UNIVERSITÉ DE MONTRÉAL

METHODOLOGY AND TOOLS TO MAKE PREDICTIONS FROM SPORADIC DELIVERY DATA

PAUL MURRAY

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR

(GÉNIE INDUSTRIEL)

MAI 2018

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

METHODOLOGY AND TOOLS TO MAKE PREDICTIONS FROM SPORADIC DELIVERY DATA

présentée par : MURRAY Paul

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment accepté par le jury d'examen constitué de :

M. FRAYRET Jean-Marc, Ph. D., président

M. AGARD Bruno, Doctorat, membre et directeur de recherche

M. BARAJAS VAZQUES Marco Antonio, Ph. D., membre et codirecteur de recherche

M. PELLERIN Robert, Ph. D., membre

M. OUHIMMOU Mustapha, Ph. D., membre externe

DEDICATION

To my wife Alison who gave constant support and encouragement every step of the way, and to my mother who has always believed that I can accomplish great things.

ACKNOWLEDGEMENTS

The authors would like to acknowledge our industrial partner (Air Liquide) and the National Sciences and Engineering Research Council of Canada (NSERC) for funding under grant RDCPJ 492021-15 and other support for this research.

RÉSUMÉ

Au cours de la révolution industrielle, les entreprises manufacturières ont vu naître la notion d’intégration verticale; elles ont acquis des matières premières qu'elles ont transformé en produits finis et livrés à leurs clients. Bien que l'intégration verticale ait été très efficace, à une certaine époque, en raison du contrôle centralisé de la qualité et de la production, elle a également conduit à la création de grandes organisations peu flexibles, qui évoluent difficilement et lentement, et souvent moins capables de tirer parti des technologies émergentes. Les technologies émergentes, les progrès en télécommunications et en transport ont permis aux entreprises de différentes régions d’améliorer leur collaboration, de produire plus efficacement et, ont finalement mené aux réseaux de production et à l'émergence de la gestion de la chaîne d'approvisionnement.

La gestion d'une chaîne d'approvisionnement nécessite une compréhension précise des exigences à tous les niveaux de la chaîne. Cependant, cette compréhension des besoins des partenaires de la chaîne d'approvisionnement dépend fortement du partage d'information entre eux. Le partage d'informations entre ces partenaires n'est pas toujours possible et le fournisseur est alors obligé de rechercher d'autres sources d'informations. Les fournisseurs peuvent par exemple disposer des données historiques provenant de leurs registres de livraison. On peut alors s'attendre à ce que ces données fournissent une bonne indication des besoins des clients. Dans la pratique, les registres de livraison sont mal adaptés pour prédire les exigences futures de la demande en raison de la relation non linéaire entre la consommation et les opérations de livraison.

Notre recherche a révélé plusieurs défis lors de la tentative d'interprétation de l'information recueillie à partir des données de livraison. Les données de livraison reflètent plus que les comportements de consommation des clients. Les décisions logistiques, telles que le calendrier, la fréquence de livraison, le volume et le nombre de camions, entre autres, sont reflétés dans les données de livraisons, malgré que ces décisions ne soient pas motivées par le client.

Une méthode pour extraire les informations de comportement de consommation à partir des données de livraison a donc été nécessaire. Un deuxième point est de savoir comment gérer des prédictions pour une large population de clients. La globalisation de tous les besoins de production présente une vue d'ensemble de l'organisation, mais peu de connaissances sont révélées sur les comportements de consommation individuels. Enfin, même lorsque les prédictions sont faites à un niveau global, il est besoin d’une méthode pour appliquer ces prédictions au niveau individuel de

chaque client. Dans cette recherche, nous proposons une méthode pour calculer des prévisions au niveau individuel de chaque client à partir d'un grand ensemble de données globales.

La littérature est unanime quant au fait que le partage d'informations collaboratif au sein d'une chaîne d'approvisionnement est bénéfique, mais les auteurs reconnaissent également que d'autres données doivent parfois être substituées, et que ces données peuvent être corrompues ou faussées par des effets de globalisation et d'amplification. Il y a une lacune dans la littérature quant à la façon d'interpréter les données et de les rendre utiles pour l'analyse. Nous répondons à cette lacune en proposant une méthode de substitution des données de livraison aux données de consommation.

Nous trouvons également une lacune dans les écrits concernant la segmentation du marché qui utilise généralement des variables descriptives pour distinguer le niveau de similitude entre les clients. Les auteurs ne traitent pas de la façon d'établir des segments lorsque les variables descriptives ne sont pas disponibles. Nous comblons cet écart en proposant une méthode qui établit des segments de marché en fonction du comportement passé démontré. La littérature sur la segmentation de marché se concentre sur le découpage d'une population en segments pour faciliter l'analyse comme la prévision. Il y a peu de conseils sur la façon de désagréger des données et d'appliquer les analyses précédentes aux clients individuels. Nous avons proposé une méthode pour cela. Enfin, pour tenter de combler le besoin d'une méthode de validation des résultats de la segmentation du marché, nous proposons une solution qui établit les segments en fonction du comportement démontré et qui vérifie ensuite si les attributs descriptifs peuvent aboutir à des résultats de segmentation similaires.

Un jeu de données réel est utilisé dans cette recherche pour tester les méthodes proposées. L'ensemble de données comprend les données de livraison d'un fournisseur pour l'ensemble de ses clients pendant plus de cinq ans; plus d'un million d'événements de livraison sont inclus. Les données ont été triées pour éliminer les valeurs aberrantes, laissant 75% des données brutes et 3000 clients uniques pour l'étude de cas.

Les composants de notre recherche sont présentés en quatre parties qui fonctionnent ensembles pour résoudre le problème général. Chaque composant a cependant des applications potentielles dans d'autres domaines et pourrait être utilisé pour résoudre d'autres types de problèmes.

Dans la première partie, les données sont préparées pour l'analyse. Les premières tentatives pour résoudre le problème de la recherche supposaient que l'ensemble de données brutes pourrait simplement être divisées en tranches mensuelles et ensuite utilisées pour élaborer une prévision. Les résultats étaient extrêmement diffus à tel point qu’aucune information n'a été révélée. Nous avons proposé une méthode pour résoudre ce problème. La deuxième partie aborde le problème du nombre trop important de clients pour permettre une analyse prévisionnelle individuelle. Nous avons proposé une méthode pour segmenter les clients en fonction de leurs comportements démontrés. La troisième partie de notre recherche est une méthode permettant de générer des prévisions par segment, puis d'appliquer ces prévisions à des clients individuels.

Dans la dernière partie de la recherche, nous tentons de valider et d'améliorer la méthode en intégrant des variables externes telles que le climat, l'emplacement et les caractéristiques propres au domaine industriel concerné. Nous pensions que les comportements étaient influencés par ces facteurs. Les résultats montrent qu'il existe en réalité très peu de corrélation entre les comportements réels des clients et ces attributs. Ceci est surprenant sachant que la segmentation des clients basée sur des attributs descriptifs est une pratique commerciale courante.

Les contributions de cette recherche sont importantes dans trois catégories : méthodologique, scientifique et pratique. La stratégie méthodologique utilisée ici démontre que les nouveaux problèmes n’impliquent pas nécessairement le besoin de nouveaux outils. Nous commençons avec un problème d'entreprise et recherchons des outils établis pour le résoudre. Bien que les outils ne soient pas nouveaux ou uniques, leur combinaison et leur application l'est.

Sur le plan scientifique, nous proposons un cadre d'étapes interconnectées pouvant être appliquées séquentiellement pour résoudre un problème métier complexe. Un ensemble de données volumineuses, globales et stochastiques est trié, interprété et transformé en une solution offrant des informations prévisionnelles. Les différentes étapes proposées peuvent également être utilisées individuellement et appliquées dans d'autres domaines pour aider à résoudre d'autres types de problèmes.

L'étude de cas qui a inspiré cette recherche est un vrai problème fourni par notre partenaire industriel. Les méthodes proposées dans cette recherche permettent de trier les données, de supprimer les informations corrompues ou faussées et d'afficher des résultats exploitables. Une fois que les modèles de comportement sous-jacents peuvent être vus, la situation de l'entreprise peut être mieux cernée, et les connaissances nouvellement disponibles peuvent aider à prendre des décisions d'affaires.

La dernière partie de la recherche est importante dans sa rupture d'un paradigme. Beaucoup d'entreprises utilisent dans la prémisse de leur planification d'entreprise, que les attributs descriptifs sont essentiels pour prédire les comportements des clients. Nos résultats montrent que ces types d'attributs ne sont pas nécessairement très clairement corrélés avec le comportement de consommation, notamment quand il y a du biais important lié au caractéristiques intrinsèques du fonctionnement de l'entreprise.

La recherche présentée ici forme un cadre pour acquérir des connaissances à partir d'un ensemble de données brutes qui sont inutilisables en l'état. L'étude de cas fournit une méthode pour mettre en œuvre le cadre proposé et un ensemble viable de résultats est produit.

ABSTRACT

Managing a supply chain requires an accurate understanding of the requirements at all levels of the chain; understanding requirements of the supply chain partners is therefore highly dependent on information sharing between partners. Information sharing, however, is not always possible and the supplier is forced to look for other sources of information. Suppliers usually have historical data from its delivery records which can be expected to provide a good indication of the customers' requirements. In practice, delivery records do not perform well for predicting future demand requirements due to the non-linear relationship between delivery transactions and consumption.

Delivery records reflect more than just the customers' consumption behaviors. Logistics decisions, such as timing, frequency, and volume of deliveries are also reflected in the delivery records. A method to extract the consumption behavior information from the noisy data is necessary. A second challenge is how to manage predictions for a large population of customers. Aggregating all production requirements together presents a high-level view of the organization, but little knowledge is revealed regarding consumption behavior. Lastly, once predictions are made at an aggregated level, a method to apply the predictions at the customer level is lacking. In this research, we propose a method for developing customer level forecasts from a large, noisy dataset.

Our research has revealed several gaps in the literature which we propose to address. The literature is unanimous in opinion that collaborative information sharing within a supply chain is beneficial, but substitute data must sometimes be used; that data may be corrupted or noisy due to aggregation and bullwhip effects. We address a gap in the literature as to how to address the noise in the data and make it useful for analysis.

We also find a gap in the literature regarding market segmentation which generally utilizes descriptive variables to distinguish the level of similarity between customers. The literature does not address how to establish segments when descriptive variables are not available. We address this gap with our proposed method that establishes market segments based on demonstrated past behavior. The literature on market segmentation all focusses on combining a population into segments to facilitate analysis such as forecasting. There is little guidance on how to de-segment and apply those subsequent analyses to the individual customers. We proposed a method for that. Finally, in attempt to address the gap of a method to validate market segmentation results, we

propose a method that establishes segments based on demonstrated behavior and then test whether descriptive attributes can achieve similar segmentation results.

A real dataset is used in this research to test the proposed methods. The dataset consists of a supplier's delivery records for all its customers for over five years; more than one million delivery events are included. The data was cleaned to remove outliers leaving 75% of the raw data and 3000 unique customers for the case study.

The components of our proposition are presented in four parts that work together for solving one specific problem. Each component has potential applications in other domains and might be utilized in solving other types of problems. Despite their individual uniqueness, the four parts are also sequentially dependent on their preceding part.

The research presented here forms a framework for gaining knowledge from an otherwise unusable dataset. The case study provides a platform for validating the proposed framework and a viable set of results is produced.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

α	Alpha (smoothing coefficient)
ADIDA	Aggregate, Disaggregate Intermittent Demand Approach
AL	Air Liquide
ANN	Artificial Neural Networks
ARIMA	Auto Regressive Integrated Moving Average
ARIMAX	Auto Regressive Integrated Moving Average (extended)
ASACT	Aggregate, Smooth, Aggregate, Convert to Time-series
CCor	Cross Correlation
CFPR	Continuous Forecasting, Planning & Replenishment
CR	Continuous Replenishment
DTW	Dynamic Time Warping
GIS	Geographical Information Systems
HAC	Hierarchical Agglomerative Clustering
K	Number of Clusters
Lin	Liquid Nitrogen
Lox	Liquid Oxygen
MAE	Mean Absolute Error
MPE	Mean Percent Error
MAPE	Mean Absolute Percent Error
ME	Mean Error
RFM	Recency, Frequency, Monetary
RMSE	Root Mean Squared Error
SC	Supply chain

SCM	Supply chain management
SES	Simple Exponential Smoothing
SOM	Self-Organizing Maps
UCM	Unobserved Component Models
VMI	Vendor Managed Inventory

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

During the industrial revolution, manufacturing firms were vertically integrated; they acquired raw materials, produced their products, and delivered to their customers. Firms such as Ford Motor Company purchased timber, iron ore, and rubber, and relied on their own internal resources to transform, fabricate, and assemble finished products, often in one large location such as Ford's Rouge River plant (Ford, 2017). While vertical integration was efficient due to centralized control of quality and production it also led to large, inflexible organizations that were slow to change and unable to leverage emerging technologies and alternate sources of labor (Lummus & Vokurka, 1999). Emergent technologies including pallets, forklift trucks, and standard-sized shipping containers enabled more efficient movement of products which allowed the distribution of production activities. Meanwhile, advances in telecommunications and reduction in cultural barriers made it possible for firms from different regions and nations to collaborate. Rather than the traditional vertical integration, firms could now look externally for their material and production requirements. Production de-integration led to the evolution of modern day supply chain management (SCM) (Lummus & Vokurka, 1999). As firms became more and more specialized in their core activities, they began to find benefit from establishing collaborative relationships with firms whose specializations were compatible. The traditional between-firm competition began to transform to a new between-supply-chain competition (Christopher, 2000).

In its most basic definition, supply chains (SC) create value by transforming and transporting goods and services that satisfy the demand requirements of downstream partners (Janvier-James, 2012). This definition omits that the success of a SC depends on its ability to leverage strengths and opportunities among a variety of partners in various locations. The successful SC quickly transforms itself to include or remove partners as necessary to suit the products being produced and the marketplace in which it competes. Managing the SC therefore requires an accurate understanding of the product requirements at all levels of the chain (Carbonneau, Laframboise, & Vahidov, 2008); understanding requirements of SC partners is therefore highly dependent on information sharing between partners. The concept of understanding SC requirements, also referred to as demand forecasting, is foundational to our research.

In their seminal paper, Angulo, Nachtmann, and Waller (2004) state that collaborative information sharing was prerequisite for establishing a successful SC, especially in a vendor managed inventory

(VMI) arrangement where the supplier is responsible for managing its customers' inventory. Subsequent research, however, found that in some VMI arrangements, the SC partners cannot or will not share information (Holweg, Disney, Holmström, & Småros, 2005). Communication technologies between SC partners can be incompatible or non-existent leading to physical barriers to communication (Hernández, Mula, Poler, & Lyons, 2014). Concern of risk, or lack of trust, confidentiality agreements, antitrust laws, and cost of information can also form a barrier to information sharing (Angulo et al., 2004; Kembro & Näslund, 2014). When collaborative information is not available, the supplier must look to other information to understand its SC requirements.

In the absence of external, down-stream SC data, a supplier normally has its own delivery records. In a VMI arrangement where exclusive, long-term supply contracts exist, the delivery records can be expected to provide a good indication of the customers' consumption behavior. In practice, delivery records are poorly suited for predicting future demand requirements due to the non-linear relationship between delivery transactions and consumption.

Our research has revealed several challenges when attempting to interpret the information gleaned from delivery records. Delivery records reflect more than just the customers' consumption behaviors. Logistics decisions, such as timing, frequency, and volume of deliveries are reflected in the delivery records; these decisions are not driven by the customer. Moreover, it is well established that as a supply chain decision point moves upstream from the point of consumption, the bullwhip effect increases (Forrester, 1958). Delivery data in its raw form contains a level of noise sufficiently high as to preclude detection of behavior patterns. A method to extract the consumption behavior information from the noisy data is necessary. A second challenge revealed is how to manage predictions for a large population of customers. Aggregating all production requirements together presents a high-level view of the organization, but little knowledge is revealed regarding consumption behavior. Lastly, even when predictions are made at an aggregated level, a method to apply the predictions at the customer level is lacking. In this research, we propose a method for developing customer level forecasts from a large, noisy dataset.

This research has provided important methodological, scientific and practical contributions. In the methodology, we show how a complex problem can be solved through the combination and reapplication of existing tools and methods. The scientific contribution is a solution to solving a complex business problem where a large stochastic data set is used to generate useful forecast information. And finally, the practical contribution is a tool that can be applied in the case study domain or other similar domains where long-range forecasts are needed, and the available data is limited.

Chapter 2 presents the state of the art relating to relevant areas of SC management and data analytics. Chapter 3 presents our research approach, explains the case study used for the research, and describes the detailed structure of the thesis. Chapters 4 through 7 are the sequential components of the research; these components have been submitted individually to peer reviewed journals for consideration and publication. Discussions, conclusions, and recommendations are offered in Chapters 8 and 9.

CHAPTER 2 CRITICAL LITERATURE REVIEW

2.1 Introduction

The research objective is to develop a method to forecast a SC's demand requirements based on limited historical data. An accurate forecast will aid the planning of the firm's resources and save money through increased efficiency. This research problem originates from a real industrial situation, it begins with a real dataset and concludes with a detailed methodology that permits to extract information that can help a firm to understand its SC requirements. The solution to the research problem involves several different steps that relate to different fields of research. SCM is an extremely broad field with many subordinate areas of research. In Section 2.2, we review those SC areas that are relevant to our research to establish context and highlight relevant gaps. In addition to understanding the relevant SC topics, it is also important to understand how the large and noisy data can be managed. Section 2.3 reviews the relevant data analytics.

Regarding VMI, the literature is unanimous in opinion that collaborative information sharing is beneficial, but they also acknowledge that other data must sometimes be substituted, and that that data may be corrupted or noisy due to aggregation and bullwhip effects. There is a gap in the literature as to how to address the noise in the data and make it useful for analysis. We address this gap in chapter 4 which proposes a method to substitute delivery data when consumption data is not available.

Market segmentation is the second step in our proposed method. The literature contains extensive information regarding segmenting the market based on a set of descriptive variables. There is acknowledgement that the accepted strategies do not always produce clusters with homogeneous behavior patterns, but it is not extensively address. Further, little research has been undertaken to develop methods to test cluster validity. The literature does not address how to establish segments when descriptive variables are not available. We address these gaps with our proposed method that establishes market segments based on demonstrated past behavior. This method, which relies on the data smoothed by the method in chapter 4 is described in chapter 5.

The literature on market segmentation all focusses on combining a population into segments to facilitate analysis such as forecasting. There is little guidance on how to de-segment and apply those subsequent analyses to the individual customers. We proposed a method for that in chapter 6 and evaluate its performance.

Finally, in attempt to address the gap of a method to validate market segmentation results, we propose a method that establishes segments based on demonstrated behavior and then test whether descriptive attributes can achieve similar segmentation results. This method and its results are presented in chapter 7.

2.2 Supply Chain Management

Supply chain management (SCM) emerged when firms recognized that they are no longer able to operate in a vertically integrated structure. Interest in SCM has steadily increased since the 1980's when firms began to recognize the benefits to between-firm collaborations (Lummus & Vokurka, 1999). There have been many attempts to provide a detailed and encompassing definition of SCM (Cooper & Ellram, 1993; Janvier-James, 2012; Lummus & Vokurka, 1999), however, we prefer a simplified definition that captures the most important components while avoiding the constraints of excessive description and details, as follows: *"SCM is the management of the flow of products (or services), money, and information among collaborating firms who move and process the products from raw materials through to the finished goods that are provided to the consumer."* The first two components, products and money, are universally measured and management of them follows well established practices. The third component, information, is not easily quantified and management of its flow can be subjective and constrained (Hernández et al., 2014). Management of information within a supply chain is a key component of understanding SC requirements.

Traditional SCM follows push / pull models (Chopra & Meindl, 2007). In "push" SCM, the supplier produces its products and then makes them available to the down-stream members of the supply chain, often placing finished goods into inventory where it awaits purchase from the consumer. In "pull" SCM, the downstream supply chain member, or consumer, determines its needs and places an order with the upstream supply chain member. The push / pull models were robust strategies in that they did not rely heavily on the information component of SCM; firms produced their products when they experienced a trigger (Chopra & Meindl, 2007). In the 1990s, supply chain collaboration began to emerge in several forms, including vendor managed inventory

(VMI); collaborative forecasting, planning and replenishment (CFPR); and continuous replenishment (CR) (Holweg et al., 2005).

2.2.1 Supply Chain Collaboration

SC partners can collaborate at many different levels including product design, promotions, inventory management, forecast planning, and risk management (Holweg et al., 2005). Since this research is focused on demand forecasting, we will consider only collaboration of inventory management and forecast planning. Holweg (2005) categorized levels of collaboration according to Figure 2.1. When neither planning nor inventory information is shared (Type 0), collaboration does not exist. With Type 1 collaboration, information is shared, and the firms use it to their mutual benefits. With Type 3 collaboration, product flow is synchronized by sharing information about inventory and planning. Type 2 collaboration, VMI, can be complex. In a VMI arrangement, several types of information may be shared between firms. Point-of-sale data, promotion planning, and inventory level information are beneficial for understanding the expected demand requirements. Regardless of how information is shared, the key feature of VMI is that responsibility for inventory management moves upstream to the supplier. The critical point here is that while planning information is beneficial, it is not necessarily shared; critical information from the point of consumption may not be visible to the supplier.

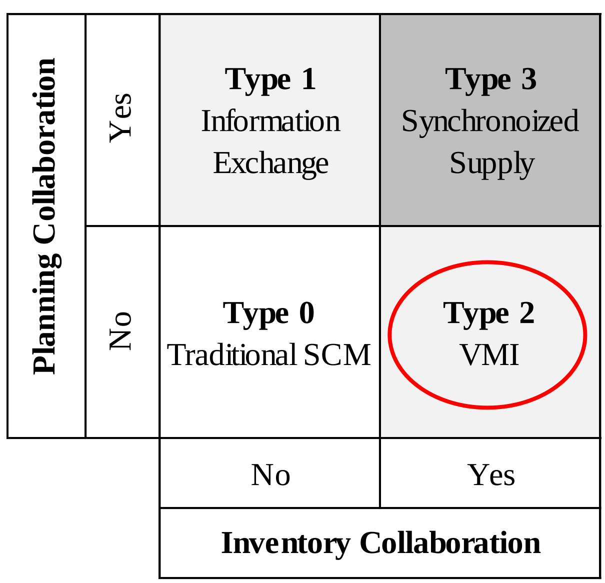


Figure 2-1: Levels of SCM Collaboration (Holweg et al., 2005)

Research has shown that VMI strategies are effective at increasing SC efficiency, reducing overall costs in the SC, and increasing SC competitiveness (Achabal, McIntyre, Smith, & Kalyanam, 2000; Jung, Chang, Sim, & Park, 2005). Despite a direct link between SC performance and information sharing (Forslund & Jonsson, 2007), SC partners are sometimes unwilling or unable to share useful forecasting information (Holweg et al., 2005; Kembro & Näslund, 2014). In VMI, the responsibility for forecasting ultimately lies with the supplier However, research has shown that supply chain performance can be improved when the information is collaboratively generated through arrangements such as collaborative planning, forecasting, and replenishment (CPFR) (Ramanathan & Gunasekaran, 2014). Ideally, the supplier in a VMI strategy has access to downstream information sources including consumption rates, inventory levels, and forecasts (Achabal et al., 2000).

In a VMI arrangement where downstream forecast information does not exist, the supplier must develop a forecast with other information, typically extracted from historical data or use qualitative information.

2.2.2 Supply Chain Information Sources

VMI requires a source of demand information that is available and timely (Angulo et al., 2004), ideally recorded at the point of consumption. Point of consumption information is not always available, downstream SC partners may be unwilling to share information due to confidentiality, trust, or regulatory concerns. Collecting and transmitting data may be costly or inconvenient, or the amount of data may be too large to process due to high number of customers and transaction events (Holweg et al., 2005).

When consumption-level data is not available, an upstream data source such as delivery records must be used. The information contained in upstream data may be diminished due to aggregation of individual transactions and temporal aggregation; the delivery record becomes a summary of all transactions that occurred since the previous delivery. Moving the data acquisition point upstream has the additional problem of induction of noise due to the bullwhip effect (Biswas & Sen, 2016; Forrester, 1958). Using upstream data for demand forecasting may exasperate the noise problem since the bullwhip effect tends to increase as the data gathering point moves upstream from the point of consumption (F. Chen, Drezner, Ryan, & Simchi-Levi, 2000). In their seminal paper, Lee, Padmanabhan, and Whang (1997) provide a more detailed explanation of the causes of the bullwhip effect including demand signal processing, order batching, price fluctuation, and rationing. Lastly, when deliveries occur on irregular time intervals and consist of irregular quantities, the resulting data when aggregated into temporal bins can appear lumpy, irregular, and intermittent (Petropoulos, Kourentzes, & Nikolopoulos, 2016), making the data difficult to work with. Despite that delivery records may be directly linked to consumption requirements; the associated noise can make them extremely difficult to work with.

2.2.3 Market Segmentation

When attempting to understand the SC requirements of a large customer population, industry and academics alike often resort to dividing the population into groups and then analyzing the groups instead of individuals (Wedel & Kamakura, 2012). A common term for this grouping is "market segmentation". Smith (1956) was one of the first to discuss the benefits of market segmentation in his seminal paper on marketing strategies, although he proposed that it should be executed through trial and error. A more advanced segmentation strategy was proposed in the 1970s where determinant attributes were scored to determine segment membership (Anderson, Cox, & Cooper, 1976). The determinant attributes were customers' responses to a series of questions; the results were subject to bias by the selection and scoring of the questions. Researchers in the 1990s were observing that the outcomes of segmentation analysis were not necessarily resulting in homogeneous clusters (Dibb & Simkin, 2001). Although the literature contains some research on methods for validating market segmentation (Huang, Cheung, & Ng, 2001), the topic has not been extensively explored. Despite some criticism, and limited research on methods to validate it, market segmentation is widely accepted in industry and academia (Rigby & Bilodeau, 2015).

2.3 Data Analytics

Data mining emerged as a popular field of research in the late 1990s (Choudhary, Harding, & Tiwari, 2008), it has since been applied in many areas of research including SCM. Data mining has many descriptors such as "big data analytics", "data mining", "data analytics", "data science", "knowledge discovery from databases", and many others including combinations of the ones listed here (Kantardzic, 2011; Manyika et al., 2011; Provost & Fawcett, 2013; G. Wang, Gunasekaran, Ngai, & Papadopoulos, 2016). Regardless of the descriptor used, the common theme is a convergence of several phenomena, including: ever expanding sources of data, powerful and readily available computers, efficient algorithms, and recognition of data analytics as a competitive advantage for business. “The convergence of these phenomena has given rise to the increasing widespread business application of data science principles and data-mining techniques. (Provost & Fawcett, 2013). In this research, we are interested in a combination of data mining tools that enable the progression from a large and noisy raw dataset to predictions that are manageable and usable in industry. More specifically, we are interested in data preprocessing, clustering, predictions, and evaluation. In some instances, we are fortunate that the necessary tools exist and may be incorporated into our framework. In other instances, we must develop new tools or apply existing tools in ways that they have not previously been tested. We discuss the relevant data analytic topics in the following sections.

2.3.1 Data Preprocessing

The first thing a data analyst must do when preparing to work on a new problem or with a new set of data is to prepare the data; a process known as data preprocessing or colloquially as data wrangling (Wickham, 2014). The important information contained in real datasets can be shielded or distorted by outliers, missing data points, or irrelevant data (Kantardzic, 2011). Additionally, the structure of the raw data may be incompatible with the intended analysis (Léger, Pellerin, & Babin, 2011). The data analysis must therefore expend considerable effort preprocessing the data to remove unwanted data, fill in missing data, and translate the data into a format that suits.

Transaction histories, such as delivery records are sometimes used when actual consumption data is unavailable, however, transaction records do not have a linear relationship to consumption behavior and can be particularly challenging due to irregular occurrence of transactions, errors and omissions in data recording, and the introduction of unknown influences such as the bull whip effect. This type of data requires careful transformation.

Transaction records normally contain a variety of data, but most importantly, date and quantity. The data is easily aggregated into temporal bins to form time-series; a format necessary for subsequent prediction analysis. The resulting time-series, however, is subject to intermittency if the temporal bin size is greater than delivery frequency (Kourentzes, 2014). Intermittent time-series are undesirable due to their difficulty to forecast (Kayacan, Ulutas, & Kaynak, 2010).

The literature contains several methods for resolving intermittency in time-series; the most famous being Croston's method (Croston, 1972) which updates only when transactions occur rather than allowing zero-activity periods to degrade the prediction as is the case with simple methods such as exponential smoothing and moving average. Croston's method is shown in formulae 2.1, 2.2, and 2.3:

$$\hat{y}_t = \hat{Z}_t / \hat{X}_t \tag{2.1}$$

where $\hat{Z}_t$ is the SES forecast for the non-zero periods and $\hat{X}_t$is the forecast for the number of inter-demand intervals. Both the demand size and the intervals use SES per formulae 3 & 4:

$$\hat{Z}_t = \alpha_z y_t + (1 - \alpha_z)\hat{y}_t \tag{2.2}$$

$$\hat{X}_t = \alpha_x x_t + (1 - \alpha_x)\hat{x}_t \tag{2.3}$$

where y_t is the non-zero demand at time t and x_t is the number of non-zero intervals.

Despite corrections and criticism (Rao, 1973; Syntetos & Boylan, 2005; Willemain, Smart, & Schwarz, 2004), the Croston method remains a standard for handling intermittent time-series (Prestwich, Tarim, Rossi, & Hnich, 2014). Figure 2.1 illustrates the application of Croston method on a simple dataset. In the example, intermittent demand (represented by blue bars) is smoothed using formulae 2.1, 2.2, and 2.3 and alpha set at 0.3. The resulting data (represented by the red line) offers a simplified interpretation of the overall demand.

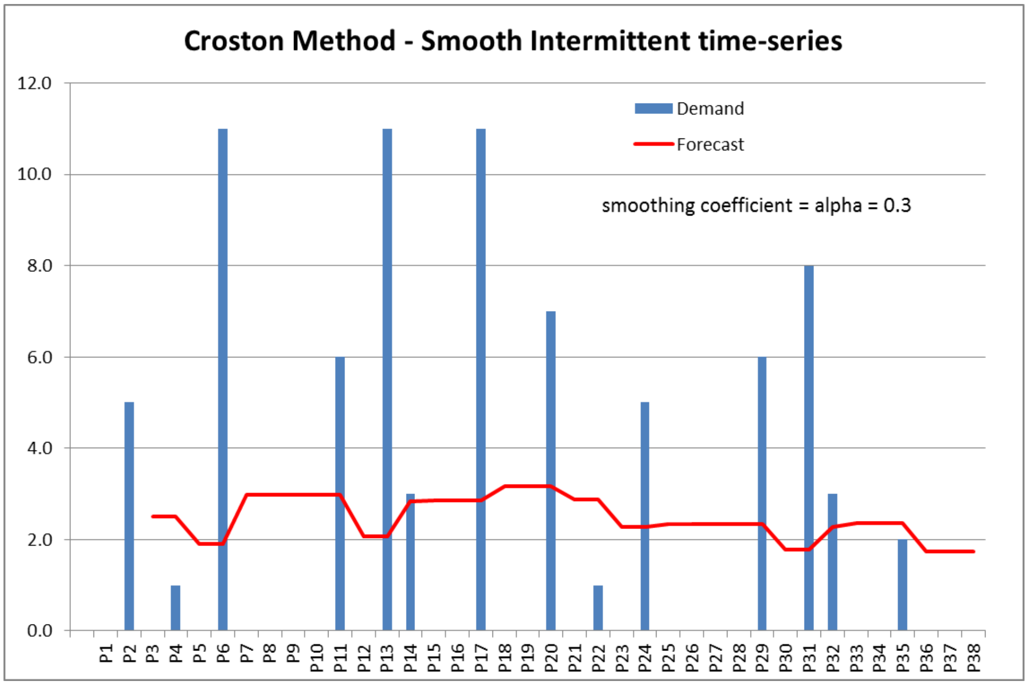


Figure 2-2 : Croston Method

A contemporary approach known as the "aggregate-disaggregate intermittent demand approach" (ADIDA) (Nikolopoulos, Syntetos, Boylan, Petropoulos, & Assimakopoulos, 2011) has gained acceptance in the literature (Kourentzes, 2014; Petropoulos et al., 2016) despite that fact that ADIDA is actually very similar to applying a simple three-period moving average. A significant improvement over Croston's method is not found in the literature. In our research, we propose a new adaptation to Croston's method.

2.3.2 Cluster Analysis

There are several methods for segmenting a population of customers. Simplistic segmentation based on geographic location or customers' industry type is sometimes used since it is relatively easy. However, the demand characteristics within these segments may be very different and therefore this is not a suitable method (Shapiro, 2007). More advanced segmentation methods are categorized as partitional, hierarchical, density-based, grid-based and model-based (Han & Kamber, 2006). Of these, the most common are hierarchical and partitional (Ahmadi Javid & Azad, 2010; Chakraborty, 2013). Partitional clustering is suitable for handling large data sets due to low computational costs. Applying data mining tools to customer segmentation was a natural progression and many methods have been tested.

By far, the most common data mining method used for clustering is the K-means algorithm, for examples see (Anil K. Jain, 2010; Krieger & Green, 1996; Kuo, Ho, & Hu, 2002). K-means is not a new technique, appearing in Fisher (1958) and further developed by MacQueen (1967). With K-means, a set of N data point are grouped into K clusters with the mean of each cluster becoming its identifying location. K-means, however, relies on Euclidean measurement and it only valid where direct, point-to-point comparisons can be made. Despite its frequent use, K-means has failings due to it ignores natural cluster pattern and rigidly assigns points to individual groups (MacKay, 2003).

An alternative clustering strategy to K-means is Artificial Neural Networks (ANN) and the related Kohonen Self-Organizing Maps (SOM) (Altintas & Trick, 2014). SOM was formally developed by Kohonen as un unsupervised alternative to traditional ANNs (Kohonen, 1990). Unlike K-means where the number of clusters must be pre-defined, ANNs are unsupervised; the number of clusters is part of the ANN's output.

Several variants of both K-means and ANNs have been developed to overcome some of their failings. These include the addition of fuzzy logic to avoid the rigid assignment of points to clusters (MacKay, 2003). Fuzzy logic allows the identification of groups with similar attributes (Barajas & Agard, 2014) and when combined with partitional clustering it allows flexibility of assigning points proportionally to more than one cluster. The variants are titled with various names that describe their content, such as soft K-means, fuzzy K-means, and fuzzy ANNs.

Hierarchical clustering algorithms are a family of unsupervised algorithms that build clusters in a progressive manner. Divisive hierarchical clustering begins with all members together and progressively divides them into separate clusters until either all clusters have a membership of one, or until the algorithm reaches a pre-defined stopping point. Conversely, hierarchical agglomerative clustering (HAC) builds clusters from bottom up, it begins by assigning one member to a single clusters and then assigning each new member to an existing cluster or a new cluster (Kantardzic, 2011). The results of hierarchical clustering are displayed in a dendrogram.

When identifying clusters based on historical data, such as delivery records, the data is normally formatted into time-series. K-means clustering uses Euclidean distance which is a point-to-point measurement. Two customers with nearly identical consumption behavior will be assessed incorrectly by K-means if their deliveries occur on different days or different frequencies; a more flexible algorithm is needed. Dynamic time warping (DTW) was initially developed for use in speech recognition, but researchers soon discovered its usefulness in comparing time-series data (Berndt & Clifford, 1994). Unlike Euclidean distance which only assesses directly aligned point, DTW is an elastic measure that is able shift the pairing of points, as illustrated in Figure 2.2, and makes a better comparison between pairs of time-series. Subsequent research supports Berndt and Clifford's findings that DTW performs well for comparing pairs of time-series (Keogh & Ratanamahatana, 2005). DTW was initially criticized as being computationally expensive, but improvements to its algorithms and increased computer speeds have resolved those concerns to the point where they are no longer a consideration (Izakian, Pedrycz, & Jamal, 2015; Ratanamahatana & Keogh, 2005). DTW is widely accepted as an effective tool for comparing time-series (Mueen & Keogh, 2016).

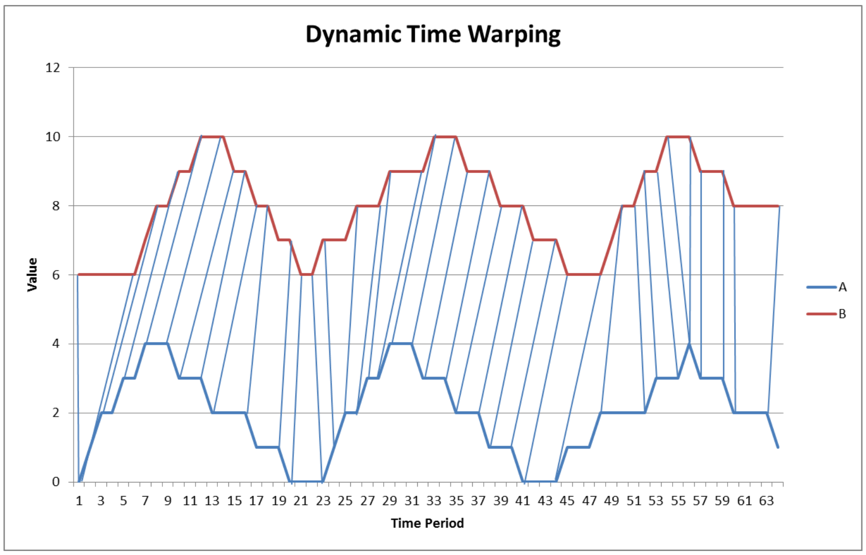


Figure 2-3: Example of DTW

2.3.3 Prediction Models

Forecasting methods are categorized as qualitative and quantitative (Chase Jr., 2013; Hoshmand, 2010; Moon, 2013). While qualitative methods are frequently used, either as a single method or in combination with other methods, they lack the rigor of quantitative methods (Hoshmand, 2010). Qualitative methods that specifically focus on forecasting do not appear in the literature; rather, the methods are based on general qualitative decision-making theory. The most common qualitative decision making practices, as applied to forecasting, are Delphi method, jury of executive opinion, sales force composite, focus groups, and panel discussions (Hoshmand, 2010). Despite the weaknesses of qualitative methods, they are essential when historical data and/or resources to develop qualitative forecasts are absent. A major distinction between qualitative and quantitative methods is that while quantitative methods all strictly rely on past event, qualitative methods incorporate forward-looking information.

In contrast to the lack of specific qualitative forecasting methods, quantitative forecasting methods appear frequently in the literature. Makridakis, Wheelwright & Hyndman (2008) list the most common methods as time series decomposition, exponential smoothing (Brown, 1956), regression, and ARIMA (Box & Jenkins, 1962). These and other qualitative methods can be sub-classified into time series methods, causal methods, machine learning methods, and hybrid methods.

Time series methods (including naïve, moving average, exponential smoothing, decomposition, and ARIMA) are built on the premise that "future sales will mimic the pattern (s) of past sales" (Chase Jr., 2013, p. 84). If demand patterns can be detected and accurately modeled, these techniques can be used to generate reasonable forecast accuracy. Time series methods are suitable for harvest brands that have sufficient historical data and steady demand (Chase Jr., 2013).

Time-series that are based on economic data such as sales histories or delivery records are typically non-stationary, so simple models such as naïve and moving average sometimes do not do a good job of representing demand (Phillips & Durlauf, 1986). Adjusting or accounting for non-stationarity can be accomplished by decomposing the information into separate components of trend, seasonality, cycle, autocorrelation, and random error (F. Robert Jacobs, Berry, Whybark, & Vollmann, 2011). Once the time series is represented by its components, it can be de-seasonalized, de-cycled, and de-trended. The resulting pattern can then be modeled by simple methods. Each method utilizes a different level of complexity and decomposition, from the very simple naïve to the more complex ARIMA. However, more complex models do not necessarily produce more accurate results. In fact, the literature generally promotes the rule of using the simplest method that produces actionable results (Clemen, 1989; F. Robert Jacobs et al., 2011; Makridakis, 1989; Maté, 2011).

Causal methods (including simple & multiple regression, ARIMAX, and Unobserved Component Models (UCM) (Harvey, 1989) are based on the assumption that demand is directly related to other internal or external variables (Chase Jr., 2013). Regression models quantify the correlation between the dependent variable (demand) and one or more independent variables (internal and/or external variables). ARIMAX and UCM both expand on the regression concept by adding features of time series analysis into the models (Chase Jr., 2013; Harvey, Ruiz, & Sentana, 1989). A criticism of causal methods is that they require the forecaster to accurately determine the weighting of the independent variables.

Machine learning methods (including neural networks, recurrent neural networks, and support vector machines) are the newest quantitative methods to be employed for forecasting (Carbonneau et al., 2008). As with time series methods, the machine learning methods are used to detect patterns in past demand behavior from which a forecast can be generated (Azadeh, Ghaderi, & Sohrabkhani, 2007). An advantage to machine learning methods it that unlike causal methods, the contribution (or relevance) of each variable does not have to be predetermined by the forecaster.

Lastly, hybrid methods have been proposed to combine the best features of multiple methods. Unlike combination forecasts that combine the results of different forecast methods (Bates & Granger, 1969), the hybrid forecast method combines the technique of different methods to create a new method. Aburto and Weber (2007) proposed a hybrid method that combines ARIMA and Artificial Neural Networks (ANNs). This appears to be a logical choice considering that ARIMA is widely accepted as an accurate forecasting method and ANNs are gaining recognition in the field. In their research, Aburto and Weber (2007) found that the hybrid model performed better than the baseline Naïve forecast and the ARIMA model.

2.3.4 Evaluation

Market segmentation is the business equivalent to cluster analysis in statistical and data mining sciences. The two differ in how the results of the process are evaluated. There are many examples in the literature for evaluating clusters through various measures of inter-cluster homogeneity and intra-cluster heterogeneity (Ding, Trajcevski, Scheuermann, Wang, & Keogh, 2008; Liao, 2005; X. Wang et al., 2013). Mathematical evaluation presumes that the variables used to calculate the results are meaningful measures. A different approach for cluster evaluation is to conduct a subjective visual assessment of a graph of the results. Visual assessment holds some merit due to human's innate ability to recognize patterns (Chellappa, Wilson, & Sirohey, 1995), however when working with big data, it can be impractical to graph the data (Hung & Tsai, 2008). A visual evaluation also lacks a structured recording of how good or bad one result is over another.

In forecasting, a widespread practice is to divide the data into training and testing datasets. The training data is used to build the forecast model and then the test data serves as a known-truth against which the model is tested. In market segmentation, there is usually no known-truth as there is no way to predetermine which clusters a member should belong. We propose a new method in Chapter 7 where the market is established based on historical consumption behavior and then segmentation strategies are tested to see if they can correctly assign segment membership.

CHAPTER 3 RESEARCH APPROACH AND STRUCTURE OF THE THESIS

3.1 Research Approach

Similar forecasting problems found in the literature tend to offer a single-step solution. For example, in Croston's seminal paper (Croston, 1972) and in contemporary applications of Croston's method (Shenstone & Hyndman, 2005), the proposed method resolves intermittent time-series and presents a forecast in the same algorithm. The same single-step strategy was pioneered by Brown (Brown, 1963) with the newly proposed, and now widely used simple exponential smoothing (SES). This research takes a different approach; the solution to the forecasting problem is broken into a series of sequential steps outlined here.

The general problem is how to create an accurate, customer-specific demand forecast based on historical data and in absence of any collaborative input from downstream SC partners. The state of the art shows that many relevant tools exist for this topic, however, attempts to simply choose the best or most suitable tool and apply it are not successful. Our research has shown that a single-step solution is not an effective method for solving this research problem; the sequential application of several different tools is necessary. We leverage existing tools that are demonstrated in the literature, but we sometimes use those tools in ways that have not previously been done. The methodology proposed to solve the overall problem is illustrated in Figure 3.1:

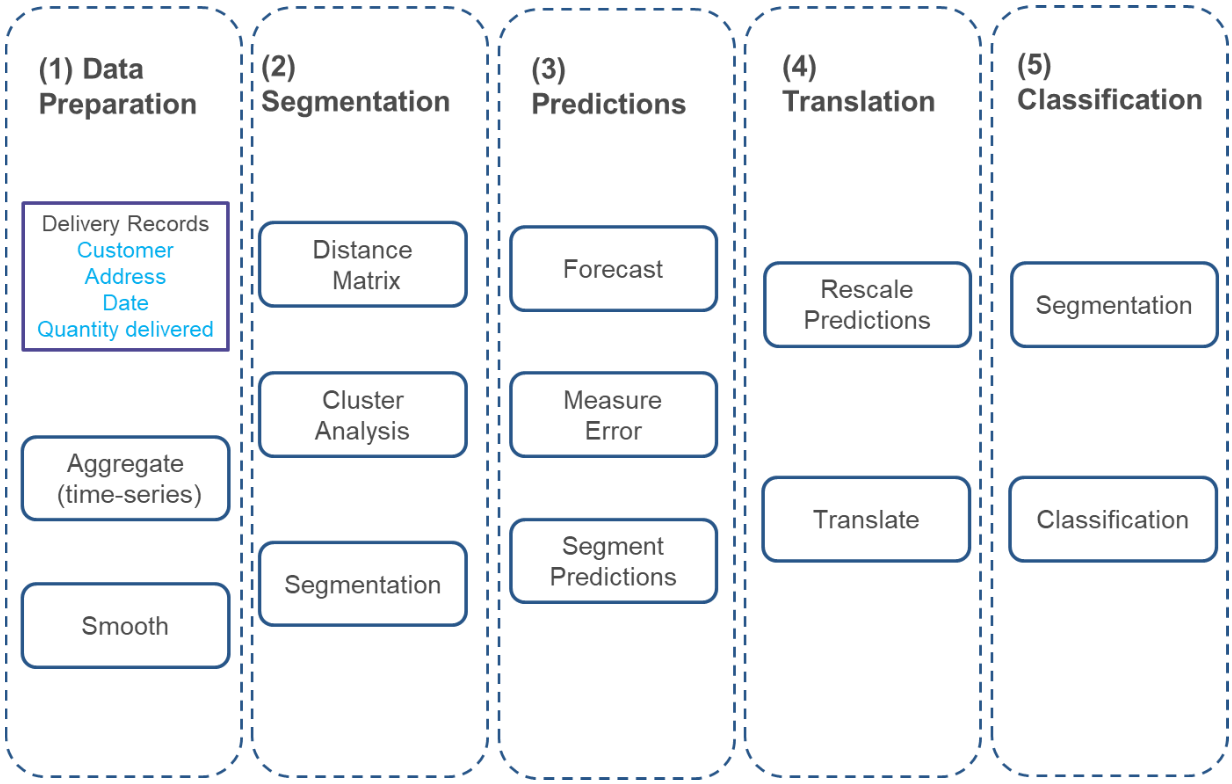


Figure 3-1: Outline of Research

Step 1 – Data preparation: The first challenge is evident when delivery records are aggregated into periodic bins to create time-series. Delivery records are a composite of information that is influenced by a combination of consumption behavior, delivery logistics decisions, inventory management, and the bullwhip effect (Forrester, 1958). The resulting time-series are intermittent, lumpy, and erratic; generating a useful prediction from them is very difficult. A method to smooth the data is necessary. Ideally, the method used to smooth the data should retain the underlying behavior patterns. We offer a method to accomplish this in Chapter 4.

Step 2 – Segment the market based on behavior patterns: The second challenge arises with the analytically intensive task of attempting to maintain individual forecasts for many customers. Although producing many individual forecasts is computationally possible, incorporating them into strategic or tactical business planning requires that they are somehow combined into segments. Forecasting customer segments has been studied in the context of electrical demand (Espinoza, Joye, Belmans, & De Moor, 2005), but it does not appear in the context of forecasting product demand. A reliable strategy for customer segmentation is necessary, it is presented in Chapter 5.

Step 3 – Create segment-level predictions: The third challenge, after the data is smoothed and customers are segmented is how to translate segment-level predictions into individual forecasts. A method to translate the information back to individual forecast is needed and evaluated. This part is presented in Chapter 6.

Step 4 - Translate the predictions to the customer level: The fourth challenge is to quantify and incorporate the effects of exogeneous factors into the forecasts. It is intuitively obvious that some external factors, such as climate, will affect the behavior of some industries, such as agriculture. However, a method to identify, quantify, and apply relevant exogeneous variables to demand forecasts is needed. Also presented in Chapter 6.

Step 5 – Classification based on attributes: The fifth challenge is first to create segments based on demonstrated behavior and then attempt to classify the customers into similar segments based on descriptive attributes. If the descriptive attributes are effective, the resulting clusters should be similar. This part is presented in Chapter 7.

We address these five challenges as individual steps in our research. Some steps are presented as an individual scientific contribution detailed in the papers that make up the body of this dissertation. Although they can stand individually, they are all necessary and propose sequential progressions in forecasting demand from delivery records. Each step is validated on a real case study that is presented in the next section.

3.2 Case Study

The real dataset used in this research was provided by our industrial partner, Air Liquide Americas (AL). AL's primary products are liquefied oxygen (Lox) and nitrogen (Lin). Point of use inventory of Lox and Lin are managed by AL in a VMI arrangement with the goal of ensuring continuous product availability for its customers while controlling its own internal operations costs (Air Liquide, 2014). In many VMI arrangements, the supplier has access to point of consumption information and is able to make its replenish decisions from that information. However, in AL's situation the point of consumption information is generally not available and replenish decisions are triggered by low-level sensors on the storage tanks at the customers' sites.

3.2.1 Industrial Context

AL produces a variety of liquefied and pressurized gasses and which it delivers to its customers via pipeline, tanker truck, and bottles. The context for this research, as illustrated in Figure 3.2 will focus on liquefied Lox and Lin, delivered via specialized tanker trucks to its customers in the continental USA, as illustrated in Fig. 4.

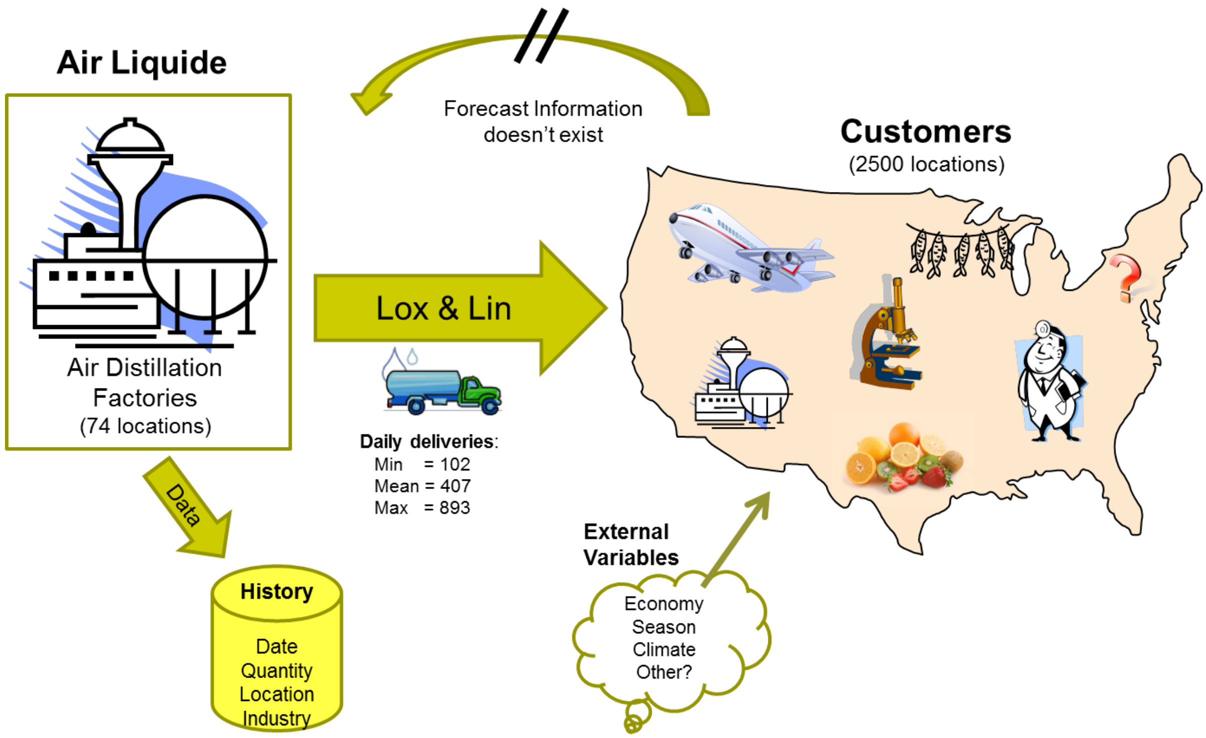


Figure 3-2: Overview of Context of Case Study

AL operates air distillation factories located throughout the continental United States. The factories extract and compress oxygen, nitrogen, and several other elements from atmosphere and store them in liquefied state for distribution to their customers. The liquefied gasses are transported via AL's specialized fleet of tanker trucks. At the customers' sites, the liquefied gasses are transferred into storage tanks. AL is responsible for maintaining the customers' tanks in a VMI arrangement. AL periodically remotely queries the tank level sensors to determine if a replenishment delivery is necessary. Continuous level sensing and recording is possible at some locations, but in practice it is not used for operations decisions; tank levels and point-of-use consumption data is not recorded. Ideally, AL will coordinate delivery schedules so that an entire tanker load will be delivered to a single customer site. However, partial deliveries are necessary for smaller customers. When partial tanker-load deliveries are necessary, the balance of the load is sometimes delivered to a second customer in effort to increase overall logistics efficiency.

AL's delivery scheduling is further complicated by its factories using the same distillation systems for more than one element and limited storage capability. Deliveries therefore depend not only on customers' needs, but also on product availability. Customer requirements, delivery logistics, and product availability all influence the information gathered in the delivery records.

3.2.2 Data Description

AL provided its delivery records for all customers and all products in the continental USA from January 2009 through September 2014. AL did not perform any data cleaning or preprocessing prior to transferring the data to us, although it did remove fields with descriptive customer information to anonymize the data. Liquid nitrogen (Lin) and liquid oxygen (Lox) make up the majority of AL's product sales and these were retained for study; records for other products were removed from the data. Each delivery event in the data includes delivery date, product quantity, product description, customer number, customer location, AL distribution center location, and an industry identifier number. Prior to cleaning, the 66 months of data contained 1.18 million delivery events and over 8000 unique customers. Considering that the research is focused on identifying and predicting customer behaviors, only customers with consistent and persisting transactions were retained. Infrequent customers, lost customers, and new customers were removed. Also, a very small number of customers with unexplainable quantities or suspected corrupted data were removed. For the data cleaning, infrequent customers are defined as having less than 12 months with deliveries, lost customers have no deliveries in the last year, and new customers have no data during the first two years. After initial data cleaning, approximately 3000 unique customers were retained for the case study. Although more than two thirds of the customers were removed, they represent only 25% of the delivery events; 880,000 of the original 1,118,000 observations are used in the case study.

3.3 Research Methodology

The research objective, as stated in section 2.1, is to develop a method to forecast a SC's demand requirements based on limited historical data. In the case study, a dataset of delivery records was the only available information. Secondary data sources, such as consumption data, invoice records, purchase contracts, or inventory audits were not available and therefore it was not possible to validate the dataset through any direct comparison. Initial inspection of the data revealed some entries that were obvious errors due to extreme or impossible values. However, deciding the level at which the records could be identified as outliers and removed from the study was subjective. Considering that the research objective did not include maximizing participation in the study, the level for identifying outliers was set low. A secondary data source for auditing the dataset would allow less outlier removal.

The literature review and testing of established forecasting methods did not reveal an existing method to produce a useful forecast from the noisy stochastic dataset. The research methodology was therefore based on a framework of existing tools and methods that were able to incrementally progress toward solution of the problem. Within each step of the framework, decisions such as selecting a smoothing coefficient value or choice of clustering method were necessary. While many tests were conducted to validate these types of decisions, the search was not exhaustive, and the overall framework is not optimized. Rather, the goal here was to demonstrate a working solution that could later be optimized depending on the domain where it is applied.

3.4 Structure of the Thesis

The thesis is presented in nine chapters, as follows:

Chapter 1 is the introduction which presents the research problem and highlights the significance of the solution to the research problem

Chapter 2 presents the state of the art in the relevant topic areas. The state of the art shows how others have managed similar problems. The gap in the research literature is highlighted.

Chapter 3 outlines the methodology for developing the proposed solution and describes the dataset that was used to test the proposed method.

Chapter 4 is an article, submitted to the International Journal of Production Economics, that describes the details of how transaction records can be translated and smoothed so that the information contained within them can provide an interpretation of the underlying consumption behavior (Murray, Agard, & Barajas, 2018a).

Chapter 5 is an article, published by the Journal of Computers and Industrial Engineering, that describes the proposed method to segment the market into groups with similar behavior patterns (Murray, Agard, & Barajas, 2017).

Chapter 6 is an article, published in the Journal of Computers and Industrial Engineering, that describes the application of segment-level predictions to individual customers. This article was submitted to the Journal of Computers and Industrial Engineering and is currently being reconsidered (Murray, Agard, & Barajas, 2018b).

Chapter 7 is an article, submitted to the Journal of Industrial Marketing Management, that tests the application of exogeneous variables to determine whether descriptive variables can be used to predict behavior patterns. This article was submitted to the Journal of the Academy of Marketing Science (Murray, Agard, & Barajas, 2018d).

Chapter 8 is a general discussion of the results of the research and highlights the scientific significance of this work.

Chapter 9 is the conclusion and recommendations. It discusses how the proposed method was able to solve the research problem and suggests future research.

CHAPTER 4 ARTICLE 1: ASACT - DATA PREPARATION FOR FORECASTING: A METHOD TO SUBSTITUTE TRANSACTION DATA FOR UNAVAILABLE PRODUCT CONSUMPTION DATA

Submitted to: International Journal of Production Economics

Murray, P., Agard, B., & Barajas, M.A., (2018a)

Abstract: *Strategic supply chain planning relies on accurate, long-range forecasts. Accurate forecasts, in turn, rely on the availability of suitable data and information from which a prediction can be made. In some domains, such as vendor managed inventory, product consumption data may not be available because of a lack of collaborative information. Delivery records are on occasion substituted for absent consumption data. This substitute information, however, can appear lumpy and intermittent due to the bullwhip effect and other logistics factors.*

The proposed ASACT (Aggregate, Smooth, Aggregate, Convert to Time-series) method applies a sequence of aggregation and smoothing to transform delivery records into time-series data that creates a good approximation of actual consumption behavior; noise in the data is reduced while behavior patterns are preserved. Synthetic data is used to test the method against traditional methods, and real data is used to demonstrate the method's application in industry. Tests have shown that ASACT improves on the results produced by traditional methods.

4.1 Introduction

In vendor managed inventory (VMI) arrangements, the supplier is responsible for its own strategic forecasting, often without the benefit of collaborative information from its customers. When point-of-consumption information is absent, a different source of information must be used. In some domains, the last available piece of supply chain data is collected when the supplier delivers to the customers' point-of-use inventory. Delivery records are easily aggregated into temporal bins to create surrogate information, which can be substituted for the missing consumption data. Interpreting consumption behavior patterns from delivery records is not a trivial task (Kourentzes, 2014). Time-series that are generated by aggregating delivery records tend to be lumpy, erratic, or intermittent and therefore difficult to use for forecasting (Petropoulos et al., 2016); as such, delivery-record derived time-series are generally referred to as noisy time-series (Ferbar Tratar, 2015; Murray et al., 2017).

There are various sources of noise in a time-series created from delivery data; here we consider three that hold relevance. With the first source of noise, it is difficult to match delivery frequency to temporal bin size when aggregating the data. The second source of noise relates to the bullwhip effect where actions of the supply chain members, such as forcast update frequency, batching, price fluctuation, and rationing may influence or distort the true demand estimates (Lee et al., 1997).

The third source of noise is due to logistics decisions. In a VMI arrangement, delivery decisions are driven partly by consumption behavior and partly by the supplier's attempts to minimize its distribution costs. In this work, the research problem will evaluate methods for mitigating the three sources of noise to allow the available delivery data to be substituted for the consumption behavior data.

In this paper, we present framework for converting delivery data into a time-series format, which more closely represents the customers' true consumption behavior. The proposed method is a significant contribution in that it provides at sequential framework that can remove the noise without loss of information as is typically experienced with the other methods, which are discussed in the literature review. The framework is a sequential combination of existing, proven tools that combine to solve a real problem that, as far as we know, has not been adequately solved in other ways. The remainder of the paper is organized as follows: Section 4.2 presents a review of the literature, Section 4.3 presents the proposed ASACT method, tests it, and compares the results to other methods using synthetic data in Section 4.4. Section 4.5 provides the results of the proposed method as tested on real data and compares it to other methods. Conclusions, limitations, and future directions are presented in Section 4.6.

4.2 Literature Review

4.2.1 Sources of Noise in Data

In a VMI context, there are many potential causes of data distortion. In this research, three specific causes are considered: temporal aggregation, bullwhip effect, and logistics decisions. Temporal aggregation (Section 4.2.1.1) holds relevance because of the necessity of converting delivery records into time-series format in an attempt to have the available data represent consumption behavior. bullwhip effect (Section 4.2.1.2) holds relevance given that the data is gathered up-stream in the supply chain from the point of interests. Logistics decisions (Section 4.2.1.3) holds relevance because of the delivery quantity and frequency being driven not only by consumption demand, but also by logistics decisions made by the supplier. The following sections detail each topic.

4.2.1.1 Temporal Aggregation of Sparse Data

Delivery records are typically in a list format which requires conversion to a time-series format to facilitate quantitative forecasting. Converting them to a time-series format is straightforward, as the delivery records generally contain quantities and dates; however, selecting an appropriate temporal bin size is non-trivial. Long bin sizes, such as quarterly or yearly, can hide important behaviors such as seasonality. Short bin sizes, such as weekly or monthly, may retain behavior patterns, but can result in lumpy and intermittent time-series. In short, a compromised bin size may result in an intermittent time-series (Kourentzes, Petropoulos, & Trapero, 2014).

4.2.1.2 Bullwhip Effect

When customer-specific forecasts are needed, and consumption data is not available, planners must use a different data source. Moving up the supply chain, the next available data is sometimes delivery data. Initially, it appears appropriate to use delivery data as a surrogate source of information about the customer's demand requirements. However, this assumption fails to account for the bullwhip effect, especially when a point-of-use inventory exists between the point of delivery and the point of consumption. Supply chains generate data at different points along the supply chain (Biswas & Sen, 2016). As the data acquisition point moves up the supply chain and away from the point of interest, information gleaned from the data becomes distorted; a well-documented phenomenon known as the bullwhip effect, originally documented by Forrester (1958) and further explained by Lee et al. (1997). Although the data may relate to the same end consumption, the flow of the product is influenced by transportation, batching, storage, and production processes at each step of the supply chain (F. Chen, Ryan, & Simchi-Levi, 2000). Hence, data is different depending on when in the supply chain it is gathered (Chopra & Meindl, 2007; Disney & Towill, 2003).

4.2.1.3 Logistics Decisions

The primary focus of logistics practitioners is to move their product from the manufacturer to the consumer as efficiently as possible; minimizing transportation costs while meeting demand requirements are the main objectives (F Robert Jacobs & Chase, 2010). A criticism of logistics, in its more traditional sense, is that logistics decisions are sometimes disconnected from the overall supply chain (Ballou, 2007) leading to decisions which are not necessarily the best choices for the overall supply chain; a phenomenon known as "localized optimization" (Goldratt, 1990). In a VMI arrangement, delivery schedules are not always directly influenced by the end use consumption; when point-of-use inventory exists, it may buffer the consumption patterns. Delivery data, therefore, can reflect both consumption behavior and logistics decisions.

Logistics practitioners have many considerations to make beyond meeting the customers' demand requirements. Known variables such as truck capacity, number of trucks, distance to customer, and production capacity must all be balanced against the needs of all customer combined; these factors all impact the timing and size of each delivery. Unknown and external variables also have an effect, for example, the inability to satisfy all customers' demands may cause the supplier to sub-contract the supply of product. These sub-contracted deliveries may not appear in normal delivery records because of different accounting systems.

4.2.2 Methods to Resolve Noise in Data

Conventional forecasting methods do not perform well on noisy data (Kayacan et al., 2010). Rather than attempting to interpret the results of forecasts based on noisy data, it is sometimes desirable to smooth the data prior to forecast analysis. Here we examine two different approaches to smoothing noisy data. The first, smoothing through data transformation, applies an algorithm that adjusts the data values while leaving the temporal bin size unchanged. The second, smoothing through data aggregation, sums the data into larger temporal bins without changing the individual values. These two approaches are explained in the following sections.

4.2.2.1 Smoothing through Data Transformation

A popular smoothing method that is simple to use and provides good results is the Holt-Winters method (Brown, 1963), otherwise known as simple exponential smoothing (SES) as per formula 1:

$$\hat{y}_t = \alpha y_t + (1 - \alpha)\hat{y}_{t-1} \tag{1}$$

where $\alpha \in (0,1)$ is the smoothing parameter, y_t is the demand and $\hat{y}_t$ is the demand forecast. A small value of α gives less weight to the most recent values and results in additional smoothing. While SES works well on irregular data, it does not perform well on intermittent demand patterns (Prestwich et al., 2014). Croston (1972) has proposed a solution for predicting intermittent demand. Croston's method proved effective and despite small corrections (Rao, 1973), improvements (de Melo Menezes, de Siqueira Braga, Hellingrath, & Buarque de Lima Neto, 2015; Willemain et al., 2004), and criticisms (Syntetos & Boylan, 2005), his remained the standard method for dealing with intermittent demand (Ghobbar & Friend, 2003; Prestwich et al., 2014).

Croston (1972) proposed using SES to update a forecast for intermittent demand where updates only occur when demand occurs; as such, during periods of zero demand, the forecast remains unchanged. Croston's method is shown in formulae 2, 3, and 4:

$$\hat{y}_t = \hat{Z}_t / \hat{X}_t \tag{2}$$

where $\hat{Z}_t$ is the SES forecast for the non-zero periods and $\hat{X}_t$is the forecast for the number of periods where no demand is recorded. Both the demand size and the intervals use SES per formulae 3 & 4:

$$\hat{Z}_t = \alpha_z y_t + (1 - \alpha_z)\hat{y}_t \tag{3}$$

$$\hat{X}_t = \alpha_x x_t + (1 - \alpha_x)\hat{x}_t \tag{4}$$

where y_t is the non-zero demand at time t and x_t is the number of non-zero intervals.

Smoothing through data transformation has the undesirable effect of not only removing noise, but also diminishing or masking the behavior patterns that could provide valuable information to a subsequent forecast analysis.

4.2.2.2 Smoothing through Data Aggregation

Rather than altering the data, a different approach smooths the data by manipulating temporal bin sizes during aggregation. Aggregating the data into larger bins can sometimes eliminate zero-demand periods, which occur when the data is in smaller bins. Several research papers conclude that manipulating temporal bin size during aggregation is very effective for improving forecast accuracy with intermittent demand data (Gansterer, 2015; Petropoulos & Kourentzes, 2015; Rostami-Tabar, Babai, Syntetos, & Ducq, 2013). Temporal aggregation is also promoted by commercial forecasting software as a sure way to reduce uncertainty (Syntetos, 2016). A recent improvement in temporal aggregation, known as the "aggregate-disaggregate intermittent demand approach" (ADIDA), smooths the data through aggregation and then disaggregates to return it to the original temporal bin size (Nikolopoulos et al., 2011). A variant to ADIDA inverts the intermittent time-series and is expected to give improved results (Petropoulos et al., 2016). Where ADIDA attempts to predict demand within pre-fixed time periods, the inverted ADIDA method attempts to predict when the next demand will occur (Petropoulos et al., 2016) A related method, known as multi aggregate prediction algorithm (MAPA) (Kourentzes et al., 2014) attempts to resolve noise problems by aggregating the data into multiple time-series, each having a different temporal bin size. Forecast combination is then employed on the group of time-series data. MAPA leverages the benefits of forecast combination to improve accuracy (Diebold & Lopez, 1996).

Despite the strong support for the method, temporal aggregation does not always perform well. A survey of early research on temporal aggregation reveals several findings in which the resulting forecast errors increased instead of decreased (Jin, Williams, Tokar, & Waller, 2015). For example, Amemiya and Wu (1972) found rogue moving average residuals; Tiao (1972) found an overall decrease in forecast accuracy; and Wei & Mehta (1980) found substantial information loss. In their empirical research, Jin et. al (2015) found that in some situations, temporal aggregation improved forecast accuracy and in other situations, there was no significant improvement.

The lack of consensus in the literature regarding the effectiveness of temporal aggregation may relate to the origin of the raw data used in the study or to preprocessing that occurred prior to analyses. Data obtained at the distributer level has a higher level of bullwhip effect than data obtained from point-of-use. Temporal aggregation aids in reducing the bullwhip effect and therefore is more beneficial to distributer level data than it is to point-of-use level data (Jin et al., 2015). Preprocessing raw data can also alter the outcome. Some research that found beneficial results with temporal aggregation appears to have begun with data that was already aggregated into monthly or weekly bins prior to the analysis. For example, see (Nikolopoulos et al., 2011; Petropoulos & Kourentzes, 2015; Rostami-Tabar et al., 2013). The variation of the point in the supply chain where data is obtained, and the various levels of pre-analysis aggregation may have contributed to the inconsistent opinions found in the literature with regards to the benefits of temporal aggregation.

4.2.3 Predicting Consumption from Delivery Data

It is possible to predict point-of-use consumption based on delivery data if the effects of using surrogate information are known and accounted for. The three causes of noise in the delivery data—temporal aggregation, bullwhip effect, and logistics decisions—can significantly impact the accuracy of the analysis. As far as we know, a method to identify and address the problem of noise in delivery data does not exist in the literature.

4.3 ASACT Method to Substitute Transactional Data for Unavailable Consumption Information

The descriptive acronym "ASACT" is used in this research to highlight the distinct sequence of steps in the proposed method and to distinguish it from other methods. ASACT addresses the three causes of noise in aggregated delivery data. The data transformation steps are: "Aggregate", "Smooth", and then "Aggregate" again.

Delivery records are generally in list format, which lacks the necessary interoperability to be useful in forecasting activities; it must first be converted into a time-series. Traditional time-series conversion, as illustrated in Figure 4.1, is frequent practice in industry and in academic research. Delivery records are directly aggregated into non-overlapping temporal bins, sometimes smoothed to reduce noise, and then employed in a time-series analysis. The selection of temporal bin size, such as weekly bins, is a compromised decision that attempts to satisfy three criteria: preservation of behavior patterns, consideration to minimize intermittency and noise, and the bin size needed for subsequent forecasting activities—a compromised bin size selection does not ensure the best results.

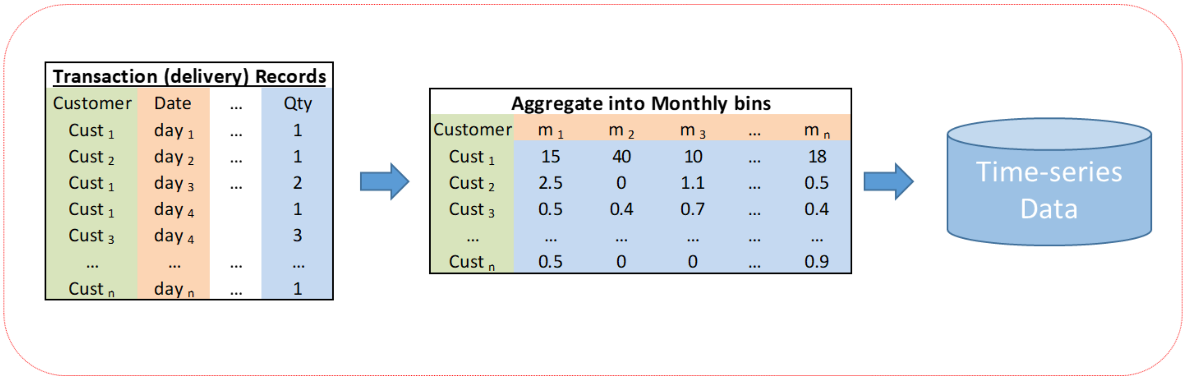


Figure 4-1: Traditional Method: Conversion to Time-series

Figure 4.2 illustrates the proposed ASACT method. The sequential steps of ASACT eliminates the need to compromise when selecting bin size. Behavior patterns are preserved by avoiding summation in the first aggregation step, intermittency and noise are resolved in the second step, and final bin size is selected in the third step. Figure 4.2 contrasts the proposed ASACT with traditional conversion methods. The first components of ASACT transform the data through aggregate, smooth, and re-aggregate steps as follows:

- **Aggregate:** For each customer i, the delivery data is first aggregated into the smallest possible temporal bin size, which is determined by the bin size where no more than one transaction occurs in each period thus preventing any summation of quantities in the first aggregation. In domains where delivery frequency varies within or between customers, the initial aggregation will result in intermittent time-series, since not every bin will have a transaction. In input, we select all transaction data for each customer i, and construct a vector of deliveries for the customer. The coordinates of the vector represent the bin size selected. For example, a bin size on one day gives a vector of 365 days for the year. For each coordinate, the data is the total volume of products delivered this day, for customer i, so 2 deliveries the same day are summed.
- **Smooth:** We follow Croston's assumption (1972) that activities between observations are uniformly distributed. Croston's method updates the demand level only after a positive demand occurs and the positive demand is distributed evenly across the preceding periods of zero demand (de Melo Menezes et al., 2015). We distribute the transaction quantities

over the transaction time period and resolve the problems of intermittency by smoothing the data with Croston's method (Croston, 1972) according to Formulas 2, 3 & 4.

- **Aggregate:** The smoothed data is then aggregated into the desired bin size for analysis. Bin size is determined based on the unit of measure that suits the subsequent forecasting activities. Unlike traditional direct-aggregation, this step of ASACT does not need to seek a compromise between intermittency and behavior pattern preservation. Since the data is smoothed and intermittency resolved in the previous steps, the data can now be aggregated to the size of bin that best suits the subsequent forecast analysis. The operation simply consists in the addition of the smooth data from previous step to the desired bin size. For example, if a daily bin was selected in the first (and then second) step, a monthly bin is simply done by adding the data for all days of the month. We then obtain 12 aggregated monthly columns.

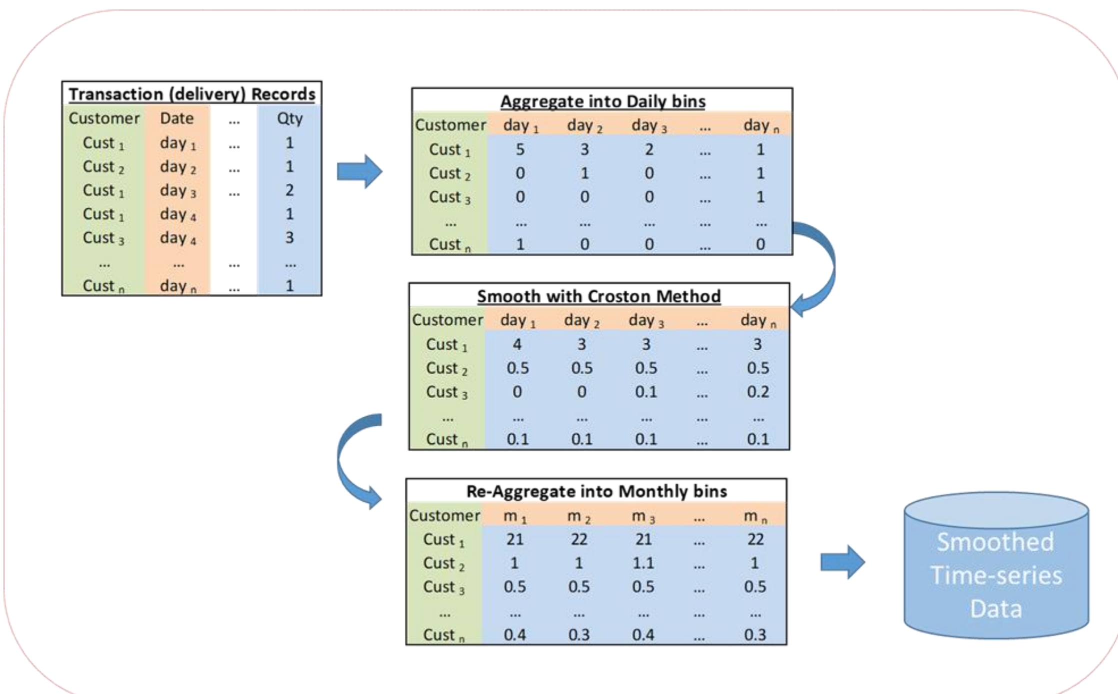


Figure 4-2: Proposed Method: Conversion to Time-series

4.4 Evaluation of ASACT with Synthetic Data

In this section, a synthetic data set is used to illustrate and evaluate the proposed method; the following section, Section 4.5 will illustrate performances on a real data set from industry. The synthetic data represents many possible situations and includes a mix of stationary, seasonal, and trend, both with and without variations. In the following sections, we will use synthetic data delivery information (including predefined noise) as an input in each method and evaluate how successful they are at representing the synthetic data's consumption information (delivery information and consumption information in the synthetic data are directly related). All calculations in the research were done with R (R Core Team, 2015).

4.4.1.1 Synthetic Data

The synthetic data contains nine types of customer *(C_i)* whose consumption patterns vary with various levels of usage including random, intermittent, seasonality, and trend, as detailed in Table 4.1. A study period of five years of daily records was generated for each customer. Customers C_1, C_2, and C_3 have consistent daily consumption with overall patterns that are stationary (no change), increasing trend, and seasonal pattern. The first three hypothetical customers were created with simple patterns to find how well each method performs on the easiest cases. Customers C_4 through C_9 have consumption patterns based on the first three customers with added variations to simulate random and intermittent patterns.

The hypothetical customers are serviced through delivery strategies, which include varying levels of noise that attempt to represent bullwhip effect and logistics decisions. Each customer type has two different delivery strategies—D_1 and D_2—for replenishing the point-of-use inventory. Delivery type 1 *(D_1)* has frequent, regularly scheduled deliveries with relatively small quantities. The absence of schedule variation in D_1 should result in high correlation between delivery quantity and consumption demand. Delivery type 2 *(D_2)* does not have regularly scheduled deliveries. Factors such as logistics decisions influence when deliveries are made, and less frequent deliveries results in larger delivery quantities. Each set of delivery records is generated directly from the hypothetical customers' consumption patterns; thus, the results of each conversion method can be quantitatively evaluated to determine how successful they are in converting delivery records into consumption behavior.

Table 4-1: Description of Hypothetical Customers

(a)	*(b)*	*(c)*	*(d)*	*(e)*	*(f)*	*(g)*	*(h)*	*(i)*
Customer	Pattern	Average Daily Consumption			Average Delivery Frequency		Average Delivery Size	
		Initial Qty	End Qty	Notes	D_1	D_2	D_1	D_2
C_1	Stationary	1	1		5.0	62.0	5.0	66.3
C_2	Trend (increasing)	1	5	Steady increase	5.0	21.3	15.0	68.0
C_3	Seasonal (quarterly)	1.5	1.5	Varies from 1 to 2	5.0	41.5	7.5	65.7
C_4	Random	1	1		5.0	57.9	5.0	62.4
C_5	Intermittent	1	1		5.0	56.1	5.4	63.8
C_6	Random & Trend	1	5	Steady increase	5.0	20.5	15.1	65.9
C_7	Intermittent & Trend	1	5	Steady increase	5.0	20.0	15.9	66.1
C_8	Random & Seasonal	1.5	1.5	Varies from 1 to 2	5.0	39.6	7.6	63.2
C_9	Intermittent & Seasonal	1.5	1.5	Varies from 1 to 2	5.0	37.9	8.1	64.2

4.4.1.2 Comparison of Methods - Customer C_2

The proposed method is compared against four other methods (direct aggregation, SES, Croston's method, and ADIDA) using the synthetic delivery records. In the first step here, we show results of "how it works" for customer C_2 and delivery type D_2 for each method. Then, in Section 4.4.1.4, an overall evaluation is done for all customer types. Some of the methods compared here include parameters such as temporal bin size and smoothing coefficient that can be adjusted to optimize their results. These parameters are set consistently for all methods in attempt to avoid bias in the results.

The first conversion method is a non-overlapping direct aggregation. The delivery data is directly aggregated into monthly bins, no other processing is performed. The results of directly aggregated data compared to actual consumption are shown in Figure 4.2.

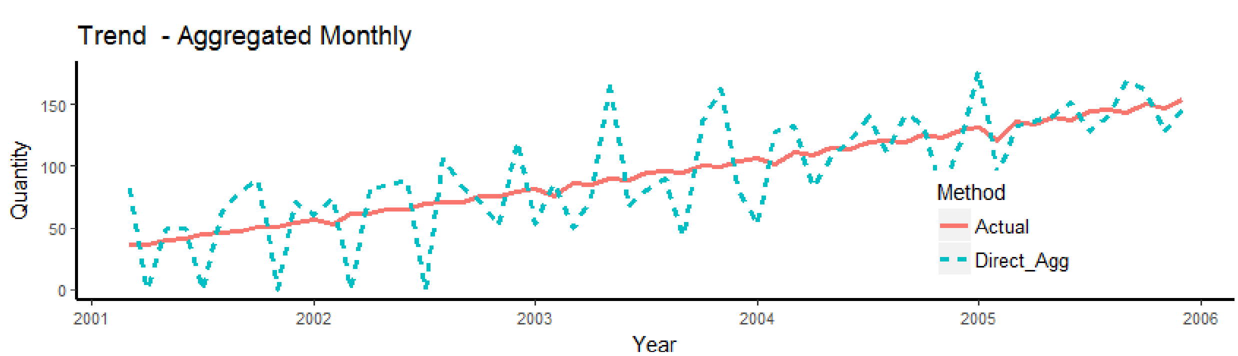


Figure 4-3: Directly Aggregated Data

From Figure 4.3, which illustrates the results of direct aggregation, we see that the data generally follows the increasing trend of the actual consumption line. There is no data transformation and, given a large enough temporal bin size, the total delivery quantity will equal the total consumption. However, with a smaller study period, it would be difficult to make a short-term prediction given the high variation in the data.

The second method compared is simple exponential smoothing (SES). SES is frequently employed in industry due to its simplicity, it is also frequently employed in literature as a comparator (Kourentzes et al., 2014). For the SES method, delivery records are first aggregated into monthly bins and then SES is applied per formula (1). In this study, the smoothing coefficient (α) is set to 0.2. (The smoothing coefficients for all methods were set the same to reduce variation in the study.) The results of the SES method are shown in Figure 4.4. While SES appears to be effective at smoothing the data, the result does not closely match the actual consumption pattern.

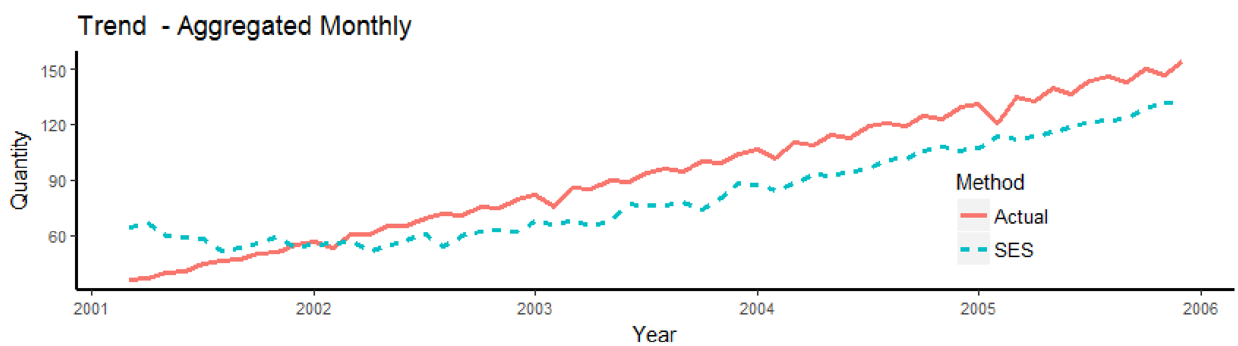


Figure 4-4: SES

The third method for comparison is based on Croston's method (1972). Croston's method is selected for comparison because of its relevance to intermittent time-series. In this method, delivery data is aggregated into monthly bins and then smoothed with Croston's algorithm. The results of the Croston method are shown in Figure 4.5. As with the evaluation of SES, the smoothing coefficient (α) for both formula 3 and 4 was set to 0.2. Even through the data from the Croston method follows the general pattern of the actual consumption, it frequently deviates and shows modest improvement over the period of the study.

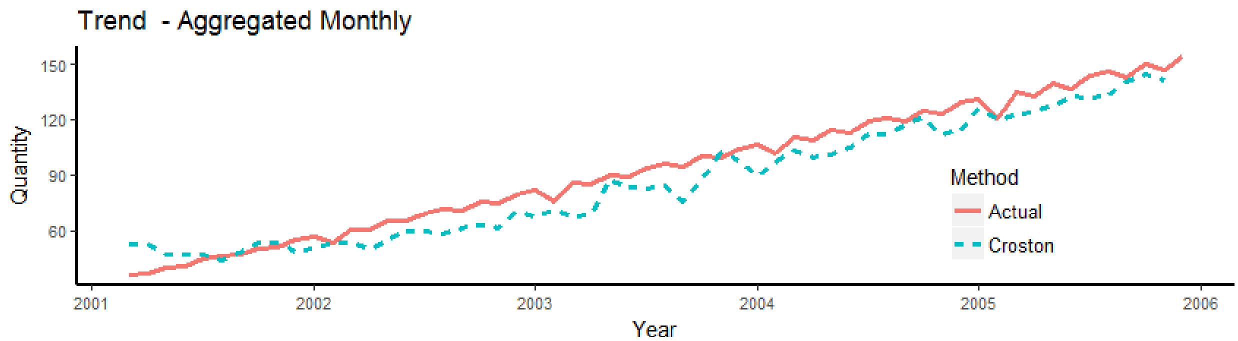


Figure 4-5: Croston's Method

The fourth method compared is a contemporary method called ADIDA, proposed by Nikolopoulos et al. (2011). For this method, data is aggregated into monthly bins, re-aggregated into quarterly bins, and then de-aggregated back into monthly bins. An example of the results of the ADIDA method is illustrated in Figure 4.6. While ADIDA data generally follows the actual data, it contains significant variability in some portions of the study.

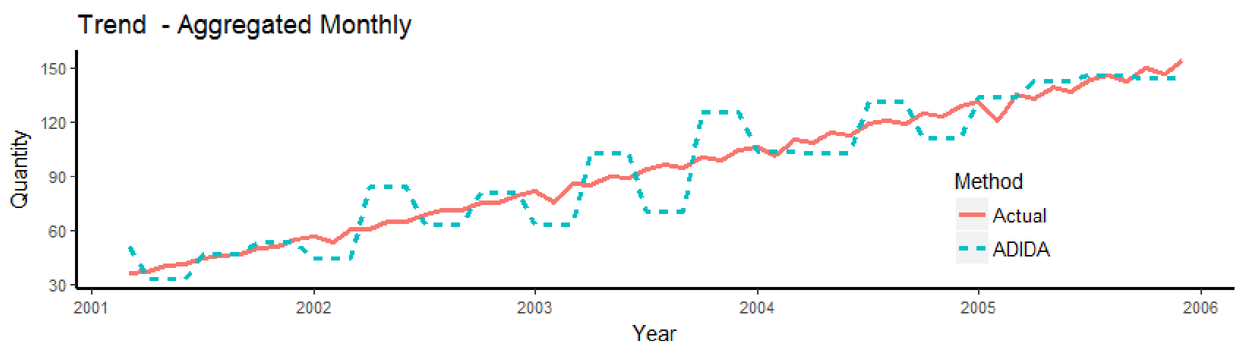


Figure 4-6: ADIDA Method

Finally, the proposed ASACT method is applied to the data. The first step of the ASACT method is to initially aggregate the delivery data into a sufficiently small temporal bin size (to prevent summation during the first aggregation). For the synthetic data, daily bins are suitable. The second step is to smooth the data using Croston's method (Croston, 1972). As with the other methods in the evaluation, the smoothing coefficient (α) was set to 0.2. The last step is to re-aggregate the smoothed data into the temporal bin size that suits the subsequent analysis; in this case, monthly bins. The results of the ASACT method are shown in Figure 4.7. After the initialization period, the data from the ASACT method closely matches the actual data.

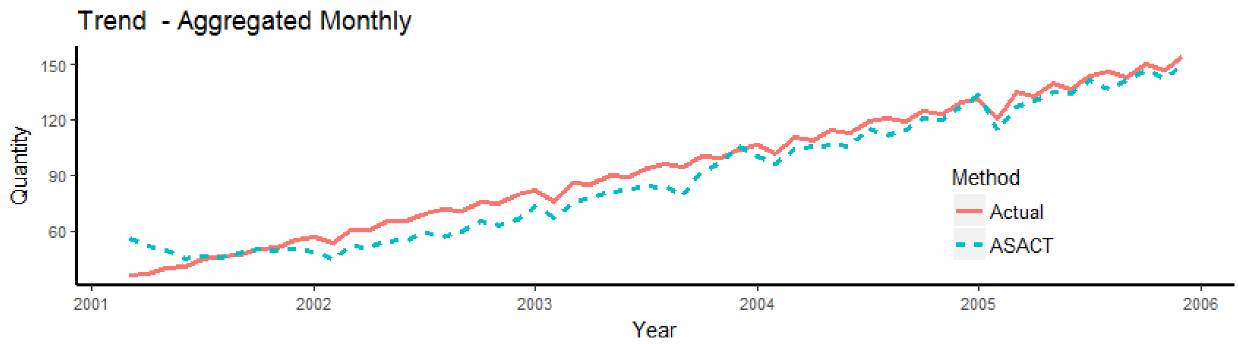


Figure 4-7: ASACT Method

4.4.1.3 Comparison of Methods - Customer C_3

Customer C_3 has a seasonal pattern and thus it can be more difficult to predict than C_2. The data conversion method would need to extract the pattern from the data to facilitate an accurate prediction. The results for each method are displayed in Figure 4.8. The directly aggregated data bears little similarity to the actual data. SES eventually found an average in 2003, but again it is not similar to the original data. Croston and ASACT both quickly fall within the boundaries of the original data although neither one does a good job of following the pattern. ADIDA occasionally follows the seasonal pattern, but its performance is detracted by significant over and under shoots and delays in pattern following. Table 4.2 in Section 4.5 quantifies the results and shows that, overall, ASACT produces the best results for this difficult pattern.

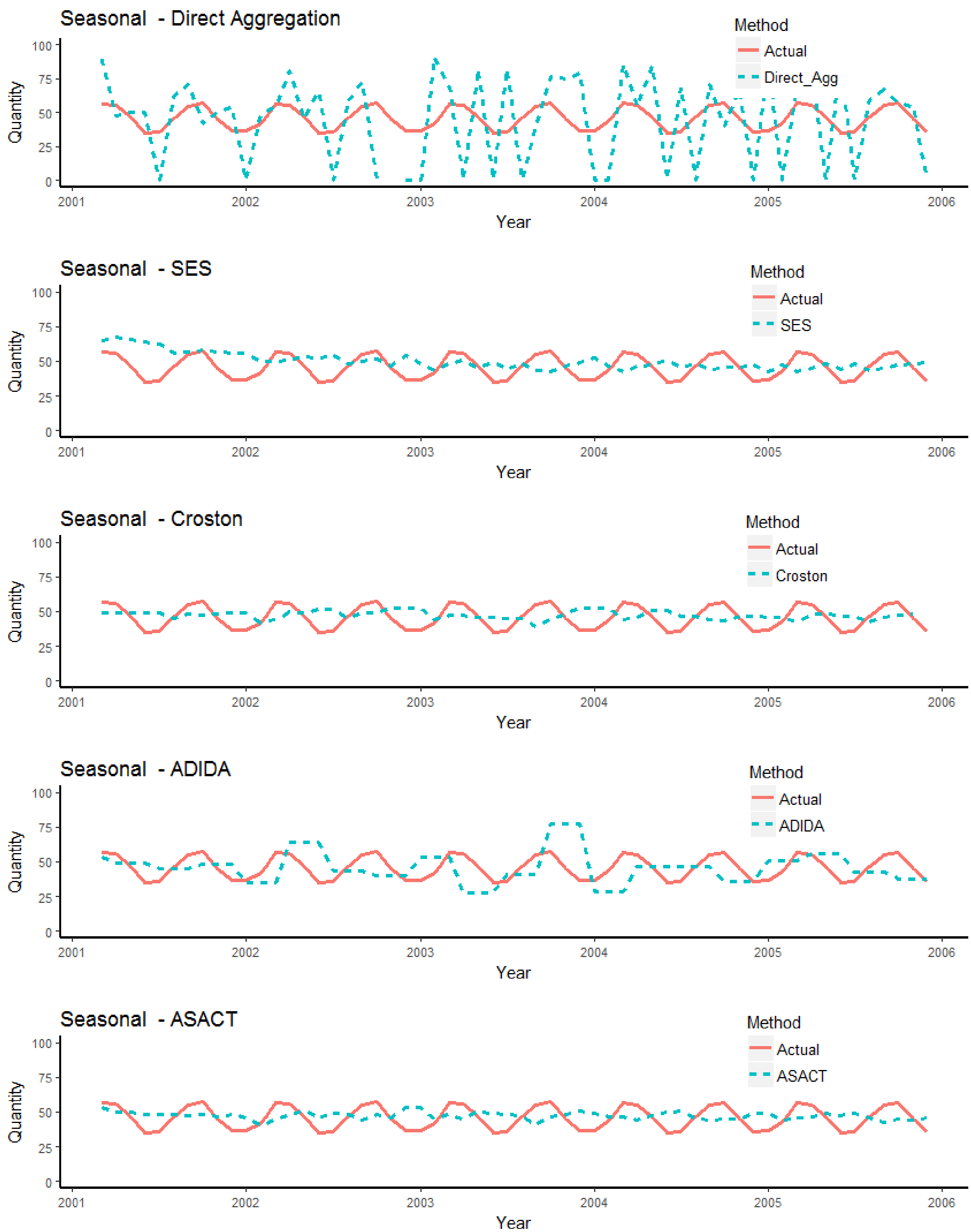


Figure 4-8: Comparitive Resuts for Customer C3 Data Conversions

4.4.1.4 Results on Synthetic Data

The performance of each conversion method is evaluated by measuring the error between the post-processed data and the ground-truth consumption. There are many different error measures available, each having positive and negative attributes. The simplest error measures such as mean error *(ME)* and mean percent error *(MPE)* may provide misleading results due to the nature of positive and negative errors cancelling out. *ME* and *MPE* are calculated as per formulas (5 & 6):

$$ME = \frac{\sum_{t=1}^{n} e_t}{n} \tag{5}$$

$$MPE = \frac{100\%}{n} \sum\nolimits_{t}^{n} \frac{e_t}{a_t} \tag{6}$$

where t = temporal bin, n = number of bins, e = error at time t, and a is the actual value at time t. In some domains, error cancellation is acceptable if the study is mostly concerned with overall magnitude over time rather than matching the demand profile. Root mean square error *(RMSE)* and mean absolute error *(MAE)* resolve the problem of error cancellation as they both effectively measure the distance from actual value to predicted value. *RMSE* and *MAE* both appear frequently in the literature. The results of these two error measures differ due to *RMSE* penalizing large errors more than *MAE*. *RMSE* and *MAE* are calculated as per formulas (7 & 8):

$$RMSE = \sqrt[2]{\frac{\sum_{t=1}^{n} (e_t{}^2)}{n}} \tag{7}$$

$$MAE = \frac{\sum_{t=1}^{n} |e_t|}{n} \tag{8}$$

where t = temporal bin, n = number of bins, and e = error at time. *ME*, *RMSE*, and *MAE* are scale-dependent and only suitable for comparing samples of similar magnitude—a very small error on a very large sample will result in a large error measure. Magnitude is normalized by basing the error on percent, as is done with *MPE* and mean absolute percent error *(MAPE)*. *MAPE* is calculated according to formula (9):

$$MAPE = \frac{100\%}{n}\sum_{t}^{n}\left|\frac{e_t}{a_t}\right| \quad (9)$$

where t = temporal bin, n = number of bins, e = error at time t, and a is the actual value at time t. Since the proposed method is intended for application in many different domains, each potentially preferring different error measures, the evaluation (Tables 4.3 & 4.4) uses all the above error measures. For a more detailed review of error measures and discussion on why some are preferred over others, the reader is directed to an excellent article by Hyndman and Koehler (2006).

Table 4.2 shows the *RMSE* results for delivery type 2 *(D_2)* for the individual customers C_1 through C_9, results with other error measures are in the appendix. For most consumption patterns, the proposed ASACT produced the best results. In the few cases where other methods produced better results, ASACT was not significantly different.

Table 4-2: RMSE for Delivery Type 2 (D2)

(a)	*(b)*	*(c)*	*(d)*	*(e)*	*(f)*	*(g)*
Customer	Pattern	RMSE				
		Direct Agg.	SES	Croston	ADIDA	ASACT
C_1	Stationary	34.20	11.06	1.97	10.96	**1.51**
C_2	Trend (increasing)	30.96	17.48	9.74	12.89	**8.11**
C_3	Seasonal (quarterly)	32.28	11.36	9.83	14.03	**9.66**
C_4	Random	32.60	11.06	2.15	8.79	**1.62**
C_5	Intermittent	31.41	12.89	**8.62**	12.90	9.16
C_6	Random & Trend	31.51	17.80	10.34	11.74	**8.42**
C_7	Intermittent & Trend	38.79	33.01	25.82	**24.21**	28.47
C_8	Random & Seasonal	29.60	11.14	9.58	13.35	**9.36**
C_9	Intermittent & Seasonal	31.59	18.09	**16.52**	17.68	17.66

Table 4.3 shows the average forecast error for all consumption patterns for each method when applied to D_1 data. Delivery type 1 *(D_1)* has consistent and frequent deliveries of at least once per week which, when aggregated into monthly buckets, avoids producing an intermittent time-series. D_1 deliveries are not far up the supply chain from the point of consumption and therefore not subject

to the effects of bullwhip and logistics noise. As a result, direct aggregation produces the most accurate results according to all error measures.

Table 4-3: Error Measures for Delivery Type 1 (D1)

(a)	*(b)*	*(c)*	*(d)*	*(e)*	*(f)*
Method	Error Measures				
	ME	MPE	RMSE	MAE	MAPE
Direct Aggregation	**0.00**	**-0.12**	**5.12**	**3.34**	**5.67**
SES	5.30	1.17	13.10	10.76	17.57
Croston	2.43	-0.50	8.99	7.27	12.76
ADIDA	0.18	-2.32	7.80	6.24	11.77
ASACT	0.49	-1.38	6.77	5.16	9.46

Table 4.4 shows the average forecast error for each method applied to delivery D_2 data. Deliveries in D_2 are less frequent than D_1 and when aggregated into monthly bins, the resulting time-series is intermittent. Direct aggregation still performs well according to *ME* and *MPE*; however, error cancellation may cause these two measures to give a false impression; other error measures all show that direct aggregation gives poor performance. The poor performance of direct aggregation correlates with the opinion that intermittent time-series are difficult to forecast (Petropoulos et al., 2016). According to *RMSE*, *MAE*, and *MAPE*, the proposed ASACT has the lowest overall error. That is, ASACT is the most successful at converting delivery data into consumption data.

Table 4-4: Error Measures for Delivery type 2 (D2)

(a)	*(b)*	*(c)*	*(d)*	*(e)*	*(f)*
	Error Measures				
	ME	MPE	RMSE	MAE	MAPE
Direct Aggregation	**-0.15**	**0.27**	32.55	28.14	65.78
SES	0.76	-9.98	15.99	12.69	24.57
Croston	1.59	-3.35	10.51	8.51	16.15
ADIDA	0.16	-1.87	14.06	11.18	23.62
ASACT	0.88	-3.83	**10.44**	**8.42**	**16.11**

4.5 Evaluation of ASACT with Real Data

In this section, ASACT is applied to a real dataset. Forecast models are usually evaluated through out-of-sample measurement; part of the available data is reserved as "test data" and the models are built on the remaining “training data”. Out-of-sample measurement is considered essential for qualitative evaluation (Chase Jr., 2013). The goal of this research is to use delivery records to predict consumption behavior, not to predict deliveries. Out-of-sample measurement would only indicate whether a method is effective at predicting deliveries and is therefore not a suitable evaluation strategy. We propose to evaluate the results based on whether behavior patterns can be detected in the data.

4.5.1 Context

The real data, provided by an industrial partner, is a series of delivery records that include a data stamp, customer identifier, and the quantity transacted. The data contains outliers due to data-entry errors, lost customers, and administrative adjustments. Outliers were detected and removed by checking for extended periods of no activity, negative delivery quantities, and delivery quantities that greatly exceed the physical limitations of the delivery capabilities.

The industrial partner is a supplier of bulk liquid products that are used in various industries including manufacturing, food packaging, and medical services. In the VMI arrangement, the supplier maintains point-of-use product inventory and is responsible for ensuring uninterrupted product availability. The bulk liquids require specialized point-of-use storage tanks and multi-year purchase contracts are normal; customers typically do not purchase from alternate suppliers during the contract period. Aside from the outlier removal, the data was not summarized, aggregated, or otherwise processed prior to the analysis.

4.5.2 Evaluation of Results with Real Data

Unlike the synthetic data, the real data does not include consumption data to evaluate the results; therefore, qualitative evaluation is necessary. According to the industrial partner, most customers have stable consumption behaviors. The industrial partner also believes that some customers have seasonal and cyclical patterns. If the conversion method is adequate, clear patterns should be evident in the resulting time-series (Murray et al., 2017). To facilitate the evaluation, after conversion, the data is normalized and segmented into four clusters based on behavior patterns using dynamic time warping and hierarchical clustering. Dynamic time warping is selected due to its good performance in measuring similarities between time-series (Keogh & Ratanamahatana, 2005), and hierarchical clustering is selected as a result of its flexibility in selecting a suitable number of clusters (Kantardzic, 2011).

4.5.3 Results with Real Data

We test the real data with two methods, the proposed ASACT method and a contemporary method known as ADIDA (Nikolopoulos et al., 2011). After smoothing, both datasets are scaled to normalize the size of the various customer and then segmented with dynamic time warping (DTW) distance measure and hierarchical agglomerative clustering (HAC). This segmentation method is selected due to its satisfactory results for segmenting customers based on historical time-series data (Rakthanmanon et al., 2013). The results of ADIDA are shown in Figure 9. Although some patterns are visible in Cluster #1 and #4, the graphs are too noisy to reveal any conclusive information.

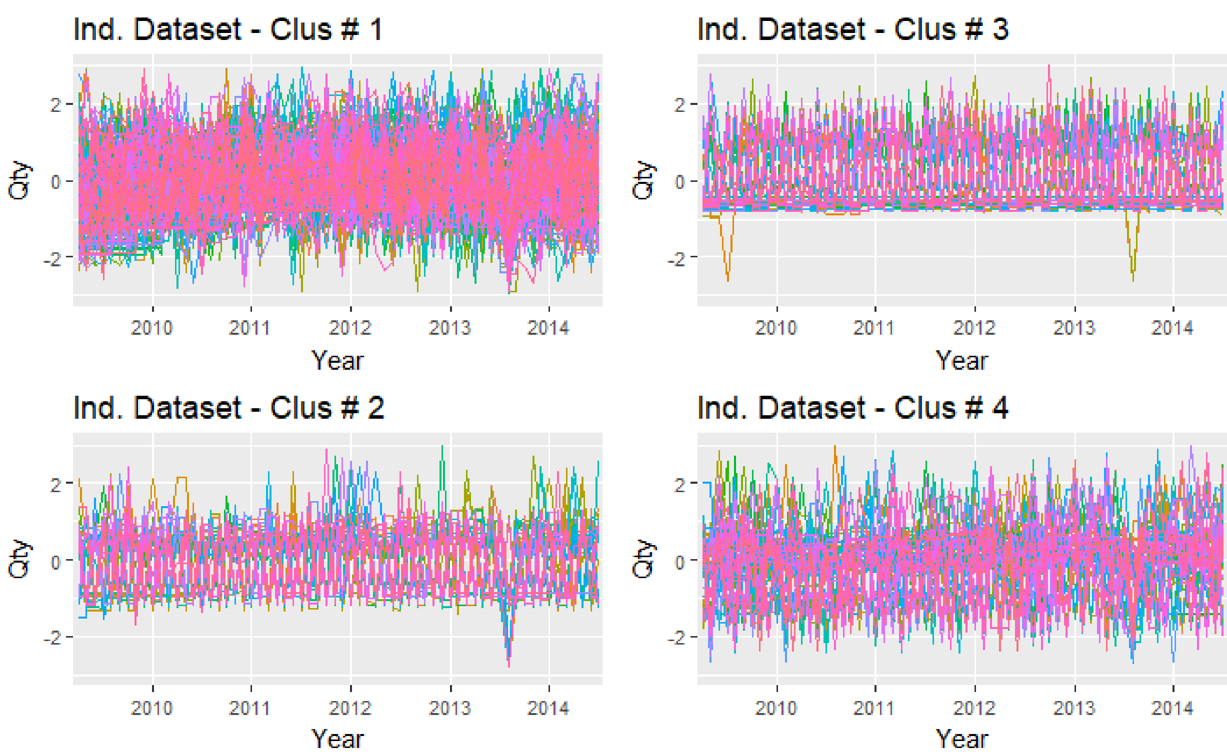


Figure 9: Customer Segments from ASACT data

Next, we apply the same clustering technique to the data smoothed by ASACT. With the ASACT method, the delivery data is first aggregated into its smallest temporal component. The Croston method is applied to smooth the data, and the data is then re-aggregated into weekly bins for analysis. For consistency with the testing of the synthetic data, the smoothing coefficient (α) is set at 0.2. This value of alpha delivers a satisfactory level of smoothing without removing too much information from the data. The clusters in Figure 10 now reveal a variety of useful behavior patterns. Cluster 1 and Cluster 4 show increasing trends occurring at different times in the time-series. Cluster 3 shows a relatively steady decreasing trend, and Cluster 2 shows a cyclical trend. The graphs show a high level of noise in some clusters; the noise could be further reduced with a higher value for α, however, the underlying behavior patterns will also be diminished.

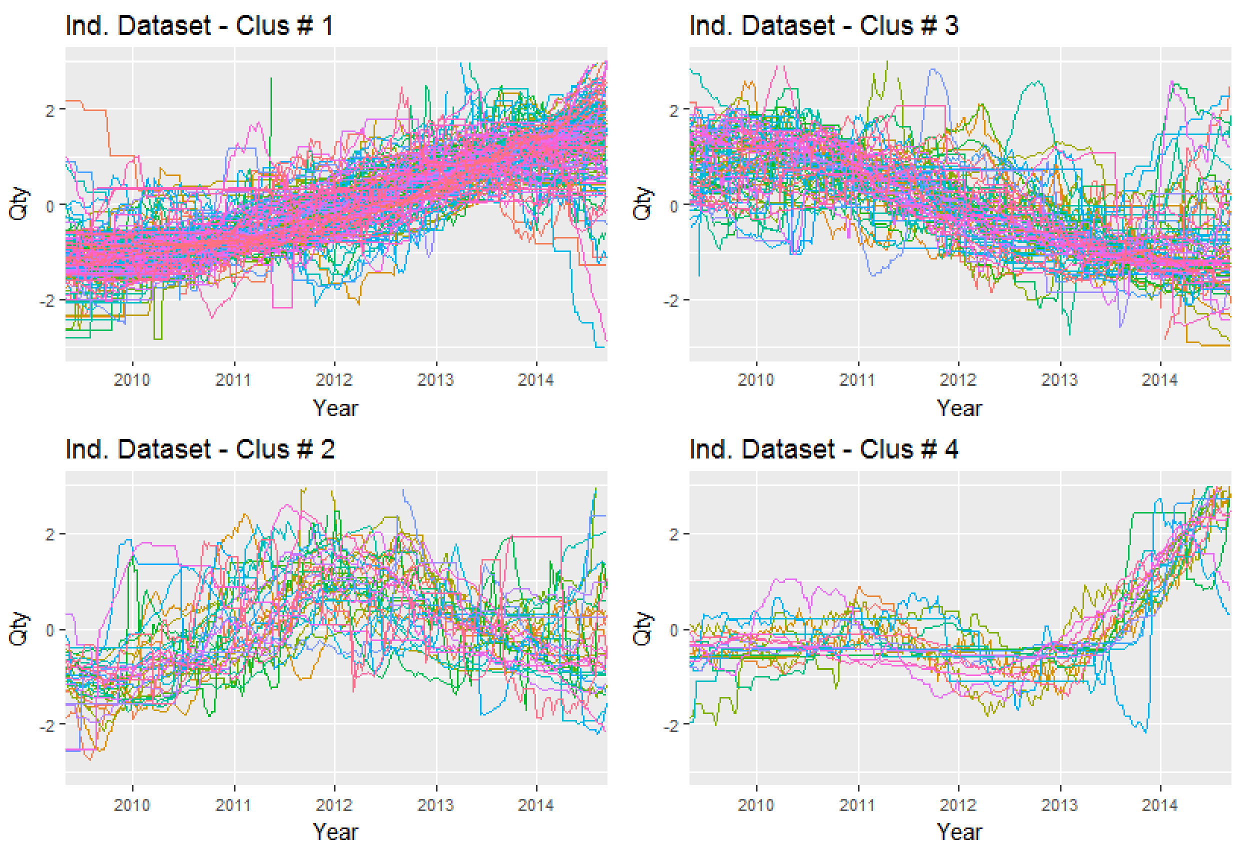


Figure 10: Customer Segments from ASACT data

4.6 Conclusion

Quantitative supply chain forecasting relies on historical consumption information which is not available in some domains. A data source upstream in the supply chain may be used, however. As the data collection point moves further up the supply chain, it is more susceptible to noise inclusion because of the bullwhip effect and logistics decisions. Aggregating delivery records into a time-series format may add an additional level of noise. Tests on delivery strategy D_1, representing a data collection point very close to the point of consumption, demonstrate that direct aggregation performs best. However, tests on delivery strategy D_2, where deliveries are infrequent and varied in size and frequency, show that a directly aggregated time-series is very noisy and does not effectively represent actual consumption behavior. Existing methods for conversing and smoothing noisy data are generally an improvement over direct aggregation; however, they also tend to remove important behavior patterns.

The proposed ASACT method, when applied to the noisy data set (D_1) is effective for removing noise while preserving the underlying behavior patterns. Quantitative evaluation shows that ASACT produces less error than other methods for most types of data. ASACT should be applicable to any domain where data is collected upstream from the point of consumption. Determining how distant the available data is from the point of consumption is an important task considering that close data, such as D_1 experiences a decrease in accuracy from all smoothing methods. Indicators such as variations in frequency and the quantity of deliveries can guide the decision, although a more robust strategy is needed.

In this research, we attempted to apply the proposed method and the compared methods in a consistent manor so that results were not biased. Most methods can be adjusted and optimized by adjusting different parameters such as bin size and the smoothing coefficient. Parameter optimization, however, may be domain specific and not necessarily valid for other domains. Further research is necessary to develop rigorous strategies for specifying the parameters. Further research is also needed to establish qualitative evaluation when applied to real data when the supporting ground-truth consumption information is not available.

The ASACT method has direct management applications in any domain where upstream data is substituted for missing consumption information.

CHAPTER 5 ARTICLE 2: MARKET SEGMENTATION THROUGH DATA MINING: A METHOD TO EXTRACT BEHAVIORS FROM A NOISY DATA SET

Published in: Journal of Computers and Industrial Engineering

Murray, P., Agard, B., & Barajas, M.A., (2017)

Abstract: *Strategic business planning requires forecasted information that contains a sufficient level of detail that reflects trends, seasonality, and changes while also minimizing the level of effort needed to develop and assess the forecasted information. The balance of information is most often achieved by grouping the customer population into segments; planning is then based on segments instead of individuals.*

Ideally, separating customers into segments uses descriptive variables to identify similar behavior expectations. In some domains, however, descriptive variables are not available or are not adequate for distinguishing differences and similarities between customers. The authors solved this problem by applying data mining methods to identify behavior patterns in historical noisy delivery data. The revealed behavior patterns and subsequent market segmentation are suitable for strategic decision-making. The proposed segmentation method demonstrates improved performance over traditional methods when tested on synthetic and real-world data sets.

5.1 Introduction

Strategic planning relies on the ability to gather information regarding the customer base and being able to process that information into predictions of future requirements. In situations where the customer base is large, a method to reduce the number of possible scenarios is necessary. Scenario reduction is commonly achieved in business through a process known as market segmentation. Early research has demonstrated that forecasting groups of customers (or products) is more accurate than aggregated individual forecasts due to the positive effects of error smoothing and error cancelation (J Scott Armstrong, 1985; Dangerfield & Morris, 1992). Market segmentation was first proposed by Smith (1956) as a method of differentiating customers based on the customers' individual preferences and desires. However, despite extensive literature regarding market segmentation, there is little guidance to managers on how to accomplish it (Dibb & Simkin, 2001; Hung & Tsai, 2008).

The traditional segmentation methods found in the literature rely on identification or creation of a set of descriptive variables from which the differences (distances) between customers are calculated, for example, see (Calvet, Ferrer, Gomes, Juan, & Masip, 2016). These methods calculate distances between members based on attribute-based variables or feature-based variables. An attribute-based distance relies on descriptive attributes such as sex, age, location, industry, or size. Attributes are sometimes converted into numerical variables and organized into standard forms prior to processing (Kantardzic, 2011). With feature-based distance, variables are created based on statistical features such as median, kurtosis, sum, or purchase frequency, from historical data. The features are weighted, summarized, and normalized to become the variables from which distance is calculated (Y. Chen, Zhang, Hu, & Wang, 2006).

Creating variables, from either attributes or features benefits the data analysis in two key areas. First, high variance and outliers are reduced or eliminated by condensing the data into a pre-determined set of variables. Second, the resulting dataset is much smaller than the original data and algorithms run much faster. Traditional segmentation methods share two fundamental assumptions: the first assumption is that the data-munging process used to create, modify, and weight the data has resulted in a set of variables that truly reflect the intentions and behaviors of the customers. The second assumption is that the information that describes the demand behavior is present within the available data (Longbing, 2014; Verdu, Garcia, Senabre, Marin, & Franco, 2006). If the assumptions are incorrect, the variables will not adequately describe the customers and the analysis will not produce robust results.

When descriptive variables are not available and cannot be adequately created from the available data, the traditional segmentation methods do not work. We propose to solve this problem by identifying behavior patterns directly from the historical delivery data and then using those patterns to segment the market. A major challenge with this approach is that delivery data can appear noisy when the delivery frequencies and quantities are driven partly by consumption demand and partly by transportation logistics decisions. The noise effect is especially evident when a point-of-use inventory exists between the delivery point and the consumption point. The level of noise in the data increases as the point of data recording moves up the supply chain (F. Chen, Ryan, et al., 2000).

The scientific relevance of this work is as follows: A method for identifying market segments in the absence of descriptive variables is developed and tested; market segments are based on consumption behavior patterns which are extracted from noisy delivery data. Testing the new market segmentation method on synthetic data demonstrates that the method can identify behavior patterns from a noisy data set and create useful groups based on those patterns. The application of the method to real-world data generates results that are useful in strategic planning applications.

The remainder of the paper is organized as follows: Section 5.2 presents a review of the state of the art. Section 5.3 presents the proposed method. Sections 5.4 and 5.5 apply the proposed method to both synthetic and real-world data and compare the proposed method to traditional segmentation method. The paper concludes in Section 5.6 with a discussion of the scientific contribution, limitations of the research, and suggestions for future research.

5.2 Literature Review

5.2.1 Market Segmentation

Market segmentation is a topic with both a long history and active contemporary investigation; its benefits were identified in a seminal paper by Smith (1956). Market segmentation for planning, marketing, and forecasting has become nearly universal practice in business; in fact, Bain & Company's "Management Tools & Trends 2015" listed it as one of the top ten executive management tools globally (Rigby & Bilodeau, 2015). The importance of market segmentation can be traced to guidance given by Aristotle to tailor your message to the audience. A more recent example is General Motors Corporation's approach to build "a car for every purpose and purse" (Gann, 1996). Armstrong (1985) found that forecasting segments rather than individual customers resulted in higher accuracy due to the reduced effect of outliers and irregularities in the data. Regardless of the specific techniques and steps used, all segmentation methods follow the same essential sequence: information or data relating to the customers is gathered, differences and similarities between customers are determined, and finally, the customer base is divided into groups with similar behavior. Figure 5.1 illustrates five segmentation strategies, including a priori, key attributes, descriptive attributes, statistical features, and behavior model. The segmentation strategies are ordered according to the level of rigor and analytical effort of each strategy. At one end, the a priori method is purely qualitative; its results depend entirely on the knowledge of the practitioner. At the other end, the behavior extraction methods are rigorous enough that they can theoretically be applied by different practitioners in different domains and produce satisfactory results. The following subsections will develop the relevant segmentation strategies in more detail.

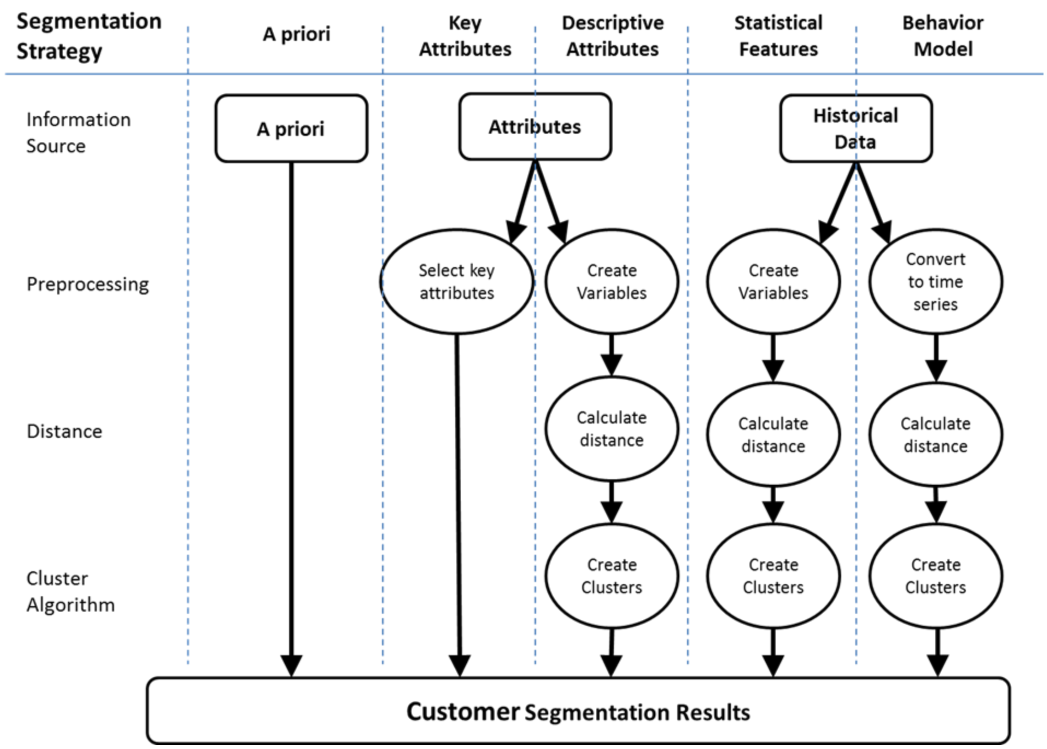


Figure 5-1: Segmentation Strategies

5.2.1.1 A Priori Segmentation

The a priori method, followed by Aristotle and General Motors, relies on the analyst instinctively knowing how to separate the groups (Gann, 1996). The flexibility of qualitative decision making makes a priori segmentation useful when the available data does not consistently support the observed results (Randle & Dolnicar, 2009). A priori method can be fast and when the analyst's knowledge is sufficient, it can be very accurate. However, since the a priori approach is entirely qualitative and depends on the level of knowledge the analyst has about the customer base; it can be slow and prone to errors (Ettl, Zadrozny, Chowdhary, & Abe, 2005).

5.2.1.2 Key Attribute Segmentation

The banking industry began to recognize in the 1970's that market segmentation solely based on demographics and socio-psychological characteristics was not accurate and began using quantifiable determinant attributes. In their seminal paper, Anderson et al. (1976) defined a list of customers' "determinant attributes" (such as sex, married, age, or education level) that should be used to calculate customers' similarity. Tsiptsis and Chorianopoulos (2011) presented several frameworks for segmentation, each beginning with identifying attributes that are used to describe the customer's expected behavior. Athanassopoulos (2000) expanded the attribute-based approach by adding simple statistical features (such as: staff size, assets, or years of operation) to improve the description of the expected customer behaviors. Key attribute segmentation is only suitable when the necessary attributes are available. The method also heavily depends on the assumption that the key attributes actually distinguish differences between groups.

5.2.1.3 Descriptive Attribute Segmentation

Descriptive attribute segmentation builds on key attributes by converting the attributes into numeric variables. The numeric variables can then be standardized and analyzed using statistical analysis techniques (Kantardzic, 2011). While descriptive attributes are better for quantitative analysis, the drawbacks of key attribute segmentation method remain.

5.2.1.4 Statistical Feature Segmentation

Feature-based distance involves extracting statistical features such as median, kurtosis, sum, or purchase frequency, from historical data (Bala, 2012). The features are normalized and become the variables from which distance is calculated (Y. Chen et al., 2006). The tasks of creating, normalizing, and importance-weighting variables—frequently referred to in literature as "data munging"— do not follow a set of rules or procedures. The accuracy of its results and of subsequent analysis depends on the skills of the analyst. Data munging is considered as much art as science (Heer & Kandel, 2012). A criticism of using statistical features is that the method is not resistant to outliers. Features such as mean and variance can be greatly compromised by outliers (Park & Leeds, 2016).

Attribute-based and feature-based segmentation share two fundamental assumptions. The first assumption is that the data munging has resulted in a set of variables that reflect the customers' behaviors. The data analyst must decide the attributes or features to include as variables and the weight to be applied to each. When deciding what variables to create, the data analyst must be cognizant of computer limitations; the computational expense of some clustering algorithms such as neural networks leads some analysts to dimensionally reduce the variables. For example, in preparation for feeding a neural network, Romdhane, Fadhel, and Ayeb (2010) calculate information entropy for each variable and then retained only the most informative attributes. Byrnes (2014) suggests that as the number of attribute variables increases the ability of an algorithm to detect useful information decreases. Variable creation is a qualitative decision process that can greatly influence the outcome of the analysis and may produce inaccurate or misleading results.

The second assumption is that the available data contains enough information to describe the behaviors adequately (Longbing, 2014; Verdu et al., 2006). Attribute information, such as the age of the customer might be essential for accurate segmentation, but if the data is not available, the missing variable might be problematic. Information about strongly influential exogenous factors may be missing from the available data—for example, in a vendor managed inventory arrangement (VMI), the supplier's logistics decisions might influence demand patterns and yet no information with regard to those decisions exists in the data.

5.2.1.5 Behavior Model Segmentation

Customer behavior can be used to classify customers based on pre-determined segments, which reflect whether a customer is increasing level of business, stable, or at risk of leaving (Ettl et al., 2005). In situations where customers do not have regular interaction, variables are necessary to measure and describe the behavior. Historical transaction data may contain patterns that can be useful to predict behaviors. A large customer base may lead to a large and noisy historical dataset from which behavior patterns are difficult to distinguish. Depending on the point in the supply chain where data is collected, there may be more or less noise incorporated. Research into the Bullwhip effect indicates that as the point of data collection moves upstream from the point of consumption, the data incorporates more noise (F. Chen, Ryan, et al., 2000). Simple aggregation of individual predictions is undesirable since higher level forecasts generally do not equal the aggregated summed forecasts (Rob J Hyndman, Lee, & Wang, 2014), and one great model for the entire group will give little information due to the noise in the data.

Chao, Fu, Lee, & Chang (2008) use historical data to measure customer behavior, employing an intermediate preprocessing step which converts the data into RFM (recency, frequency, monetary) variables and then weight those variables. The conversion to RFM and weighting involves subjective decisions and hence has similar limitations to the variable creation strategies discussed previously. As with key attribute segmentation, behavior model segmentation relies on the availability of variables or other information that adequately describes the behavior.

5.2.2 Measures of Similarity

With most traditional segmentation methods, variables are created and then distances between customers are calculated to create a distance matrix. Once a distance matrix has been calculated, general-purpose clustering algorithms can be applied to identify clusters (Liao, 2005). The most popular distance metric in the literature is Euclidean straight-line distance. Simplicity, robustness, and relatively good performance make Euclidean a frequent comparison baseline or foundation for other methods (for examples, see Chang, Tay, & Lim (2015), Hung & Tsai (2008), or Serra & Zárate (2015)). While Euclidean distance performs well with static data, it does not do well when directly applied to time-series data. Euclidean distance is constrained to point-to-point evaluation; similar patterns such as examples "C5" and "C8" in Figure 5.2, which have temporal shifts, are not recognized. Euclidean distance treats each time-period as a separate, unrelated variable.

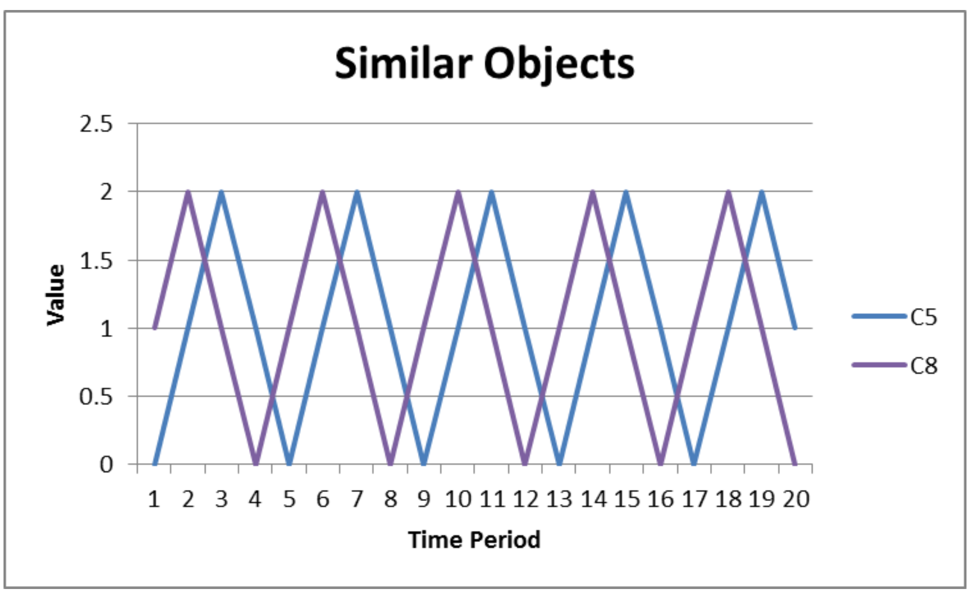


Figure 5-2: Similar Objects with Temporal Shifted Patterns

Evaluating time-series with temporal shift is possible when the distance is calculated with cross correlation (CCor) distance. The CCor distance has the characteristic that noise in the data does not significantly impact the results (Golay et al., 1998). The CCor method is limited in that it makes global adjustments in an attempt to find the best correlation between time-series. In effect, it is similar to Euclidean distance, but with added constants (Agrawal, Faloutsos, & Swami, 1993).

Variations in time-series data are never as simple as the examples shown in Figure 5.2. Typical variations include changes in frequency, magnitude, and duration. While shifting dissimilar time-series with CCor is often quite successful (Höppner & Klawonn, 2009), there are other techniques, such as dynamic time warping, which perform well on time-series datasets (X. Wang et al., 2013).

Dynamic time warping (DTW) was first developed as a technique for speech recognition (Sakoe & Chiba, 1978); it was introduced in the 1990's into data mining research for comparing time-series (Keogh & Pazzani, 1999). DTW is different from other distance measures as it uses non-linear mapping to compare pairs of time-series. While an excellent description of the details of the DTW algorithm can be found in Keogh and Ratanamahatana (2005), a brief illustration is presented in Figure 5.3. Although the patterns of objects "A" and "B" in Figure 5.3 are very similar, the relationship is not linear. The DTW algorithm warps the time series to achieve a best match between objects and then calculates the distances.

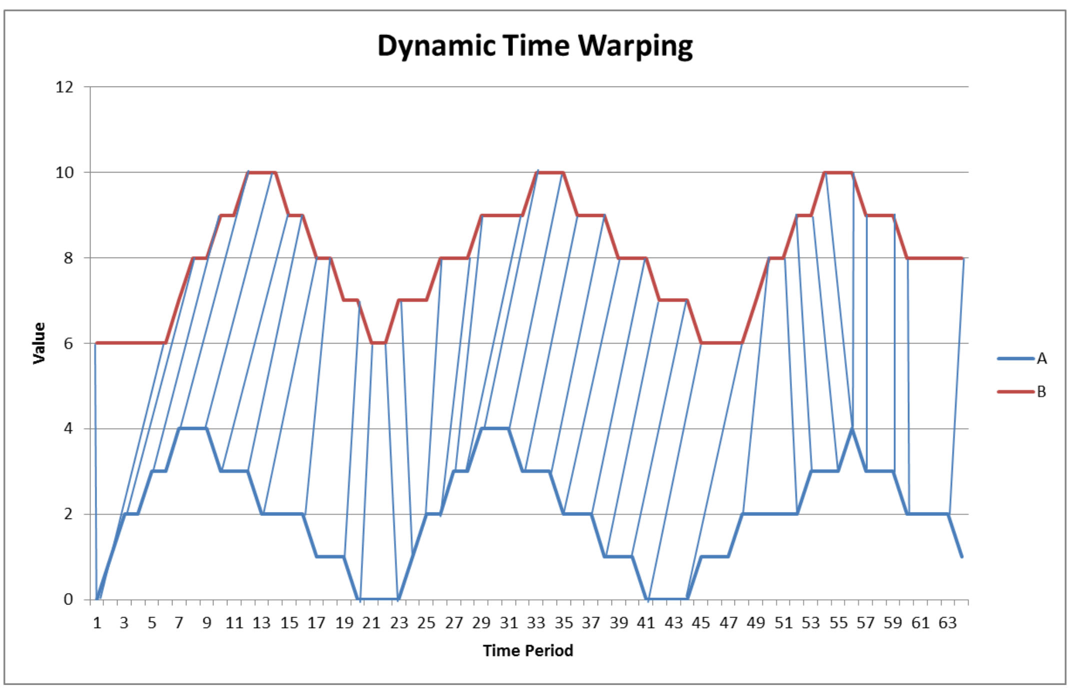


Figure 5-3: Example of DTW

While DTW delivers good results, some researchers are reluctant to use it. A major criticism of DTW is its high computational cost which sometimes results in its elimination as a viable distance measure (Z. Zhang, Kaiqi, & Tieniu, 2006). Two developments offset DTW's high computational cost. First, as computers become faster and less expensive, computational cost becomes more of an abstract consideration. Second, the continual refinement and development of DTW algorithms make the distance calculations faster (Ratanamahatana & Keogh, 2005). Extensive comparison testing of DTW against other distance methods has demonstrated that criticisms that DTW is exceedingly slow are mostly unfounded (X. Wang et al., 2013). The authors believe that despite DTW's slow performance—compared to Euclidean distance calculation—the concern is not great enough to prevent its use.

5.2.3 Cluster Methods

Obtaining accurate segmentation results depends on the selection of a good clustering method (Kashwan & Velu, 2013); of the many available clustering methods, some are too complex for practical applications (Agard, Partovi-Nia, & Trepanier, 2013) and not all are suitable for time-series data. A review by Liao (2005) lists the most common time-series related clustering methods, including partitional, artificial neural networks, and hierarchical. Partitional clustering methods, such as K-means are the most commonly used and often employed as a baseline for comparison of other clustering methods, see for example, Hautamaki et al. (2008). Although partitional clustering methods create clusters efficiently, the results do not always provide useful information (Murray, Agard, & Barajas, 2015). Additionally, K-means algorithm requires a pre-determination on the number of clusters, which is a challenge in most applications.

Artificial neural networks (ANNs), also known as self-organizing maps (SOMs) (Altintas & Trick, 2014) do not need the number of clusters pre-defined and often give good cluster results. ANNs have generated interest in the forecasting field recently (Co & Boosarawongse, 2007), however, they have drawbacks. One criticism of ANNs is that the algorithms are computationally expensive (Ye & Li, 2002) leading some analysts to employ dimensional reduction of variables prior to computation. See, for example, Bala (2012) or Romdhane (2010). While the dimensional reduction allows the algorithm to operate more efficiently, it also incorporates a subjective decision process into the analysis, which may reduce the accuracy of the results. ANNs characteristic of not needed the number of clusters pre-defined makes the method suitable for the first state of a two stage clustering analysis where the first stage defines the number of clusters (Kuo et al., 2002).

A third type of clustering method is hierarchical clustering (Ward, 1963). In hierarchical clustering, the first step is the calculation of similarity (distance) between customers. The distance matrix is used to construct a hierarchy that begins with all customers in individual clusters and ends with all customers in one cluster. The hierarchy is often displayed in a dendrogram as illustrated in Figure 5.4. The dendrogram facilitates visual assessment of the cluster results and enables the incorporation of a "cut-line" which established the number of clusters in the result. Unlike K-means, the number of clusters does not have to be pre-determined.

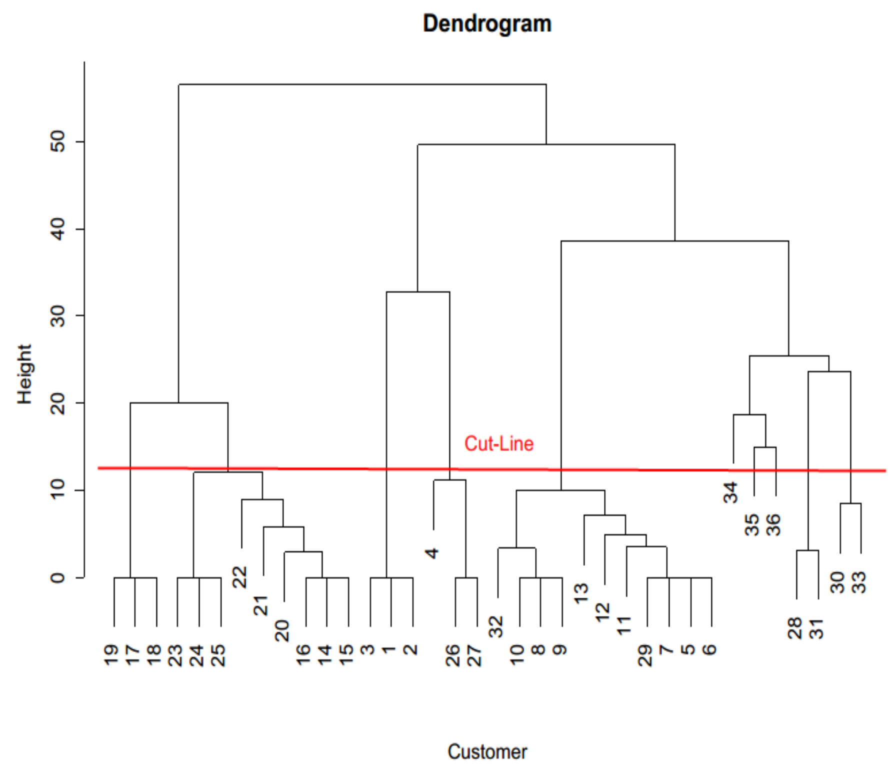


Figure 5-4: Dendrogram of Hierarchical Clustering Results

5.2.4 Cluster Evaluation

There are many methods to create clusters, ranging from simple a priori to complex algorithms, and nearly every cluster method shares the common trait that clusters (good or bad) are created. No algorithm guarantees that genuine clusters are created and different algorithms often create different cluster results for no discernable reason (Huang et al., 2001). An important step in the segmentation process is to evaluate the clusters' goodness of fit. The literature generally agrees that segmentation results can be evaluated through comparison of inter-cluster homogeneity and intra-cluster heterogeneity (Ding et al., 2008; Liao, 2005; Seret, Maldonado, & Baesens, 2015). Wang et al. (2013) propose using a one-nearest neighbor classifier on labeled data to generate quantifiable results. The evaluation methods share a common limitation in that they compare one cluster method to others; there is an underlying assumption that the best clusters are "good" clusters. Keogh et al. (2003) conclude that due to data and implementation biases, "many of the results claimed in the literature have very little generalizability to real world problems". The lack of accepted evaluation criteria and ambiguity in testing confuses the selection of a "good method" for an evaluation of cluster methods.

Humans' natural ability to detect patterns visually is excellent, and in some cases superior to computers' abilities (Chellappa et al., 1995; Esling & Agon, 2012; Van Wijk & Van Selow, 1999). It would appear then that visual evaluation of cluster results should prove effective—provided that the data is presented in a format humans can visualize (Huang et al., 2001). Unfortunately, high dimensional datasets often contain so many points that it becomes impractical to display graphically precluding visual evaluation (Hung & Tsai, 2008; Kantardzic, 2011).

5.2.5 Market Segmentation through Data Mining

Market segmentation is both an important part of business management and an active area of contemporary research. Traditional methods employ a variety of strategies with varying degrees of a priori knowledge necessary for successful application. Data mining breaks from traditional paradigms and explores the discovery of knowledge without the preconceptions pre-established hypothesis (Agard & Kusiak 2004). Successful application of data mining, however, requires careful selection of data mining tools that are able to measure differences between objects (or customers) and then arrange those objects into sensible groups. Finally, a means to evaluate the results of the analysis is required. Unfortunately for the data mining practitioner, there is no tool or prescription for selecting the best technique at each step of the data mining process—rather, it is an experimental and iterative process that includes subjective decisions at different points.

Market segmentation through data mining relies not only on selection of suitable algorithms to analyze the data, but also on suitable inputs to feed into the algorithms. Extracting behaviors from the data requires careful consideration of how the data should be processes so that it actually reflects the behavior (Kantardzic, 2011).

5.3 Methodology

Market segmentation is traditionally based on descriptive attributes of the customers or on a set of variables created from various data sources. Our method, illustrated in Figure 5.5 utilizes only historical data and does not convert the data into variables. This section describes the method to import the raw data, clean it, and extract behavior patterns from it. To validate the methodology, we use a synthetic dataset in section four and a real dataset in Section 5.5. We compare our results with a traditional method (presented in Section 5.5.4).

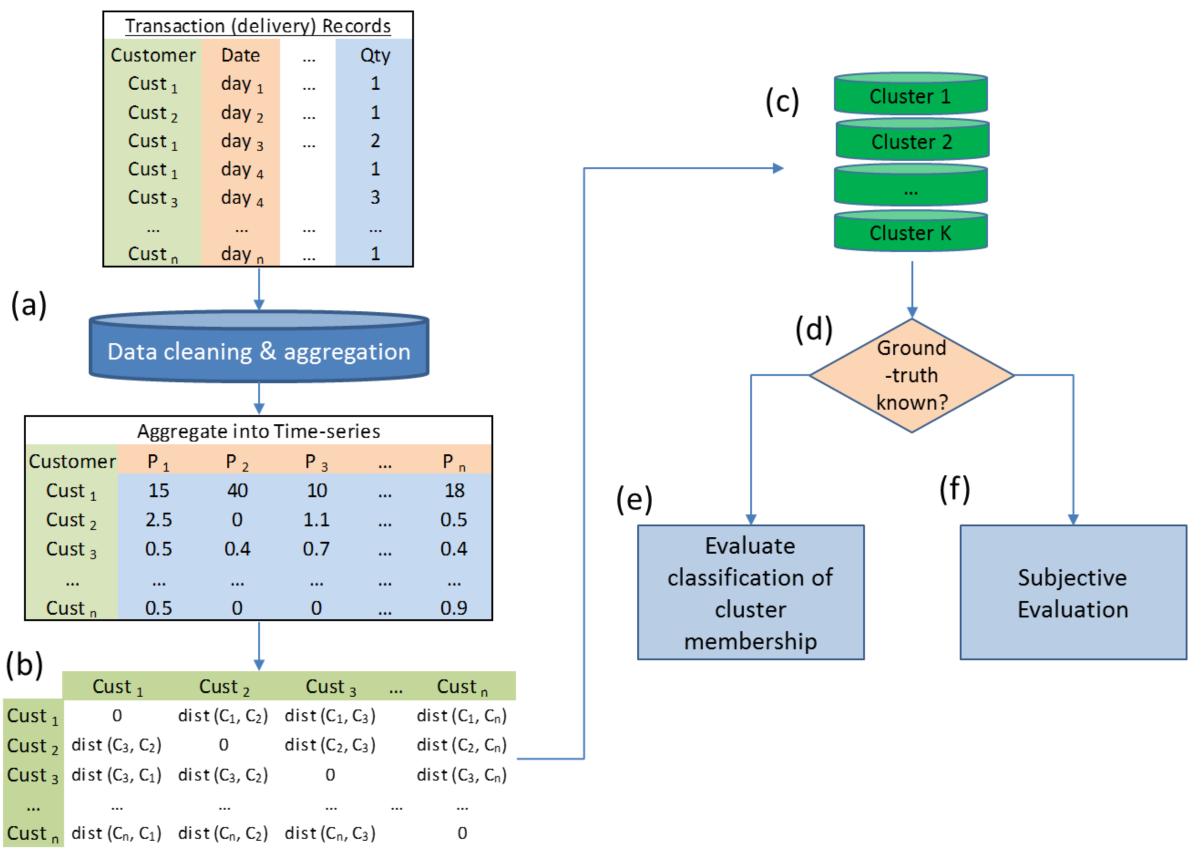


Figure 5-5: Methodology

5.3.1 Data Preprocessing

The industrial data in its raw format is a set of delivery records; each record contains a date stamp, quantity of product delivered, origin & destination locations, and a customer identifier. In preparation for analysis, the data is first cleaned per Figure 5.5(a). Outliers exist in the data due to administrative adjustments (account debits & credits) and errors in data recording. Negative and unreasonably large values are considered outliers and removed from the dataset. Delivery records are then aggregated into monthly bins (monthly bins suit the industrial partner's planning activities). The resulting sets of customer specific time-series contain a significant level of noise due to irregular delivery frequencies and quantities. Customers with less than monthly deliveries exhibit intermittent patters due to months with no activity. Newly acquired and lost customers exhibit long periods of zero activities. And, casual or low-consumption customers have very infrequent and unpredictable patterns. Although the analysis would be much easier if all the difficult to predict customers were removed, it would not produce useful real-world results. Therefore, outlier removal was limited and the resulting dataset appears very noisy.

5.3.2 Segmentation

The literature contains many different segmentation methods, which are summarized in the literature review above. Nearly all of these segmentation methods depend on the availability of several attributes that when analyzed together can lead to useful groupings. In the context of this research, the industrial data contains only time and quantity; there is not enough information for most segmentation methods. The available data can be interpreted to reflect how much product was consumed during each period and therefore is suitable for behavior model segmentation method.

Distance calculation, per Figure 5.5(b), is a critical component to the segmentation process. Once the distances are calculated, segments are identified via a hierarchical clustering algorithm per Figure 5.5(c). Hierarchical clustering generates good results and enables visualization of the relationship between clusters via the dendrogram output (Barirani, Agard, & Beaudry, 2013). Some researchers have found that hierarchical clustering does not give accurate results, but they were unable to determine the extent that variable selection and distance calculations affected their results (Maqbool & Babri, 2007). Liao (2005) suggests that the actual clustering calculation method does not significantly have an effect on the outcome. Therefore, hierarchical clustering is used for comparing all distance methods.

The segmentation tests were conducted with “R” language (R Core Team, 2015) using Euclidean, ANN (Wehrens & Buydens, 2007), DTW (Giorgino, 2009) and CCor (Mori, Mendiburu, & Lozano, 2015). The most common methods are extensively reviewed in Mori et al.(2015) and Wang et al. (2013).

Artificial neural networks (ANN) and K-means were also tested to validate the proposed methodology. Unlike the methods discussed above, ANN and K-means do not have separate steps to calculate a distance matrix and then build clusters; rather, they produce clusters directly from the raw data. Lastly, a statistical feature-based distance method is included. The feature-based method is a traditional variable creation segmentation method.

5.3.3 Cluster Evaluation and Patterns Extraction

When deciding how to best evaluate the cluster results, the analysist must determine whether the ground-truth cluster assignment exists, as illustrated in Figure 5.5(d). Although there are many statistical methods available for evaluating cluster results, the goal of this research is not to compare and conclude whether the clusters satisfy some arbitrary statistical measure. Rather, the research goal is to determine if one (or many) method can produce “good” clusters in terms of extracted behaviors. The problem of visualizing high-dimensional data is resolved by using a synthetic dataset to test various clustering methods; the synthetic dataset is simple enough to enable visual evaluation of the resulting clusters. While simple, the synthetic data is sufficiently complex that not all algorithms will return a satisfactory solution. We leverage humans’ natural ability for pattern recognition in the evaluation.

Determining a good number of clusters is a challenge yet to be solved and remains as much of an art as a science (Kantardzic, 2011). Many distance and cluster options were tested in this research to evaluate the effects of different cluster sizes while retaining the ability to display the results. For this research, the synthetic data described in the following section is developed and evaluated with eight clusters. For the industrial data, 8 and 24 clusters appeared pertinent in the dendrogram; this level of segmentation suits the managerial needs of the industrial partner. This research uses cluster homogeneity to evaluate the effectiveness of the cluster methods.

5.3.3.1 Cluster Evaluation – Synthetic Dataset

For the synthetic data, the ground-truth cluster assignment is known, so the analysis can proceed according to Figure 5.5(e). We developed the synthetic dataset with the intention that clusters can be readily distinguished by human visual evaluation. Predetermined clusters establish the ground-truth of the expected cluster results from which the results of each clustering method are compared. For the second step of evaluation, the results from each segmentation method are graphically displayed for visual evaluation. The two-step evaluation established not only which segmentation method performs best, but also whether the methods produce "good" clusters.

5.3.3.2 Cluster Evaluating – Industrial Dataset

Unlike the synthetic dataset, the industrial dataset has no ground-truth segmentation and the analysis must rely on a subjective assessment, per Figure 5.5(f). The subjective evaluation is enhanced by utilizing a baseline for comparison. For the baseline, variables are extracted from the raw data using simple statistical analysis; sum, mean, kurtosis, skew, frequency, and standard deviation are used as variables. Euclidean distance is applied to calculate a distance matrix based on the baseline's variables. The baseline and the proposed method both begin with the same raw data; segmentation for both is done with hierarchical clustering. Once clusters are created, the resulting graphical output are inspected for evidence of patterns in the data. Although a visual evaluation lacks the rigor of quantifiable measure, it offers a valid assessment of whether or not an industrial practitioner can detect patterns such as seasonality, increasing trends, or lost customers. The kind of results expected are explained in Table 5.2.

5.4 Synthetic Dataset

Selecting a suitable distance measure and clustering algorithm are critical to the overall method in this research. A synthetic dataset was created to enable easy and effective evaluation of the results of each candidate distance measure.

5.4.1 Research Dataset

The synthetic dataset was developed by plotting customers with eight different types of simple behaviors patterns. Within each group, the customer exhibit behaviors that are similar, but occur at different magnitude levels, time lag, or have slight variations and noise. When viewed as a single group in Figure 5.6, the synthetic data appears complex; visually, it is difficult to detect more than two or three distinct patterns in the data.

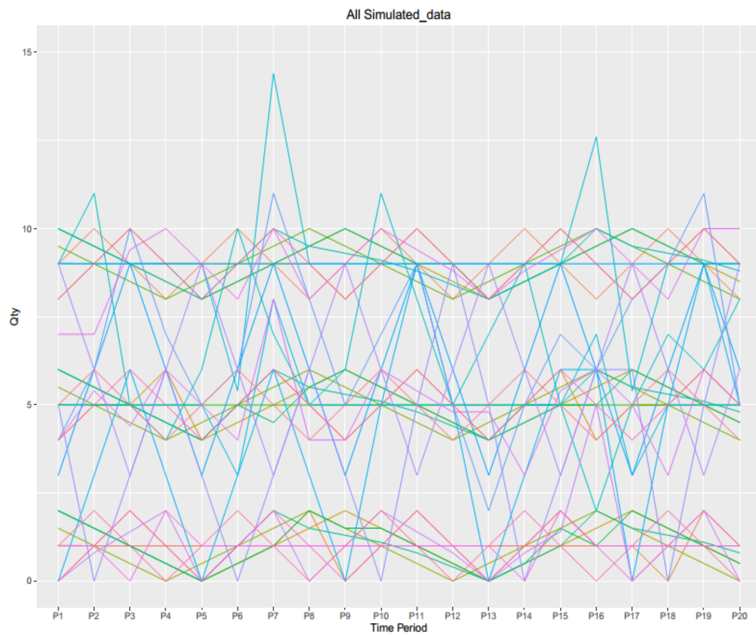


Figure 5-6: Synthetic Data Aggregated

It is expected that a good clustering method will detect the similar behaviors and not be incorrectly influenced by where or when the behavior occurred. Figure 5.7 shows the synthetic data manually separated into the expected clusters based on behavior similarities—we consider the cluster assignment in Figure 5.7 as the ground-truth for evaluating segmentation strategies. Although the synthetic data appears extremely simple, some distance methods and clustering algorithms preformed surprisingly poorly. The test results are in the following section.

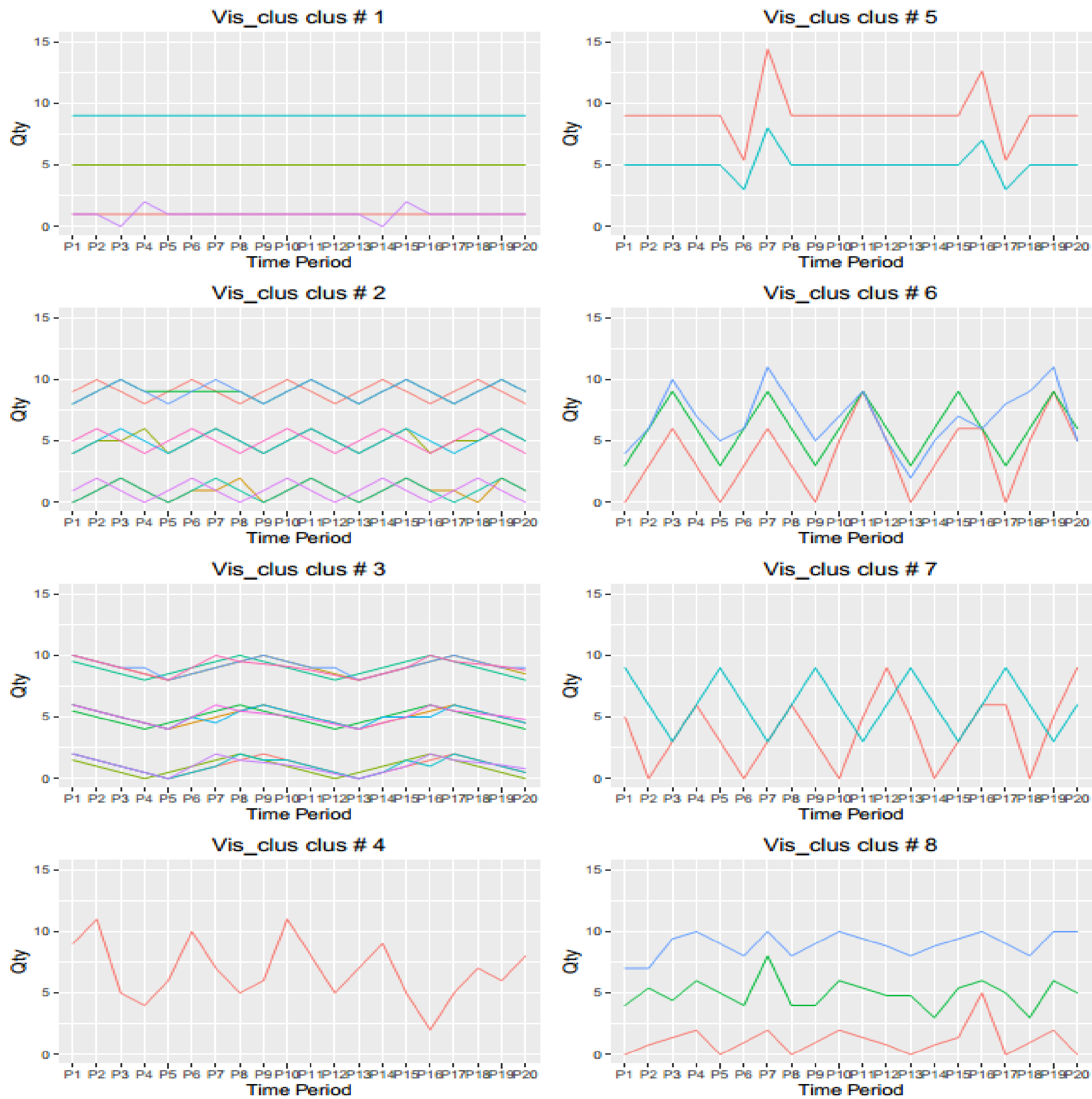


Figure 5-7: Expected Clusters in Synthetic Data

5.4.2 Segmentation – Synthetic Dataset

The goal of the research is to find a method that leads to logically matched cluster members. The evaluation of the different methods was based on how closely the results are to the expected results from Figure 5.7. The synthetic data was passed to the distance calculation and clustering algorithms and cluster assignments were generated by each method. For each test, the number of clusters (K) was set to eight (the expected number as illustrated in Figure 5.7). Table 5.1 shows the results of each method when applied to the synthetic data. The maximum value for the results in Table 5.1 is 36—indicating a complete match to the visual baseline; lower values indicate that the strategy does not produce an expected result. The data is normalized to remove the undue effect of the size of each customer; behavior patterns are preserved while the effect of magnitude is eliminated. The results in Table 5.1 show that, with the exception of CCor, every method performs significantly better with normalized data.

Table 5-1: Segmentation Results

Number of Customers segmented as expected		
Method	Raw data	Normalized data
Statistical Features	11	19
K-means	12	27
Euclidean	13	29
ANN	6	27
CCor	30	30*
DTW	17	31
Visual (expected results)	36	na

* Normalized data does not affect CCor results

Some data mining methods require normalization of the data to calculation. Normalization prevents overweighting features that have larger average values (Kantardzic, 2011). In the tests conducted, most methods performed better with normalized data. CCor distance produces the same results with or without normalized data due to its inherent scaling within the correlation calculations.

5.4.3 Discussion on Synthetic Dataset

The synthetic dataset is intentionally simple to ensure human visual evaluation could be used to conclude whether clusters were "good" or not. The simplicity enabled the establishment of the baseline expected cluster results. We observed that DTW with scaled data produce the best results; DTW produced 31 correct results when compared to the baseline. However, DTW detected similarities that were not expected. For example, the bottom member (blue) of Cluster 2 in Figure 5.8 was expected to be included in Cluster 1 due to its nearly flat pattern. The small variation is visually similar to other members of Cluster 2, although it does not follow the expected results, the solution appears to be valid. We conclude from the results that DTW with scaled data is able to generate acceptable clusters.

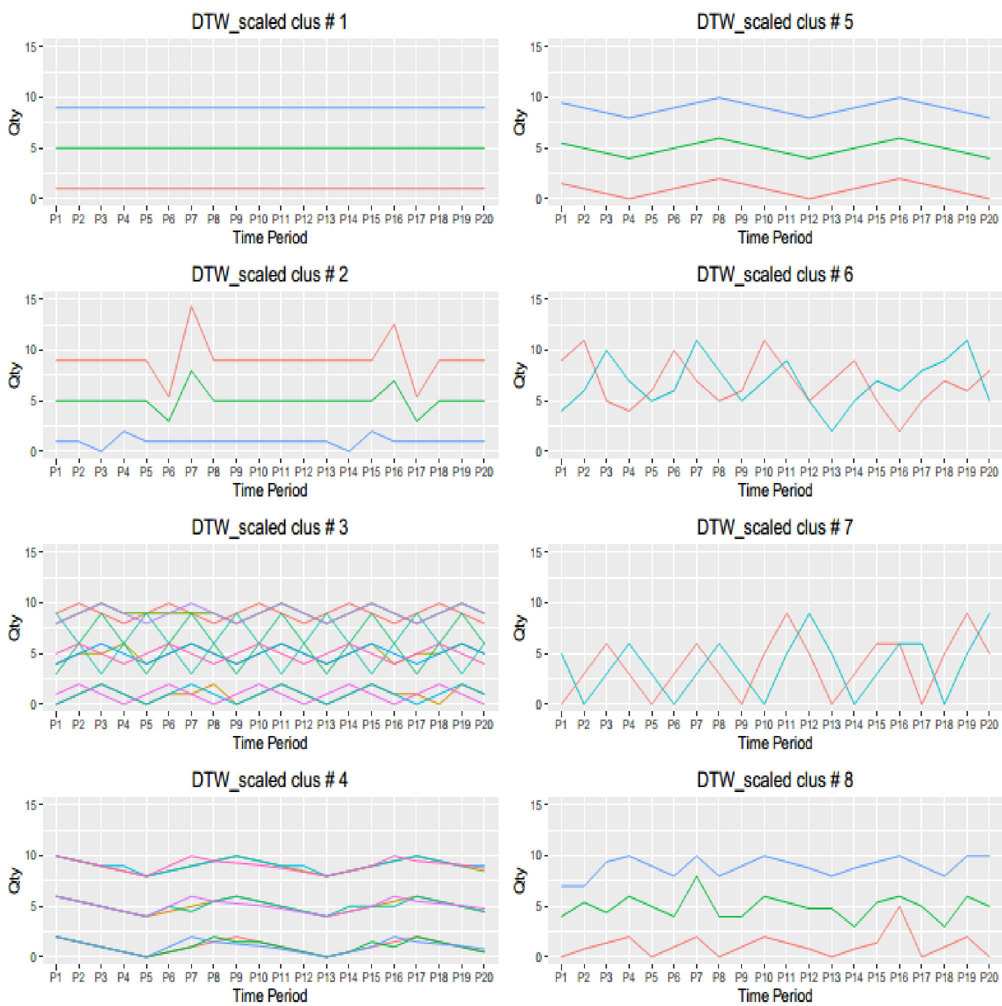


Figure 5-8: DTW with Normalized Data

5.5 Industrial Dataset

5.5.1 Context

A supplier of bulk liquid materials—used for manufacturing, food packaging, and medical services—provided the industrial dataset used in our research. The supplier maintains point-of-use inventories of the bulk liquid and is responsible for ensuring uninterrupted product availability. The bulk liquids require specialized supplier-specific point-of-use storage tanks and multi-year purchase contracts are the norm; customers typically do not purchase from alternate suppliers during the contract period. Delivery dates and quantities for a five-year period are available for the study; actual customer-level consumption data is not available.

Figure 5.9 represents the original data set aggregated into monthly bins. Each line represents delivery pattern for one specific customer (units on the y-axis are suppressed for confidentiality). We can observe a large variation between customers, in terms of quantities, number of deliveries, and delivery dates, yet no patterns appear between customers—this is the data that is actually available to the company to make prediction models, with more or less success.

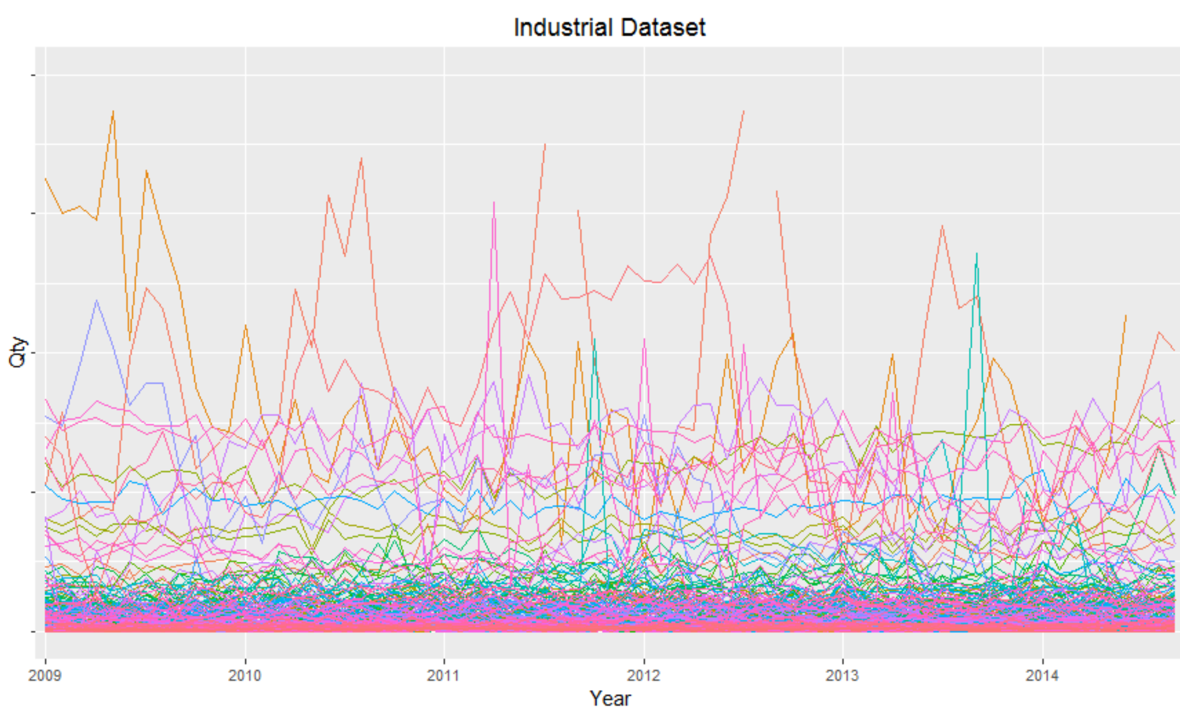


Figure 5-9: Industrial Dataset

The customers represent many different industries that use the product for different purposes. Various internal and external factors influence each customer's operation and in turn affect their specific product consumption. The customers' location and industry type are known; however, segmentation based on these attributes does not result in any similarities in consumption behavior patterns. Therefore, knowing who the customers are and their industry type does not provide useful information to predict future consumption rates.

In the VMI arrangement of the case study, delivery requirements are triggered when sensors in the point-of-use storage tanks indicate a low level. The supplier attempts to both minimize delivery frequency and avoid point-of-use inventory stock-outs. Although the ability to adjust delivery quantities and frequency affords the supplier flexibility in its logistics, the resulting historical data, based on deliveries, does not necessarily reflect the customers' actual consumption behavior. For example, while two customers may have identical consumption behavior, the supplier might use different delivery strategies due to distance from warehouse or proximity to other customers. The resulting historical data can indicate different demand behavior for customers with identical consumption behavior.

The industrial partner views its production and delivery of products as a continuous process with monthly production volumes as the preferred measure. The available data, however, is not a measure of a continuous process, but instead it is a set of delivery records. While some customers have frequent and consistent deliveries, others have very infrequent deliveries and/or highly varied delivery quantities. Additionally, the total demand quantity varies from very small customers to very large When viewed as a monthly aggregated time-series format in Figure 5.9, the data reveals little information; customers with relatively small consumption quantities are all banded together in the lower region of the graph.

Preprocessing the industrial data began with normalizing to remove the effect of overall magnitude; normalization preserves behavior patterns, and produces superior cluster results. The normalized data in Figure 5.10 is no longer influenced by customer size leaving only behavior patterns, however, the normalized data reveals no discernable patterns.

The following normalization formula is used:

$$x_{norm} = \frac{(x - \bar{x})}{\sqrt{\frac{\sum (x - \bar{x})^2}{(n - 1)}}} \qquad (1)$$

where x is the value of the delivery quantity, $\bar{x}$ is the mean of all delivery quantities, and x_{norm} is the normalized value. The chosen normalization method is well suited to this type of data due to its resistance to the distorting effect of outliers.

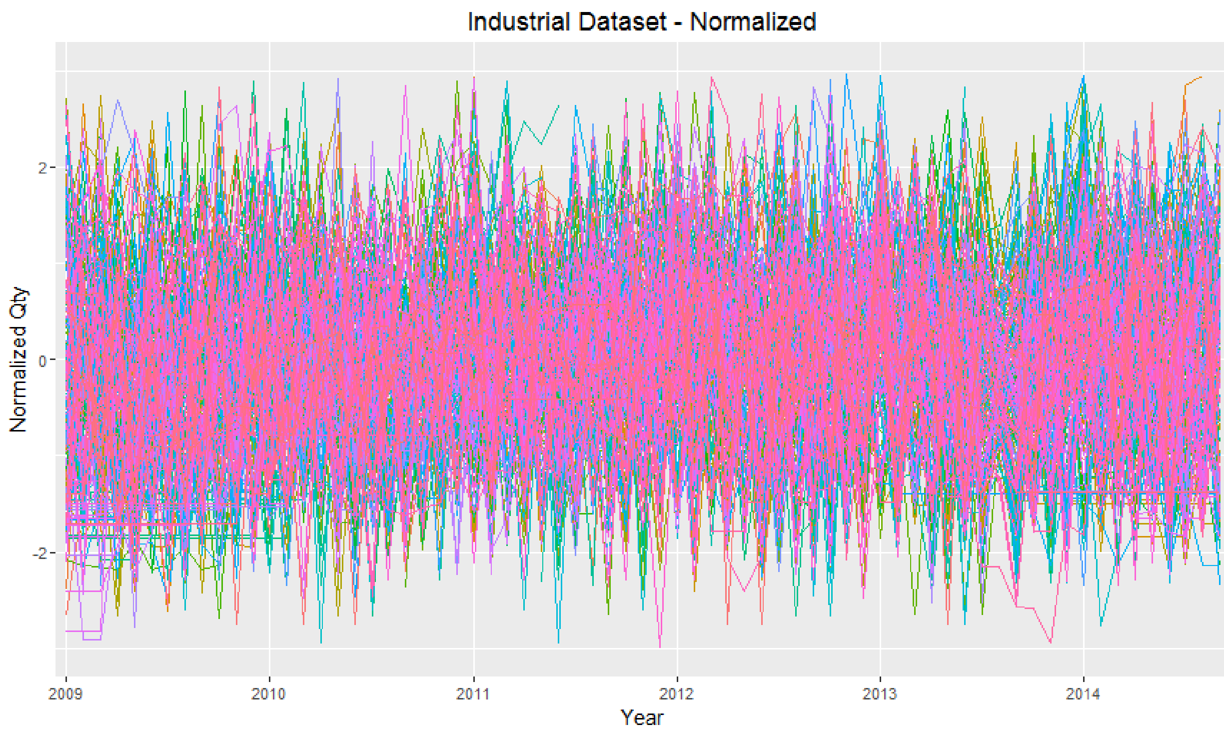


Figure 5-10: Industrial Dataset - Normalized

5.5.2 Calculating a Distance Matrix

In order to group the customers into segments with similar behavior patterns, a measure of similarity is needed; DTW is employed to calculate the distances. The industrial dataset comprises a matrix of many time-series data; a wrapper algorithm enables direct application of the DTW calculations to the time-series matrix (Mori et al., 2015). Within the DTW algorithm, the time-series are elastically stretched (warped) to create a best fit between time-series (Keogh & Ratanamahatana, 2005). The warping is constrained by a window of two to prevent over-manipulation and loss of important behaviors such as seasonality (When unrestrained, DTW will search the entire time-series in attempt to find the closes match between. A warping window is used to constrain the algorithm and prevent it from overly distorting the timeseries (Keogh & Ratanamahatana, 2005)[1]). After warping, the algorithm uses Euclidean distance to calculate the pair-wise distances.

5.5.3 Segmentation – Industrial Dataset

We applied the clustering methodology (using DTW and hierarchical clustering) to the industrial dataset to separate it into sensible groups. Unfortunately, there is no automatic method for determining the best number of clusters; selecting number of clusters is one of the most difficult problems in data mining (Anil K. Jain, 2010). The industrial partner in this research is interested in limiting the number of clusters to a relatively small and easily managed number while still having enough separations to enable detection of different behaviors. To enable an objective selection of the number of clusters, we calculated within cluster homogeneity and in-between cluster heterogeneity.

- Within cluster homogeneity is the mean value of all pairwise distances between all elements within the same cluster
- In-between cluster heterogeneity between clusters A and B is the mean distance of all pairs of elements a_i and b_j where a_i belongs to A and b_j belongs to B.

Homogeneity is a measure of similarity of members of a cluster, and heterogeneity is a measure of dissimilarity between clusters. When distance is used as the measure, it is desirable to minimize heterogeneity distance and homogeneity distance (Fisher, 1958). The cluster evaluation displayed in Figure 5.11 indicates that for less than eight clusters (K<8) the results do not look stable. Additionally, for K>25, there is only little change. Therefore, considering the industrial partner's

[1] The addition of this note in the published paper is on the recommendation of the Jury

preferences and the quality of information that we anticipate, we tested “K” equal to 8 and 24 for this research.

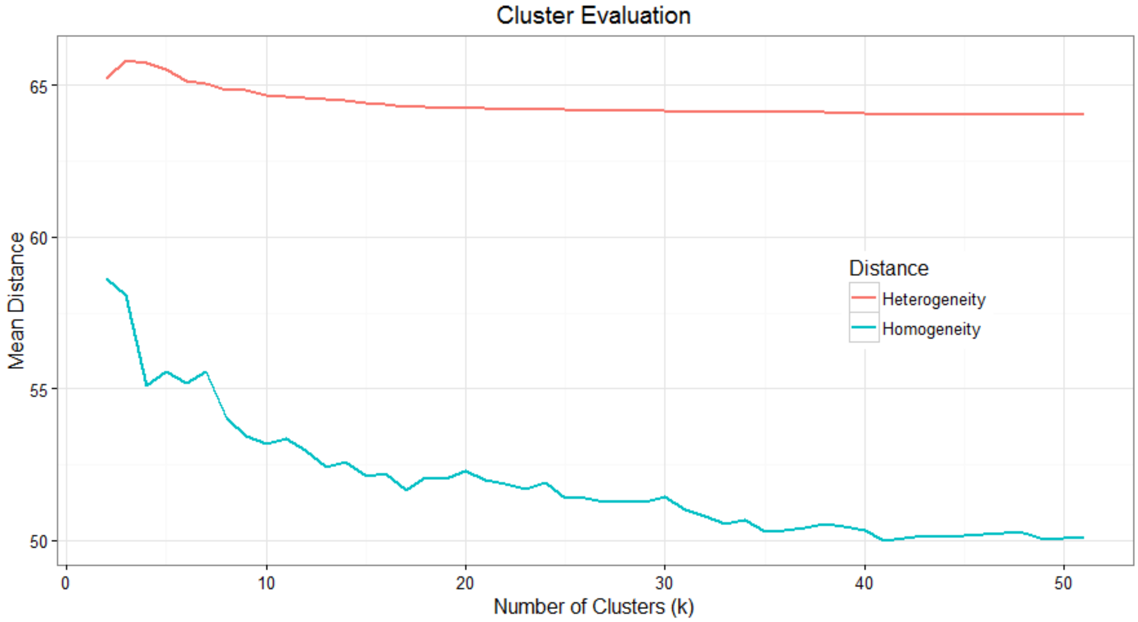


Figure 5-11: Cluster Evaluation

Figure 5.12 is a dendrogram that shows the relationship between elements of the industrial data. The red rectangles illustrate the makeup of the clusters when split into 8 clusters.

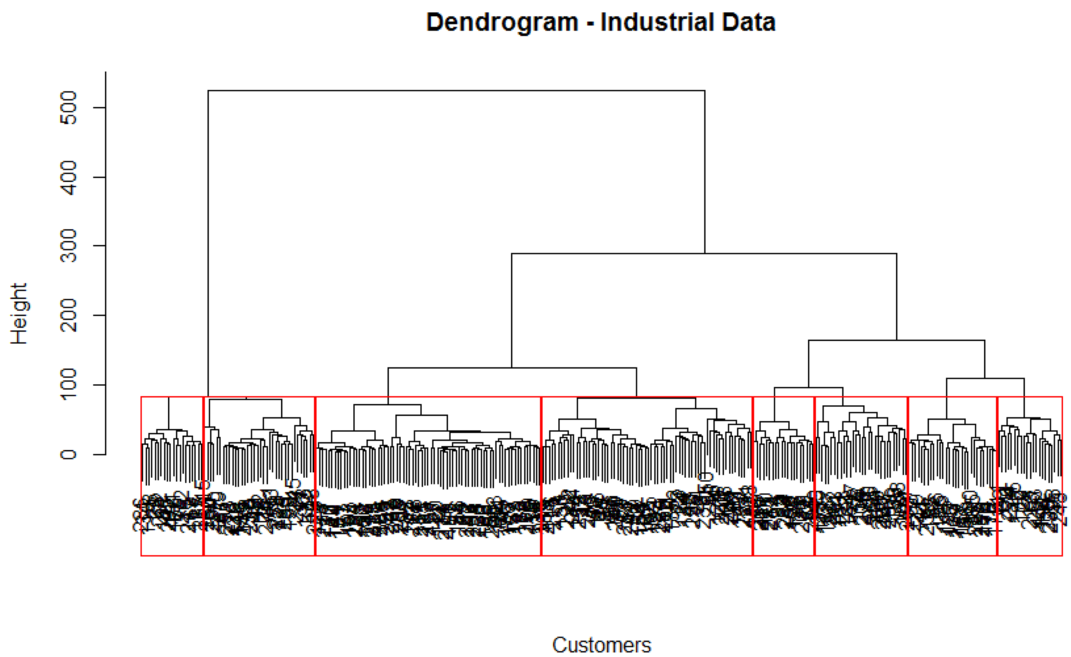


Figure 5-12: Dedrogram of Industrial Data

5.5.4 Results of Segmentation Attempts

Segmentation is performed using both a traditional segmentation method and the proposed method. The results in Figure 5.13 are for K=8 clusters using the proposed method. Although the data remains quite noisy, the sub-groups begin to show some limited information in the behavior patterns. Most clusters indicate stable demand, although Cluster 5 and Cluster 8 appear more stable than the others. Cluster 3 shows an increasing trend over the entire period of the data; Clusters 2 and 4 also show an increasing trend, although it is not as evident as in Cluster 3 and only occurs during the first few years. Cluster 7 contains customers with a mixture of no activity and spikes in activity.

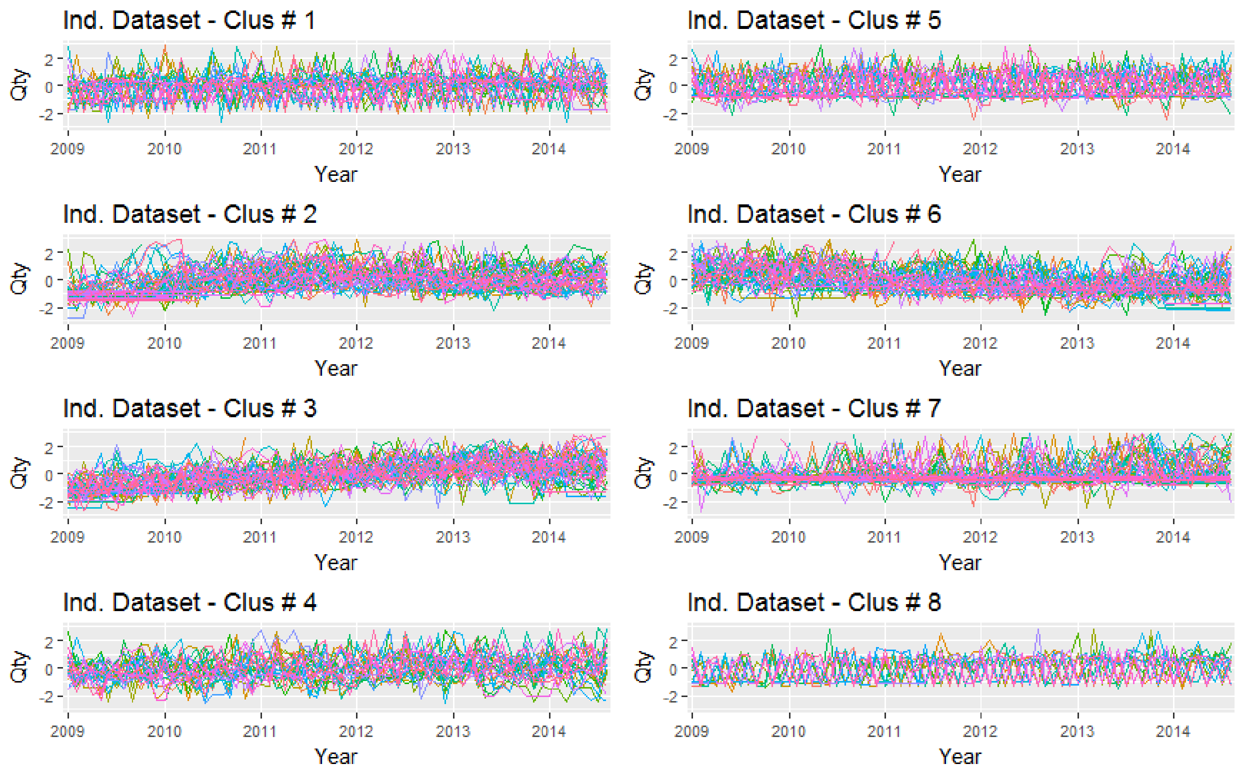


Figure 5-13: Proposed Segmentation Method, 8 Clusters

The same dataset is segmenting using a traditional segmentation method; variables are created by extracting statistical information from the data. The results of the traditional method are shown in Figure 5.14. Unlike the results from the proposed method, the traditional segmentation method produces clusters that offer no apparent indication of behavior pattern. The traditional method has assigned most customers to only three clusters while leaving Cluster 7 and Cluster 8 with only one customer in each.

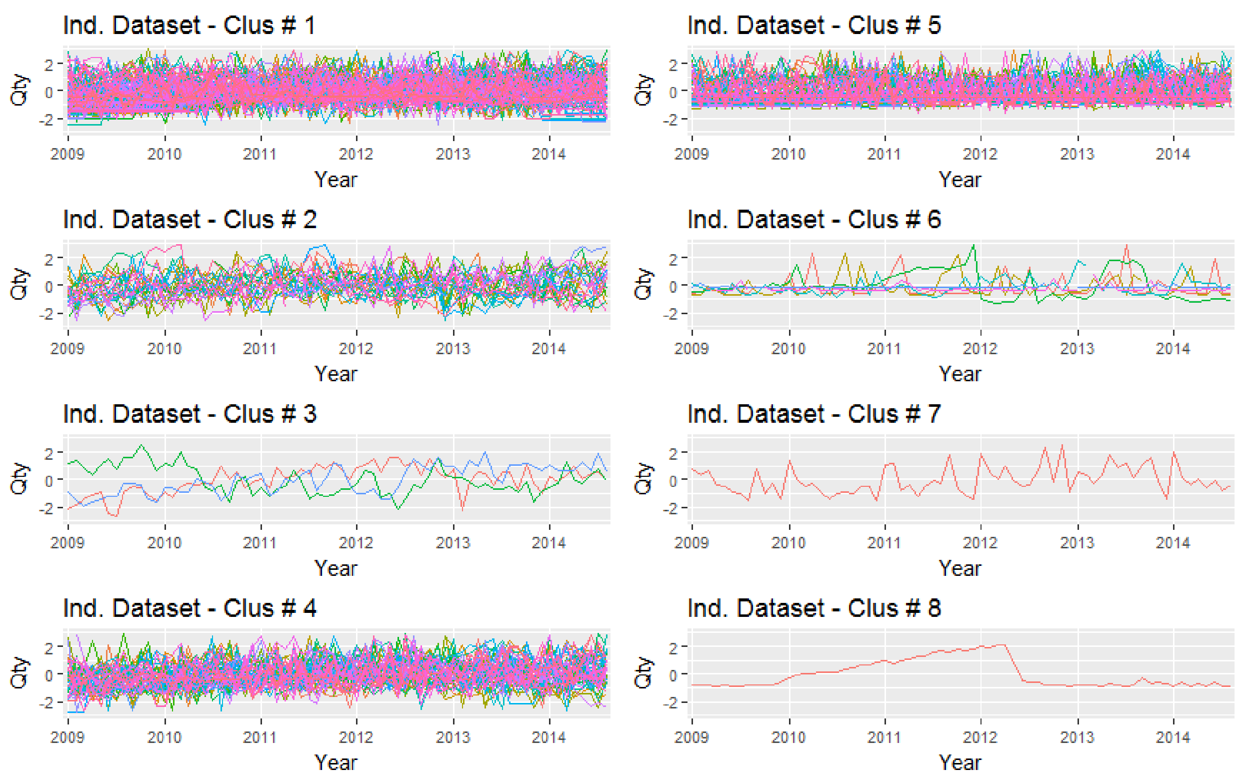


Figure 5-14: Traditional Segmentation Method, 8 Clusters

Both segmentation methods are again tested, this time using K=24 clusters. With the proposed method, the clusters become clearer and more patterns emerge, illustrated in Figure 5.15. Clusters 3 & 17 both indicate increasing trends, however, they have slightly different patterns. Cluster 2 shows a multi-year cycle; a similar cycle may also be evident in Cluster 7. Cluster 6 and 24 both appear stationary over time (not increasing or decreasing), however, Cluster 24 has less variation and should produce predictions with a higher level of confidence.

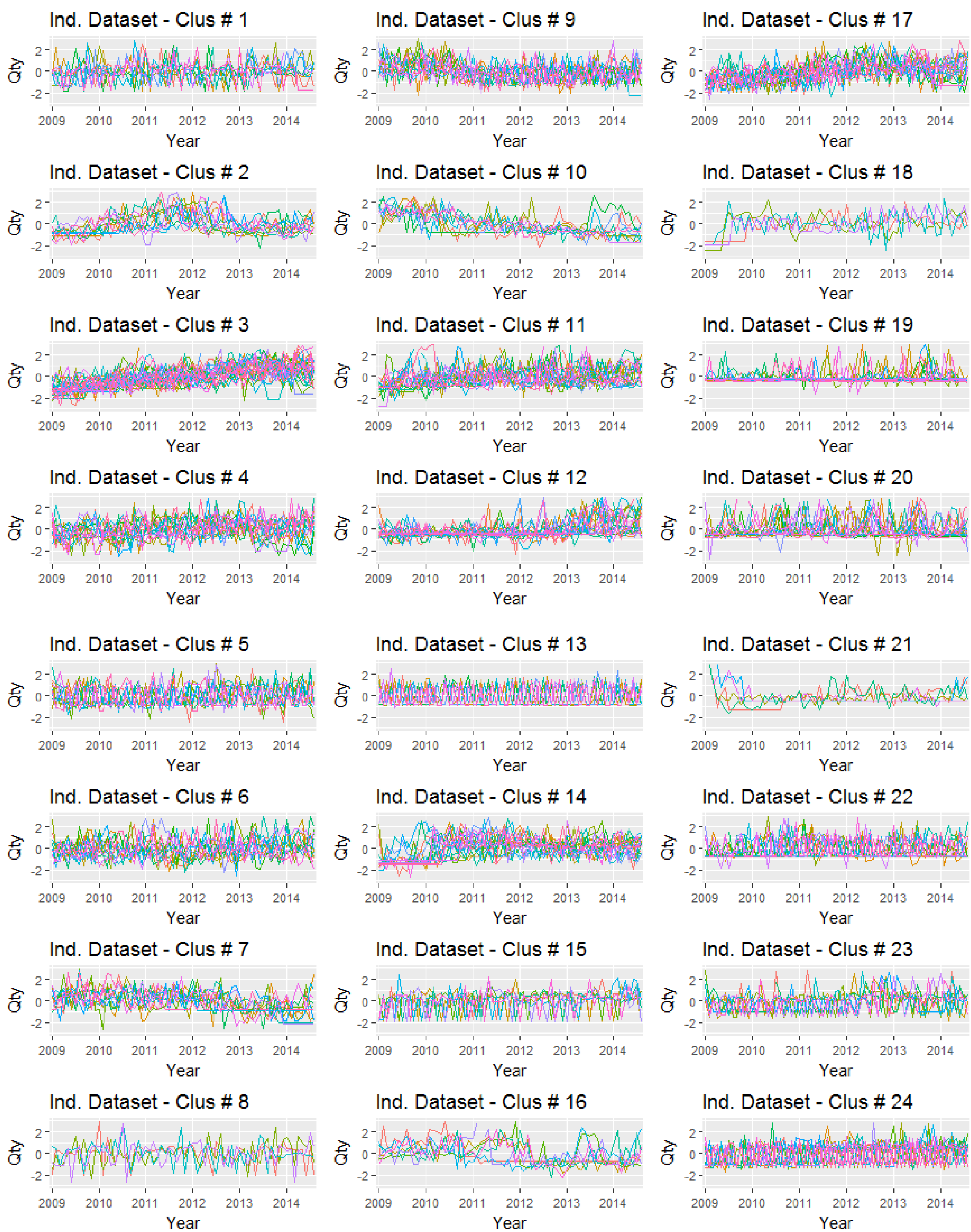


Figure 5-15: Proposed Segmentation Method, 24 Clusters

The traditional method of extracting statistical information to create variables is used to produce 24 clusters , as shown in Figure 5.16. The traditional method does not give good results. With a few exceptions, none of the clusters produced by the traditional segmentation method indicate any useful patterns. As with the attempt to produce eight clusters, the tradtional method assigns most customers to a few clusters and leave other clusters with a single member.

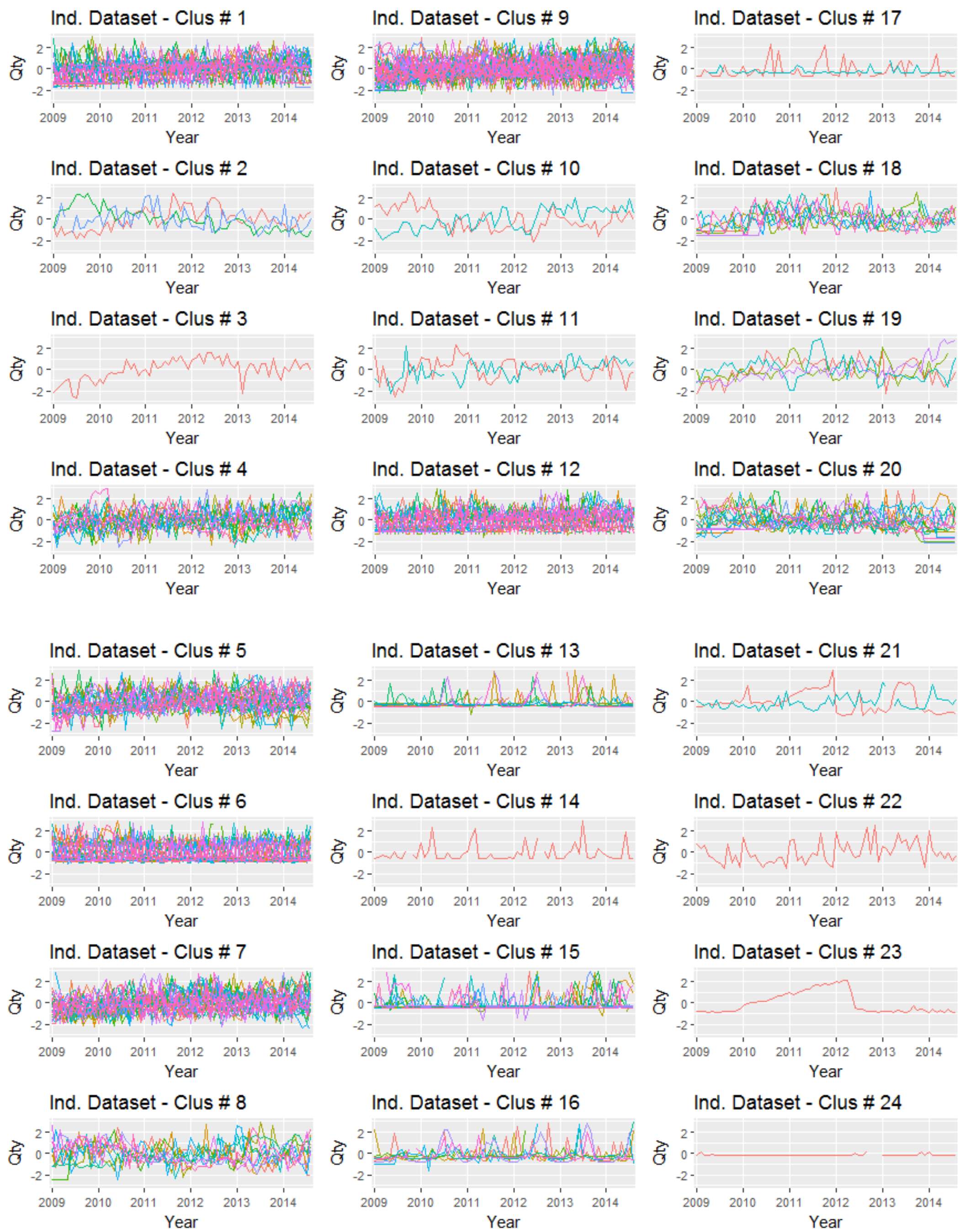


Figure 5-16: Traditional Segmentation Method, 24 Clusters

5.5.5 Discussion on Industrial Dataset

While several clusters in Figure 5.15 exhibit very distinct patterns, not all clusters are easy to interpret. Aggregating the original delivery data into monthly periods has created a set of very noisy time-series, which are difficult to visualize. Since some customers have less than monthly delivery frequencies, the monthly aggregation has resulted in intermittent time-series for those customers. We compensate for the noise in the data by applying a classical smoothing technique prior to segmentation. A commonly applied correction for intermittent time-series data is Croston's method (Croston, 1972). When we apply Croston's method to smooth the data, the resulting clusters reveal much more information than before. Figure 5.17 shows the 24 clusters with smoothed data.

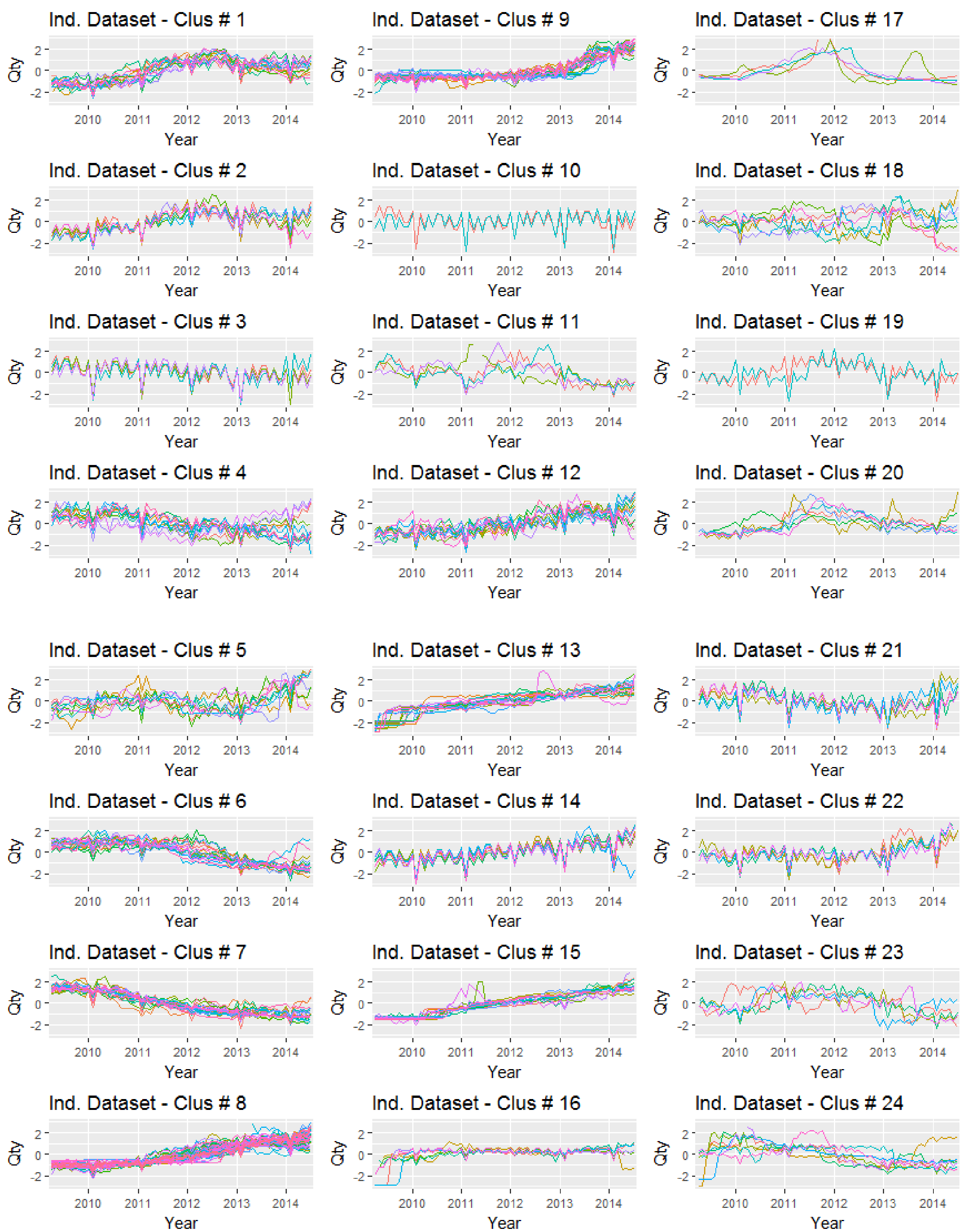


Figure 5-17: Data Smoothed with Croston’s Method

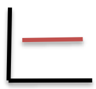
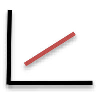
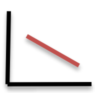
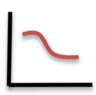
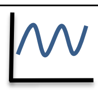
Most clusters in Figure 5.17 reveal now several clear behavior patterns. For example, seasonal changes are evident at the beginning of the year for most segments, but not in Clusters 13, 15, 17 and 24. Other changes in demand indicate either monthly or quarterly seasonality or possible a cyclic behavior. Changes in overall trends are also evident; some customers have stable consumption, some increasing, and others decreasing. Table 5.2 summarizes the segments based on trend patterns. From Table 5.2, we observe that a significant portion of the customer base is either stable or increasing. Conversely, we also observe that 23% of the customers are decreasing; the decreasing customers account for nearly a third of the historic delivery quantities; the decreasing trend indicates a significant population of customers at risk of decreasing or ceasing business. Table 5.2 also reveals more subtle details such as the point in time when a change occurred and the magnitude of the change.

Column k of Table 5.2 lists the homogeneity of cluster based on the distance between the clusters members calculated per formula 2

$$H_{homogeniety} = \sum \sqrt{(x_i - y_i)^2} \quad (2)$$

where H is the sum of distances between observations and x_i and y_i are numerical values of two observations. The values of H are all similar in magnitude, ranging from 40 to 52. The similarity of homogeneities indicates that all clusters have similar density and uniformity.

Table 5-2: Summary of Behavior Patterns

(a)	*(b)* Type of Behavior Pattern	*(c)* Cluster	*(d)* Noise in Data	*(e)* Amount of Change	*(f)* When change occurs	*(g)* % of Customers	*(h)* % of Customers	*(i)* % of Total Qty	*(j)* % of Total Qty	*(k)* Cluster Homogeneity (H)
	Stable	3	Low	-	-	1%	7%	0.1%	2%	27
		10	Medium	-	-	1%		0.1%		27
		16	Low	-	-	2%		0.1%		21
		18	High	-	-	2%		2.0%		31
		19	High	-	-	1%		0.1%		36
	Increasing	5	Medium	Large	Late	3%	55%	6.0%	53%	15
		8	Low	Medium	Continuous	21%		12.0%		12
		9	Low	Large	Late	10%		5.0%		14
		12	Medium	Small	Continuous	5%		24.0%		24
		13	Low	Small	Continuous	6%		3.0%		15
		14	Medium	Small	Continuous	3%		1.0%		24
		15	Low	Medium	Continuous	5%		2.0%		11
		22	Medium	Small	Late	2%		0.1%		28
	Decreasing	4	Medium	Small	Continuous	5%	23%	2.0%	15%	16
		6	Low	Medium	Late	7%		4.0%		12
		7	Low	Large	Continuous	8%		8.0%		15
		24	High	Small	Continuous	3%		1.0%		26
	Behavior Change	1	Low	Small	Middle	6%	13%	9.0%	22%	20
		2	Low	Small	Middle	2%		0.1%		26
		17	Medium	Large	Continuous	1%		11.0%		14
		20	Medium	Large	Middle	2%		2.0%		22
		21	Medium	Small	Late	2%		0.1%		33
	No Pattern	11	High	-	-	1%	3%	4.0%	5%	27
		23	High	-	-	2%		1.0%		28

The methodology that we proposed permits to extract patterns in the data that show distinctive behaviors of the customers. These patterns are not visible in the combined data in Figure 5.9. Once visible, past behaviors can be used to make predictions for future behavior to support the development of strategic forecasts. In some cases, additional information is evident and may promote other management activities, such as identifying at-risk customers before they become lost customers. The initial dataset has a high level of noise that reveals little information. From the patterns in Figure 5.17, we can detect new customers, lost customers, growing customers, and decreasing customers. Although some clusters, such as Cluster 7 and Cluster 9 in Figure 5.17 reveal distinct patterns, others are more difficult to assess. Column *(d)* of Table 5.2 reflects how much noise or variation exists within the cluster. The assessment of noise in each cluster will help to establish the level of confidence in subsequent analysis performed on each cluster. Clusters 11 and 23 contain customers with a high level of noise, these observations share only the common trait that they do not fit well into any other clusters.

5.6 Conclusion

The engineering challenge was to provide information to facilitate the strategic forecasting process. The available data was limited to historical delivery data that initially revealed little information due to high noise in the data. It was proposed to create sub-groups of customers with similar behavior patterns. Traditional market segmentation methods rely on availability of descriptive variables; they could not work given a lack of information about the customers. The situation was further complicated due to vendor managed inventory arrangement whereby customers' consumption behavior and the supplier's logistics decisions both influence delivery patterns.

The proposed methodology for market segmentation relied on detecting patterns of behavior in customers' historical data. By using dynamic time warping, we were able to measure similarities (distance) between customers. Sub-groups were subsequently established based on the distances. Behavior patterns, indistinguishable in the overall population of customers, were revealed in the sub-groups. We resolved some of the problems of high noise level in the data by smoothing it by applying Croston's method (Croston, 1972); the resulting market segments revealed a variety of information including seasonality, trends, and cycles. Information produced by the method proposed in this research has direct management applications for identifying segments of customers with similar behaviors—from the identified behaviors, management can make informed decisions on future actions within the segments.

Market segmentation is a commonly applied management tool, and while there is a growing abundance of data available regarding industrial, economic, and social trends, there is little guidance to managers on how to interpret and convert the data into actionable information. Our proposed method extracts useful information using only historic delivery data. This method could be successfully applied to any domain that has historic delivery data available. Analysis of intermittent time-series is an active area of research that attempts to address noise in data due to aggregating delivery or transaction data into time-series format. Our proposed method should be effective on any dataset that originates from transaction data.

While this study gives insight into a method to segment customers, it has limitations in that the number of segments has not been optimized. Further research is necessary to determine a suitable method to define the best number of segments and an analytical method to evaluate the resulting clusters. Future research should also investigate ways to incorporate external data information so that new customers could be classified into existing segments based on available characteristics. The proposed method is also limited to application in domains where ongoing customer/supplier relationships are stable. Being able to detect and predict a customer's behavior through historical data is predicated on availability of the historic data. If customers utilize multiple suppliers the necessary data will not be available for analysis.

CHAPTER 6 ARTICLE 3: FORECAST OF INDIVIDUAL CUSTOMER'S DEMAND FROM A LARGE AND NOISY DATASET

Published in: Journal of Computers and Industrial Engineering

Murray, P., Agard, B., & Barajas, M.A., (2018b)

Abstract: *Optimization of the supply chain relates on data that describes actual or future situation. Besides in many situations available data may not correspond directly to what is expected for the different models because of too large quantity and imprecision of the data that may lead to suboptimal or even bad decisions. Actual problem considers the availability of a large and noisy dataset concerning historical information about each customer that will be used to make improved prediction models, that may fit models to optimize the supply chain.*

When dealing with large datasets, market segmentation is frequently employed in business forecasting; many customers are grouped based on some measure of similarity. Segment-level forecasting is then employed to represent the population within each segment. Challenges with successfully applying market segmentation include how to create segments when descriptive customer information is lacking and how to apply the segment-level demand forecasts to individual customers. This research proposes a method to create customer segments based on noisy historical transaction data, create segment-level forecasts, and then apply the forecasts to individual customers. The proposed method utilizes existing data mining and forecasting tools but applies them in a unique combination that results in a higher level of customer-level forecast accuracy than other traditional methods. The proposed forecasting method has significant management applications in any domain where forecasts are needed for a large population of customers and the only available data is delivery data.

6.1 Introduction

Supply chains create value by transforming and transporting goods and services that satisfy the demand requirements of downstream partners; understanding demand requirements is necessary for strategic planning (Carbonneau et al., 2008). Forecasting methods all share a common prerequisite; they require suitable input information from which demand behaviors can be extracted, interpreted, and predicted. In ideal situations, strategic forecasts incorporate collaborative information shared by downstream partners, however, supply chain partners are not always able or willing to share information (Holweg et al., 2005). When collaborative information is not available, the supplier must rely on other information such as historical data to build its forecasts; however, the available data is not always the correct data to use (Makridakis, 1989) and may not truly represent actual consumption behaviors. Historic data collected at the point of consumption is preferred, but sometimes not available. Data obtained upstream in a supply chain is often distorted due the Bullwhip effect (Forrester, 1958); it can become noisy, lumpy, and intermittent, making it difficult to use for segmentation and forecasting.

In addition to obtaining suitable input data, the forecaster must also consider how to manage large populations of customers. In domains with many customers, it is advantageous to divide the market into analogous segments so that a manageable number of forecasts can be created & maintained (Calvet et al., 2016; Wind, 1978). Market segmentation has the additional benefit of improving forecast accuracy due to forecast errors off-setting each other (J. Scott Armstrong, 2006). A critical component of market segmentation is to identify customers' distinguishing attributes from which similarities and differences can be measured (Calvet et al., 2016; Huber, Gossmann, & Stuckenschmidt, 2017). When distinguishing attributes are absent, historical data in the form of transaction records is often the only information available—just as forecasts can be built from historical data, markets can be segmented from the same data.

A challenge with forecasting by market segments is applying the segment-level forecasts to individual customers and validating the results. While the literature is rich with methods for creating customer segments, there is little guidance on how to apply segment-related predictions to individual customers and evaluate the results.

In this research, we present a method that produces customer-specific forecasts using noisy delivery data as the only input. This method is based on several steps that permit segmenting the customer's consumption behaviors based on the delivery data, generating forecasts for each segment, scaling those forecasts to individual customers, and finally, evaluating the accuracy of the predictions. This new method is a significant contribution in that it produces segment-based forecasts using noisy data and then applies the forecasts to individual customers. Rather than create new tools, we apply proven effective tools in a way that they have not been used before to solve a persistent problem. The proposed method can be effectively applied in many domains and provide significant value in many applications.

The remainder of the paper is organized as follows: Section 6.2 presents a review of the literature followed by the proposed method in Section 6.3. Section 6.4 is a case study in which the method is applied to a real industrial dataset and compared with other methods. The conclusions are presented in Section 6.5.

6.2 Literature Review

In the context of this research, we propose a new method that combines several sequential steps including treatment of noisy data, market segmentation, forecasting, and forecast evaluation. The state of the art for these steps are summarized in the following sections of the literature review.

6.2.1 Treatment of Noisy Data

When transactional data (such as delivery records) is substituted for absent consumption demand information, it must first be transformed into a suitable format for the analysis. Transaction data normally contains a time-stamp that facilitates aggregation into temporal bins, however, selecting an appropriate bin duration is not trivial (Petropoulos & Kourentzes, 2015). High aggregation (longer periods) removes noise and reduces or eliminates zero-quantity periods, however, it also carries the risk of excessive smoothing and loss of information (Spithourakis, Petropoulos, Nikolopoulos, & Assimakopoulos, 2012). Small aggregation reduces smoothing and loss of information, but it tends to create periods of zero-quantity resulting in intermittent time-series. After transformation, the dataset is still difficult to interpret due to noise incorporated from the Bullwhip effect (Forrester, 1958) and logistics decisions.

Croston (1972) proposed a method to update a forecast for intermittent demand where updates only occur when demand occurs; during periods of zero-demand, the forecast remains unchanged. Croston's method is calculated as per formulae (1), (2), and (3):

$$\hat{y}_t = \hat{Z}_t / \hat{X}_t \qquad (1)$$

where $\hat{Z}_t$ is the SES forecast for the non-zero periods and $\hat{X}_t$is the forecast for the number of inter-demand intervals. Both the demand size and the intervals use SES per formulae (2) & (3):

$$\hat{Z}_t = \alpha_z y_t + (1 - \alpha_z)\hat{y}_t \qquad (2)$$

$$\hat{X}_t = \alpha_x x_t + (1 - \alpha_x)\hat{x}_t \qquad (3)$$

where y_t is the non-zero demand at time t and x_t is the number of non-zero intervals. The smoothing coefficient α is set to 0.02.

An improved method for resolving noise in time-series data is the aggregate-disaggregate intermittent demand approach (ADIDA), proposed by Nikolopoulos et al. (2011). ADIDA reduces noise by aggregating into high-level temporal bins and then reverting to the original bin sizes. This approach is appealing due to its effectiveness and simplicity.

6.2.2 Market Segmentation

Market segmentation is comprised of two distinct steps: measuring the difference between customers (known as a distance measure), and then creating analogous segments (a process known as clustering). Many methods exist for dividing a market into segments (Chakraborty, 2013; Han & Kamber, 2006; Le, Agard, & Deveault, 2009), but in general, the methods are based either on descriptive attributes or behavioral attributes (Murray et al., 2017). Descriptive attributes (such as size, location, sex, age or transaction frequency) are commonly used since this type of variable is easy to quantify. Success in segmentation based on descriptive attributes relies heavily on two assumptions: that data to support the variables is available, and that the descriptive attributes are truly relevant for creating the segments. There are many external factors, such as economy, competitors' actions, and social perception, that influence customer behavior (Sevlian & Rajagopal, 2018), however, selecting appropriate factors can be challenging. Segmenting based on behavioral attributes requires historical data that reflects the behavior and a method for extracting and identifying the behavior within the data (Barragan, Fontes, & Embiruçu, 2016; Kashwan & Velu, 2013). In the context of this research, descriptive attributes are not available; the historical delivery transactions, converted into time-series format, are the input.

6.2.2.1 Time-series Distance Measures

The advent of data mining and the ability to process vast amounts of data has made it practical to detect differences between multiple time-series. Early efforts to index time-series relationships did not produce meaningful results due to the focus on exact matches rather than similarities (Agrawal et al., 1993). In order to measure similarity rather than sameness, a method for elastically shifting the time indexes was needed (Keogh & Ratanamahatana, 2005). The ability to identify similar customers rather than exact matches enables the grouping of customers by similar behavior. Planning can then be based on groups rather than on individuals.

Dynamic time warping (DTW) was developed in the 1970s for speech recognition (Sakoe & Chiba, 1978) and is a popular method for indexing time-series data through elastic manipulation of the time axis (Giorgino, 2009). DTW is considered an expensive algorithm with regards to computing time (Z. Zhang et al., 2006), however, faster computer hardware and continual refinement of DTW algorithms have diminished the concern of computational expense. Exhaustive literature reviews have concluded that DTW is the best method for comparing time-series similarities (Ding et al., 2008; Rakthanmanon et al., 2013).

6.2.2.2 Clustering Time-Series

Once customer similarity has been quantified with a suitable distance measure, groups can be established. Clustering is the process of grouping a population such that the inter-group homogeneity is maximized and the intra-group heterogeneity is also maximized (Esling & Agon, 2012). Ideally, group members are similar to each other and dissimilar from members of other groups. Unfortunately, there is a vast collection of clustering algorithms in the literature and no definitive means of selecting a best method (A. K. Jain, Murty, & Flynn, 1999). According to Liao (2005), clustering time-series data is most commonly accomplished through partitional clustering, artificial neural networks (ANNs), or hierarchical clustering. Partitional clustering, such as the popular k-means, are generally easy to use and have low computational expense (Kantardzic, 2011). Partitional cluster algorithms commonly create groups by minimizing the total squared error for a giving number of clusters *(K)*. Time-series data does not necessarily provide an adequate description of the differences within a population and subsequently, direct comparison through squared error measurement does not always result in useful information (Murray et al., 2015).

Partitional clustering is also hindered by the non-trivial task of pre-defining the number of clusters *(K)*.

Hierarchical clustering, first introduced by Ward (1963), creates a hierarchical structure of grouping based on similarity measures. The hierarchical structure can be graphically displayed as a dendrogram, which facilitates selecting the number of clusters *(K)* after the calculations have completed.

6.2.3 Quantitative Forecasting

6.2.3.1 Traditional Forecasting

Traditional forecasting deals directly with the population, either as a group or individually. A quantitative approach to strategic forecasting could follow several strategies found in the literature. For this research, we explore two well established methods—top-down forecasting and attribute-based segment forecasting—and compare them to the proposed behavior-based segment forecasting.

The first method, top-down forecasting, applies a prediction model to the aggregated total of a product, the top level, and then applying the prediction down to individual customers, segments, or products (Dangerfield & Morris, 1992). The top-down approach is often able to generate accurate results—at the top level—due to the cancelation of errors when the data is aggregated. Error cancelation is generally most effective when the time-series have similar patterns (Lapide, 2006). Despite its tendency to produce accurate results, top-down forecasting has a fundamental flaw in that it is unable to detect underlying demand signals, and does not lead the analysis to incorporate triggers to respond to the underlying information (Chase Jr., 2013; van Donselaar, Peters, de Jong, & Broekmeulen, 2016).

The second method, attribute-based segment forecasting, divides the overall population of customers into analogous segments based on attributes associated with the customers. Prediction models are applied to individual segments and the resulting forecasts are then applied to individual customers within the segment. Market segmentation has a long history of informal application; its benefits were first formally presented in the literature in 1956 (Smith, 1956). Attribute-based segmentation is heavily dependent on the availability of suitably descriptive variables. In some cases, descriptive variables such as age, sex, and income are used (Anderson et al., 1976). In other cases, statistical features, such as mean, sum, and kurtosis are extracted from historical data (Bala, 2012). The effectiveness of attribute-based segment forecasting depends on whether the variables used effectively describe underlying behaviors (Murray et al., 2015).

Like attribute-based segment forecasting, behavior-based segment forecasting divides the overall customer population into segments, predicts the segments, and then applies the forecasts to individual customers within the segments. Behavior-based segment forecasting differs in that it identifies behavior patterns in historical data and uses those patterns to establish customer segments. Behavior-based segmentation can generate superior results over attribute-based segmentation in that customers with different descriptive attributes sometimes exhibit similar behavior (Vlckova, Lostakova, & Patak, 2014).

6.2.3.2 Time-Series Forecasting

Time-series forecasting methods (including naïve, moving average, exponential smoothing, decomposition, and ARIMA) are "built on the premise that future sales will mimic the pattern (s) of past sales" (Chase Jr., 2013, p. 84). If demand patterns can be detected and accurately modeled, these techniques can be used to generate reasonable forecast accuracy. Time-series methods are suitable for harvest brands that have sufficient historical data and steady demand.

Economic time-series are typically non-stationary, so simple models such as naïve and moving average sometimes do not do a good job of representing demand (Phillips & Durlauf, 1986). Adjusting or accounting for non-stationarity can be accomplished by decomposing the information into separate components of trend, seasonality, cycle, autocorrelation, and random error (F. Robert Jacobs et al., 2011). Once the time series is represented by its components, it can be de-seasonalized, de-cycled, and de-trended. The resulting patter can then be modeled by simple methods. Each method utilizes a different level of complexity and decomposition, from the very simple naïve to the more complex ARIMA. However, more complex models do not necessarily produce more accurate results. In fact, the literature generally promotes the rule of using the simplest method that produces actionable results (F. Robert Jacobs et al., 2011; Makridakis, 1989; Maté, 2011). Despite significant research in time-series forecasting, the ARIMA method remains popular due to its flexibility and good accuracy (Kourentzes & Petropoulos, 2016; G. P. Zhang, 2003).

6.2.4 Forecast Evaluation

Once predictions are generated, a method to validate them is needed. Validation is performed by measuring the variation between predictions and actual data. Simple measures, such as mean absolute percent error *(MAPE)* and root mean squared error *(RMSE)* are suitable when the context is comparing between forecasts rather than comparing a forecast to actual (Kim, Dekker, & Heij, 2017), although these measures are not necessarily optimum when applied to intermittent data (Kolassa, 2016). *MAPE* is suitable when comparing forecasts for different data sets (as is the case when we forecast clusters separately from each other). However, *MAPE* gives biased results depending on how the calculation is set up, so although it is useful for comparing between multiple forecasts, it is not as good for evaluating overall performance results. *MAPE* also performs poorly when actual demand approaches zero, resulting in infinite values calculated (Rob J. Hyndman & Koehler, 2006). Lastly, it does not scale the contribution of different magnitude data (Chase Jr., 2013). For these reasons, *MAPE* is a useful indicator of performance, but it cannot be used exclusively. Formula (4) shows the calculation for *MAPE* where *A* is actual value, *F* is forecast value, and *n* is the number of periods. Formula (4) is slightly altered from the traditional format. Using *A+1* in the denominator prevents divide-by-zero when *A=0*. This alteration to the formula is necessary when evaluating intermittent time-series and does not measurably affect the calculation when actual values are sufficiently large.

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|A - F|}{A + 1}}{n} * 100 \tag{4}$$

Another common statistical measure is *RMSE*. Formula (5) shows the calculation for *RMSE* where *A* is actual value, *F* is forecast value, and *n* is the number of periods. The error measure *RMSE* has nearly opposite attributes to *MAPE*. It handles values approaching zero and provides results representative of magnitude. However, *RMSE* is not as useful for comparing performance between clusters because error size is skewed by data magnitude.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A - F)^2}{n}} \tag{5}$$

The basic premise in most forecast evaluation strategies is that the forecast errors should have zero mean (Diebold & Lopez, 1996). Absolute error is usually preferred over total error to avoid errors cancelling out. Percentage-error measure removes overall magnitude from the evaluation, and scaled errors are beneficial when zero values might hinder calculations (Kourentzes & Petropoulos, 2016).

Despite several improvements proposed in the literature, traditional forecast error calculation methods all follow a common assumption—that the predicted value should directly match the actual value in each time-period (Chase Jr., 2013). This assumption is valid in domains where the transaction occurrence (the delivery of the product) and consumption demand coincide in time. However, the assumption is less valid in domains where the timing of peaks or transactions is not critical (Haben, Ward, Vukadinovic Greetham, Singleton, & Grindrod, 2014), such as when the delivery is made to point-of-use inventory storage. A point-wise evaluation does not give relevant information if it measures the forecast against the deliveries and not against the customers' consumption needs. Point-wise evaluation has a double-penalty effect when the transaction prediction is slightly missed (Haben et al., 2014), when in reality the requirements to maintain the point-of-use inventory are satisfied.

The forecast evaluation should therefore focus on the similarity between the predicted time-series and the actual time-series; precise matching between demand and actual, especially with regards to timing, is not important. It is well established that DTW is a superior method for evaluating similarities between time-series (Keogh & Ratanamahatana, 2005; Rakthanmanon et al., 2013; X. Wang et al., 2013). DTW, therefore, is a better method to compare a forecasted time-series against an actual time-series when the precise timing of transactions is not critical (Haben et al., 2014).

6.3 Proposed Method for Predicting Customers' Consumption from Delivery Data

In preparation for forecasting, the time-series data is first separated into training data and test data, as illustrated in Figure 6.1, step *a*. The training data is used to build the model which will be compared to the test data to evaluate the model's performance. The training data is fed to a DTW algorithm to calculate the pair-wise distances between customers, as illustrated in Figure 6.1, step *b.* DTW is used due to its superior performance for comparing time-series data (Keogh & Ratanamahatana, 2005; Murray et al., 2015).

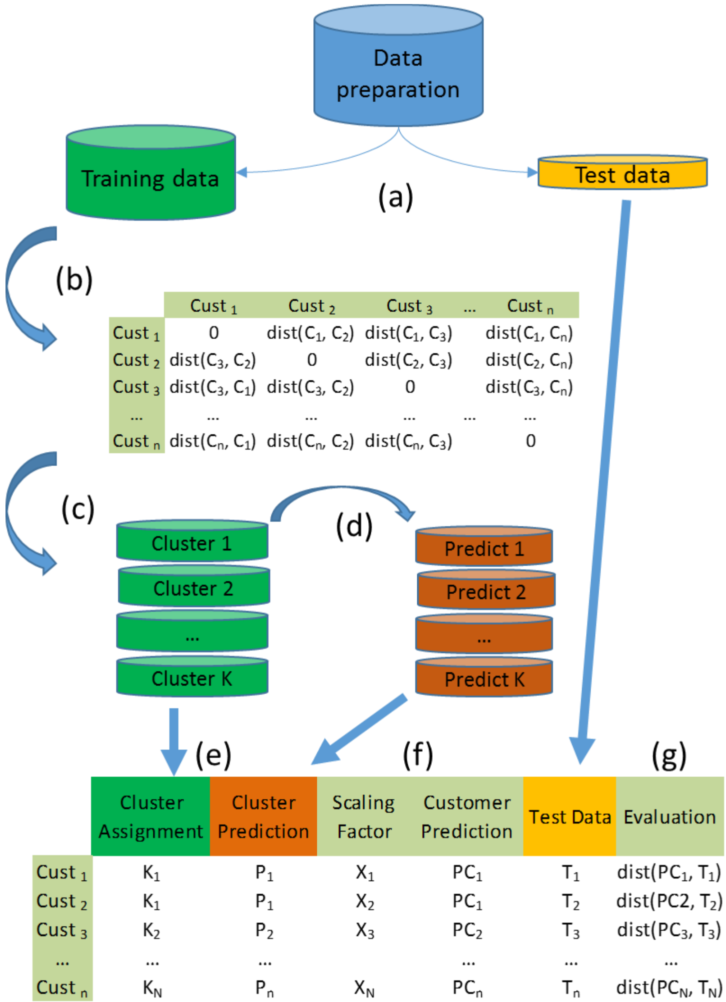


Figure 6-1: Proposed Method

Once pair-wise distances are known, the individual customers can be grouped into clusters with similar behavior patterns. We use hierarchical agglomerative clustering (HAC) with Ward's linkage to identify the cluster membership for each customer, as illustrated in Figure 6.1 step (c).

After segmentation, it is necessary to develop segment-level predictions. The ARIMA forecasting method is used to predict each segment because of ARIMA's good level of performance and wide use in forecasting (Pai & Lin, 2005), per Figure 6.1 step (d). This research utilizes the auto.arima function in the forecast package for R (Rob J. Hyndman, 2017). The segment-level predictions are then assigned to the individual customers based on their cluster membership, per Figure 6.1, step (e). Predictions are made for each segment based on the mean values of the members of the segment.

During the data preprocessing step, each customer's data was normalized to prevent large customers' behavior from dominating the analysis. However, prediction accuracy must be measured against the original un-scaled data—it would serve little purpose if the prediction performed very well with small customers but performed poorly with large customers. In Figure 6.1, step (f), the predictions are scaled per the customer's individual normalization scaling factor. This puts the predictions into the same scale as the test data. Finally, DTW is used to evaluate the forecast based on the distance between the prediction and the test data Figure 6.1, step *(g)*.

6.4 Results – Industrial Dataset

A real dataset, provided by an industrial partner, is used to test the proposed method. To validate, we solve a real problem using the proposed method (Section 6.4.4), classical methods (Section 6.4.5), and we compare the performances of everything (Section 6.4.6). The evaluation method is illustrated in Figure 6.2. Calculations on the real data are conducted using R programming (R Core Team, 2015).

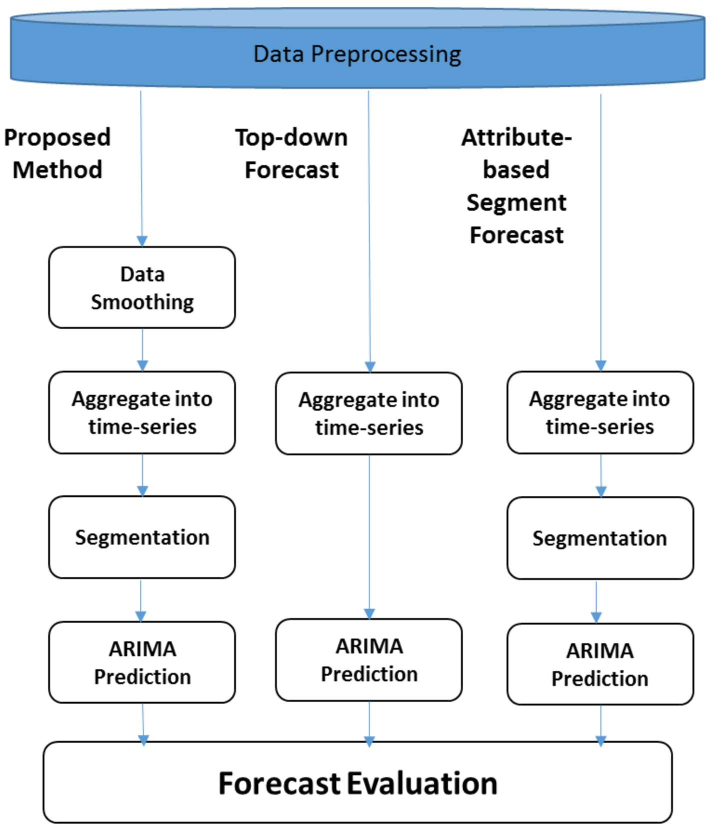


Figure 6-2: Comparison of Methods

6.4.1 Context of Industrial Data

The industrial partner is a supplier of bulk liquid materials that are used in manufacturing, food packaging, and medical services. The supplier maintains its customers' point-of-use inventories of the bulk liquid in a vendor managed inventory (VMI) arrangement; the supplier is responsible for ensuring uninterrupted product availability at its customers' sites. Customers do not normally purchase from alternate suppliers due to multi-year purchase contracts. The dataset consists of delivery records for approximately five years. Each record contains the delivery date, delivery quantity, customer location, customer identifier, and distribution source location. The data was not cleaned, summarized, aggregated, or otherwise processed prior to the research analysis.

The original delivery data is easily aggregated into time-series format. However, the result, as illustrated in Figure 6.3, does not reveal any useful patterns. Customers can have significantly different magnitudes of consumption leading to a band of customers at the lower level of the Y-axis.

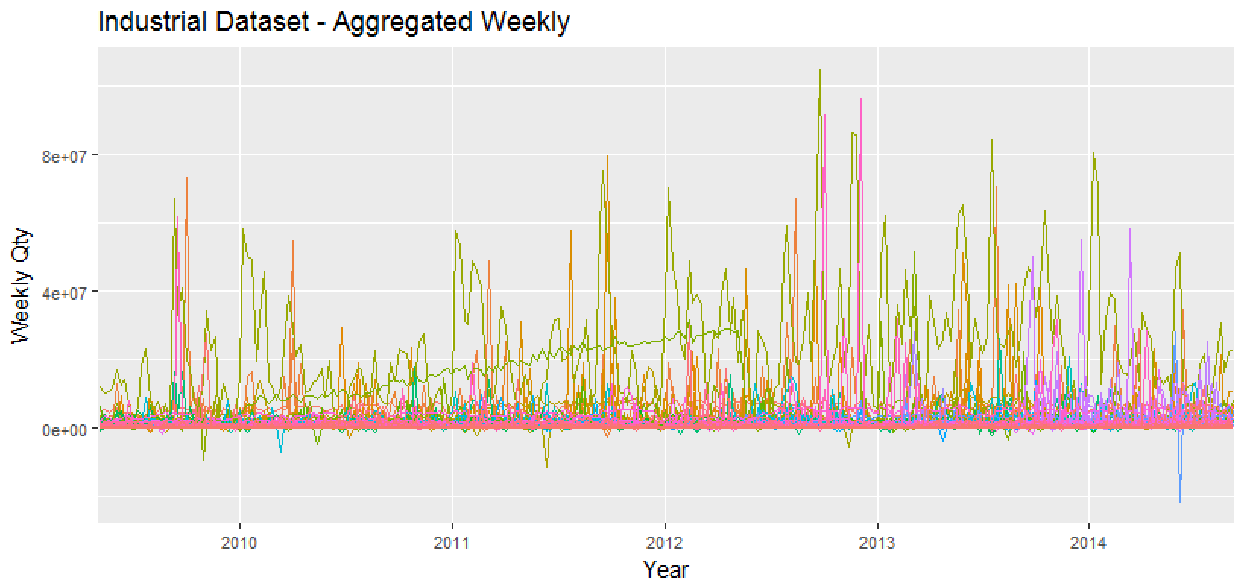


Figure 6-3: Original Customer Data

From Figure 6.3, we conclude that simply aggregating the data into time-series format and then attempting to develop forecasts, regardless of how sophisticated or complex the forecast model, would prove fruitless.

6.4.2 Data Preprocessing

Outliers exist in the data. In Figure 6.3, several negative values are evident. Due to the nature of the product, returns are not possible and therefore negative values should not exist. In addition to outliers, some customers are not suitable for inclusion in the analysis due to lack of sufficient history. To identify outliers and customers with insufficient data, the dataset was temporarily aggregating into monthly bins. Customers with negative monthly quantities are considered outliers and removed. New customers, with no data during the first two years, and lost customers, with no data in the last year are considered unsuitable and are also removed. Applying a minimalist approach to outlier detection and removal avoids removing customer whose data, while unexpected or unusual, nevertheless consists of actual valid observations. Figure 6.4 illustrates the distribution of customers removed due to outlier detection. The customers removed tend to have a high number of periods with no deliveries and low overall delivery quantity.

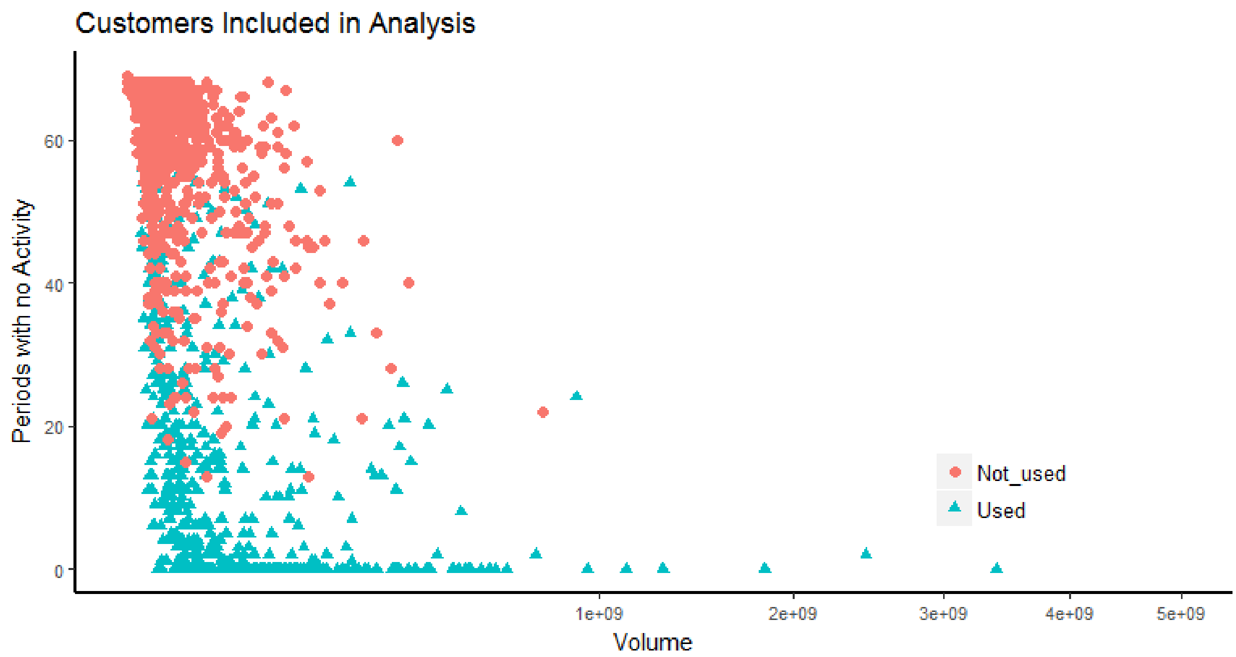


Figure 6-4: Outlier Detection & Removal

The final data preparation step is to normalize the data so that meaningful comparisons can be made. Normalization prevents large customers' data from overshadowing smaller customers. There are several ways to normalize a dataset, the most common approach in the literature is Z-normalization (Rakthanmanon et al., 2013) as per formula (7):

$$x_{norm}=\frac{(x-\bar{x})}{\sqrt{\frac{\sum(x-\bar{x})^2}{(n-1)}}} \qquad (7)$$

where x is the value of the transaction quantity, $\bar{x}$ is the mean of all transaction quantities, n is the number of observations, and x_{norm} is the normalized value. Z-normalization is appropriate for this type of data due to its resistance to the distorting effect of outliers (Murray et al., 2015). After preprocessing, the effects of outliers, logistics decisions, and relative size of the customer are removed. The resulting data can reveal behaviors and is suitable for subsequent analysis.

After outlier removal, the data is normalized to ensure that small customers are not over-shadowed by larger ones. Data normalization is a necessary step for all types of forecasting methods when applied to multiple time-series (Kalpakis, Gada, & Puttagunta, 2001). The first four years of data, spanning 2009 through mid-2013, are used as training data for each method. The final year is reserved as test data to evaluate the results.

6.4.3 Determining the Number of Segments *(K)*

One of the challenges for any market segmentation analysis is determining the number of segments *(K)*. The industrial partner is interested in maintaining a small number of forecasts, so the number of segments should be no more than *K=12*. Conversely, there needs to be enough segments that useful distinctions between segments can be made. A common technique for selecting number of clusters is based on measures of inter-cluster similarity (homogeneity) and intra-cluster dissimilarity (heterogeneity); the point where improvements diminish is used to determine number of clusters (Chakraborty, 2013). In Figure 6.5, K was incremented from K=2 to K=25 and heterogeneity and homogeneity measured at each increment. It is evident in the graph that selecting K above 10 provides only small improvements. Selection of *K=10* suits the requirements of the industrial partner and is supported by the evaluation in Figure 6.5. To have unbiased comparisons, the same *K=10* is used for all methods in the following evaluations.

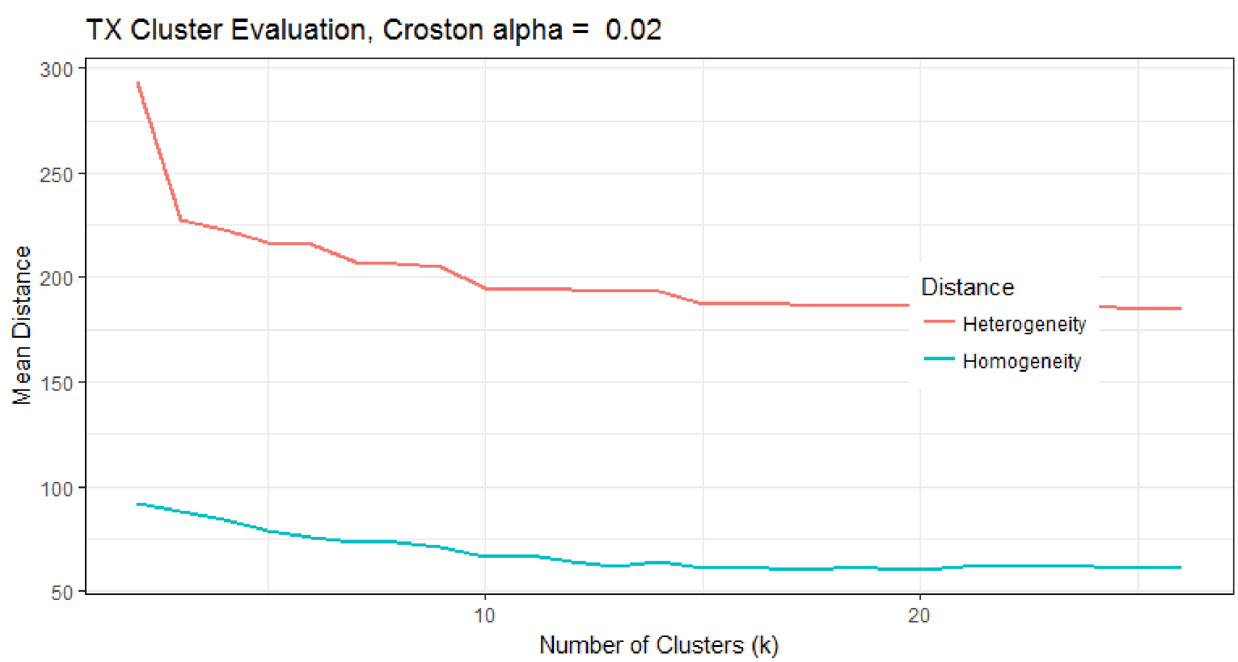


Figure 6-5: Cluster Evaluation

6.4.4 Application of the Proposed Method

The proposed method is applied to the real data per the steps outlined in Section 6.3. DTW is employed to calculate the similarity distances between customers per Figure 6.1(b). The resulting distance matrix is then fed to a hierarchical agglomerative clustering (HAC) algorithm per Figure 6.1(c). The output of the HAC algorithm is displayed in a dendrogram, which is then cut into segments of customers with similar behaviors. For the industrial data, ten segments are used. The results of clustering the smoothed dataset into ten segments are illustrated in Figure 6.6. The clusters from the proposed method reveal distinct patterns—Clusters 3,4, and 6 all exhibit decreasing trends while Clusters 2,5,7, and 8 all exhibit increasing trends. Clusters 1,9, and 10 also exhibit cyclic patterns.

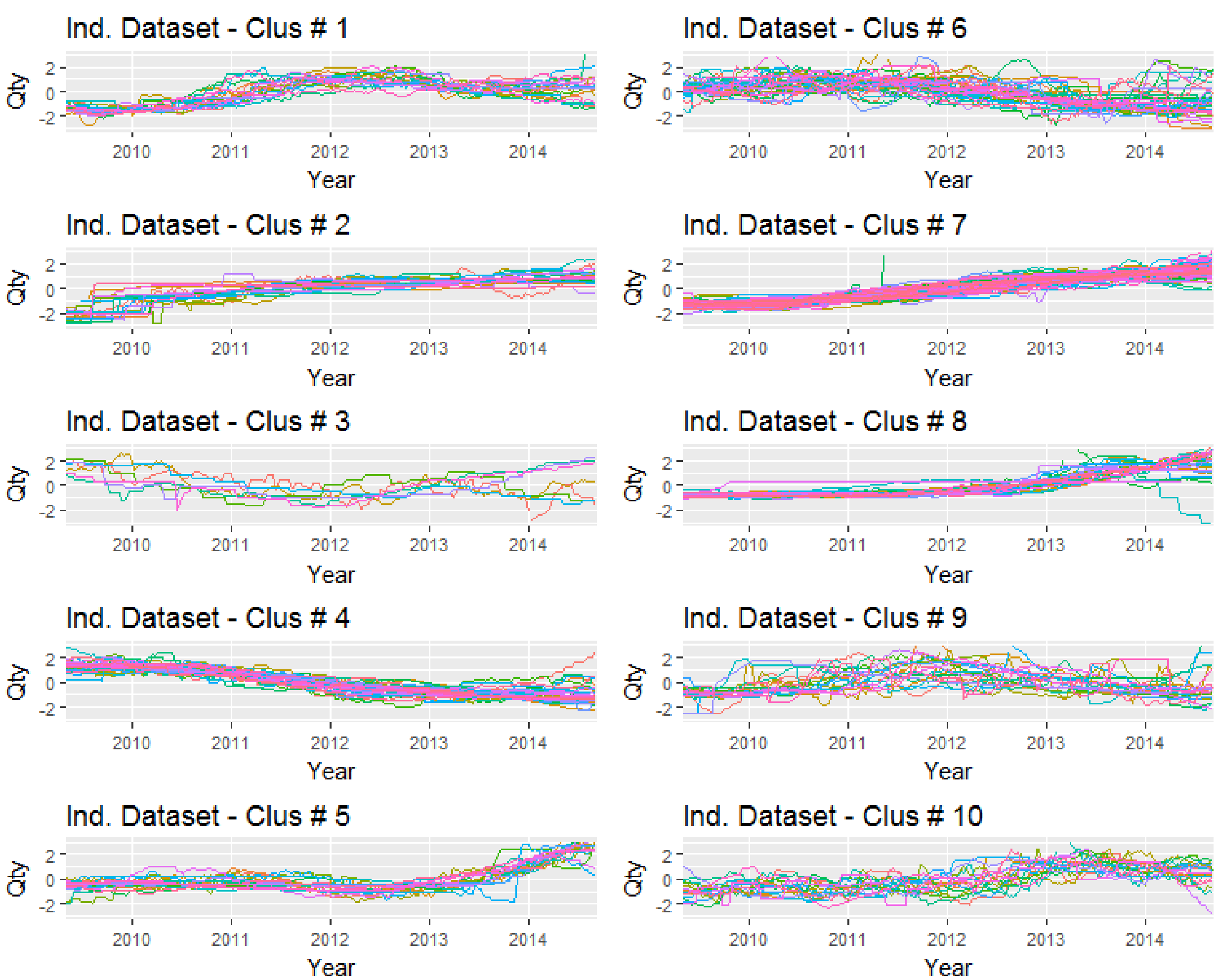


Figure 6-6: Proposed Method Segments

Once the customers are clustered based on behavior similarity it is possible to aggregate the clusters; research has shown that forecasts for aggregated groups of customers can lead to more accurate forecasts (Sevlian & Rajagopal, 2018). The mean value of all customers within each segment is used to create a prediction for the segment per Figure 6.1(d). Figure 6.7 shows the prediction for each segment along with 80% and 95% confidence bands shaded. The prediction for each segment is assigned to each customer within that segment. The customers' predictions are then re-scaled using the same factors that were used earlier to normalize the data. Re-scaling the prediction is necessary so that the resulting forecast can be used in real-world applications. Forecast errors are presented in Table 6.1 in Section 6.4.6.

6.4.4.1 Cluster Predictions

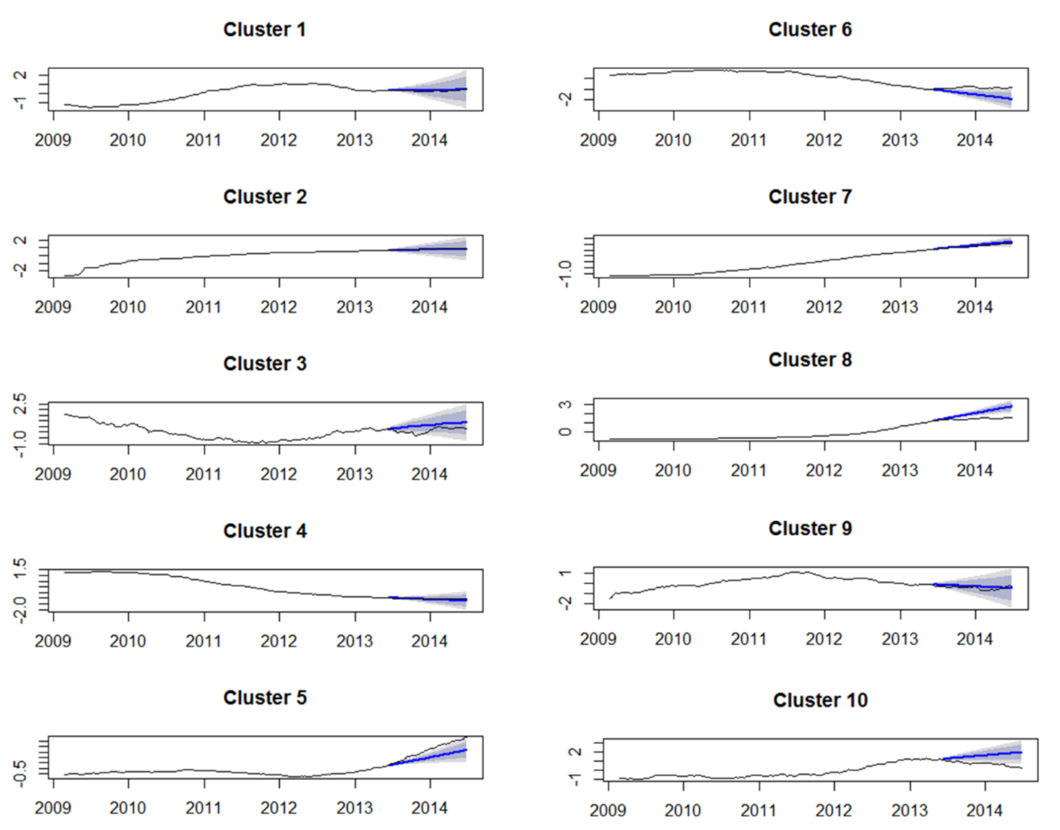


Figure 6-7: Segment Forecasts

6.4.5 Comparison to "Classical" Methods

Two classical methods of quantitative forecasting, as described in the literature review, are applied to the real data to compare to the proposed method. The classical methods offer a good benchmark due to their common application in industry. The outlier removal steps described in Section 6.4.2 are applied for all methods tested.

6.4.5.1 Top-Down Forecast

The top-down forecast uses an aggregate of all customer data to prepare a forecast. For this method, segments are not appropriate since the customer population is viewed as a whole. Figure 6.8 provides an overview of the population of customers.

The data of large customers no longer dominates the graph as it does in Figure 6.3; however, it is still not possible to see any useful patterns in the data. Despite the over-plotting effect in the center portion of Figure 6.8, viewing the upper and lower portions leads to an obvious conclusion that the data is extremely noisy and would be difficult to analyze.

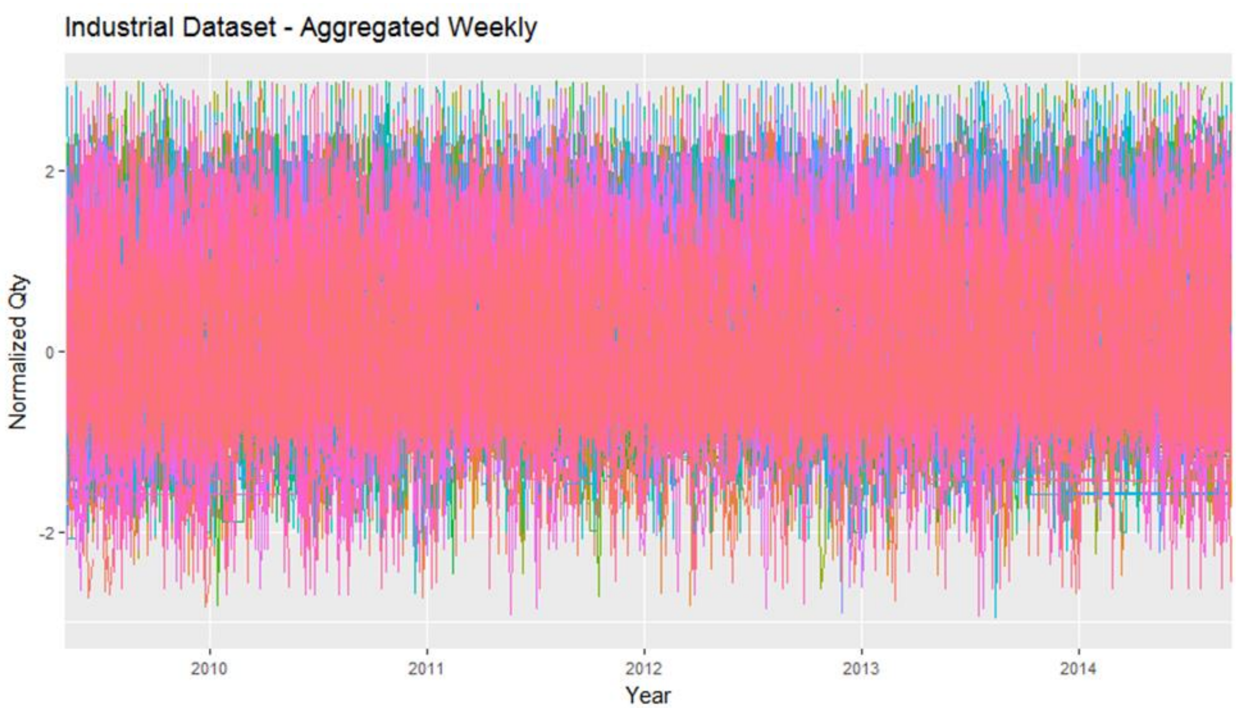


Figure 6-8: Data for Top-Down Forecast

The forecast for the top-down method, based on the mean value of the training data, is illustrated in Figure 6.9. The dark grey area on the right side of Figure 6.9 represents an 80% confidence level and the light grey area represents 95% confidence. For the top-down forecast, the auto.arima function (Rob J. Hyndman, 2017) has produced an over-fitted forecast that attempts to interpret the noisy data in the training set. Forecast errors are presented in Table 6.1 in Section 6.4.6.

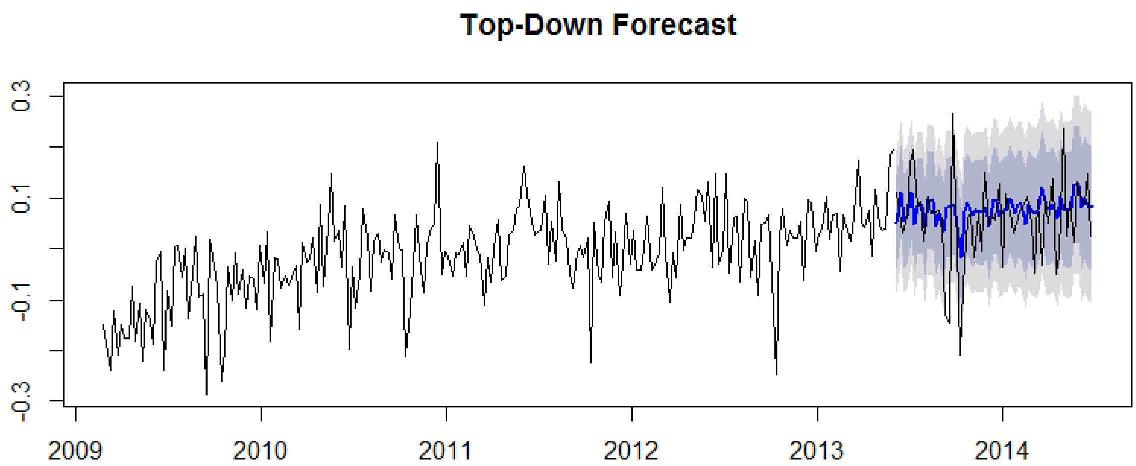


Figure 6-9: Top-Down Forecast

6.4.5.2 Attribute-based Segment Forecasting

The second classical method tested is an attribute-based segment forecast; based on the premise that customer behavior is influenced by or can be described based on a set of attributes (Bag, Tiwari, & Chan, 2017). Attribute-based segment forecasting uses descriptive attributes to divide the market into segments. In the context of the case study, the attributes are statistical descriptions of the historic data. The attributes include sum, standard deviation, median, kurtosis, and skewness. Bias in the comparison of methods is minimized by adopting components of the proposed method where possible. The input data is the same, the number of segments is set to K=10, and auto.arima is used to generate the forecasts.

The clusters generated by the classic method are illustrated in Figure 6.10. Most customers are assigned to Clusters 1, 4 & 5, while Cluster 7 & 8 have only single members. There are no distinguishable patterns in any of the clusters created by this method.

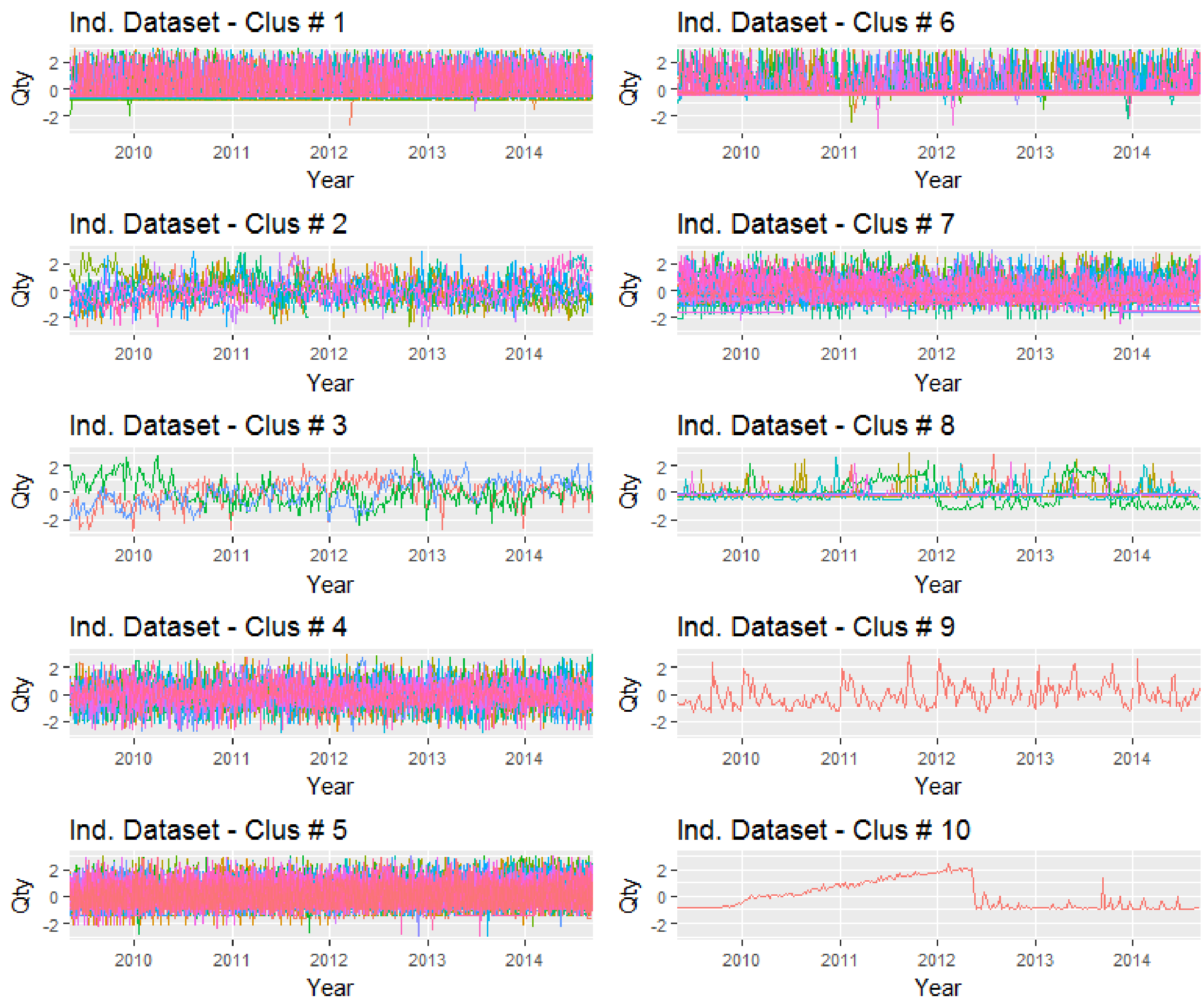


Figure 6-10: Attribute-Based Segments

The forecasts for the attribute-based segment method, illustrated in Figure 6.11, are based on the mean value of the training data for each segment. As with the top-down method, the training data spans from 2009 through mid-2013 and the test data consists of the remaining year. Forecasts for most clusters are over-fitted due to the prominent level of noise in the training data. Forecast errors are presented in Table 6.1 in Section 6.4.6.

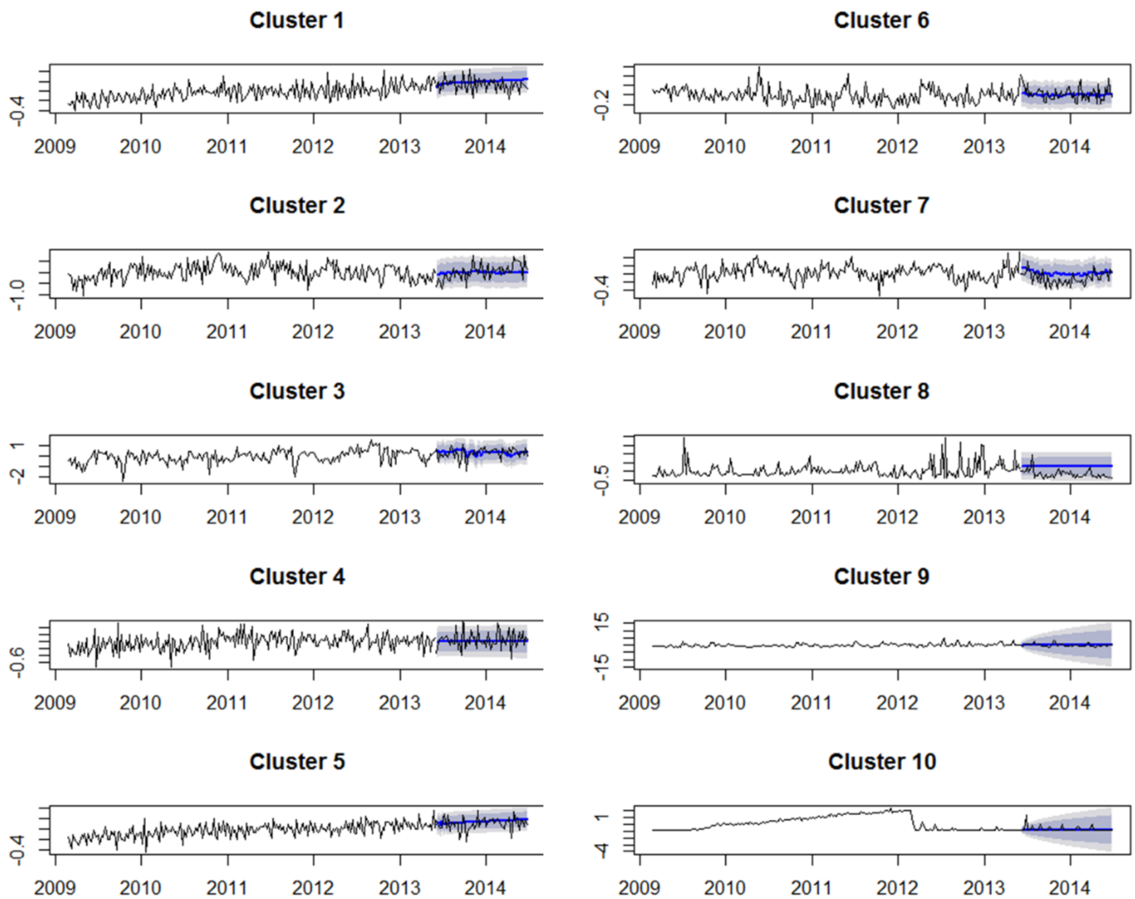


Figure 6-11: Attribute-based Segment Forecasts

6.4.6 Analysis of the Results

The results of the proposed method and the classical methods are evaluated for accuracy using both traditional and non-traditional forecast accuracy measures. The test data for the classic methods is the final year of the original data that was aggregated into weekly bins. The test data for the proposed method is the final year of the smoothed data, also aggregated into weekly bins. Forecast evaluation uses three measurement strategies. The first two, *MAPE* and *RMSE*, are traditional measures very commonly used both in industry and in academic research. These measures share a common trait with most forecast accuracy measures in that they are Euclidean in nature: every point in the forecasted time-series is assessed against its coinciding point in the test data time-series. In a domain such as VMI, where point-of-use inventory exists, exact matching of time and quantity is not relevant. Considering this, DTW is also employed as an accuracy measure due to its suitability for measuring similarities between time-series. The forecast evaluation results are shown in Table 6.1.

Table 6-1: Forecast Evaluations

	Top-Down Forecast	Segment Forecast	Proposed Method
Description	Predict from Original dataset (no clusters)	Predict from Original dataset (with clusters)	Predict from smoothed dataset
RMSE	1.38 e8	1.33 e8	0.33 e8
MAPE	3.66 e9	3.27 e8	2.89 e3
DTW	6.55 e9	6.18 e9	2.11 e9

Table 6.1 presents the total amount of error measured; lower values are preferred. The proposed method outperforms the two traditional methods when evaluating with MAPE and DTW. According to RMSE, the proposed method seems to have the worst performance of all methods. Besides, RMSE measures point-to-point error, which it is not a suitable measure for time-series comparisons (Haben et al., 2014). RMSE is included solely due to its wide acceptance in forecasting practices in order to make it visible for anyone that would like to reproduce similar results. MAPE is preferred due to its suitability to compare between time-series of different magnitudes. Lastly, we include DTW as a non-traditional error measure which is better suitable for measuring the distance between time-series (Ratanamahatana & Keogh, 2005). The results indicate that the proposed method, when evaluated with a suitable error measure, produces more accurate forecasts than traditional methods.

6.5 Conclusion

This research presented a original method that permit to make improved individual prediction, for each customer, from a large dataset of noisy dataset, using a pertinent combination of proven existing tools. The method utilizes delivery data that, in its original format, is noisy, intermittent, and does not accurately represent demand requirements at the point of consumption. Customers are grouped into segments of similar consumption behavior patterns and then forecasts are made for the segments. In the last step, the segment-based forecasts are applied to individual customers. Forecast evaluations show that this new method produces more accurate forecasts than actual traditional methods.

The proposed forecasting method has significant management applications in any domain where forecasts are needed for a large population of customers with noisy consumption information. This is particularly true for vendor managed inventory applications where collaborative information is not available. The proposed method could also be adapted to domains such as mass customization where highly aggregated information must be applied to individual cases or customers.

In this research, we employ traditional forecast error measures (RMSE and MAPE) due to their wide acceptance in the forecasting domain, however, these measures are not optimal in actual case study due to their point-to-point measures. We supplement the evaluation with DTW, which is a method of measuring the distance between two series of data. DTW is widely accepted in the literature as an excellent measure of differences between two time-series, however, more research is necessary to explore the performance of DTW as a forecast evaluation metric.

Proposed method is of great interest for the optimization of various supply chains situations. Effective performance of most optimization models are highly dependent to the data that describes actual or future situation. Besides, in many situations available data may not correspond directly to what is expected for the different models because of too large quantity and imprecision of the data that may lead to suboptimal or even bad decisions. Actual problem considers the availability of a large and noisy dataset concerning historical information about each customer that will be used to make improved prediction models, that may fit models to optimize the supply chain. Impact of the balance between the number of clusters (which gives the number of prediction models) and the overall performance of the prediction models deserves some supplementary interest.

CHAPTER 7 ARTICLE 4: PREDICTING CUSTOMER PURCHASE BEHAVIOR PATTERNS BASED ON DESCRIPTIVE ATTRIBUTES

Submitted to: The Journal of Industrial Marketing Management

Murray, P., Agard, B., & Barajas, M.A., (2018d)

Abstract: Market segmentation based on descriptive attributes is used in business to aid in forecasting, product planning, and strategic decision making. Segmentation generally begins with identifying a set of descriptive attributes such as age, sex, or location and then using those attributes as variables on which segmentation decisions are based. It is assumed that the information contained in the attributes will help to predict behavior. In this research we test the widely accepted segmentation strategy against segmentation based on purchase behaviors. We first segment a group of customers based on their actual historical behavior patterns. We then attempt to explain the groupings using classical descriptive attributes. Our research demonstrates that in some domains, market segmentation based on descriptive attributes does not produce good clusters.

7.1 Introduction

Market segmentation, a staple of the business environment, is considered one of the most important business tools available to managers (Boejgaard & Ellegaard, 2010) and has become so ubiquitous that its validity is seldom questioned (Sarabia, 1996). The benefits of market segmentation were presented half a century ago in a seminal paper by Smith (1956). Various methods to divide a population into segments range from highly subjective to highly analytical. A priori knowledge is applied when the analyst has sufficient knowledge of the domain, lacks resources or skills to conduct a more sophisticated approach, or lacks the necessary data to apply an analytical approach. Similarly, a priori knowledge is utilized when market members are asked to self-identify with a specific group; in this case, the customer decides its own membership. Analytical approaches require data that represents either the customers or their behaviors (Murray et al., 2017). Regardless of the approach taken to identify segments, the base assumption is that the available data or knowledge is relevant for establishing similarities and differences between customers.

Given that market segmentation is a well-established and researched field of study (Rezaei & Ortt, 2013) and given its prevalence, it is reasonable to assume that it is a reasonably accurate and thoroughly tested method. This assumption, however, is not necessarily correct. Some research suggests that businesses tend to struggle with implementing segmentation strategies due to lack of similarities of segment members (Dibb & Simkin, 2001). Inconsistent results for market segmentation is compounded by a lack of research on methods to evaluate segmentation results (Sarabia, 1996).

The context for this research is a vendor managed inventory (VMI) situation where one supplier provides a product to multiple customers. In the VMI arrangement, the customers obtain their product from a sole source thus eliminating the need to consider the effects of changing market shares, promotional activities, and other influences that might alter the demand for the product.

In our research, we employ two different strategies to segment a population of customers and evaluate whether one strategy can be used to validate the other. We first segment the customers based on their demonstrated purchasing behaviors. The homogeneity of the segments is visualized by graphing the behavior patterns and witnessing seasonal, cyclic, and trend patterns that are present in some segments and absent in others. Next, we attempt to classify the same population into their segments based on their descriptive attributes. We have found that the results are contrary to what was expected. Attempting to classify customers based on descriptive attributes does not result in groups with similar behaviors. The findings in this research are significant in that they demonstrate that, contrary to common business practice, market segmentation is not universally applicable. The remainder of the paper is organized as follows: Section 2 presents a review of the state of the art. Section 3 presents the methodology and the results are presented in Section 4. The paper concludes in Section 5 with a discussion of the scientific contribution, limitations of the research, and suggestions for future research.

7.2 Literature Review

The literature review begins with a brief review of the state of the art in market segmentation, a more complete review is offered by Boejgaard and Ellergaard (2010). Additional details regarding descriptive attributes are reviewed due to their relevance to the research. Segment evaluation is also an important topic in this research.

7.2.1 Market Segmentation

The goal of market segmentation is to establish groups of customers who are similar within the group (homogeneous), and dissimilar to other groups (heterogeneous); the underlying assumption is that similar customers within a segment will exhibit similar purchasing behaviors (Sommers & Barnes, 1982). Establishing segments, also known as clustering, requires a set of variables that reflect similarities and differences. Once variables are established, a method to measure similarities based on those variables is needed. Many methods exist for dividing a market into segments (Chakraborty, 2013; Han & Kamber, 2006; Le et al., 2009), but in general, the methods are based either on descriptive attributes or behavioral attributes (Murray, Agard, & Barajas, 2018c). Descriptive attributes (such as size, location, sex, age or transaction frequency) are commonly used since these variables are easy to quantify. Success in segmentation based on descriptive attributes relies heavily on two assumptions: that data to support the variables is available, and that the descriptive attributes are truly relevant for creating the segments. Segmenting based on behavioral attributes requires historical data that contains information relevant to the behavior and a method for extracting that behavior information from the data (Barragan et al., 2016; Kashwan & Velu, 2013).

.

7.2.2 Descriptive Attributes

Descriptive attributes are obtained through a variety of primary and secondary sources. Banking industries gather detailed customer information when accounts are opened and update their records when transactions are conducted (Moro, Cortez, & Rita, 2015). Domains such as banking, where customer-related data is abundant, may suffer from too many variables creating high-dimensional datasets which are challenging to analyze. When a domain has multiple sources of data on the same subjects, it is prudent to cross check the information in attempt to increase the potential accuracy of the data (Dan & Zondag, 2016) At the other extreme, some domains have limited information about their customers. The limited information source needs to be expanded and quantified into variables.

Knowing a customer's location, even if only described by a zip code, enables identification and quantification of many potentially useful variables (Miller & Han, 2009). Location allows the development of variables such as climate, urban density, travel time between warehouse and customer, population & economic influences, and proximity to other customers. While location linked information is available through different databases, it has been found that overlaying information layers in geographical information system (GIS) software reduces the level of error that might be undetected in direct database matching (Caviglia-Harris & Harris, 2005). GIS analysis adds more dimension to variables. For example, the travel time between warehouse and customer site is more than a simple distance measure; topographical features, urban density, and proximity to other customers also plays a role in calculating travel time (Cheng, Li, & Yu, 2007).

Some data produced through GIS analysis is too detailed and customer specific to be used in its original form as descriptive attributes. For example, every customer has a unique number of minutes of travel time between warehouse and customer site. Inducing the detailed data into descriptive variables is a process known as generalization, often accomplished through hierarchy categorization (Petry & Zhao, 2009).

Through GIS analysis, economic analysis, and statistical calculations, a broad variety of descriptive variables can be developed and quantified. The descriptive variables are then available for segmentation analysis.

7.2.3 Segment Evaluation

Sarabia (1996) identified that little attention in research had addressed how to evaluate the results of market segmentation. Dibb (1999) suggested several criteria upon which segments could be evacuated, but did not offer a qualitative method to accomplish the evaluation. Statisticians generally approach cluster evaluation by measuring homogeneity and heterogeneity. The homogeneity of each cluster is measured to determine the level of similarity of members within the cluster and heterogeneity between clusters to determine how different each cluster is from the other clusters (Seret et al., 2015). Ideally, all members of a cluster will have similar behavior patterns and the composite pattern of each cluster will be uniquely different from the patterns of other clusters. The concept of segment evaluation via measurement of homogeneity and heterogeneity is straightforward, and there are well established statistical methods that are suitable (Aggarwal & Reddy, 2013). However, the statistical measures critically depend on the availability of qualitative variables upon which the calculations are applied, if the variables are absent, similarity measures cannot be calculated.

Market segmentation is pertinent due to its wide acceptance in business. Dividing a population into segments requires some type of measure that can distinguish one group from another. The measure can be as rudimentary as "girls and boys" or some advanced measurement system. Regardless of the measure selected, it is critical that it represents a useful distinction and not merely a difference. Tools exist for evaluating clusters, but they also depend on the adequacy of the available variables. The following section proposes a methodology to establish market segments based on demonstrated behavior and then test the effectiveness of attribute-based segmentation on the same population.

7.3 Methodology

To evaluate this problem, we proposed the following methodology which consists of three high level steps, as presented in Figure 7.1. In Step 1, historical data is used to build a quantitative description of the customers' past behavior. Once behaviors are quantified, a measure can be employed to build a similarity matrix which in turn can be used to divide the population into segments; a process commonly referred to as "clustering". Behavior-based segmentation is robust in that it ignores potential (assumed) causes of behavior types—which may be incorrect or misleading—and relies only on actual demonstrated behavior. In Step 2 the same population of customers are segmented based on descriptive attributes. The results of the two different approaches to segmentation are evaluated in Step 3.

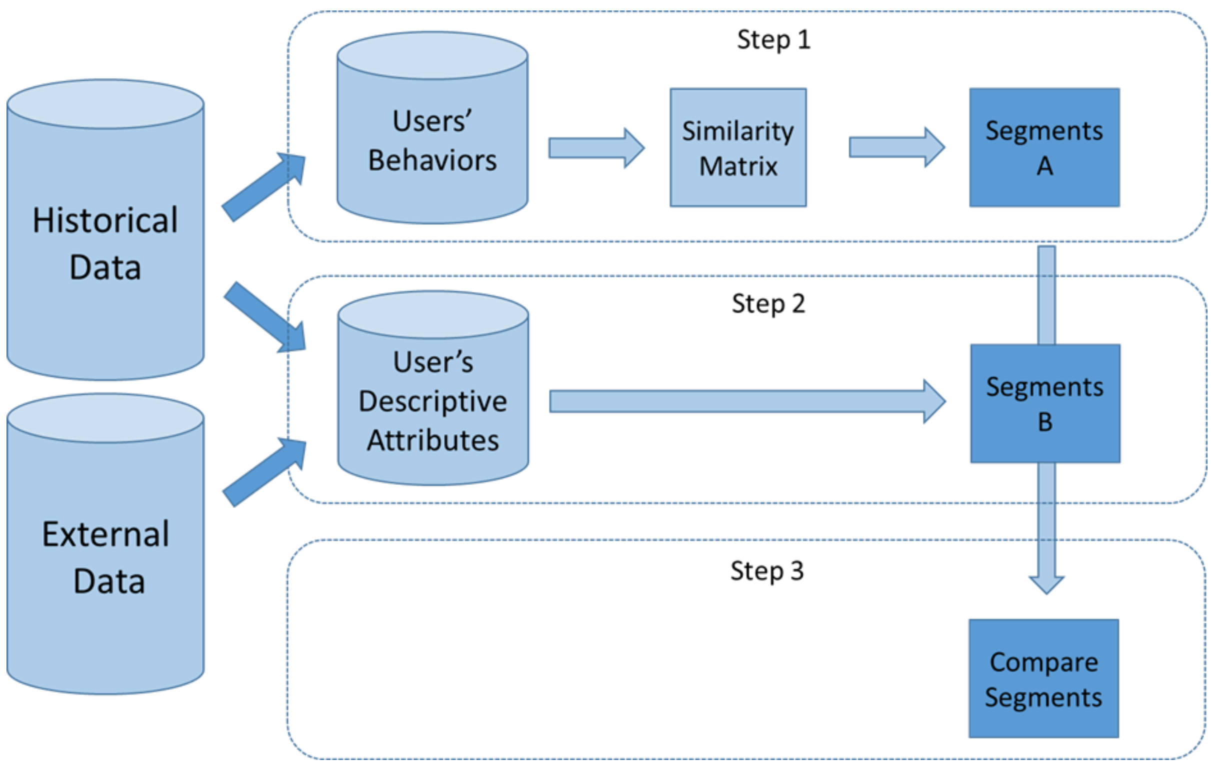


Figure 7-1: Methodology Overview

7.4 Results with Real Data

In this section, we apply the proposed methodology in a case study using a real dataset to evaluate its effectiveness and suitability for real world application.

7.4.1 Context of the Case Study

Simulated data is often used in research to demonstrate the application of new theories. In this research, the goal is not to propose a new method, but rather to evaluate whether an existing method—market segmentation based on descriptive attributes—produces clusters of customers with similar behavior patterns. The evaluation must therefore be conducted using real data. The case study data was provided by an industrial partner that supplies bulk liquid products that are used in various industries including manufacturing, agriculture & food processing, medical research, and mining. This diverse group of customer types are located throughout the continental United States. Point-of-use inventories are managed through a vendor managed inventory (VMI) arrangement; the supplier is responsible for replenishment decisions to ensure uninterrupted product availability. Long-term purchasing contracts and specialized point-of-use storage systems normally preclude customers from purchasing from alternate suppliers. Delivery records for all customers for a period of approximately five years are available.

The case study should be an ideal subject for market segmentation due to its diverse mix of industry types and locations. For example, it is expected that both climate and location will directly influence agricultural customers' behaviors. Conversely, manufacturing customers should be less influenced by climate since they normally operate in-doors. Other factors, such as economy may have a greater influence on manufacturing since manufacturers' output is driven by economic demand.

Prior to attempting to analyze a real dataset, preprocessing is necessary to produce a set of relatively clean time-series from which patterns may be detected and analyzed (Kantardzic, 2011). Customers with unexplainable data histories, such as a delivery quantity that greatly exceeds delivery truck capacity are considered outliers. Lost customers and newly acquired customers are also considered outliers due to insufficient historical data.

Preprocessing results in a study group is comprised of customers who have continuous operations and do not exhibit erratic behavior patterns. Limiting the study group in this fashion ensures that the analysis is not compromised by attempting to interpret behavior patterns of customers with highly erratic behaviors, insufficient data points, or delivery events that cannot reasonably be explained (such as delivery quantities that are negative or that exceed actual tanker size). The results of the preprocessing and study group selection are shown in Figure 7.2.

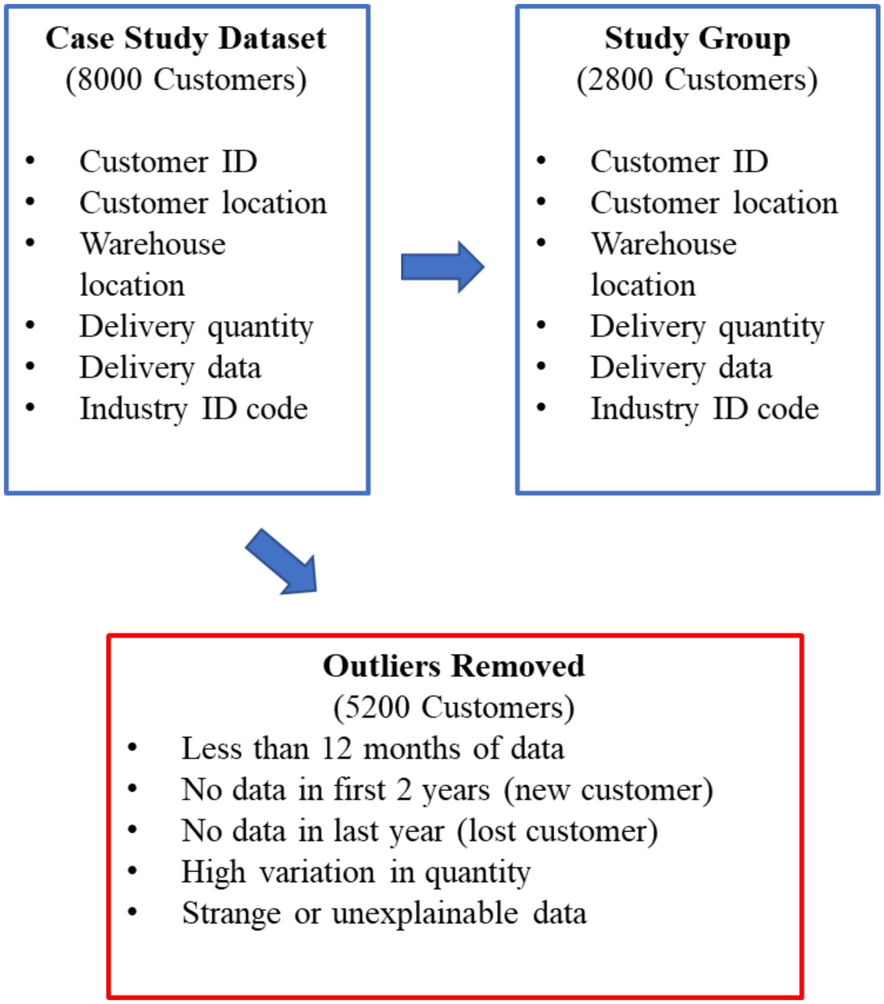


Figure 7-2: Data Preprocessing

7.4.2 Step I – Behavior-based Segmentation

Step I of the proposed methodology consists of four sub-steps as illustrated in Figure 7.2. The dataset is a list of unique transaction events that must be aggregated into time-series format prior to segmentation analysis per Figure 7.3(a). The conversion to time-series format also includes normalization to put all customers on the same scale. Normalization is necessary to prevent differences in magnitude from overshadowing behavior patterns. Aggregating delivery records often produces intermittent time-series due to delivery occurrences less frequent than the bins size selected for aggregation (Kourentzes et al., 2014). Intermittent time-series are very difficult to predict and therefore need to be smoothed prior to subsequent analysis to resolve the intermittency.

Smoothing is accomplished per Figure 7.3(b) by applying a simple technique known as the "aggregate-disaggregate intermittent demand approach" (ADIDA) (Petropoulos et al., 2016). With ADIDA, the intermittent time-series is aggregated to a level sufficiently high to remove the intermittency and then reaggregated back to the original bin sizes.

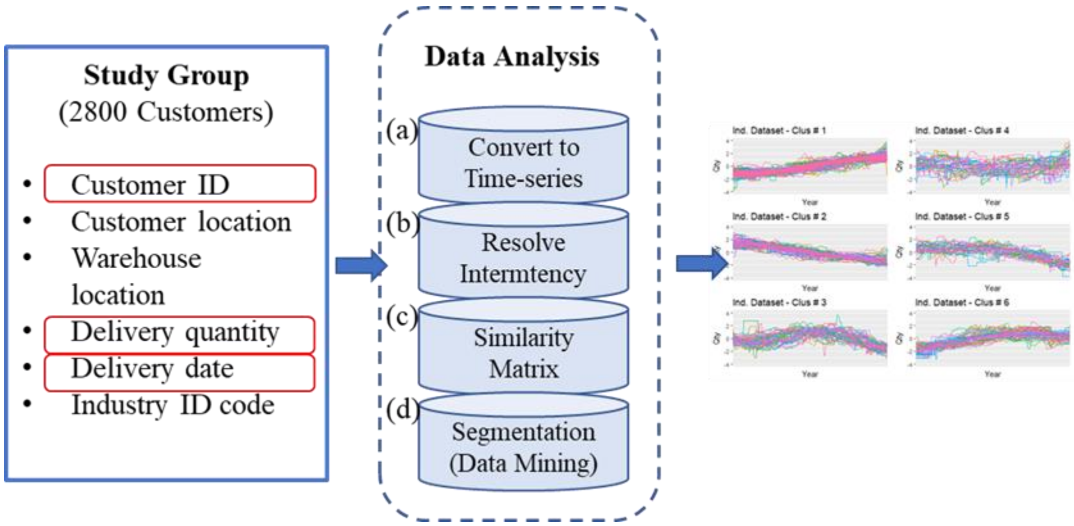


Figure 7-3: Behavior-Based Segmentation

A sample population of customers are clustered according to the steps outlined in Figure 7.3. The results are shown in Figure 7.4. The members of Cluster #1 all demonstrate an increasing trend over the period of the study. Members of Cluster #2 and #4 show distinct seasonal patterns, and members of Cluster #3 show both decreasing trend and some seasonal pattern. In addition to the more obvious behavior patterns, other subtler patterns are evident; Cluster #4 appears to either decline at the end of the study period, or perhaps has a cyclical period that is much longer than the period of the study data. Some interpretation of the results requires more a priori input than others.

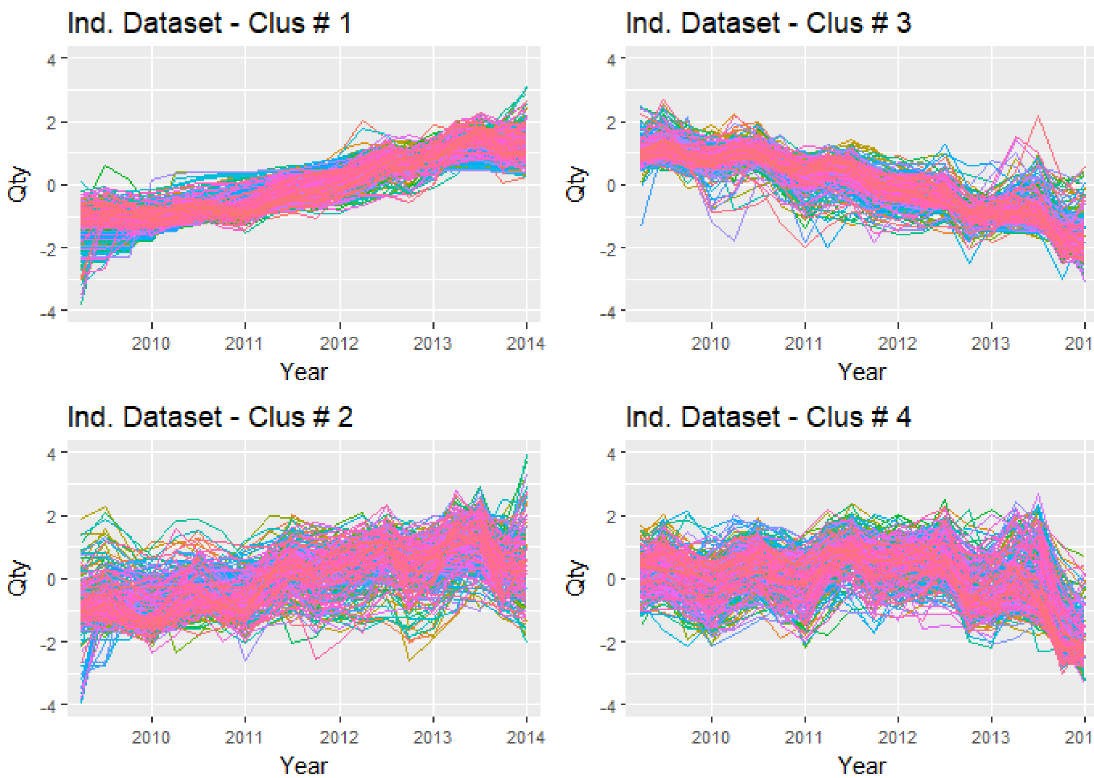


Figure 7-4: Behavior-based Segmentation

7.4.3 Step II – Attribute-based Segmentation

Attribute-based segmentation requires the identification or development of descriptive attributes from which the classification can be based. The case study dataset is sparse regarding descriptive attributes; however, the available information can be converted into variables, as illustrated in Figure 7.5, that are expected to describe the customers and according to the practices employed in business environments, these descriptive variables should allow us to segment the population into groups of customers with similar behaviors. The customers' cluster membership is known from Step I; assigning them to clusters based on the descriptive variables is a task known in data mining as classification.(Witten, Farnk, & Hall, 2011). There are many algorithms available for classification; one of earliest and simplest is the divide and conquer style decision tree (Witten et al., 2011). Unlike Step I where only date and time are considered; this step of the methodology utilizes all available information from the dataset. The location of the customer and its designated warehouse are analyzed with ArcMap (ESRI, 2011) in conjunction with GIS databases made available from Esri (ESRI, 2011) and US government sources (United States Census Bureau, 2015). The GIS analysis produces variables relating to climate, transportation logistics, and population.

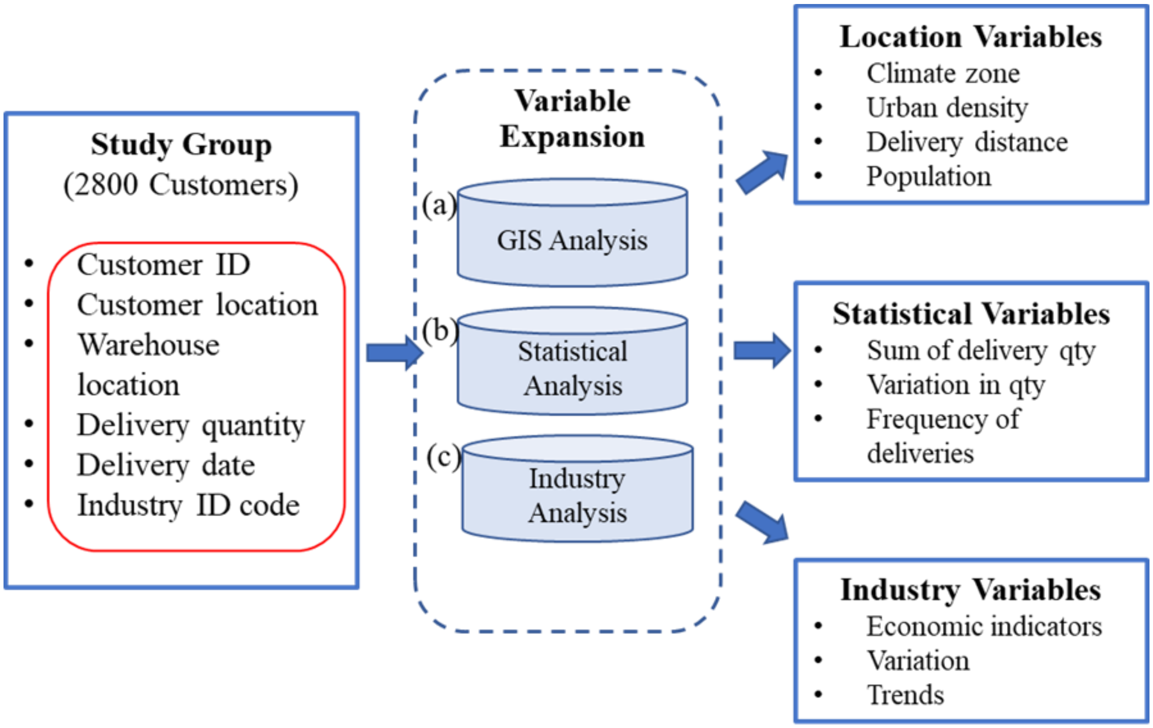


Figure 7-5: Step II – Variable Expansion

The statistical analysis produces several simple measures including sum of quantity, variation in delivery quantity, and delivery frequency. There are many other statistical measures that could be produced, however, we limit the analysis to the most basic ones to avoid masking the actual knowledge gained from them. Like the statistical analysis, the industry analysis is virtually unlimited in the number of variables that might be produced.

After variable expansion, some variables must be generalized to convert a continuous variable into a descriptive attribute. For example, a simple hierarchical approach as illustrated in Figure 7.6 is used to convert the distance variable. Customers are generalized as “Near”, “Typical”, or “Remote” based on their distance from the warehouse. Generalization may not evenly distribute the variables; most observations might fall into a single category. In this example, the distance attribute “Typical” is further split into “100” mile range and “250” mile range.

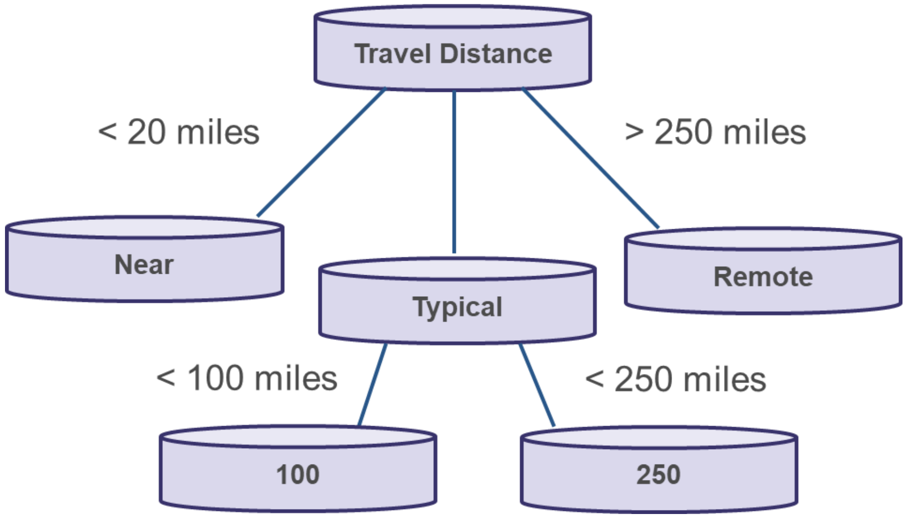


Figure 7-6: Generalization Hierarchy

Table 7.1 illustrates an example of the variables available for classification. They include cluster assignment, industry type, climate zone, urban location, distance from warehouse, size (based on total quantity delivered), quantity variance between deliveries, and delivery frequency.

Table 7-1: Variables (sample shown for illustration)

Cluster	Industry	Climate Zone	Urban	Distance	Size	Qty Variance	Deliv. Frequency
1	unknown	2	Y	250	xl	xl	xl
1	manuf	4	Y	250	tiny	tiny	tiny
1	trans	3	N	near	tiny	small	tiny
1	manuf	3	Y	100	small	tiny	small
…	…	…	…	…	…	…	…
6	unknown	2	Y	near	tiny	small	tiny
6	trans	2	Y	near	small	small	xl

The new dataset containing the descriptive attributes is divided into training and test datasets, which will permit to evaluate if over fitting occurs. The training data is passed to a decision tree algorithm that builds a decision tree based on the data presented, as illustrated in Figure 7.7.

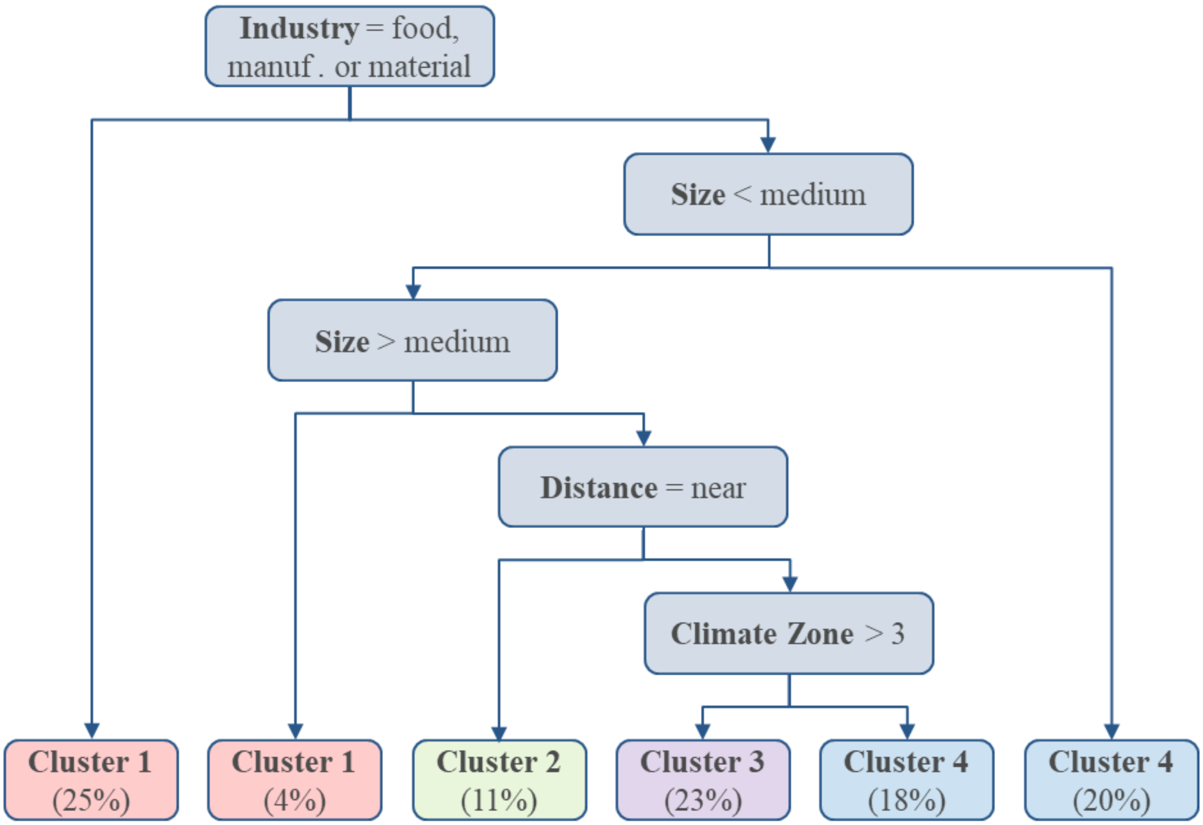


Figure 7-7: Decision Tree

7.4.4 Step III – Segment Evaluation

After building a decision tree, the test data is processed by the tree to determine whether it is capable of correctly classifying the test data. The confusion matrix illustrated in Table 7.2 shows that only 25% of the test data is correctly classified. The values outside of the red dashed boundary represent customers who were assigned to incorrect clusters.

Table 7-2 - Confusion Matrix

	Predicted			
Actual	**1**	**2**	**3**	**4**
1	5	2	10	8
2	10	4	8	12
3	11	2	14	9
4	11	0	12	9

The mediocre performance of the decision tree is partly explained by inspecting some of the exogeneous variables. Figure 7.8 illustrated the distribution of the variable “climate zone”. Rather than different behaviors being associated with different climate zones, we see that the behavior types are nearly uniformly distributed throughout the climate zones. Comparable results were experienced with the other variables.

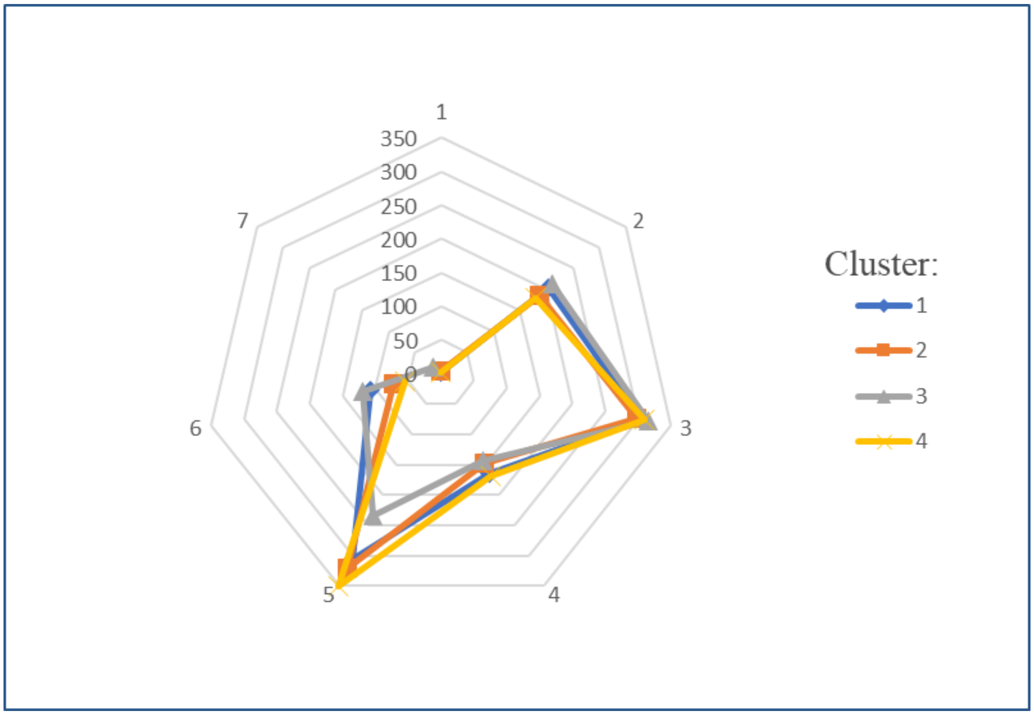


Figure 7-8: Distribution of Behavior Types in all Climate Zones

7.5 Conclusion

The foundational assumption behind market segmentation is that customers with similar attributes will have similar behaviors. Our results do not support that assumption In the case study, customers were segmented based on clearly demonstrated behavior patterns such as seasonal effect and increasing or decreasing trends.

In preparation for classification, the available information from the case study data was expanded into descriptive attributes which in turn were fed to a decision tree algorithm. An important and perhaps inaccurate assumption for the test case was that although a nearly unlimited set of descriptive attributes could be created for the available data, the most basic ones such as climate zone or overall customer size should provide useful behavior predictors. Classification based on descriptive attributes, however, was not able to group customers with similar behavior patterns. These findings are important considering that little research and application has been done to develop methods to evaluate the results of market segmentation. This research is limited in that the case study was limited to one domain. The research is also limited in that no attempt was made to produce an exhaustive set of exogenous variables which are nearly limitless given the huge amount of data that could be readily generated through statistical, GIS, and economic analysis. The number of variables was kept small in interest of keeping the experiment simple, however, a greater number of variables may have produced better results. This research is a significant contribution in that it reveals that the nearly universally accepted business tool known as market segmentation may not be applicable in all domains. These findings should inspire future research on the neglected topic of methods to validate market segmentation

CHAPTER 8 GENERAL DISCUSSION

8.1 Comments on the Methodology and Results

The methodology employed in this research consists of a series of steps to progress from a noisy dataset to a specific and actionable conclusion. Following the overall framework presented herein, there are many steps in the execution where decision points are necessary. Selection of smoothing coefficients, distance calculations, data mining algorithms, and other similar decisions are sometimes guided by comparable research found in the literature. At other points, the decisions required trial and error approaches and while experimentation was extensive, it was by no means exhaustive. Each decision point has the potential to contribute error into the overall framework and while in some domains errors tend to cancel each other, in other domains they may compound and magnify.

We did not attempt to minimize the overall error in the framework due to a second potential problem with the research. The experimentation used to test and demonstrate the proposed methodology is based on a single case study that we assume is representative of other similar VMI situations. Due to the limited case study, we did not attempt to optimize the models since those adjustments may not prove effective in other domains or with other case study data. The methodology and the framework that it is presented in does provide a functioning solution that due to its generic structure, should be easily applied to other data in another domain.

The methodology is developed and presented as a set of related, yet separate components (presented in Chapters 4 through 7). When applied to a comparable situation, the overall method should prove useful for solving the problem of interpreting and forecasting based on noisy transaction data. The separate components may also be applied in other domains as a portion of a solution to a different type of problem.

8.2 Research Limitations

The research presented herein has limitations due to the outlier removal strategy, selection of parameters within algorithms, and testing in a single domain. The outlier removal strategy was structured such that customers with corrupted data were removed from the study. One example is a customer whose identifier number changed would appear in the data as two separate customers;

one lost and another acquired at the same point in the data. The two customers would have insufficient data duration to include in the study and would be removed. Errors in data entry could make an individual customer's data appear erratic and it would be identified as an outlier and removed. The goal of the research was to prove a concept, however, a method to repair the data rather than remove all outliers is needed prior to industrial application.

The proposed method includes many individual steps and where available, published algorithms are employed within those steps. Many of the steps and algorithms have decision points such as selecting the number (K) or clusters or specifying a smoothing coefficient (alpha). These decision points were mostly selected through trial and error to obtain satisfactory results; a concerted effort to optimize the decision points could increase the accuracy. The decision points were not optimized due to the case study being limited to one domain; optimization might not hold true for other domains. Employing published algorithms within the method holds the advantage of using proven tools, however, this incorporates two risks. First, there are often more than one published algorithm to accomplish the same task, selecting the best one is not always intuitive. Second, this research often employs tools that were developed for other purposed; problems that are inconsequential in other domains might become more significant after repurposing.

The proposed method has been tested on a single domain. There is no reason to doubt its validity in other domains where similar conditions exist. However, additional testing would be beneficial both for proof of concept and to assess where decision points should be future optimized.

Lastly, the overall intent of this research is to solve a real industrial problem. We believe we have presented a working solution that, at least in one domain, can be applied and should lead the firm to improved understanding of its supply chain requirements. The method, however, is not structured as a user-friendly tool. Decision points are hard coded within the software and are not intuitively obvious for their locations and/or purpose. Prior to industrial application, the software for the proposed method needs to be converted into a user-friendly tool.

CHAPTER 9 CONCLUSION AND RECOMMENDATIONS

9.1 Findings

Vendor managed inventory is a well-established form of supply chain management; its success is highly dependent on collaborative information. Barriers to sharing collaborative information, whether technical, political, or trust-related sometimes force the supply chain planners to look to other data sources to substitute into their planning activities. Substitute data, however, may contain noise from other influences such as logistics decisions and the bull whip effect. The ability to see through the noise and find the underlying information is paramount for converting a large and noisy dataset into actionable knowledge.

The research presented here forms a framework for gaining knowledge from an otherwise unusable dataset. The case study provides a platform for validating the proposed framework and a viable set of results is produced. The application on the case study is focused on demonstrating that the methods work as proposed. However, the framework is sub-optimal in that there are many small decision points throughout the method, such as selecting a smoothing coefficient value, that will influence the overall outcome. Application of the proposed method will require further testing to determine the overall effect of adjusting the various decision points.

The components of our proposition are presented in four parts that work together for solving one specific problem. Each component has potential applications in other domains and might be utilized in solving other types of problems. Despite their individual uniqueness, the four parts are also sequentially dependent on their preceding part.

In the first part, Chapter 4, the data is prepared for the analysis. Initial attempts to solve the research problem assumed that the raw dataset, which consists of a list of delivery transactions, could simply be aggregated into monthly buckets and then a forecasting tool such as ARIMA (Newbold, 1983) could be applied to develop useful forecasts. Monthly bins appeared suitable due to having five years of data to analyze in the case study. The initial results were extremely noisy to the point where no information was revealed. Changing to large bin size, quarterly instead of monthly, made almost no improvement. The case study contains a variety of customers with delivery frequencies ranging from daily to others with only a few deliveries per year. Aggregating the data into small temporal bins resulted in intermittent time-series and large bins resulted in loss of behavior patterns. We found that applying Croston's method at the smallest possible aggregation level (daily bins) allowed us to resolve the intermittency without removing any behavior patterns. The data could then be re-aggregated into a bin size (such as monthly) to facilitate the subsequent analysis. This part is described in detail in Chapter 4.

Once the data was preprocessed and suitable for forecast analysis, we had to address the second challenge—there are too many customers to allow individual forecast analysis. In chapter 5, we proposed a method to segment (cluster) the population of customers based on their demonstrated behavior patterns. Much of the literature regarding cluster analysis relies on K-means. K-means is popular due to often producing satisfactory results. However, we use time-series, generated from the original transaction data to represent each customer. K-means utilized direct, point-to-point Euclidean distance measure and it not suitable for comparing sets of time-series. We apply dynamic time warping (DTW) and can create segments with high inter-cluster similarity. This part is described in detail in Chapter 5.

In Chapter 6, we proposed a method to generate segment-based forecasts and then apply those forecasts to individual customers. The segment-level forecasts exhibit good accuracy due to the cluster members having similar behavior patterns. Those forecasts are then applied to the individual customers. In doing this, a relative small number of forecasts can represent a large population of customers.

In the last part of the research, we attempt to validate and improve the method by incorporating external variables. It is expected that behavior patterns should be influenced by exogenous factors such as climate, urban density, and distance between the customer location and the distribution point. In Chapter 7, we utilize the case study data to identify locations and industry types. Those two items are expanded through GIS and economic analysis to create a set of variables that could represent descriptive attributes for each customer. This part of the research show that there is actually very little correlation between the customers' actual behavior patterns and their descriptive attributes. This is surprising considering that customer segmentation based on descriptive attributes is a common business practice. We concluded in this part that either our selected set of attributes was not good, or that the segmentation practice is significantly flawed. More research on this topic is warranted.

9.2 Research Contributions

The contributions of this research are important in three categories, methodological, scientific, and practical. These are discussed in the following sections:

9.2.1 Methodological Contribution

The focus of many research topics is often to create new tools even if the need for the tool is not yet recognized or fully defined. The approach considers what can be done and then what can it be used for. Certainly, the "create and then apply" approach has led to important breakthroughs in many domains. This research takes a more mechanical approach. Here, we begin with a real and somewhat well-defined business problem and then search for an established set of tools to solve the problem. The tools we use are known to work in other domains, but a new application and combination with other tools requires testing and evaluation.

This methodological strategy follows the well documented example of the re-application of dynamic time warping (DTW). DTW was initially developed in the domain of speech recognition; the predecessor of the algorithms used in most automated telephone systems. Researchers in other domains quickly discovered that DTW's ability to measure similarity between speech patterns could be applied to measuring similarities between other types of time-series data. DTW has essentially become the standard method to compare time-series.

The methodological strategy employed here demonstrates that new problems do not necessarily require new tools.

9.2.2 Scientific Contribution

This research presents a scientific contribution on two levels. First, when taken as a whole as presented in the thesis, a method of solving a complex business problem is proposed as a series of interconnecting steps. A set of large, noisy, and stochastic data is cleaned, interpreted, and transformed into a solution that offers forecast information for the domain. Second, when taken as individual parts, as presented in the chapters, several different stand-along solutions are available. While the domain of our research is vendor managed inventory, the individual components may be applied in other domains for solving other types of problems.

9.2.3 Practical Contribution

The case study that inspired this research is a real problem that our industrial partner had difficulties to solve to solve with other methods. AL was fortunate to have access to several years of delivery records for its continental USA customers. However, extracting any useful knowledge from the data was still a challenge. The data was far too big and noisy to interpret using traditional graphical or statistical tools. The methods proposed in this research allows AL to clean the data, remove the noise, and view the results in a manageable format. Once AL can "see" the underlying behavior patterns, it can better comprehend its situation and use the newly available knowledge to aid with business decisions and planning.

The practical contributions begin with a method to clear the data so that it might be utilized in analysis. The first three parts comprise a set of computer codes that AL can directly apply to manage its medium and long-term forecast planning. These codes could be easily adapted to any domain where historical transaction records are available, and customers have an ongoing repeat relationship. The extremely noisy, big dataset is cleaned and processed adequately such that a firm can use the results to understand its customers' behaviors and develop plans based on those behaviors.

The final portion of the research, presented in Chapter 7 is important in its disruption of a paradigm. Many companies use the premise in their business planning that descriptive attributes are useful for predicting customers behavior patterns. For example, forecasts for new customers' consumption may partly be based on industry type and geographic location. Chapter 7 demonstrates that these types of attributes are not necessarily correlated with consumption behavior. Knowing that the assumed correlation may not be strong will aid AL in deciding how much it can rely on this type of planning.

9.3 Perspectives, Future Work

The research presented herein begins with the imprecise step of converting a set of delivery records into a representation of consumption behaviors. A workable solution is proposed in Chapter 4, but more research is needed to refine this conversion. Specifically, comparison of the converted data to actual consumption records is necessary so that refinements to the conversion methods can be made. A refined method may also lead to better correlation with exogenous variables.

In Chapter 7, we demonstrated that there was no detectable correlation between descriptive variables and consumption behavior. This is significant due to the widely accepted views, in both academia and industry, that this type of correlation exists, is strong, and is useful as the foundation for market segmentation. Further research is necessary to determine when and how these correlations exist, and (more importantly) whether the widely accepted attribute-based segmentation method is valid.

Lastly, for industrial application, our methods require more testing and refinements. We have presented a methodology that produces an effective result. However, application in industry would benefit from optimizing the parameters of the decision points. Additionally, the algorithms selected, such as DTW for distance calculation, may be replaced with another algorithm that better suits the domain where the method is applied.

BIBLIOGRAPHY

Aburto, L., & Weber, R. (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing, 7*(1), 136-144. doi:http://dx.doi.org/10.1016/j.asoc.2005.06.001

Achabal, D. D., McIntyre, S. H., Smith, S. A., & Kalyanam, K. (2000). A decision support system for vendor managed inventory. *Journal of Retailing, 76*(4), 430-454. doi:10.1016/S0022-4359(00)00037-3

Agard, B., & Kusiak , A. (2004). Data-mining-based methodology for the design of product families. *International Journal of Production Research, 42*(15), 2955-2969. doi:10.1080/00207540410001691929

Agard, B., Partovi-Nia, V., & Trepanier, M. (2013). *Assessing public transport travel behaviour from smart card data with advanced data mining techniques*. Paper presented at the 13thWCTR - World Conference on Transportation Research, Rio de Janeiro, Brazil.

Aggarwal, C. C., & Reddy, C. K. (2013). *Data clustering: Algorithms and applications*: CRC press.

Agrawal, R., Faloutsos, C., & Swami, A. (1993). Efficient similarity search in sequence databases *Foundations of Data Organization and Algorithms*. Chicago, IL: Springer.

Ahmadi Javid, A., & Azad, N. (2010). Incorporating location, routing and inventory decisions in supply chain network design. *Transportation Research Part E: Logistics and Transportation Review, 46*(5), 582-597. doi:http://dx.doi.org/10.1016/j.tre.2009.06.005

Air Liquide. (2014). *Air Liquide context*. Retrieved from Air Liquide:

Altintas, N., & Trick, M. (2014). A data mining approach to forecast behavior. *Annals of Operations Research, 216*(1), 3-22. doi:http://dx.doi.org/10.1007/s10479-012-1236-9

Amemiya, T., & Wu, R. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American statistical association, 67*(339), 628-632.

Anderson, W. T. J., Cox, E. P., & Cooper, J. G. (1976). Bank selection decisions and market segmentation. *Journal of Marketing, 40*(1).

Angulo, A., Nachtmann, H., & Waller, M. A. (2004). Supply chain information sharing in a vendor managed inventory partnership. *Journal of Business Logistics, 25*(1), 101-120.

Armstrong, J. S. (1985). Long-range forecasting: From crystal ball to computer. *Journal of Accountancy, 146*(5), 87-88.

Armstrong, J. S. (2006). Findings from evidence-based forecasting: Methods for reducing forecast error. *International Journal of Forecasting, 22*(3), 583-598. doi:http://dx.doi.org/10.1016/j.ijforecast.2006.04.006

Athanassopoulos, A. D. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of Business Research, 47*(3), 191-207. doi:http://dx.doi.org/10.1016/S0148-2963(98)00060-5

Azadeh, A., Ghaderi, S., & Sohrabkhani, S. (2007). Forecasting electrical consumption by integration of neural network, time series and ANOVA. *Applied Mathematics and Computation, 186*(2), 1753-1761.

Bag, S., Tiwari, M. K., & Chan, F. T. (2017). Predicting the consumer's purchase intention of durable goods: An attribute-level analysis. *Journal of Business Research*.

Bala, P. K. (2012). Improving inventory performance with clustering based demand forecasts. *Journal of Modelling in Management, 7*(1), 23-37. doi:http://dx.doi.org/10.1108/17465661211208794

Ballou, R. H. (2007). The evolution and future of logistics and supply chain management. *European Business Review, 19*(4), 332-348. doi:http://dx.doi.org/10.1108/09555340710760152

Barajas, M., & Agard, B. (2014). A methodology to form families of products by applying fuzzy logic. *International Journal of Interactive Design and Manufacturing*, 1-15. doi:10.1007/s12008-014-0230-7

Barirani, A., Agard, B., & Beaudry, C. (2013). Competence maps using agglomerative hierarchical clustering. *Journal of Intelligent Manufacturing, 24*(2), 373-384. doi:10.1007/s10845-011-0600-y

Barragan, J. F., Fontes, C. H., & Embiruçu, M. (2016). A wavelet-based clustering of multivariate time series using a multiscale SPCA approach. *Computers & Industrial Engineering, 95*, 144-155. doi:http://dx.doi.org/10.1016/j.cie.2016.03.003

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Journal of Operational Research, 20*(4), 451-468. doi:10.2307/3008764

Berndt, D. J., & Clifford, J. (1994). *Using dynamic time warping to find patterns in time series.* Paper presented at the KDD Workshop.

Biswas, S., & Sen, J. (2016). A proposed architecture for big data driven supply chain analytics. *Journal of Supply Chain Management, 13*(3), 7-33.

Boejgaard, J., & Ellegaard, C. (2010). Unfolding implementation in industrial market segmentation. *Industrial Marketing Management, 39*(8), 1291-1299. doi:10.1016/j.indmarman.2010.06.004

Brown, R. (1963). *Smoothing, forecasting and prediction of discrete time series*. Englewood Cliffs, NJ: Prentice-Hall.

Byrnes, P. (2014). *A three-stage approach to customer profiling.* (Ph.D.), Rutgers, The State University of New Jersey, Newark, NJ.

Calvet, L., Ferrer, A., Gomes, M. I., Juan, A. A., & Masip, D. (2016). Combining statistical learning with metaheuristics for the Multi-Depot Vehicle Routing Problem with market segmentation. *Computers & Industrial Engineering, 94*, 93-104. doi:http://dx.doi.org/10.1016/j.cie.2016.01.016

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research, 184*(3), 1140-1154. doi:http://dx.doi.org/10.1016/j.ejor.2006.12.004

Caviglia-Harris, J. L., & Harris, D. W. (2005). Examining the reliability of survey data with remote sensing and geographic information systems to improve deforestation modeling. *The Review of Regional Studies*, *35*(2), 187-205.

Chakraborty, G. (2013). *Customer segmentation using SAS Enterprise Miner*. Cary, NC: SAS Institute Inc.

Chang, W. L., Tay, K. M., & Lim, C. P. (2015). Clustering and visualization of failure modes using an evolving tree. *Expert Systems with Applications, 42*(20), 7235-7244. doi:http://dx.doi.org/10.1016/j.eswa.2015.04.036

Chao, P., Fu, H.-P., Lee, H.-H., & Chang, Y.-C. (2008). Identifying the customer profiles for 3C-prpduct retailers: A data mining approach. *International Journal of Electronic Businenss Management, 6*(4), 195-202.

Chase Jr., C. W. (2013). *Demand-driven forecasting: a structured approach to forecasting* (Second edition ed.): John Wiley & Sons.

Chellappa, R., Wilson, C. L., & Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE, 83*(5), 705-741. doi:10.1109/5.381842

Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, *46*(3), 436-443.

Chen, F., Ryan, J., & Simchi-Levi, D. (2000). The impact of exponential smoothing forecasts on the bullwhip effect. *Naval Research Logistics, 47*(4), 269-286.

Chen, Y., Zhang, G., Hu, D., & Wang, S. (2006). Customer segmentation in customer relationship management based on data mining. In K. Wang, G. Kovacs, M. Wozny, & M. Fang (Eds.), *Knowledge Enterprise: Intelligent Strategies in Product Design, Manufacturing, and Management* (Vol. 207, pp. 288-293). Boston, MA: Springer.

Cheng, E. W. L., Li, H., & Yu, L. (2007). A GIS approach to shopping mall location selection. *Building and Environment, 42*(2), 884-892. doi:http://dx.doi.org/10.1016/j.buildenv.2005.10.010

Chopra, S., & Meindl, P. (2007). *Supply chain management. Strategy, planning & operation* (5th ed.): Springer.

Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2008). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing, 20*(5), 501. doi:10.1007/s10845-008-0145-x

Christopher, M. (2000). The Agile Supply Chain: Competing in Volatile Markets. *Industrial Marketing Management, 29*(1), 37-44. doi:https://doi.org/10.1016/S0019-8501(99)00110-8

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559-583. doi:http://dx.doi.org/10.1016/0169-2070(89)90012-5

Co, H. C., & Boosarawongse, R. (2007). Forecasting Thailand's rice export: Statistical techniques vs. artificial neural networks. *Computers & Industrial Engineering, 53*(4), 610-627. doi:http://dx.doi.org/10.1016/j.cie.2007.06.005

Cooper, M. C., & Ellram, L. M. (1993). Characteristics of supply chain management and the implications for purchasing and logistics strategy. *International Journal of Logistics Management, 4*(2), 13-24.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly, 23*(3), 289-303. doi:10.2307/3007885

Dan, S. M., & Zondag, M. M. (2016). Drivers of alliance terminations: An empirical examination of the bio-pharmaceutical industry. *Industrial Marketing Management, 54*, 107-115. doi:https://doi.org/10.1016/j.indmarman.2015.06.006

Dangerfield, B. J., & Morris, J. S. (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting, 8*(2), 233-241. doi:Doi 10.1016/0169-2070(92)90121-O

de Melo Menezes, B. A., de Siqueira Braga, D., Hellingrath, B., & Buarque de Lima Neto, F. (2015, 13-16 Oct. 2015). *An evaluation of forecasting methods for anticipating spare parts demand.* Paper presented at the 2015 Latin America Congress on Computational Intelligence, Piscataway, NJ, USA.

Dibb, S. (1999). Criteria guiding segmentation implementation: Reviewing the evidence. *Journal of Strategic Marketing, 7*(2), 107-129. doi:10.1080/096525499346477

Dibb, S., & Simkin, L. (2001). Market segmentation: Diagnosing and treating the barriers. *Industrial Marketing Management, 30*(8), 609-625. doi:http://dx.doi.org/10.1016/S0019-8501(99)00127-3

Diebold, F. X., & Lopez, J. A. (1996). 8 Forecast evaluation and combination *Handbook of Statistics* (Vol. 14, pp. 241-268): Elsevier.

Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., & Keogh, E. (2008). Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment, 1*(2), 1542-1552.

Disney, S. M., & Towill, D. R. (2003). Vendor-managed inventory and bullwhip reduction in a two-level supply chain. *International Journal of Operations & Production Management, 23*(6), 625-651. doi:10.1108/01443570310476654

Esling, P., & Agon, C. (2012). Time-series data mining. *ACM Computer Surveys (CSUR), 45*(1), 1-34. doi:10.1145/2379776.2379788

Espinoza, M., Joye, C., Belmans, R., & De Moor, B. (2005). Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series. *IEEE Transactions on Power Systems, 20*(3), 1622-1630. doi:10.1109/TPWRS.2005.852123

ESRI. (2011). ArcGIS Desktop, (Version 10). Redlands, CA: Environmental Systems Research Institute.

Ettl, M., Zadrozny, B., Chowdhary, P., & Abe, N. (2005, 22-26 Aug. 2005). *Business performance management system for CRM and sales execution.* Paper presented at the Database and Expert Systems Applications, 2005. Proceedings. Sixteenth International Workshop on.

Ferbar Tratar, L. (2015). Forecasting method for noisy demand. *International Journal of Production Economics, 161*, 64-73. doi:http://dx.doi.org/10.1016/j.ijpe.2014.11.019

Fisher, W. (1958). On grouping for maximum homogeneity. *Journal of the American statistical association, 53*(284), 789-798.

Ford, T. H. (2017). Henry Ford's Rouge. Retrieved from https://www.thehenryford.org/visit/ford-rouge-factory-tour/history-and-timeline/fords-rouge/

Forrester, J. W. (1958). Industrial dynamics-a major breakthrough for decision makers. *Harvard business review, 36*(4), 37.

Forslund, H., & Jonsson, P. (2007). The impact of forecast information quality on supply chain performance. *International Journal of Operations & Production Management, 27*(1), 90-107. doi:doi:10.1108/01443570710714556

Gann, D. M. (1996). Construction as a manufacturing process? Similarities and differences between industrialized housing and car production in Japan. *Construction Management and Economics, 14*(5), 437-450. doi:10.1080/014461996373304

Gansterer, M. (2015). Aggregate planning and forecasting in make-to-order production systems. *International Journal of Production Economics, 170, Part B*, 521-528. doi:http://dx.doi.org/10.1016/j.ijpe.2015.06.001

Ghobbar, A. A., & Friend, C. H. (2003). Evaluation of forecasting methods for intermittent parts demand in the field of aviation: a predictive model. *Computers & Operations Research, 30*(14), 2097-2114. doi:http://dx.doi.org/10.1016/S0305-0548(02)00125-9

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: The dtw package. *Journal of Statistical Software, 31*(7), 1-24.

Golay, X., Kollias, S., Stoll, G., Meier, D., Valavanis, A., & Boesiger, P. (1998). A new correlation-based fuzzy logic clustering algorithm for FMRI. *Magnetic Resonance in Medicine, 40*(2), 249-260. doi:10.1002/mrm.1910400211

Goldratt, E. M. (1990). Theory of constraints: What is this thing called the theory of constraints and how should it be implemented. *Croton-on-Hudson, North River, New York*.

Haben, S., Ward, J., Vukadinovic Greetham, D., Singleton, C., & Grindrod, P. (2014). A new error measure for forecasts of household-level, high resolution electrical energy consumption. *International Journal of Forecasting, 30*(2), 246-256. doi:http://dx.doi.org/10.1016/j.ijforecast.2013.08.002

Han, J., & Kamber, M. (2006). *Data mining, concepts and techniques* (2nd ed.). San Francisco, CA: Morgan kaufmann.

Harvey, A., Ruiz, E., & Sentana, E. (1989). Unobserved component time series models with ARCH disurbances. *London School of Economics*, 1-29.

Hautamaki, V., Nykanen, P., & Franti, P. (2008, 8-11 Dec. 2008). *Time-series clustering by approximate prototypes.* Paper presented at the The 19th International Conference on Pattern Recognition.

Heer, J., & Kandel, S. (2012). Interactive analysis of big data. *XRDS, 19*, 50-54.

Hernández, J. E., Mula, J., Poler, R., & Lyons, A. C. (2014). Collaborative planning in multi-tier supply chains supported by a negotiation-based mechanism and multi-agent system. *Group*

Decision and Negotiation, 23(2), 235-269. doi:http://dx.doi.org/10.1007/s10726-013-9358-2

Holweg, M., Disney, S., Holmström, J., & Småros, J. (2005). Supply Chain Collaboration: Making Sense of the Strategy Continuum. *European Management Journal, 23*(2), 170-181. doi:http://dx.doi.org/10.1016/j.emj.2005.02.008

Höppner, F., & Klawonn, F. (2009). Compensation of translational displacement in time series clustering using cross correlation. In N. Adams, C. Robardet, A. Siebes, & J.-F. Boulicaut (Eds.), *Advances in Intelligent Data Analysis VIII* (Vol. 5772, pp. 71-82): Springer Berlin Heidelberg.

Hoshmand, A. R. (2010). *Business forecasting: A practical approach*: Routledge.

Huang, Z., Cheung, D. W., & Ng, M. K. (2001, 21-21 April 2001). *An empirical study on the visual cluster validation method with Fastmap.* Paper presented at the Database Systems for Advanced Applications, 2001. Proceedings.

Huber, J., Gossmann, A., & Stuckenschmidt, H. (2017). Cluster-based hierarchical demand forecasting for perishable goods. *Expert Systems with Applications, 76*, 140-151. doi:10.1016/j.eswa.2017.01.022

Hung, C., & Tsai, C.-F. (2008). Market segmentation based on hierarchical self-organizing map for markets of multimedia on demand. *Expert Systems with Applications, 34*(1), 780-787. doi:http://dx.doi.org/10.1016/j.eswa.2006.10.012

Hyndman, R. J. (2017). forecast: Forecasting functions for time series and linear models: R package Version 8.0. Retrieved from http://github.com/robjhyndman/forecast.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679-688. doi:http://dx.doi.org/10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., Lee, A., & Wang, E. (2014). *Fast computation of reconciled forecasts for hierarchical and grouped time series*. Monash University.

Izakian, H., Pedrycz, W., & Jamal, I. (2015). Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence, 39*, 235-244. doi:10.1016/j.engappai.2014.12.015

Jacobs, F. R., Berry, W. L., Whybark, D. C., & Vollmann, T. E. (2011). *Manufacturing planning & control for supply chain management* (6th ed.): McGraw-Hill Irwin.

Jacobs, F. R., & Chase, R. B. (2010). *Operations and supply management: The core*: McGraw-Hill Irwin New York, NY.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, 31*(8), 651-666. doi:http://dx.doi.org/10.1016/j.patrec.2009.09.011

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *Acm Computing Surveys, 31*(3), 264-323. doi:Doi 10.1145/331499.331504

Janvier-James, A. M. (2012). A new introduction to supply chains and supply chain management: Definitions and theories perspective. *International Business Research, 5*(1), 194-207.

Jin, Y., Williams, B. D., Tokar, T., & Waller, M. A. (2015). Forecasting with temporally aggregated demand signals in a retail supply chain. *Journal of Business Logistics, 36*(2), 199-211. doi:10.1111/jbl.12091

Jung, S., Chang, T., Sim, E., & Park, J. (2005). Vendor managed inventory and its effect in the supply chain. In D.-K. Baik (Ed.), *Systems Modeling and Simulation: Theory and Applications* (Vol. 3398, pp. 545-552): Springer Berlin Heidelberg.

Kalpakis, K., Gada, D., & Puttagunta, V. (2001, 2001). *Distance measures for effective clustering of ARIMA time-series.* Paper presented at the Proceedings 2001 IEEE International Conference on Data Mining.

Kantardzic, M. (2011). *Data mining : Concepts, models, methods, and algorithms* (Second edition ed.). Hoboken, NJ: Wiley-IEEE Press.

Kashwan, K. R., & Velu, C. M. (2013). Customer segmentation using clustering and data mining techniques. *International Journal of Computer Theory and Engineering, 5*(6), 856-861. doi:10.7763/ijcte.2013.v5.811

Kayacan, E., Ulutas, B., & Kaynak, O. (2010). Grey system theory-based models in time series prediction. *Expert Systems with Applications, 37*(2), 1784-1789. doi:10.1016/j.eswa.2009.07.064

Kembro, J., & Näslund, D. (2014). Information sharing in supply chains, myth or reality? A critical analysis of empirical literature. *International Journal of Physical Distribution & Logistics Management, 44*(3), 179-200.

Keogh, E., & Kasetty, S. (2003). On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery, 7*(4), 349-371. doi:Doi 10.1023/A:1024988512476

Keogh, E., & Pazzani, M. (1999). Scaling up dynamic time warping to massive datasets *Principles of Data Mining and Knowledge Discovery* (pp. 1-11): Springer.

Keogh, E., & Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems, 7*(3), 358-386. doi:10.1007/s10115-004-0154-9

Kim, T. Y., Dekker, R., & Heij, C. (2017). Spare part demand forecasting for consumer goods using installed base information. *Computers & Industrial Engineering, 103*, 201-215. doi:https://doi.org/10.1016/j.cie.2016.11.014

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78*(9), 1464-1480.

Kolassa, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting, 32*(3), 788-803. doi:https://doi.org/10.1016/j.ijforecast.2015.12.004

Kourentzes, N. (2014). On intermittent demand model optimisation and selection. *International Journal of Production Economics, 156*, 180-190. doi:http://dx.doi.org/10.1016/j.ijpe.2014.06.007

Kourentzes, N., & Petropoulos, F. (2016). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics, 181, Part A*, 145-153. doi:http://dx.doi.org/10.1016/j.ijpe.2015.09.011

Kourentzes, N., Petropoulos, F., & Trapero, J. R. (2014). Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting, 30*(2), 291-302. doi:https://doi.org/10.1016/j.ijforecast.2013.09.006

Krieger, A. M., & Green, P. E. (1996). Modifying cluster-based segments to enhance agreement with an exogenous response variable. *Journal of marketing research, 33*(3), 351.

Kuo, R. J., Ho, L. M., & Hu, C. M. (2002). Cluster analysis in industrial market segmentation through artificial neural network. *Computers & Industrial Engineering, 42*(2–4), 391-399. doi:http://dx.doi.org/10.1016/S0360-8352(02)00048-7

Lapide, L. (2006). Top-down & bottom-up forecasting in S&OP. *The journal of business*.

Le, T., Agard, B., & Deveault, S. (2009). *Application du data mining à la segmentation du marché des meubles aux États-Unis.* Paper presented at the Eighth Congres International de Genie Industriel.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). The bullwhip effect in supply chains. *Sloan management review, 38*(3), 93.

Léger, P. M., Pellerin, R., & Babin, G. (2011). Chapter 15: The analytical organization *Readings on Enterprise Resource Planning* (pp. 332). Montreal, QC: HEC.

Liao, W. T. (2005). Clustering of time series data—a survey. *Pattern Recognition, 38*(11), 1857-1874. doi:http://dx.doi.org/10.1016/j.patcog.2005.01.025

Longbing, C. (2014). Behavior informatics: A new perspective. *Intelligent Systems, IEEE, 29*(4), 62-80. doi:10.1109/MIS.2014.60

Lummus, R. R., & Vokurka, R. J. (1999). Defining supply chain management: A historical perspective and practical guidelines. *Industrial Management & Data Systems, 99*(1), 11-17.

MacKay, D. (2003). An example inference task: Clustering. *Information Theory, Inference and Learning Algorithms*, 284-292.

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Paper presented at the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.

Makridakis, S. (1989). Why combining works? *International Journal of Forecasting, 5*(4), 601-603. doi:http://dx.doi.org/10.1016/0169-2070(89)90017-4

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications*: John Wiley & Sons.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity.

Maqbool, O., & Babri, H. A. (2007). Hierarchical clustering for software architecture recovery. *Ieee Transactions on Software Engineering, 33*(11), 759-780. doi:10.1109/Tse.2007.70732

Maté, C. G. (2011). A multivariate analysis approach to forecasts combination. Application to Foreign Exchange (FX) markets. *Revista Colombiana de Estadística, 34*(2), 347-375.

Miller, H. J., & Han, J. (2009). *Geographic data mining and knowledge discovery*. Boca Raton, Fla.: CRC.

Moon, M. A. (2013). *Demand and Supply Integration: The Key to World-Class Demand Forecasting*: FT Press.

Mori, U., Mendiburu, A., & Lozano, J. (2015). *TSdist: Distance measures for time series data*. Retrieved from https://CRAN.R-project.org/package=TSdist

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent dirichlet allocation. *Expert Systems with Applications, 42*(3), 1314-1324. doi:10.1016/j.eswa.2014.09.024

Mueen, A., & Keogh, E. (2016). *Extracting optimal performance from dynamic time warping.* Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Murray, P., Agard, B., & Barajas, M. (2015). *Forecasting supply chain demand by clustering customers*. Paper presented at the 2015 IFAC Symposium on Information Control in Manufacturing - INCOM 2015, Ottawa, Canada. http://www.sciencedirect.com/science/article/pii/S2405896315005923

Murray, P., Agard, B., & Barajas, M. (2017). Market segmentation through data mining: A method to extract behaviors from a noisy data set. *Computers & Industrial Engineering, 109*, 233-252. doi:https://doi.org/10.1016/j.cie.2017.04.017

Murray, P., Agard, B., & Barajas, M. (2018a). ASACT - Data preparation for forecasting: A method to substitute transaction data for unavailable product consumption data. *Submitted to: International Journal of Production Economics*, 26.

Murray, P., Agard, B., & Barajas, M. (2018b). Forecast of individual customer's demand from a large and noisy dataset. *Computers & Industrial Engineering, 118*, 33-43. doi:10.1016/j.cie.2018.02.007

Murray, P., Agard, B., & Barajas, M. (2018c). Forecast of individual customer's demand from a large and noisy dataset. *Computers & Industrial Engineering, In Press*. doi:https://doi.org/10.1016/j.cie.2018.02.007

Murray, P., Agard, B., & Barajas, M. (2018d). Predicting customer purchase behavior patterns based on descriptive attributes. *Submitted to: The Journal of Industrial Marketing Management*, 19.

Newbold, P. (1983). ARIMA model building and the time series analysis approach to forecasting. *Journal of Forecasting, 2*(1), 23.

Nikolopoulos, K., Syntetos, A. A., Boylan, J. E., Petropoulos, F., & Assimakopoulos, V. (2011). An aggregate-disaggregate intermittent demand approach (ADIDA) to forecasting: An empirical proposition and analysis. *Journal of the Operational Research Society, 62*(3), 544-554. doi:10.1057/jors.2010.32

Pai, P.-F., & Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega, 33*(6), 497-505. doi:http://dx.doi.org/10.1016/j.omega.2004.07.024

Park, C., & Leeds, M. (2016). A highly efficient robust design under data contamination. *Computers & Industrial Engineering, 93*, 131-142. doi:http://dx.doi.org/10.1016/j.cie.2015.11.016

Petropoulos, F., & Kourentzes, N. (2015). Forecast combinations for intermittent demand. *Journal of the Operational Research Society, 66*(6), 914-924. doi:10.1057/jors.2014.62

Petropoulos, F., Kourentzes, N., & Nikolopoulos, K. (2016). Another look at estimators for intermittent demand. *International Journal of Production Economics, 181, Part A*, 154-161. doi:http://dx.doi.org/10.1016/j.ijpe.2016.04.017

Petry, F. E., & Zhao, L. (2009). Data mining by attribute generalization with fuzzy hierarchies in fuzzy databases. *Fuzzy Sets and Systems, 160*(15), 2206-2223. doi:https://doi.org/10.1016/j.fss.2009.02.014

Phillips, P. C. B., & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies, 53*(4), 473-495. doi:10.2307/2297602

Prestwich, S. D., Tarim, S. A., Rossi, R., & Hnich, B. (2014). Forecasting intermittent demand by hyperbolic-exponential smoothing. *International Journal of Forecasting, 30*(4), 928-933. doi:http://dx.doi.org/10.1016/j.ijforecast.2014.01.006

Provost, F., & Fawcett, T. (2013). *Data science for business*. Sebasyopol, CA: O'Reilly Media, Inc.

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org

Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Qiang, Z., . . . Keogh, E. (2013). Addressing big data time series: mining trillions of time series subsequences under dynamic time warping. *ACM Transactions on Knowledge Discovery from Data, 7*(3), 10 (31 pp.). doi:10.1145/2500489

Ramanathan, U., & Gunasekaran, A. (2014). Supply chain collaboration: Impact of success in long-term partnerships. *International Journal of Production Economics, 147, Part B*(0), 252-259. doi:http://dx.doi.org/10.1016/j.ijpe.2012.06.002

Randle, M., & Dolnicar, S. (2009). Not just any volunteers: Segmenting the market to attract the high contributors. *Journal of Nonprofit & Public Sector Marketing, 21*(3), 271-282. doi:10.1080/10495140802644513

Rao, A. V. (1973). A comment on: Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society, 24*(4), 639-640.

Ratanamahatana, C. A., & Keogh, E. (2005). *Three myths about dynamic time warping data mining.* Paper presented at the Proceedings of SIAM International Conference on Data Mining.

Rezaei, J., & Ortt, R. (2013). Supplier segmentation using fuzzy logic. *Industrial Marketing Management, 42*(4), 507-517. doi:https://doi.org/10.1016/j.indmarman.2013.03.003

Rigby, D., & Bilodeau, B. (2015). *Mangement tools & trends 2015*. Retrieved from http://www.bain.com/publications/articles/management-tools-and-trends-2015.aspx

Romdhane, L. B., Fadhel, N., & Ayeb, B. (2010). An efficient approach for building customer profiles from business data. *Expert Systems with Applications, 37*(2), 1573-1585. doi:10.1016/j.eswa.2009.06.050

Rostami-Tabar, B., Babai, M. Z., Syntetos, A., & Ducq, Y. (2013). Demand forecasting by temporal aggregation. *Naval Research Logistics, 60*(6), 479-498.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on, 26*(1), 43-49. doi:10.1109/TASSP.1978.1163055

Sarabia, F. J. (1996). Model for market segments evaluation and selection. *European Journal of Marketing, 30*(4), 58-74.

Seret, A., Maldonado, S., & Baesens, B. (2015). Identifying next relevant variables for segmentation by using feature selection approaches. *Expert Systems with Applications, 42*(15–16), 6255-6266. doi:http://dx.doi.org/10.1016/j.eswa.2015.01.070

Serra, A. P., & Zárate, L. E. (2015). Characterization of time series for analyzing of the evolution of time series clusters. *Expert Systems with Applications, 42*(1), 596-611. doi:http://dx.doi.org/10.1016/j.eswa.2014.08.012

Sevlian, R., & Rajagopal, R. (2018). A scaling law for short term load forecasting on varying levels of aggregation. *International Journal of Electrical Power & Energy Systems, 98*, 350-361. doi:https://doi.org/10.1016/j.ijepes.2017.10.032

Shapiro, J. F. (2007). *Modeling the supply chain* (2nd ed.). Belmont, CA: Thompson Brooks/Cole.

Shenstone, L., & Hyndman, R. J. (2005). Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting, 24*(6), 389-402.

Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing, 21*(1), 3-8. doi:10.2307/1247695

Sommers, M., & Barnes, J. (1982). *Marketing* (10th ed.). Canada: McGraw-Hill Ryerson.

Spithourakis, G. P., Petropoulos, F., Nikolopoulos, K., & Assimakopoulos, V. (2012). A systemic view of the ADIDA framework. *IMA Journal of Management Mathematics, 25*(2), 125-137.

Syntetos, A. A. (2016). Daily, weekly, monthly, yearly; Why forecasting at the right time makes all the difference.

Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting, 21*(2), 303-314. doi:http://dx.doi.org/10.1016/j.ijforecast.2004.10.001

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. *Biometrika, 59*(3), 525-531.

Tsiptsis, K. K., & Chorianopoulos, A. A. (2011). *Data mining techniques in CRM: Inside customer segmentation*: John Wiley & Sons.

United States Census Bureau. (2015). Geography. Retrieved from https://www.census.gov/geo/maps-data/data/tiger.html van Donselaar, K., Peters, J., de Jong, A., & Broekmeulen, R. (2016). Analysis and forecasting of demand during promotions for perishable items. *International Journal of Production Economics, 172*, 65.

Van Wijk, J. J., & Van Selow, E. R. (1999, 1999). *Cluster and calendar based visualization of time series data.* Paper presented at the Information Visualization, 1999 IEEE Symposium on.

Verdu, S. V., Garcia, M. O., Senabre, C., Marin, A. G., & Franco, F. J. G. (2006). Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. *IEEE Transactions on Power Systems, 21*(4), 1672-1682. doi:10.1109/Tpwrs.2006.881133

Vlckova, V., Lostakova, H., & Patak, M. (2014). *Selected problems arising within cluster analysis usage for market segmentation.* Paper presented at the 23rd International Conference on Metallurgy and Materials, METAL 2014, Brno, Czech republic.

Wang, G., Gunasekaran, A., Ngai, E. W. T., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics, 176*, 98-110. doi:http://dx.doi.org/10.1016/j.ijpe.2016.03.014

Wang, X., Mueen, A., Ding, H., Trajcevski, G., Scheuermann, P., & Keogh, E. (2013). Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery, 26*(2), 275-309. doi:10.1007/s10618-012-0250-5

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association, 58*(301), 236-244.

Wedel, M., & Kamakura, W. A. (2012). *Market segmentation: Conceptual and methodological foundations* (Vol. 8): Springer Science & Business Media.

Wehrens, R., & Buydens, L. M. (2007). Self-and super-organizing maps in R: The Kohonen package. *Journal of Statistical Software*, *21*(5), 1-19.

Wei, W. W. S., & Mehta, J. (1980, Oct. 1979). *Temporal aggregation and information loss in a distributed lag model.* Paper presented at the Analysing Time Series. International Conference, Amsterdam, Netherlands.

Wickham, H. (2014). Tidy data. *Journal of Statistical Software, 59*(1). doi:10.18637/jss.v059.i10

Willemain, T. R., Smart, C. N., & Schwarz, H. F. (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting, 20*(3), 375-387. doi:http://dx.doi.org/10.1016/S0169-2070(03)00013-X

Wind, Y. (1978). Issues and advances in segmentation research. *Journal of marketing research, 15*(3), 317-337. doi:10.2307/3150580

Witten, I., Farnk, E., & Hall, M. (2011). *Data mining* (Third edition ed.). Burlington, MA: Morgan Kaufmann / Elsevier Inc.

Ye, N., & Li, X. (2002). A scalable, incremental learning algorithm for classification problems. *Computers & Industrial Engineering*, *43*(4), 677-692. doi:http://dx.doi.org/10.1016/S0360-8352(02)00132-8

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175. doi:http://dx.doi.org/10.1016/S0925-2312(01)00702-0

Zhang, Z., Kaiqi, H., & Tieniu, T. (2006, 0-0 0). *Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes*. Paper presented at the The 18th International Conference on Pattern Recognition.

APPENDICES

APPENDIX A – R PACKAGES

The following list are packages and their version number used in this research:

"assertthat 0.2.0"
"base 3.4.1"
"bindr 0.1"
"bindrcpp 0.2"
"C50 0.1.0.24"
"cellranger 1.1.0"
"chron 2.3.50"
"class 7.3.14"
"clue 0.3.53"
"cluster 2.0.6"
"clv 0.3.2.1"
"clValid 0.6.6"
"codetools 0.2.15"
"colorspace 1.3.2"
"compiler 3.4.1"
"data.table 1.10.4"
"DataCombine 0.2.21"
"datasets 3.4.1"
"DEoptimR 1.0.8"
"digest 0.6.12"
"diptest 0.75.7"
"dplyr 0.7.1"
"dtw 1.18.1"
"dtwclust 4.0.1"
"flexclust 1.3.4"
"flexmix 2.3.14"
"foreach 1.4.3"
"forecast 8.1"
"Formula 1.2.2"
"fpc 2.1.10"
"fracdiff 1.4.2"
"ggdendro 0.1.20"
"ggplot2 2.2.1"
"glmnet 2.0.10"
"glue 1.1.1"
"graphics 3.4.1"
"grDevices 3.4.1"
"grid 3.4.1"
"gridExtra 2.2.1"
"gtable 0.2.0"

"hms 0.3"
"htmltools 0.3.6"
"htmlwidgets 0.8"
"hts 5.1.4"
"httpuv 1.3.3"
"ifultools 2.0.4"
"iterators 1.0.8"
"jsonlite 1.5"
"kernlab 0.9.25"
"KernSmooth 2.23.15"
"knitr 1.16"
"kohonen 3.0.2"
"lattice 0.20.35"
"lazyeval 0.2.0"
"lmtest 0.9.35"
"locpol 0.6.0"
"longitudinalData 2.4.1"
"lubridate 1.6.0"
"magrittr 1.5"
"MAPA 2.0.2"
"MASS 7.3.47"
"Matrix 1.2.10"
"matrixcalc 1.0.3"
"mclust 5.3"
"methods 3.4.1"
"mime 0.5"
"misc3d 0.8.4"
"modeltools 0.2.21"
"munsell 0.4.3"
"mvtnorm 1.0.6"
"nloptr 1.0.4"
"nnet 7.3.12"
"parallel 3.4.1"
"partykit 1.1.1"
"pdc 1.0.3"
"pkgconfig 2.0.1"
"pkgmaker 0.22"
"plyr 1.8.4"
"prabclus 2.2.6"
"proxy 0.4.17"
"quadprog 1.5.5"
"quantmod 0.4.10"
"R6 2.2.2"
"RColorBrewer 1.1.2"
"Rcpp 0.12.11"
"readr 1.1.1"
"readxl 1.0.0"

"registry 0.3"
"reshape2 1.4.2"
"rgl 0.98.1"
"rlang 0.1.1"
"rngtools 1.2.4"
"robustbase 0.92.7"
"rpart 4.1.11"
"rpart.plot 2.1.2"
"RSpectra 0.12.0"
"scales 0.4.1"
"shiny 1.0.3"
"smooth 1.9.9"
"SparseM 1.77"
"splines 3.4.1"
"splus2R 1.2.2"
"stats 3.4.1"
"stats4 3.4.1"
"stringi 1.1.5"
"stringr 1.2.0"
"survival 2.41.3"
"tibble 1.3.3"
"tidyr 0.6.3"
"timeDate 3012.100"
"tools 3.4.1"
"trimcluster 0.1.2"
"TSclust 1.2.3"
"TSdist 3.4"
"tseries 0.10.42"
"tsintermittent 1.9"
"TTR 0.23.1"
"useful 1.2.3"
"utils 3.4.1"
"wmtsa 2.0.2"
"xtable 1.8.2"
"xts 0.9.7"
"zoo 1.8.0"