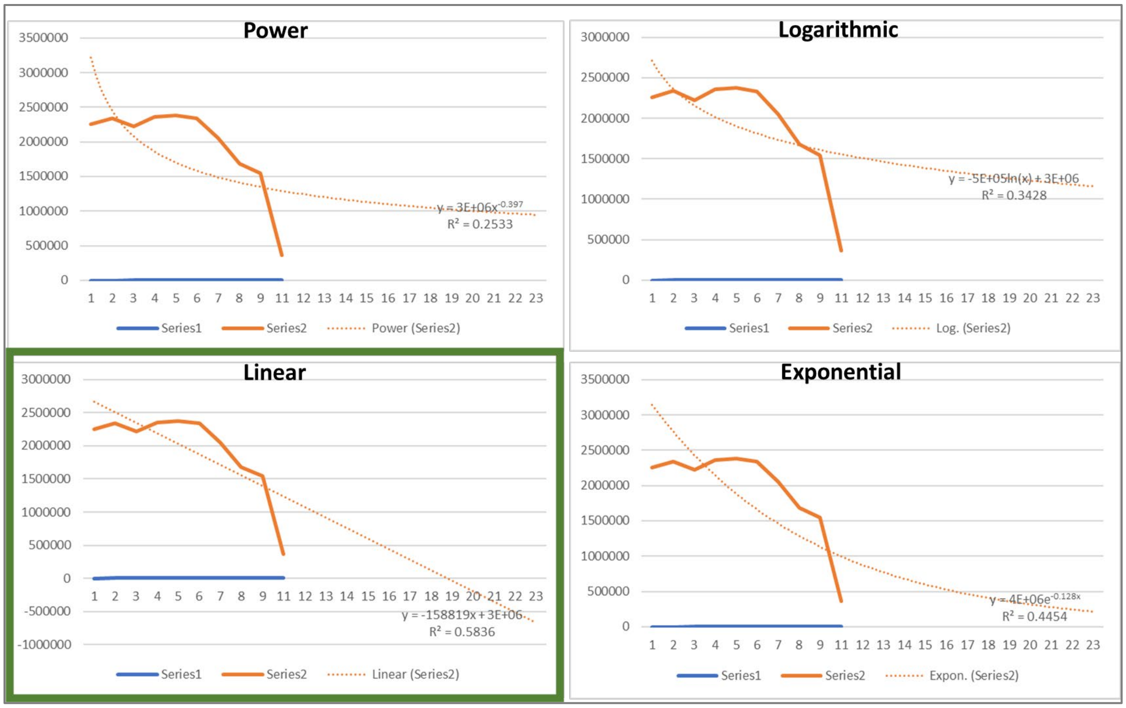


Figure 3.14 Fitting different functions to added population rates after removing outlier

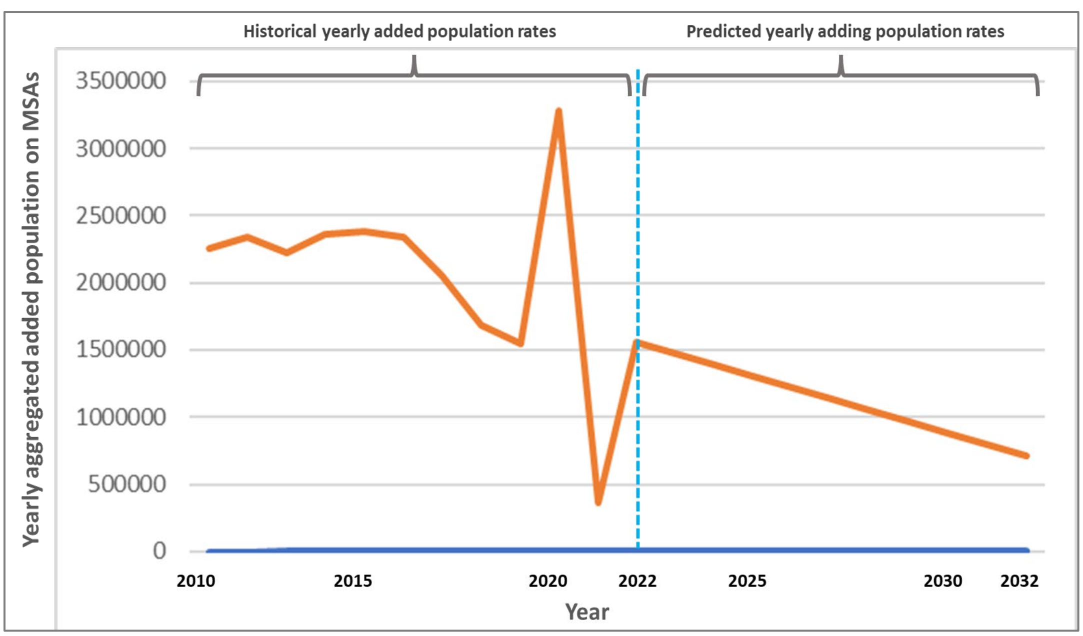


Figure 3.15 The aggregated rate of predicted adding population using best fitting function on each MSAs in the USA

3.3.2 MSAs' GDP per capita prediction

The key macroeconomic variables such as GDP are only available at high time granularity, such as quarterly, and are generally only published with a significant delay. In the U.S., the advance estimates of GDP and its components are only available a month after the reference quarter, and in some countries, the delays are longer. Traditionally, three sources of data have been considered for macroeconomic nowcasting[2] (Bantis et al., 2022): (i) hard indicators, such as retail sales and industrial production, (ii) surveys of opinions and intentions, and (iii) high-frequency financial market data.

However, in recent years, due to computer technology advancements and the advent of online information-gathering services, alternative data sources have become available, such as a big data. A popular source of big data for short-term macroeconomic forecasting is Google Trends, which provides information about the frequency with which a particular term is searched. Google Search data may contain insights into consumers' and other agents' plans and intentions, and perhaps especially consumer spending. Consumers may seek information on Google's search engine before making economic decisions regarding purchases. Consequently, Google Search data may constitute a valuable source of information for nowcasting macro-variables. Recently, Bantis et al. (2022) and Bouayad et al. (2022) developed approaches for forecasting GDP growth rates using Google Trends data. Moreover, special disruptive events such as the COVID-19 pandemic have caused global health impacts, and governments have restricted movements to a certain extent. Such restrictions have led to disruptions in

[2] Nowcasting in economics is the prediction of the present, the very near future, and the very recent past state of an economic indicator.

economic activities. Jena et al. (2021) have proposed a multilayer artificial neural network model to evaluate the impact of COVID-19 on GDP.

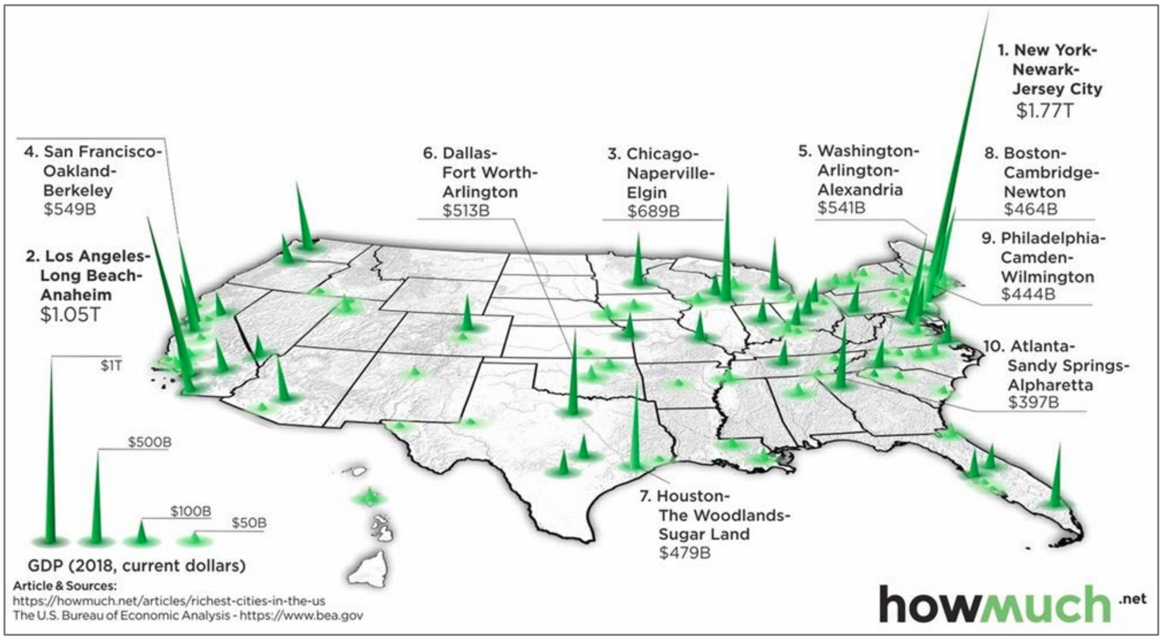


Figure 3.16 Visual comparison of MSAs on their GDP (2018 reported by howmuch.net)

On December 12, 2018, the Bureau of Economic Analysis (BEA) released prototype measures of gross domestic product (GDP) by county (see Figure 3.16). GDP by county is a measure of the market value of final goods and services produced within a county area in a particular period. While other measures of county economies rely mainly on labor market data, these statistics are the first of their kind to incorporate multiple data sources that capture trends in labor, revenue, and value of production (Panek et al. 2019).

In addition to all above approaches available in literature, we can estimate the predicted MSAs' GDP rates as one of the model's scenario assumptions by using ETS algorithm or fitting functions over historical rates of each MSA.

3.3.3 MSAs' Modular Attractiveness Weights

This section presents how to model so as to differentiate MSAs in terms of modular construction attractiveness. In general, the essence of the challenge is summarized through the following question: If we have a general scenario assumption on which the total percentage of aggregate demand will be modular, how can this proportion of demand be distributed on MSAs based on their potentials and characteristics which may be impactful on having modular preference?

For this purpose, we have developed the Geographically Weighted Modular Construction Attractiveness (GWMCA) algorithm. It uses MSAs features such as GDP per capita, labor/expert cost, Incentive / restrictive regulations for modular construction, access to materials, and distribution of demand volume (demand share among MSAs). Between these features, GDP, labor/expert cost, and incentive regulation have positive impact, and the rest of features access to material and restrictive regulations have negative impact on absorbing modular style of construction.

Algorithm (GWMCA): For each of given future years, the following steps result with the vector of MSAs' modular construction attractiveness weights.

Step 1: Build a matrix of corresponding MSAs' feature values (M) that the columns represent features and rows assigned to MSAs. Thus, M_{ij} is the value of the metropolitan statistical area *i* on feature *j* ($1 \le i \le l \ \ and \ 1 \le j \le k$ where *l* is equal to 384 here, and *k* is the number of features included). For the qualitative values such as low, high, etc., they have been replaced with numbers, For instance 1, 2, 3 instead of low, moderate, high.

Step 2: Build the cost matrix C from matrix M by normalizing it with below formula: if the feature j has a positive impact on demand for MC:

$$C_{ij} = \frac{M_j^{\max} - M_{ij}}{M_j^{\max} - M_j^{\min}} \tag{18}$$

And, if the feature j has a negative impact on demand for MC:

$$C_{ij} = \frac{M_{ij} - M_j^{\min}}{M_j^{\max} - M_j^{\min}} \tag{19}$$

Step 3: For each column of matrix C, compute the inverse of its entropy (Anand & Bianconi, 2009). σ_j is the jth kernel scale parameter which describes the substitution influence.

$$\forall_j \quad e_j = -\frac{1}{ln(l)} \sum_{i=1}^{l} C_{ij} ln(C_{ij}) \tag{20}$$

$$\forall_j \quad \sigma_j = \frac{1}{e_j} \tag{21}$$

Step 4: For each row of matrix C, calculate the density function which results from a Gauss Influence function (Hinneburg & Keim, 1998).

$$\forall_{1 \le i \le l} \quad r_i = \sum_{j=1}^{k} \exp(-\frac{\left(C_{ij}\right)^2}{2\sigma_j^2}) \tag{22}$$

Step 5: Normalize the vector of $R = [r_i]$, $1 \le i \le l$, and return it as vector W that is called the MSAs' modular construction attractiveness weights.

$$\forall_{1\le i\le l} \qquad w_i = \frac{r_i}{\sum_{i=1}^{l} r_i} \tag{23}$$

For a given future year (y), the demand rate of MSAs is indicated with d_i, $1 \le i \le l$, (l is equal to 384 metropolitan area here) and α is the percentage of total demand at this given year y which is likely to be modular (as one of scenario assumptions). The modular proportion of demand for metropolitan area *i* is d_i^{mod} and can be computed by:

$$\forall_{1\le i\le l} \qquad d_i^{\mathrm{mod}} = \beta * w_i * d_i \tag{24}$$

Where, the β is an adjustment coefficient computed by below equation:

$$\beta = \frac{\alpha * \sum_{i=1}^{l} d_i}{\sum_{i=1}^{l} w_i * d_i} \tag{25}$$

3.4 Implementation and results

In this study, the proposed demand model has been implemented for predicting modular construction demand over the US metropolitan statistical areas during the period of the year 2023 to 2032. For this purpose, we have used some historical data bases from the resources below:

- U.S. Census Bureau, Population Division (Population source 2010 to 2021)
- U.S. Census Bureau, Monthly seasonality of permits/starts of multifamily housing

- U.S. Census Bureau, Historical yearly permits rates over MSAs (2017-2021) for multifamily housing, hotel/motel, clinic/nursing, and dormitories, with characteristics of area (ft^2), value ($), number of units, number of stories, number of projects. Figures 3.16 and 3.17 present the aggregated level of these rates.
- Bureau of Economic Analysis (GDP source 2010 to 2021)
- Construction Laborers (bls.gov)[3], the US Bureau of Labor Statistics

Moreover, we had access to the experienced expertise from the modular team of MiTek[4] who provided their estimated values of the following scenario assumptions over MSAs:

- Incentive/Restrictive regulations
- Access to the materials
- Yearly percentage of the US potential construction demand, which is modular
- Set of module types and probability distribution for usage of each one on different types of building.

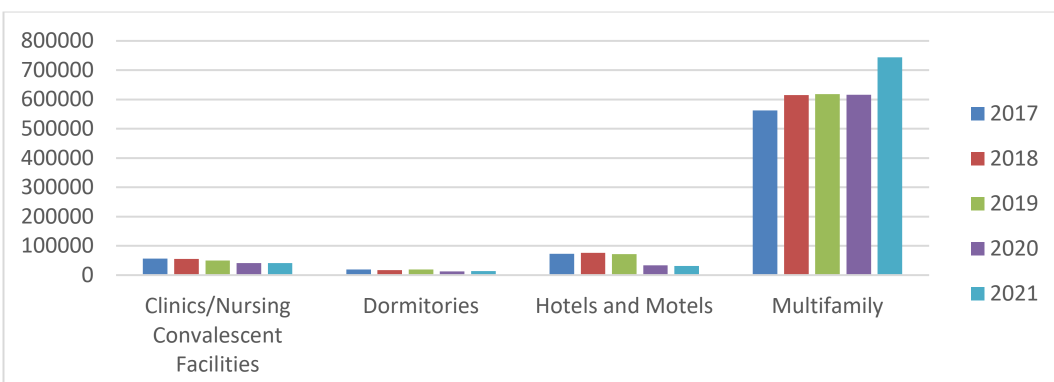


Figure 3.17 Historical construction area (1000 square feet) aggregated yearly level over MSAs

[3] https://www.bls.gov/oes/current/oes472061.htm#st
[4] https://www.mitek-us.com/

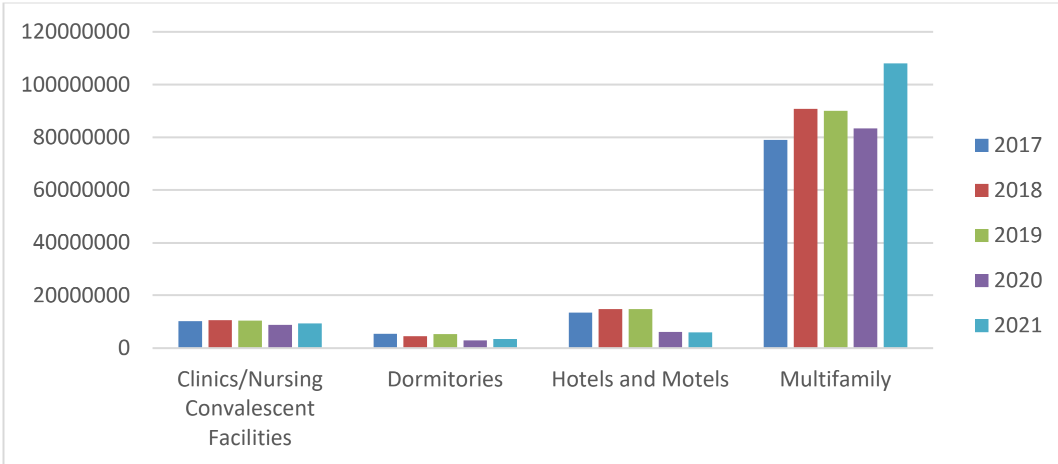


Figure 3.18 Historical construction value (1000 $$) aggregated yearly level over MSAs

A scenario-based Application (in PYTHON - TKinter package) with user interface, able to make scenarios, has been developed and its outputs are including:

- Scenario-based Potential/modular demand logs for yearly/monthly aggregated level in csv format (Table 3.2 shows its format)
- Scenario-based Potential/modular demand logs in daily level (see Table 3.3)
- Tableau Dashboard for visualization of scenario-based results on maps and graphs.

Table 3.2 A sample of yearly scenario-based predicted demand on MSAs resulted from proposed application (output in CSV format)

Year	MSA_State	State name	State_code	Type of building	StoryClass	PotentialDemand_Area_1000sqf	PotentialDemand_NumUnits	PotentialDemand_NumProjects	PotentialDemand_Value_1000Dollars	ModularpartofDemand_NumProjects
2023	New York-Northern New Jersey-Long Island, NY-NJ-PA	New York	NY	Multifamily	3-7 Strys	26619.05137	861133	1070	4739225.735	13.91
2023	Dallas-Fort Worth-Arlington, TX	Texas	TX	Multifamily	3-7 Strys	24108.33638	808526	357	2435421.4	4.641
2023	Austin-Round Rock, TX	Texas	TX	Multifamily	3-7 Strys	18139.73303	668963	397	1783940.815	5.161
2023	Unassigned	Unassigned	Unassigned	Multifamily	3-7 Strys	19854.84	606603	671	2244952.854	8.723
2023	Houston-Baytown-Sugar Land, TX	Texas	TX	Multifamily	3-7 Strys	12259.94651	444806	228	1219100.312	2.964
2023	Denver-Aurora, CO	Colorado	CO	Multifamily	3-7 Strys	11158.36371	411338	210	1538190.552	2.73
2023	Los Angeles-Long Beach-Santa Ana, CA	California	CA	Multifamily	3-7 Strys	15882.50561	355136	318	2762775.992	4.134
2023	Minneapolis-St. Paul-Bloomington, MN-WI	Minnesota	MN	Multifamily	3-7 Strys	12077.2715	345350	141	1487218.001	1.833
2023	Orlando, FL	Florida	FL	Multifamily	3-7 Strys	11879.7094	314617	258	1314129.067	3.354
2023	Miami-Fort Lauderdale-Miami Beach, FL	Florida	FL	Multifamily	8+ Strys	13414.74919	294712	47	2370401.775	0.611
2023	Seattle-Tacoma-Bellevue, WA	Washington	WA	Multifamily	3-7 Strys	10706.83723	294098	249	1439967.536	3.237
2023	Washington-Arlington-Alexandria, DC-VA-MD-WV	District Of Columbia	DC	Multifamily	3-7 Strys	12798.08229	280322	310	1759449.941	4.03
2023	Washington-Arlington-Alexandria, DC-VA-MD-WV	District Of Columbia	DC	Multifamily	8+ Strys	8258.356148	265878	32	1659549.594	0.416
2023	Boston-Cambridge-Quincy, MA-NH	Massachusetts	MA	Multifamily	3-7 Strys	11124.1053	261998	261	2256897.042	3.393
2023	Atlanta-Sandy Springs-Marietta, GA	Georgia	GA	Multifamily	3-7 Strys	12161.43753	219594	236	1426125.815	3.068
2023	Jacksonville, FL	Florida	FL	Multifamily	3-7 Strys	4673.547426	206568	97	502818.4383	1.261
2023	Phoenix-Mesa-Scottsdale, AZ	Arizona	AZ	Multifamily	3-7 Strys	11050.03472	199158	249	1205599.973	3.237
2023	San Francisco-Oakland-Fremont, CA	California	CA	Multifamily	3-7 Strys	8423.917329	197016	188	1463940.555	2.444
2023	New York-Northern New Jersey-Long Island, NY-NJ-PA	New York	NY	Multifamily	8+ Strys	21730.80007	196570	156	8208148.131	2.028
2023	Nashville-Davidson--Murfreesboro, TN	Tennessee	TN	Multifamily	3-7 Strys	7362.603395	192770	200	741726.8185	2.6
2023	Charlotte-Gastonia-Concord, NC-SC	North Carolina	NC	Multifamily	3-7 Strys	8407.819194	183129	252	851951.7508	3.276
2023	San Antonio, TX	Texas	TX	Multifamily	3-7 Strys	6871.26836	168929	209	673756.918	2.717
2023	Portland-Vancouver-Beaverton, OR-WA	Oregon	OR	Multifamily	3-7 Strys	6068.306527	168072	151	801594.2364	1.963
2023	Kansas City, MO-KS	Missouri	MO	Multifamily	3-7 Strys	4279.399208	160712	72	570173.0656	0.936
2023	Miami-Fort Lauderdale-Miami Beach, FL	Florida	FL	Multifamily	3-7 Strys	12016.90609	155250	271	1275778.898	3.523
2023	Tampa-St. Petersburg-Clearwater, FL	Florida	FL	Multifamily	3-7 Strys	5958.556393	151203	131	622647.0994	1.703
2023	Unassigned	Unassigned	Unassigned	Multifamily	1-2 Strys	7649.62	149977	799	789886.62	10.387
2023	Cincinnati-Middletown, OH-KY-IN	Ohio	OH	Multifamily	3-7 Strys	2379.7062	135051	47	287178.9572	0.611
2023	Columbus, OH	Ohio	OH	Multifamily	3-7 Strys	4717.563656	133975	120	458868.3309	1.56
2023	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	Pennsylvania	PA	Multifamily	3-7 Strys	10060.26715	132899	339	1369716.768	4.407
2023	Seattle-Tacoma-Bellevue, WA	Washington	WA	Multifamily	8+ Strys	4112.504877	131602	19	1003623.488	0.247
2023	San Diego-Carlsbad-San Marcos, CA	California	CA	Multifamily	3-7 Strys	4348.164991	107551	145	694342.3167	1.885
2023	Chicago-Naperville-Joliet, IL-IN-WI	Illinois	IL	Multifamily	3-7 Strys	6772.172073	81512	152	965778.5119	1.976
2023	Las Vegas-Paradise, NV	Nevada	NV	Multifamily	3-7 Strys	2620.116218	79135	42	249408.9484	0.546
2023	Salt Lake City, UT	Utah	UT	Multifamily	3-7 Strys	4001.814086	75139	57	507626.2167	0.741

Table 3.3 A sample of daily-level scenario-based predicted demand logs (projects) on MSAs resulted from proposed application (output in CSV format)

Project_ID	Date	Metropolitan_Statistical_Area	Zip_code	Type_of_building	Num_Units	Num_Floor	Area_1000sqf	Total_Mods	mod1:mod2:mod3	Expected_Start_Erection	Expected_durationdays_Erection
P_1	8/1/2023	Atlanta-Sandy Springs-Alpharetta, GA MSA	30307	Multifamily	53	5	100	159	30:70:59	11/1/2023	20
P_2	8/1/2023	Grand Rapids-Kentwood, MI MSA	70444	Multifamily	54	5	87	147	58:0:89	11/1/2023	19
P_3	8/1/2023	Baltimore-Columbia-Towson, MD MSA	21013	Clinics/Nursing Convalescent Facilities	140	7	500	386	306:80:0	11/1/2023	49
P_4	8/1/2023	New York-Newark-Jersey City, NY-NJ-PA MSA	56567	Dormitories	89	8	200	274	56:104:114	11/1/2023	34
P_5	8/1/2023	New York-Newark-Jersey City, NY-NJ-PA MSA	56567	Hotels and Motels	102	10	200	428	200:200:28	11/1/2023	54
P_6	8/1/2023	Eugene-Springfield, OR MSA	97401	Multifamily	20	2	18	65	30:35:0	11/1/2023	10
P_7	8/1/2023	Atlanta-Sandy Springs-Alpharetta, GA MSA	30307	Clinics/Nursing Convalescent Facilities	33	2	27	89	43:46:0	11/1/2023	12

3.5 Conclusion

Modular construction companies are seeking ways to upgrade their manufacturing and logistics systems to improve their production capability and maintain a competitive edge over traditional site-based construction. This study presents a framework to determine where a modular building construction company should invest its capital in maximizing profit and demand share using a novel mathematical model and machine learning techniques. This model considers all the MSA and states within the United States and multivariate datasets from the U.S. Census Bureau has been used to predict the demand in the future 10 years (2023-2032) based on economic and demographic features.

Using both the predicted potential total demand and the projected modular part of demand growth rates, a modular construction company has the basis for narrowing the list of potential MSAs for optimal investment and advertisement/incentive planning and contracting policies. These results also help to find the best location for setting the production hubs and storages, in the way to catch more market share with a lower cost of transportation and supply erection sites. Moreover, the predicted future number of potential projects and their size/value helps with long-term resource and capacity assignments planning. In this study, GDP per capita, labor/expertise cost, access to materials, and regulation have been considered to compute the attractiveness weight of MSA's markets for modular construction. The prediction of property tax rates and building price movements across MSAs in the United States are two other important features that can impact the attractiveness of a market for modular construction and can be added to the model in further research.

Qualitative considerations such as what the business knows about MSA, competitors in the area, existing vendor and builder relationships, and additional insider knowledge would be needed to improve the full operationalization of our proposed model. Once done, however, the data-driven results and generated scenario-based synthetic demand logs for the next 10 years (2023-2032) would provide a reliable means to identify where to invest and give the company a competitive advantage compared to competitors. It is also expected to support stakeholders in the modular construction industry to obtain the maximum benefits from adopting a modular approach for high-rise buildings.

3.6 References

Anand, K. and G. Bianconi, (2009) "Entropy measures for networks: Toward an information theory of complex topologies," Physical Review E, 80(4): p. 045102.

Arcadis. (2018). Tackling costs in the digital age: international construction costs. Amsterdam, Netherlands.

Bantis, E., Clements, M. P., & Urquhart, A. (2022). Forecasting GDP growth rates in the United States and Brazil using Google Trends. International Journal of Forecasting.

Barker, J., Ramos E., Rodgers J., and Saqib Shahzad. (2021) "Machine Learning Forecasting of Construction Demand in the USA as Indicated by Economic, Geographic and Demographic Cues." SMU Data Science Review 5, no. 2: 4.

Bouayad, I., Zahir, J., & Ez-zetouni, A. (2022). Nowcasting and Forecasting Morocco GDP growth using Google Trends data. IFAC-PapersOnLine, 55(10), 3280-3285.

Construction Industry Council. (2018). About modular integrated construction.p.1–7.

Fullerton, T., Luevano, J., & West, C. (2000). Accuracy of Regional Single-Family Housing Start Forecasts. Journal of Housing Research, 11(1), 109-120. Retrieved January 31, 2021, from http://www.jstor.org/stable/24833749

Gibb, A.G., (1999). Off-site fabrication: prefabrication, pre-assembly and modularisation. John Wiley & Sons.

Goodier C, Gibb A. (2007). Future opportunities for offsite in the UK. Constr Manag Econ. 25(6):585–595.

Gupta, R., Gupta, R., Miller, S., & Miller, S. (2012). The Time-Series Properties of House Prices: A Case Study of the Southern California Market. The Journal of Real Estate Finance and Economics, 44(3), 339–361. https://doi.org/10.1007/s11146-010-9234-7

Gupta, R., Marfatia, H.A., Pierdzioch, C. et al.(2021). Machine Learning Predictions of Housing Market Synchronization across U.S. States: The Role of Uncertainty. J Real Estate Finan Econ. https://doi-org.proxy.libraries.smu.edu/10.1007/s11146-020-09813-1

Hinneburg, A., & Keim, D. A. (1998) "An efficient approach to clustering in large multimedia databases with noise". KDD, Vol. 98, pp. 58-65).

Jena, P. R., Majhi, R., Kalli, R., Managi, S., & Majhi, B. (2021). Impact of COVID-19 on GDP of major economies: Application of the artificial neural network forecaster. Economic Analysis and Policy, 69, 324-339.

Krylovas, A., Kosareva, N., & Laura Gudelytė. (2011). Construction of social indicators using information measuring principles. case study of real estate prices simulation model. Lietuvos Matematikos Rinkinys, 52 doi:10.15388/LMR.2011.mt03

Linné, M., & Cirincionne, J. (2010). Integrating geographic information and valuation modeling for real estate. The Appraisal Journal, 78(4), 370-378.

Little, Jani S. 2004. "Goodness-of-Fit Measures." In The Sage Encyclopedia of Social Science Research Methods, ed. Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao. Thousand Oaks, California: Sage Publications, 435-437.

Mao, Y., Yao, N., & Zhang, M. (2014). Hangzhou Housing Demand Forecasting Model Based on B.P. Neural Network of Genetic Algorithm Optimization. Applied Mechanics and Materials, 587-589, 37–41.
https://doi.org/10.4028/www.scientific.net/AMM.587-589.37

McGibany, J.M., (1991). THE EFFECT OF PROPERTY TAX RATE DIFFERENTIALS ON SINGLE-FAMILY HOUSING STARTS IN WISCONSIN, 1978–1989. Journal of Regional Science, 31(3), pp.347-359.

McGraw Hill Construction. (2013). Safety Management in the Construction Industry: Identifying Risks and Reducing Accidents to Improve Site Productivity and Project ROI. Bedford, MA.

McKinsey Global Institute. (2017). Reinventing construction: a route to higher productivity. New York, United States.

McKinsey Global Infrastructure Institute. (2019), Modular construction: From projects to products. https://www.mckinsey.com/business-functions/operations/our-insights/modular-construction-from-projects-to-products#

Pan W, Hon CK. (2018). Modular integrated construction for high-rise buildings. Proc Inst Civ Eng - Munic Eng. 1–12.

Pan, W., Yang, Y., & Yang, L. (2018, April). High-rise modular building: ten-year journey and future development. In Construction Research Congress (pp. 523-532). New Orleans.

Panek, S., Rodriguez, R., & Baumgardner, F. (2019). Research spotlight new county-level gross domestic product. Survey of Current Business. https://apps. bea. gov/scb/2019/03-march/pdf/0319-county-level-gdp. pdf.

Raftery, A. E., & Ševčíková, H. (2021). Probabilistic population forecasting: Short to very long-term. International Journal of Forecasting.

Rentschler C., Mulrooney M., Shahani G. (2016). Modularization: the key to success in today's market. Hydrocarb Process. 95(12):27–30.

Talkhabi, H., Ghalehteimouri, K. J., Mehranjani, M. S., Zanganeh, A., & Karami, T. (2022). Spatial and temporal population change in the Tehran Metropolitan Region and its consequences on urban decline and sprawl. Ecological Informatics, 70, 101731.

Vespa, J., Armstrong, D. M., & Medina, L. (2018). Demographic turning points for the United States: Population projections for 2020 to 2060. Washington, DC: US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Wuni, I. Y., & Shen, G. Q. (2022). Towards a decision support for modular integrated construction: An integrative review of the primary decision-making actors. International Journal of Construction Management, 22(5), 929-948.

Wuni I.Y., Shen GQP, Mahmud AT. (2019). Critical risk factors in the application of modular integrated construction: a systematic review. Int J Constr Manag.:1–15.

Yang, Y., Pan, W., & Pan, M. (2017). Manufacturing of modular buildings: A literature review. Modular and Offsite Construction (MOC) Summit Proceedings.

Yang, Y., NIU, S., & Pan, W. (2019). Supply chains of high-rise modular buildings in high-density cities. In CIB World Building Congress 2019. Hong Kong.

APPENDIX A. RESULTS OF BIBLIOMETRIC ANALYSIS CORRESPONDING TO CHAPTER 1

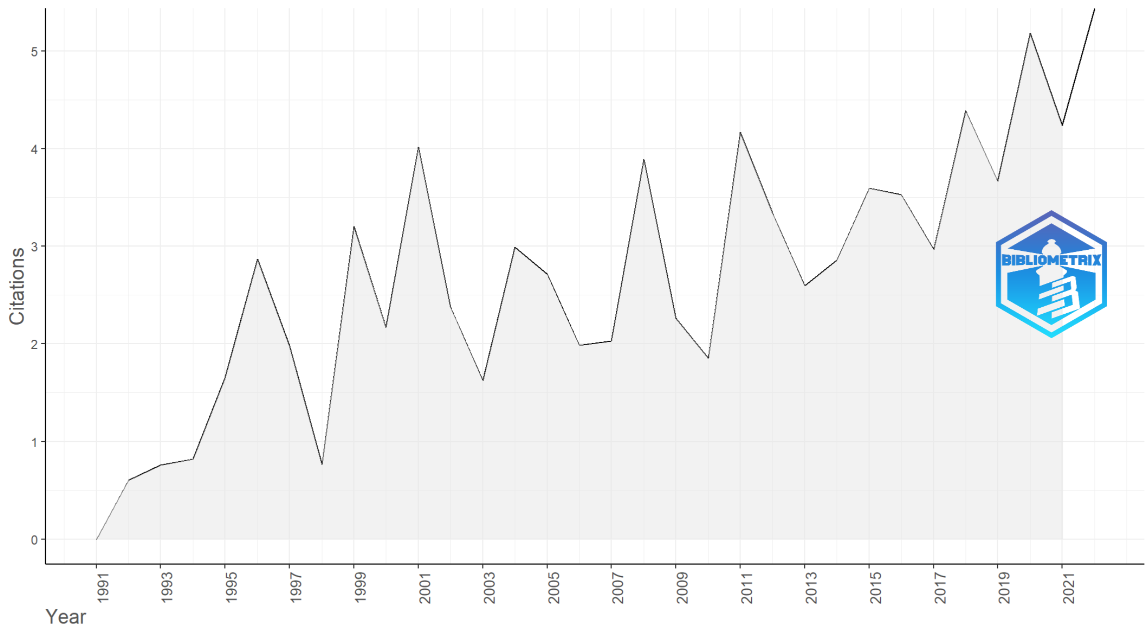


Figure A.1 Average article citations per year (result of bibliometric in chapter 1)

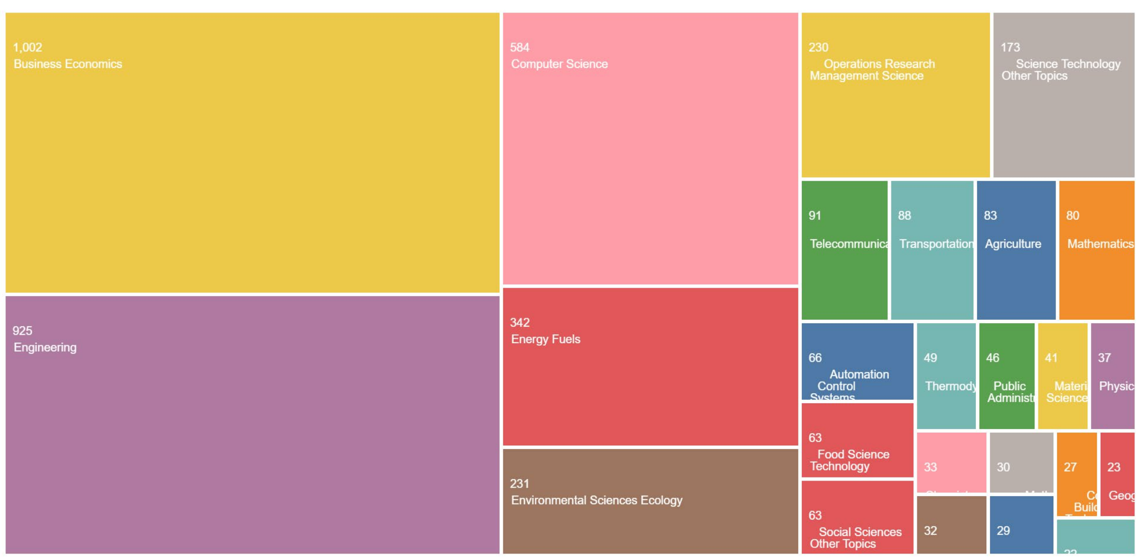


Figure A.2 Map of publications area (result of bibliometric in chapter 1)

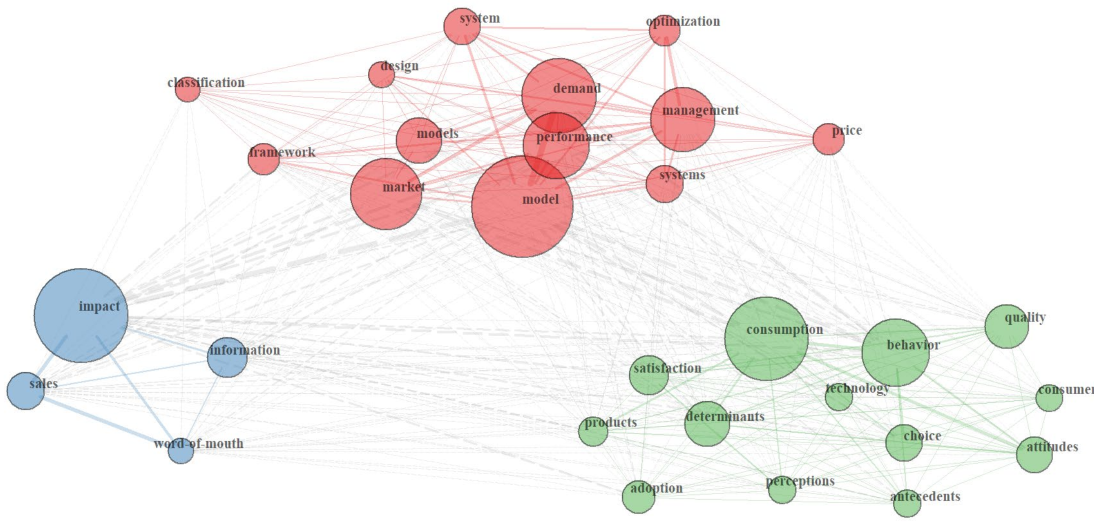


Figure A.3 Keyword co-occurrences (result of bibliometric in chapter 1)

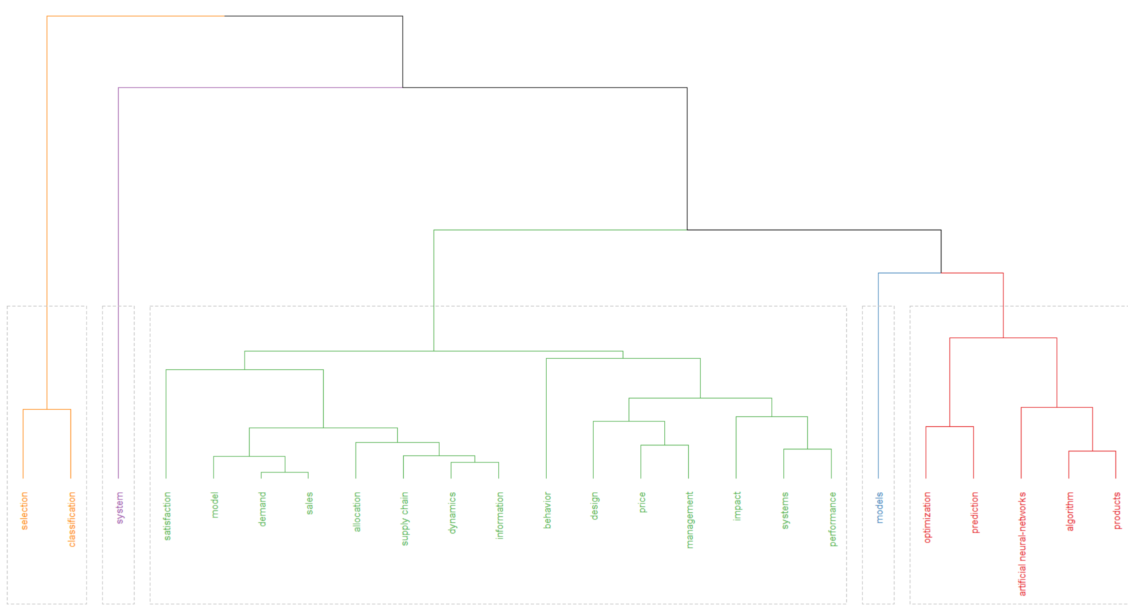


Figure A.4 Topic dendrogram (result of bibliometric in chapter 1)

APPENDIX B. RESULTS OF BIBLIOMETRIC ANALYSIS CORRESPONDING TO CHAPTER 2

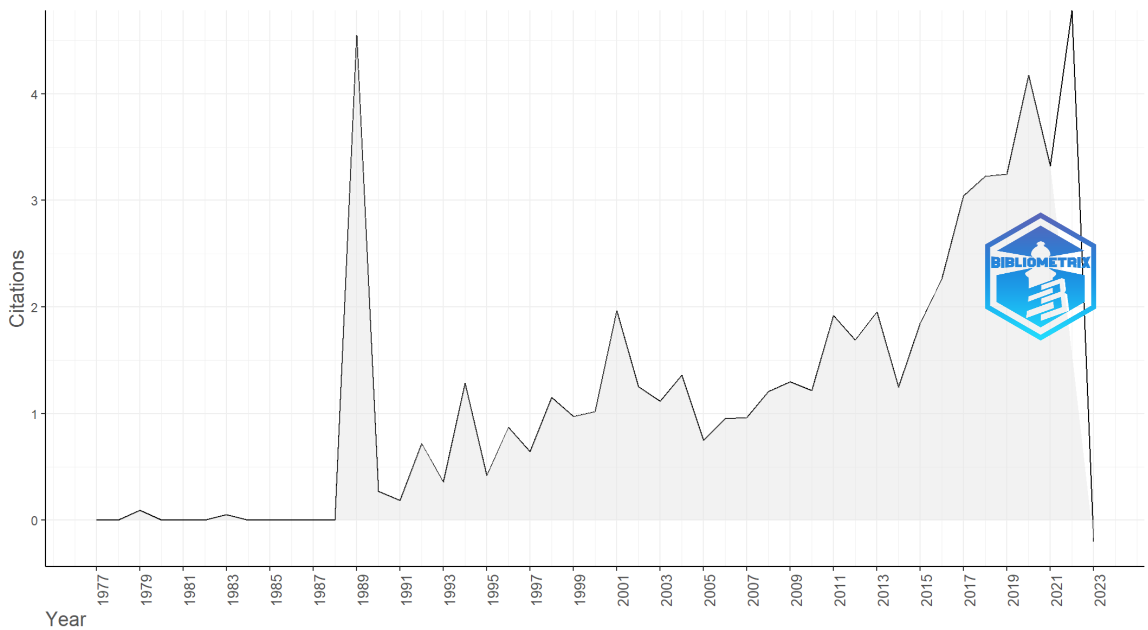


Figure B.1 Average article citations per year (result of bibliometric in chapter 2)

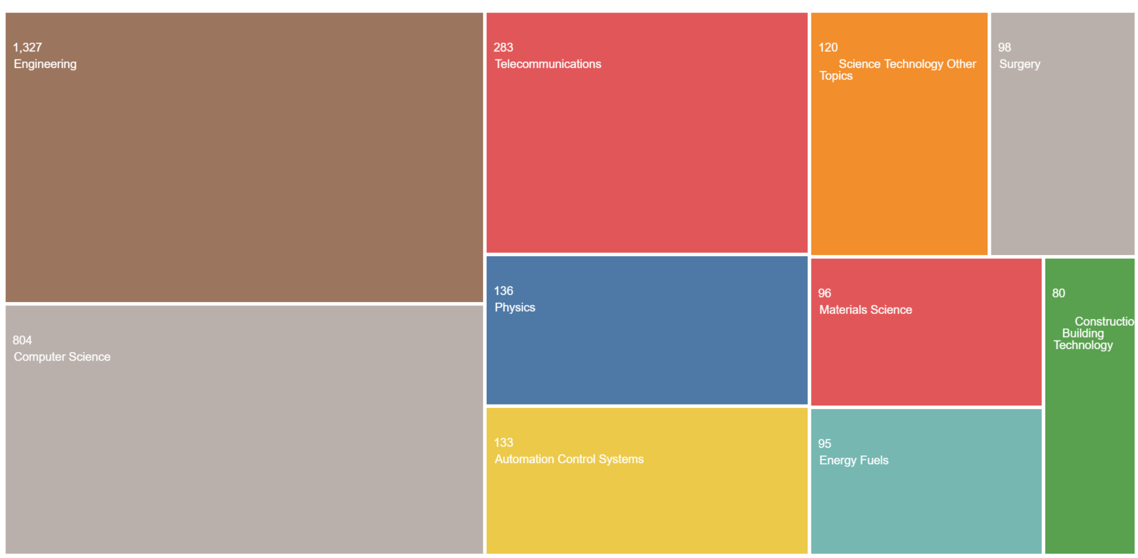


Figure B.2 Map of publications area (result of bibliometric in chapter 2)

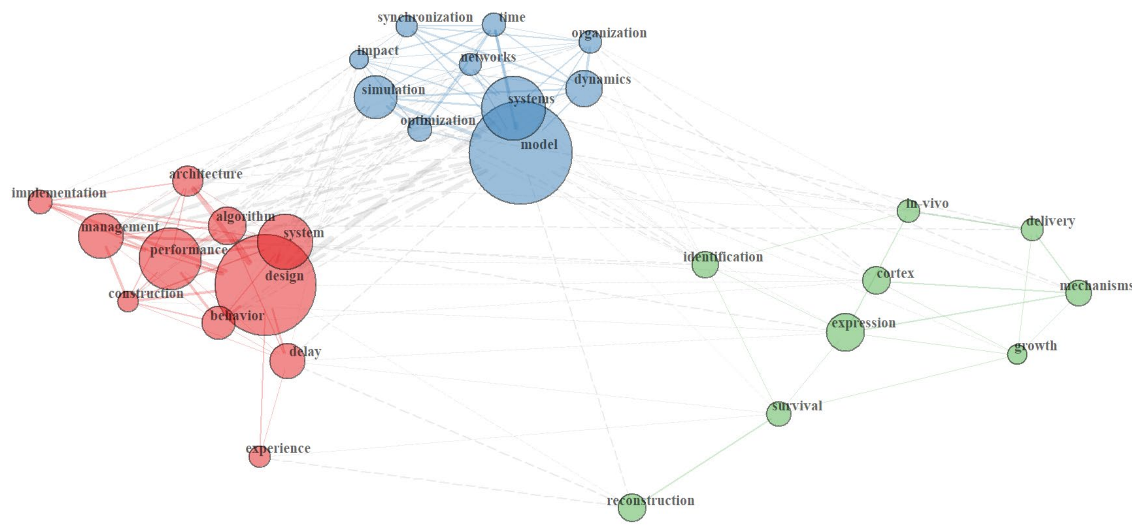


Figure B.3 Keyword co-occurrences (result of bibliometric in chapter 2)

APPENDIX C. RESULTS OF BIBLIOMETRIC ANALYSIS CORRESPONDING TO CHAPTER 3

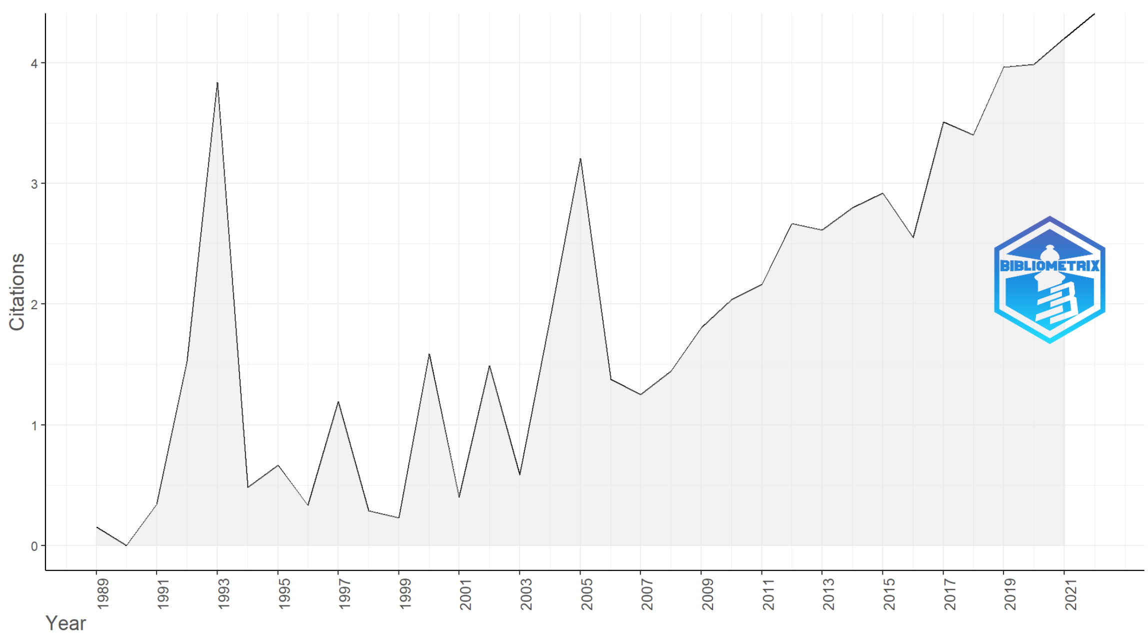


Figure C.1 Average article citations per year (result of bibliometric in chapter 3)

Figure C.2 Map of publications area (result of bibliometric in chapter 3)

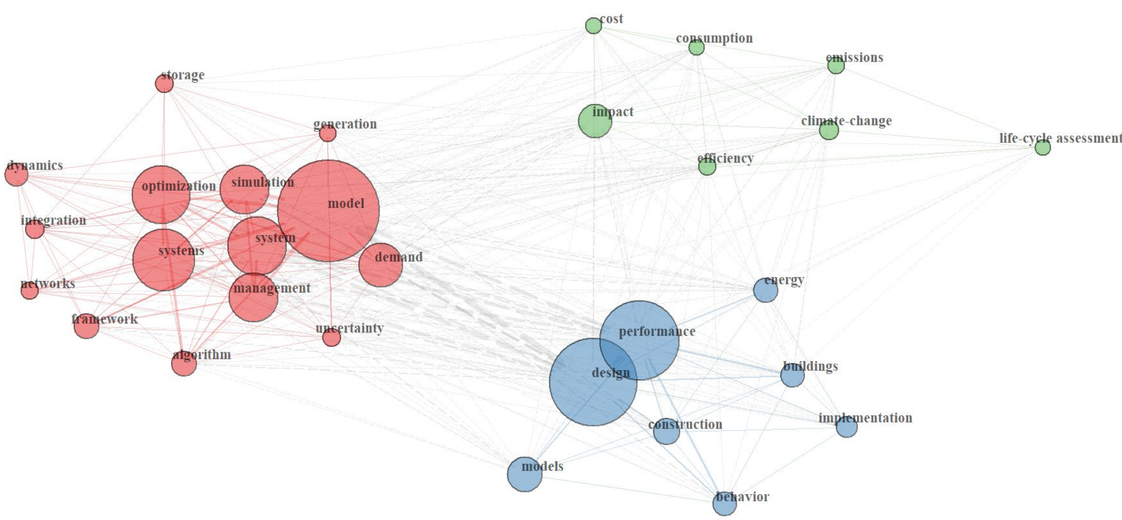


Figure C.3 Keyword co-occurrences (result of bibliometric in chapter 3)